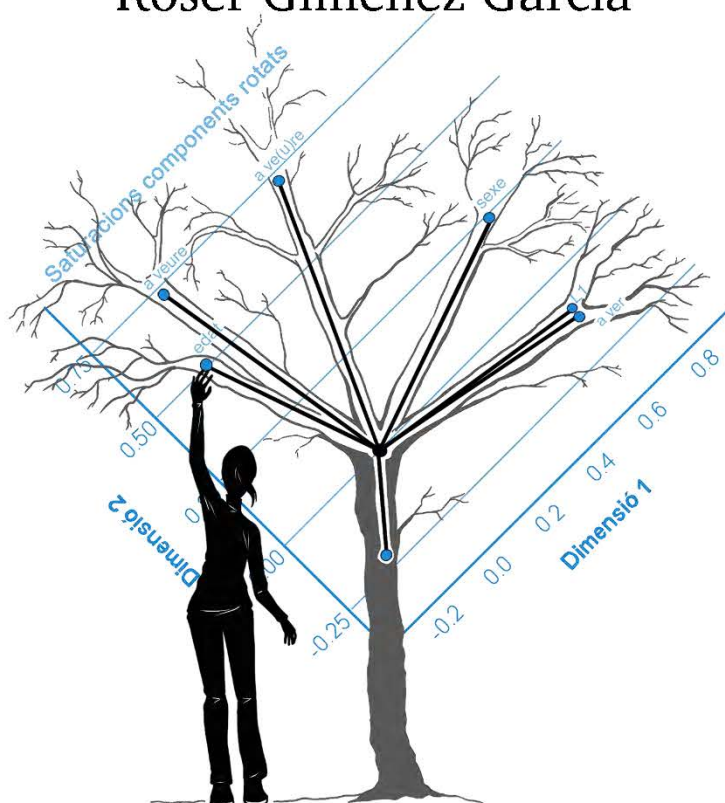


Programa de doctorat en
Estudis Avançats en Llengua i Literatura Catalanes

Edat, sexe i llengua inicial en l'elaboració de perfils lingüístics forenses d'adolescents en català

Roser Giménez García



Directors:

F. Xavier Vila i Moreno

Sheila Queralt Estevez

2022



UNIVERSITAT DE
BARCELONA



Edat, sexe i llengua inicial en l'elaboració de perfils lingüístics forenses d'adolescents en català

Roser Giménez García

Directors:

F. Xavier Vila i Moreno

Sheila Queralt Estevez

Tutor:

Lluís Payrató Giménez

Tesi doctoral

Programa de doctorat en

Estudis Avançats en Llengua i Literatura Catalanes

Facultat de Filologia i Comunicació

Universitat de Barcelona

2022



UNIVERSITAT DE
BARCELONA

Als meus avis
i a les moltes altres persones que
mai podrem acabar d'entendre
que la meva mare ja no hi sigui.

Encarna

Su partida nos partió.
Profundamente injusto,
su sufrimiento nos fundió
un enfriamiento eterno dentro. Y dolor.
Arrasaron las lágrimas la ilusión.

Siguió el tiempo.
Ni más rápido ni más lento.

Sin inmutarse, sigue.
Y su ausencia persiste.

Pero también el sol insiste.
Y nos riega su color. Dice:
Encarna la vida. Encarna pasión.

Una de tus últimas mañanas te prometí llorando que terminaría el doctorado y tú solo pudiste acariciarme. Desde el 3 de enero del 2021 hasta ahora no he dejado de preguntarme cómo iba a poder hacerlo sin ti. Pero ahora lo entiendo: lo he hecho *sintigo*.

AGRAÏMENTS

Al treball de fi de grau, vaig condensar tot el meu agraïment en tres línies. Al treball de fi de màster, vaig aconseguir explicar-me millor, però no vaig omplir ni una pàgina. Així que espero que per valorar l'agraïment que sento, però em temo que no sé expressar gaire bé, sigui aplicable allò que el que compta és la intenció, perquè és molt difícil abocar-lo tot en unes paraules. Jo ho intento i ja em direu què m'he deixat o què no calia. Això sí, al cos de la tesi, tot és en “primer, segon, tercer lloc...” i “primer això i a continuació allò altre”, però aquí no tingueu gaire en compte l'ordre, perquè podia haver estat qualsevol altre.

Sant Pau, Barcelona, El Pont, Sant Joan, Manresa. Gràcies per ser-hi sempre, per continuar sent-hi en els moments més terribles i malgrat les ganes d'engegar-ho tot a rodar, per la perseverança aquesta nostra (tan seva) i pels bons moments, que són impagables i incontables (almenys per a mi, que ja sabeu que més de dos no...). Sense vosaltres, res.

El Catllar, Manresa, Sant Joan, Canet. Visca tota la comprensió, tota la paciència, tots els riures i els plors, les abraçades, les alegries, les paraules i els silencis plens d'emocions que hem compartit i que hem de compartir encara.

Amistats a distància, de contacte més o menys intermitent, però interès genuï recíproc i genuïnament desinteressat, en català, castellà and in English: *moltes gracias you for reaching out, giving some of your time to me across mad times zones, busy schedules, unstable restrictions and, generally, crazy lives. Thank you for being real good friends. Soc molt afortunada. Recordo berenars, passejades, viatges i moments i sento la vostra escalfor. Every time we meet it's like we were never apart, and if you write this in English it sounds like a cheesy love song, but it doesn't make it any less true. Nenes maques, FrAb, Aman(i)da, patos and Euan [ˈjuːən], TodasLasMujeresAmanLasVentas idò, Kelsie, Albin.*

A la Sheila. Un dia, en una habitació d'una casa de quatre habitacions, cadascuna habitada per un jove d'un origen ètnic diferent que compartir, compartir, el que es diu compartir casa no la compartien, en un barri qualsevol de la segona ciutat més poblada del Regne Unit, vaig decidir-me a enviar-te un correu. Mai a la vida (però mai a la vida, eh?) m'hauria imaginat tot el que em portaria aquell correu. Moltes gràcies per la sinceritat, el saber fer, la paciència, l'entusiasme i, sobretot, la generositat en les oportunitats d'aprenentatge i en tot. I per les provisions.

Al Xavier. Un dia, vam tenir una reunió en què no recordo haver dit gaire i, de sobte, estava fent els tràmits per fer una cosa que havia dit que jo no faria, en una universitat a les classes de la qual mai no havia entrat. Ara sí que entenc allò que deies dels doctorands i els esportistes d'elit. I tenies raó que és una etapa en

la qual passen coses a “la vida” (aquella cosa de fora de la tesi) que un no s’espera. Però el més important és he après a cada pas que has suggerit i, al final del viatge, estic molt contenta del resultat. Moltes gràcies per acceptar dirigir una persona desconeguda, per fer temps on no n’hi havia i per tot l’encoratjament que m’has regalat des del principi.

A la Cristina Illamola. Moltes gràcies per haver-me escoltat, ajudat, encoratjat, i ajudat mil vegades. Tu vas ser la primera “persona UB” que vaig conèixer, em vas ensenyar el CUSC, em vas donar accés al corpus de RESOL. Gràcies pels Juliols. No hauria pogut fer res d’això sense la teva ajuda. Et dec i t’admiro molt.

Al Lluís Payrató, tutor *in extremis*. Hem pogut parlar en poques ocasions, però en totes n’he après molt. Gràcies de tot cor per tots els llibres i els gestos, per adaptar-te als canvis de manera tan àgil i per posar les coses fàcils.

A Laura Barrios, por su pronta disposición, su amable paciencia y su ayuda inestimable (“tan valiosa”, como dice el DRAE, “que no puede ser estimada como corresponde”). De verdad, he aprendido mucho con tus explicaciones y esta tesis es mucho mejor de lo que habría sido sin tu ayuda. Muchas gracias.

Al Laboratorio, als sèniors (la Núria, el Jordi, la Montse), pel privilegi d’aprendre de vosaltres, i als que encara no som sèniors (però doneu-nos un temps): Garazi (a quien deseo lo mejor), l’Ariadna, la Novella, els estudiants virtuals i de pràctiques que hem tingut. Som-hi, que queda molt camí per recórrer! (I en tinc moltes ganes!).

Al CUSC, i especialment l’Avel·lí, el Natxo i la Marina, amb qui he tingut més tracte, moltes gràcies per acollir-me, respondre tots els meus dubtes i tractar-me com una investigadora més. Als altres investigadors predoctorals: Montse, Laia, Eloi, Pau, Jesús, moltes gràcies per tot, malgrat que no hàgim pogut compartir tantes estones com m’havia imaginat al principi. Tinc moltes ganes de veure els resultats finals de les vostres recerques. A les coordinadores de la secció, la Neus Nogué i la Núria Alturo, perquè sempre han estat molt eficients i amables. A la Lídia Solé, l’Encarna Hospital i les altres secretàries amb qui no he tingut tant de tracte. A la resta de companys del Departament, moltes gràcies per l’acollida (més o menys digital) i per les respostes pacients a les meves moltes preguntes.

Als investigadors, col·laboradors i informants dels projectes que van fer possible la creació del corpus RESOL, moltes gràcies per tota la feina.

A l’Andreu Gallofré per la generositat d’ajudar-me amb els comandaments de PowerShell, que van suposar un abans i un després a la tesi.

A la Cristina Garriga Turró per la il·lustració flipant de la portada i la seva gran eficiència. A l’Elena per posar-m’hi en contacte i per mil altres coses boniques.

Als professors del grau i del màster. Cada cita a algun dels vostres treballs és un recordatori afectuós tant a vosaltres com a la resta de professorat del qual he tingut la sort de poder aprendre. També aprofito per agrair les companyes i els companys tant de la Pompeu com d'Aston totes les estones que hem compartit reflexionant en veu alta, rient, fent treballs en grup i, en general, vivint al màxim una etapa de formació que sabíem que era molt especial.

Molta força a la Núria, la Rut i l'Alba, que aviat es trobaran davant d'aquesta mateixa pàgina en blanc i segur que se'n sortiran amb molta més elegància que jo. Moltes gràcies per compartir aquesta etapa tan estranya i els interrogants i les dificultats, per les reflexions brillants i plenes de seny, per no rendir-vos, per llegir cartes de motivació i per les referències bibliogràfiques, per venir a mostrar suport a diversos actes, per les passejades amb mascareta i les que farem parlant de quins llocs tindran la sort de tenir-vos quan hàgim acabat el doctorat.

A la colla dels diumenges, per les rutes marinades amb riures i aprenentatges. Cati, Tomás, Teresa, Montses, Pep, Joan: gràcies.

A José, Cristina y los demás amigos de Swansea por la oportunidad de aportar algo a grandes causas y por todo su apoyo, que significa mucho para mí.

A les moltes altres persones que són importants per a mi i que m'han ajudat en aquests anys llargs i curts alhora, bojos, desesperants, privilegiats, inimaginables. Em posaria a llistar noms i raons i tindriem unes altres sis-centes pàgines. No cal; sabeu que us estic agraïda i, si no, ja us ho recordaré.

Als gegants, perquè m'han deixat pujar a les seves espatlles i des d'aquí el panorama es veu ple de possibilitats.

I, a més, perquè ballen molt bé a la festa major.

I a les persones amb sentit de l'humor, perquè la vida ja és prou seriosa.

RESUM

La lingüística forense s'ha definit, en sentit ampli, com la interfície entre la llengua i el dret i, en sentit restrictiu, com l'aplicació de la sociolingüística a processos judicials i investigacions policials i privades. Actualment, aquesta branca de la lingüística aplicada es troba en expansió a molts països, inclosos els territoris de parla catalana. La perfilació sociolingüística de mostres dubitades és una de les aplicacions de la lingüística a contextos forenses que es poden sol·licitar en el marc de processos judicials o d'investigacions policials i privades. Per a l'elaboració de perfils sociolingüístics fiables, calen estudis empírics sobre les varietats contemporànies de la llengua. Aquesta tesi sorgeix d'una necessitat específica de la perfilació sociolingüística de mostres dubitades en català: comptar amb estudis que descriguin les varietats col·loquials utilitzades per parlants adolescents i que correlacionin els trets lingüístics d'aquestes varietats amb variables socials, com l'edat, el sexe i la llengua inicial dels parlants. La revisió bibliogràfica realitzada en aquesta investigació mostra que, en el cas del català, encara no existeixen estudis que cobreixin aquesta necessitat, ja que tant els que s'ocupen de la variació com els que tracten l'ús de la llengua que fan els parlants adolescents presenten diverses limitacions importants per a la seva aplicació en la perfilació lingüística relacionades amb les dades utilitzades (per exemple, per la seva antiguitat, mida o localització geogràfica, massa específica). En aquest estudi s'analitza un corpus de parla adolescent de 69 informants de dues localitats catalanes des del plantejament metodològic de la sociolingüística de la variació, al qual es dona l'enfocament pràctic de la lingüística forense per investigar la pregunta de recerca ¿hi ha variables lingüístiques que permeten identificar l'edat, el sexe i la llengua inicial més probable en el discurs d'un parlant adolescent en català? Per respondre-la, s'analitzen 22 variables lingüístiques en relació amb l'edat, el sexe i la llengua inicial dels informants mitjançant diverses tècniques estadístiques: l'Anàlisi de Components Principals, la prova dels rangs amb signe de Wilcoxon, la prova U de Mann-Whitney i la prova de Kruskal-Wallis. En total, catorze variables lingüístiques es correlacionen de forma significativa amb l'edat, deu amb la llengua inicial i sis amb el sexe. Des de la perspectiva inversa, s'utilitza una tècnica inferencial de mineria de

dades (arbres de classificació) per caracteritzar els grups d'edat, sexe i llengua inicial a partir de les variables lingüístiques. El percentatge de classificació correcta per als grups d'edat arriba al 85,5 %, per als grups de sexe, al 78,3 % i, en el cas de la llengua inicial, al 70,3 %. Per tant, es conclou que aquests resultats són aplicables a la perfilació sociolingüística de produccions anònimes en català comparables al corpus analitzat i que l'explotació de la sociolingüística de la variació ofereix un gran ventall d'oportunitats per a l'impuls de la lingüística forense catalana.

Paraules clau: sociolingüística, lingüística forense, perfilació lingüística, edat, sexe, llengua inicial, català.

RESUMEN

La lingüística forense se ha definido, en sentido amplio, como la interfaz entre la lengua y el derecho y, en sentido restrictivo, como la aplicación de la sociolingüística a procesos judiciales e investigaciones policiales y privadas. Actualmente, esta rama de la lingüística aplicada está en expansión en muchos países, incluidos los territorios de habla catalana. La perfilación sociolingüística de muestras dubitadas es una de las aplicaciones de la lingüística a contextos forenses que se pueden solicitar en el marco de procesos judiciales o de investigaciones policiales y privadas. Para la elaboración de perfiles sociolingüísticos fiables, son necesarios estudios empíricos sobre las variedades contemporáneas de la lengua. Esta tesis surge de una necesidad específica de la perfilación sociolingüística de muestras dubitadas en catalán: contar con trabajos que describan las variedades coloquiales utilizadas por hablantes adolescentes y que correlacionen los rasgos lingüísticos de estas variedades con variables sociales, como la edad, el sexo y la lengua inicial de los hablantes. La revisión bibliográfica realizada en esta investigación muestra que, en el caso del catalán, todavía no existen estudios que cubran esta necesidad, ya que tanto los que se ocupan de la variación como los que tratan del uso de la lengua que hacen los hablantes adolescentes presentan varias limitaciones importantes para su aplicación en la perfilación lingüística relacionadas con los datos utilizados (por ejemplo, por su antigüedad, tamaño o localización geográfica, demasiado específica). En este estudio se analiza un corpus de habla adolescente de 69 informantes de dos localidades catalanas desde el planteamiento metodológico de la sociolingüística de la variación, al que se da el enfoque práctico de la lingüística forense para investigar la pregunta de investigación ¿hay variables lingüísticas que permitan identificar la edad, el sexo y la lengua inicial más probable en el discurso de un hablante adolescente en catalán? Par responderla, se analizan 22 variables lingüísticas en relación con la edad, el sexo y la lengua inicial de los informantes mediante varias técnicas estadísticas: el Análisis de Componentes Principales, la prueba de los rangos con signo de Wilcoxon, la prueba U de Mann-Whitney y la prueba de Kruskal-Wallis. En total, catorce variables lingüísticas se correlacionan de forma significativa con la edad, diez

con la lengua inicial y seis con el sexo. Desde la perspectiva contraria, se utiliza una técnica inferencial de minería de datos (árboles de clasificación) para caracterizar los grupos de edad, sexo y lengua inicial a partir de las variables lingüísticas. El porcentaje de clasificación correcta para los grupos de edad alcanza el 85,5 %, para los grupos de sexo, el 78,3 % y, en el caso de la lengua inicial, el 70,3 %. Por tanto, se concluye que estos resultados son aplicables a la perfilación sociolingüística de producciones anónimas en catalán comparables al corpus analizado y que la explotación de la sociolingüística de la variación ofrece un amplio abanico de oportunidades para el impulso de la lingüística forense catalana.

Palabras clave: sociolingüística, lingüística forense, perfilación lingüística, edad, sexo, lengua inicial, catalán.

ABSTRACT

Forensic linguistics has been broadly defined as the intersection of language and the law and, in a more restricted sense, as the application of sociolinguistics to legal processes and police or private investigations. Currently, this area of applied linguistics is expanding in many countries, including the Catalan-speaking areas. Linguistic profiling of questioned samples is one of the applications of linguistics to forensic contexts that may be requested as part of legal processes for investigative purposes. The preparation of reliable linguistic profiles relies on preexisting empirical studies on contemporary language varieties. This research arises from a specific need in the linguistic profiling of questioned samples in Catalan, namely, to have studies which describe informal varieties used by adolescent speakers and which correlate the linguistic features of these varieties with social variables, such as speakers' age, sex and first language. The literature reviewed carried out in this study shows that no works meet this need for the Catalan language yet, since both studies on language variation and those on adolescent speech present limitations in their applicability to linguistic profiling due to sampling issues (e.g., the samples were collected too long ago, their size is insufficient, or the generalizability of the results is limited to a very specific geographical area). This study analyzes a sample of the speech of 69 adolescent speakers from two Catalan cities through the lens of variationist sociolinguistics and the hands-on approach of forensic linguistics and examines the research question 'are there linguistic variables that may be used to identify the most likely age, sex and first language in adolescent speech in Catalan?'. To answer this question, this study analyzes 22 linguistic variables in relation to the informants' age, sex and first language by means of several statistical techniques: Principal Component Analysis, the Wilcoxon signed-rank test, the Mann-Whitney U test and the Kruskal-Wallis test by ranks. In total, fourteen linguistic variables show a statistically significant correlation with age, ten with first language and six with sex. From the opposite perspective, a data mining technique (classification trees) is used to characterize the groups of informants by age, sexe and first language on the basis of their use of the linguistic variables. The model reaches a correct classification rate of 85.5% for

Abstract

speakers' age, 78.3% for sex, and 70.3% for first language. Thus, it is concluded that the results obtained are applicable to the linguistic profiling of questioned samples comparable to the one used in this study and that variationist sociolinguistics offers a wide range of possibilities to boost forensic linguistics in Catalan.

Keywords: sociolinguistics, forensic linguistics, linguistic profiling, age, sex, first language, Catalan.

ÍNDIX

AGRAÏMENTS	ii
RESUM	v
RESUMEN	vii
ABSTRACT	ix
ÍNDIX	1
1 INTRODUCCIÓ.....	7
2 MARC TEÒRIC	15
2.1. La lingüística forense	15
2.1.1. Àmbits d'actuació de la lingüística forense.....	15
2.1.2. Història de la lingüística forense.....	17
2.1.3. La lingüística forense a Catalunya.....	20
2.1.4. L'anàlisi forense del discurs dels menors i l'elaboració de perfils	24
2.1.4.1. L'anàlisi del discurs en àmbits forenses.....	24
2.1.4.1.1. L'anàlisi del discurs produït fora de l'àmbit forense	25
2.1.4.1.2. L'anàlisi del discurs produït dins de l'àmbit forense	33
2.1.4.2. Els perfils sociolingüístics.....	35
2.1.5. La lingüística forense: síntesi	37
2.2. La sociolingüística	37
2.2.1. Història de la sociolingüística	38
2.2.1.1. El sorgiment de la sociolingüística	38
2.2.1.2. La interdisciplinarietat de la sociolingüística.....	42
2.2.2. Divergències teòriques entre la sociolingüística i la lingüística formal.....	44
2.2.3. Els enfocaments sociolingüístics.....	47
2.2.4. El variacionisme.....	49
2.2.4.1. Les bases teòriques i metodològiques del variacionisme.....	50
2.2.4.1.1. L'aplicació del mètode hipoteticodeductiu a l'estudi de la variació.....	50
2.2.4.1.2. Les nocions de variable i regla variable.....	52
2.2.4.1.3. La revolució variacionista del temps	53
2.2.4.1.4. L'oposició variacionista entre els canvis des de dalt i des de baix.....	55
2.2.4.2. L'evolució del variacionisme.....	56
2.2.4.3. Els estudis variacionistes de la llengua catalana.....	58
2.2.5. La sociolingüística com a camp de recerca: síntesi.....	62

Índex

2.3.	Les principals variables independents dels estudis sociolingüístics	62
2.3.1.	La variable sexe	63
2.3.1.1.	La rellevància de la variable sexe.....	63
2.3.1.2.	L'anàlisi variacionista de les diferències per sexe.....	64
2.3.1.3.	Diferències entre els sexes en contextos interaccionals.....	66
2.3.1.4.	Explicacions i crítiques a l'anàlisi del sexe i llengua	68
2.3.1.5.	L'evolució de l'estudi de la variable sexe	70
2.3.1.6.	La variable sexe: síntesi.....	71
2.3.2.	La variable edat: la llengua dels adolescents.....	72
2.3.2.1.	La importància de la variable edat	72
2.3.2.2.	L'adquisició lingüística durant l'adolescència	73
2.3.2.3.	Explicacions de la variació lingüística relacionada amb l'edat	77
2.3.2.4.	La variable edat: síntesi.....	77
2.3.3.	Les variables classe social i nivell educatiu.....	78
2.3.4.	Llengua inicial i varietat dialectal	81
2.3.5.	La interacció entre les variables i els biaixos en la recerca existent	86
2.3.6.	Les principals variables independents dels estudis sociolingüístics: síntesi	91
3	OBJECTIUS I HIPÒTESIS.....	93
4	CONTEXT DE LA RECERCA	95
4.1.	La realitat demolingüística dels joves catalans	95
4.2.	La realitat demolingüística de Manlleu i Mataró.....	107
4.3.	Estudis sociolingüístics previs sobre els adolescents de Manlleu i Mataró provinents de RESOL	110
4.4.	El català dels joves.....	119
4.5.	Context de la recerca: síntesi	122
5	METODOLOGIA	123
5.1.	El projecte RESOL	123
5.2.	El corpus.....	125
5.3.	Disseny de l'estudi	129
5.3.1.	Primera etapa: preparació del corpus.....	131
5.3.2.	Segona etapa: les variables.....	132
5.3.2.1.	Les variables independents.....	132
5.3.2.1.1.	Edat	132

Índex

5.3.2.1.2.	Sexe	133
5.3.2.1.3.	Llengua inicial.....	134
5.3.2.2.	Altres variables independents.....	134
5.3.2.2.1.	Gènere.....	135
5.3.2.2.2.	Classe social.....	136
5.3.2.2.3.	Nivell educatiu.....	137
5.3.2.2.4.	Varietat dialectal	137
5.3.2.3.	Les variables dependents.....	138
5.3.2.3.1.	Les variables pragmàtiques.....	144
5.3.2.3.2.	Les variables morfosintàctiques	153
5.3.2.3.3.	Les variables lèxiques.....	163
5.3.2.3.4.	Les variables dependents: síntesi	168
5.3.2.4.	Prova pilot de les variables	171
5.3.3.	Tercera etapa: l'anàlisi	177
5.4.	La codificació de les variables independents	178
5.5.	La codificació de les variables dependents	179
5.5.1.	Doblets de partícules discursives (V3)	179
5.5.2.	Obertura de la interacció (V5).....	182
5.5.3.	Heterorepeticions (V6).....	183
5.5.4.	Posició dels pronoms clítics en construccions perifràstiques (V14).....	184
5.5.5.	Nombre de preguntes no retòriques (V16)	186
5.5.6.	Tipus de preguntes (V17)	187
5.5.7.	Construccions condicionals (V18).....	188
5.5.8.	Perífrasis d'obligació (V22).....	189
5.5.9.	Determinants demostratius (V23).....	190
5.5.10.	Pronoms clítics de datiu plural (V25).....	191
5.5.11.	Fenòmens de contacte de llengües (V38).....	192
5.5.12.	Colors (V41).....	193
5.5.13.	Nombre d'adverbis acabats en <i>-ment</i> (V45).....	193
5.6.	Les tècniques d'anàlisi estadística.....	194
5.6.1.	Exploració de les relacions globals i caracterització dels informants	198
5.6.1.1.	Anàlisi de Components Principals	199
5.6.1.2.	Arbres de classificació	202

Índex

5.6.2.	Significació estadística de les dades.....	203
5.6.2.1.	U de Mann-Whitney.....	204
5.6.2.2.	Kruskal-Wallis.....	205
5.6.2.3.	Wilcoxon.....	206
6	RESULTATS	207
6.1.	Estudi exploratori descriptiu.....	207
6.1.1.	Estudi estadístic exploratori del doblat <i>a ver / a veure</i> (V3b).....	208
6.1.2.	Estudi estadístic exploratori del doblat <i>bueno / bé</i> (V3c).....	208
6.1.3.	Estudi estadístic exploratori del doblat <i>después / després</i> (V3f).....	209
6.1.4.	Estudi estadístic exploratori del doblat <i>hombre / home</i> (V3h).....	209
6.1.5.	Estudi estadístic exploratori del doblat <i>o sea / o sigui</i> (V3i).....	210
6.1.6.	Estudi estadístic exploratori del doblat <i>pero / però</i> (V3j).....	210
6.1.7.	Estudi estadístic exploratori del doblat <i>pues / doncs</i> (V3k).....	211
6.1.8.	Estudi estadístic exploratori del doblat <i>vale / d'acord</i> (V3l).....	212
6.1.9.	Estudi estadístic exploratori de les fórmules d'obertura de la interacció (V5).....	212
6.1.10.	Estudi estadístic exploratori de les heterorepeticions (V6).....	213
6.1.11.	Estudi estadístic exploratori de la posició dels pronoms clítics en construccions perifràstiques (V14).....	213
6.1.12.	Estudi estadístic exploratori del nombre (V16) i el tipus de preguntes (V17).....	214
6.1.13.	Estudi estadístic exploratori de les construccions condicionals (V18).....	215
6.1.14.	Estudi estadístic exploratori de les perífrasis d'obligació (V22).....	216
6.1.15.	Estudi estadístic exploratori dels determinants demostratius (V23).....	217
6.1.16.	Estudi estadístic exploratori dels pronoms clítics de datiu plural (V25).....	218
6.1.17.	Estudi estadístic exploratori dels fenòmens de contacte de llengües (V38).....	218
6.1.18.	Estudi estadístic exploratori dels colors (V41).....	219
6.1.19.	Estudi estadístic exploratori del nombre d'adverbis acabats en <i>-ment</i> (V45).....	220
6.2.	Anàlisi de Components Principals: Exploració de les relacions globals entre les variables.....	221
6.2.1.	Anàlisi de Components Principals de l'estudi univariant	221
6.2.2.	Anàlisi de Components Principals del doblat <i>a ver / a veure</i> (V3b).....	226
6.2.3.	Anàlisi de Components Principals del doblat <i>bueno / bé</i> (V3c).....	230
6.2.4.	Anàlisi de Components Principals del doblat <i>después / després</i> (V3f).....	234
6.2.5.	Anàlisi de Components Principals del doblat <i>hombre / home</i> (V3h).....	237
6.2.6.	Anàlisi de Components Principals del doblat <i>o sea / o sigui</i> (V3i).....	241
6.2.7.	Anàlisi de Components Principals del doblat <i>pero / però</i> (V3j).....	245

Índex

6.2.8.	Anàlisi de Components Principals del doblet <i>pues / doncs</i> (V3k).....	249
6.2.9.	Anàlisi de Components Principals del doblet <i>vale / d'acord</i> (V3l).....	252
6.2.10.	Anàlisi de Components Principals de les fórmules d'obertura de la interacció (V5) 256	
6.2.11.	Anàlisi de Components Principals de la posició dels pronoms clítics en construccions perifràstiques (V14)	260
6.2.11.1.	Anàlisi de Components Principals de la posició proclítica (V14a)	260
6.2.11.2.	Anàlisi de Components Principals de la posició enclítica (V14b)	265
6.2.11.3.	Anàlisi de Components Principals de la duplicació de pronoms clítics (V14c) 269	
6.2.12.	Anàlisi de Components Principals del tipus de preguntes (V17).....	273
6.2.13.	Anàlisi de Components Principals de les construccions condicionals (V18).....	277
6.2.14.	Anàlisi de Components Principals de les perífrasis d'obligació (V22).....	281
6.2.15.	Anàlisi de Components Principals dels determinants demostratius (V23).....	285
6.2.16.	Anàlisi de Components Principals dels pronoms clítics de datiu plural (V25).....	289
6.2.17.	Anàlisi de Components Principals dels fenòmens de contacte de llengües (V38)	293
6.2.18.	Anàlisi de Components Principals dels colors (V41).....	297
6.3.	Anàlisi de significació estadística de les dades.....	302
6.3.1.	Significació estadística de les variables pragmàtiques.....	303
6.3.2.	Significació estadística de les variables morfosintàctiques.....	305
6.3.3.	Significació estadística de les variables lèxiques	307
6.4.	Arbres de classificació	309
6.5.	Anàlisi estadístiques: síntesi.....	326
7	DISCUSSIÓ DELS RESULTATS	332
7.1.	Discussió dels resultats de les variables pragmàtiques	333
7.1.1.	Doblets de partícules discursives: V3b <i>a ver / a veure</i>	334
7.1.2.	Doblets de partícules discursives: V3c <i>bueno / bé</i>	336
7.1.3.	Doblets de partícules discursives: V3f <i>después / després</i>	339
7.1.4.	Doblets de partícules discursives: V3h <i>hombre / home</i>	341
7.1.5.	Doblets de partícules discursives: V3i <i>o sea / o sigui</i>	343
7.1.6.	Doblets de partícules discursives: V3j <i>pero / però</i>	348
7.1.7.	Doblets de partícules discursives: V3k <i>pues / doncs</i>	352
7.1.8.	Doblets de partícules discursives: V3l <i>vale / d'acord</i>	357
7.1.9.	Obertura de la interacció: V5.....	360
7.1.10.	Heterorepeticions: V6.....	363

Índex

7.1.11.	Discussió dels resultats de les variables pragmàtiques: Síntesi.....	364
7.2.	Discussió dels resultats de les variables morfosintàctiques	365
7.2.1.	Posició dels pronoms clítics en construccions perifràstiques: V14.....	366
7.2.2.	Nombre de preguntes no retòriques: V16	367
7.2.3.	Tipus de preguntes: V17	369
7.2.4.	Construccions condicionals: V18.....	370
7.2.5.	Perífrasis d'obligació: V22.....	373
7.2.6.	Determinants demostratius: V23.....	377
7.2.7.	Pronoms clítics de datiu plural: V25.....	379
7.2.8.	Discussió dels resultats de les variables morfosintàctiques: Síntesi.....	381
7.3.	Discussió dels resultats de les variables lèxiques	382
7.3.1.	Fenòmens de contacte de llengües: V38	382
7.3.2.	Colors: V41.....	387
7.3.3.	Nombre d'adverbis acabats en <i>-ment</i> : V45	391
7.3.4.	Discussió dels resultats de les variables lèxiques: Síntesi	395
7.4.	Discussió dels resultats: síntesi.....	396
8	CONCLUSIONS	398
9	BIBLIOGRAFIA	410
10	ANNEXOS	468
	Annex 1. Informants del corpus d'anàlisi per població, sexe i llengua inicial.	468
	Annex 2. Pautes de transcripció.	469
	Annex 3. Realitzacions de V38 al corpus d'anàlisi.	470
	Annex 4. Estudi estadístic exploratori.	486
	Annex 5. Matriu de correlacions entre les variables de l'estudi univariant.	578
	Annex 6. Matriu de correlacions entre variables per a V23.....	579

«I permeteu-m'ho de dir-ho ben clar: els números, tot i que només s'utilitzen en la recta final d'aquest marc analític i metodològic, expliquen moltes coses, sempre que es sigui un bon lingüista i que s'hagin formulat les hipòtesis d'un determinat estudi d'una forma correcta.» (Turell, ed., 1995: 12).

1 INTRODUCCIÓ

Bona part de la lingüística es dedica a comprendre i explicar el complex procés que seguim, majoritàriament sense ser-ne conscients, des que naixem (i fins i tot abans) i que ens converteix en usuaris experts de les llengües, uns sistemes de codis, normes i excepcions, tan complicats com fascinants, que ens permeten expressar i rebre idees, és a dir, comunicar-nos¹. Les primeres etapes d'aquest procés són les més espectaculars, perquè en un període molt curt de temps (uns mesos, uns pocs anys) incrementem exponencialment la nostra capacitat d'interactuar amb l'entorn i, gradualment, incorporem al repertori de recursos que utilitzem per comunicar-nos elements que els altres humans que ens envolten poden descodificar molt més clarament que els senyals a què estàvem limitats al naixement (com el plor o moviments corporals poc específics), les mirades, els gestos, les vocalitzacions. I, un cop els hem incorporat a la llista de les habilitats que tenim, a base de pràctica (assaig i error), arribem a dominar-los, tant per separat com en combinació, i comencem així, a comunicar de forma efectiva les diverses necessitats que experimentem. Aquestes primeres etapes són, doncs, molt atractives per als científics que es pregunten com adquirim el llenguatge, perquè els resultats de les nombroses fases que superem són perceptibles a ull nu (fet que no exclou que es puguin mesurar i analitzar mitjançant procediments empírics escrupolosos i complexos).

A mesura que creixem i avancem en el procés d'adquisició del llenguatge, però, les incorporacions que fem al nostre repertori comunicatiu es van tornant més subtils. Per exemple, després que ens adonem que un gest o una cadena de sons determinats estan vinculats a certs objectes o fets del món (*naming insight*; du Plessis, 2017; Samuelson i McMurray, 2017) i que experimentem el que s'ha

¹ En aquest treball, ens centrem en individus amb un desenvolupament de les habilitats lingüístiques típic.

conegut com l'explosió del lèxic (aproximadament, entre l'any i mig i els dos anys), en què passem, de forma admirable, de conèixer i emprar unes desenes d'unitats a uns centenars en qüestió de setmanes (cf. Galián Conesa, Ato Lozano i Carranza Carnicero, 2010), seguim aprenent el significat i l'ús convencional de noves unitats lèxiques. De fet, mai no deixem de fer-ho al llarg de la vida, però el ritme al qual les adoptem disminueix amb el temps. Així doncs, les modificacions i els refinaments que fem a la nostra capacitat comunicativa després de la primera infància són també molt importants, en tant que necessàries per assolir el domini de les llengües que aprenem, però, en comparació, són menys evidents i, potser per això, el volum de treballs empírics que s'ha centrat a explicar-los i intentar comprendre'ls és més reduït que la literatura que tracta les etapes inicials del procés d'adquisició lingüística.

No obstant això, cal tenir en compte, a l'hora de valorar el gruix de publicacions disponible sobre les diverses etapes del procés d'adquisició del llenguatge, que el coneixement que genera la literatura especialitzada en aquest tema s'aplica de forma directa a diversos àmbits (de recerca i professionals), com el de trastorns del llenguatge i el de l'ensenyament i aprenentatge d'idiomes. D'aquesta manera, els treballs empírics sobre, entre altres qüestions, què diferencia la nostra capacitat per codificar i descodificar missatges de la que tenen altres espècies, com aconseguim desenvolupar aquesta capacitat a partir dels estímuls que rebem i de quins recursos ens servim per fer-nos sentir i per comprendre els altres en contextos diferents han proliferat des de l'antiguitat (Comajoan, 2003).

Coneixem, doncs, com continua aquest aprenentatge i refinament de les habilitats de comprensió i producció del llenguatge al llarg de la infantesa, l'adolescència i l'adultesa, si bé amb graus diferents de precisió, en funció de les llengües, les edats i els nivells lingüístics en què ens interressi centrar-nos. Per tant, és possible descriure, de manera general i fins a cert punt, quines habilitats lingüístiques s'han desenvolupat i quines encara no a les diferents edats per les quals passem els parlants. Aquest coneixement resulta clau, a més dels àmbits que ja hem mencionat, en contextos relacionats amb l'aplicació de la legalitat en els quals es tracten qüestions lingüístiques.

Aquests contextos poden ser molt diversos, de la mateixa manera que les qüestions lingüístiques que poden resultar-hi rellevants formen un ventall molt ampli i variat, ja que ens servim de la llengua per realitzar un nombre incalculable d'accions (Austin, 1962), moltes de les quals poden incidir en el desenvolupament d'investigacions privades, policials o judicials de possibles activitats delictives. El coneixement científic sobre el llenguatge pot resultar crucial, a tall d'exemple, en entrevistes a persones involucrades en fets delictius dutes a terme per agents policials o altres professionals, ja que pot servir per valorar l'existència de coacció, si una confessió és falsa o si en el transcurs de la interacció es produeixen deficiències comunicatives, entre altres (Coulthard, May, Sousa-Silva, eds., 2020). De la mateixa manera, l'aplicació del coneixement científic disponible sobre el llenguatge i la comunicació també pot tenir un paper destacat en el context de processos judicials en els quals, com a part de les proves aportades per les parts que hi intervenen, hi hagi proves lingüístiques, és a dir, gravacions o produccions escrites que resultin rellevants per a la decisió judicial sobre la possible comissió de delictes com frau, incitació a l'odi o a la violència, injúria, amenaça, assetjament, violència de gènere, etc. (Shuy, 1993; Solan i Tiersma, 2005).

Ara bé, en molts d'aquests casos, (per exemple, quan es tracta de dilucidar si un parlant amb unes característiques concretes pot haver produït certa mostra lingüística) resulta necessari conèixer quins recursos lingüístics específics (segments discursius, unitats lèxiques, construccions sintàctiques, eleccions ortotipogràfiques, etc.) tendeixen a utilitzar els parlants d'una comunitat de parla concreta que comparteixen certs trets (com un rang d'edat determinat). Una de les disciplines que ha generat, de manera més sistemàtica, coneixement empíric que pot contribuir a l'anàlisi de mostres lingüístiques per a aquest tipus de finalitats forenses és la sociolingüística de la variació, ja que el seu objectiu és precisament relacionar els usos lingüístics de la població amb factors interns i externs al llenguatge. Així doncs, juntament amb els treballs sobre el procés d'adquisició lingüística i altres parcel·les de la lingüística (com la pragmàtica i l'anàlisi del discurs, per exemple), els estudis variacionistes constitueixen

recursos valuosos per als lingüistes que es dediquen a aquesta àrea de la lingüística aplicada.

Malgrat tot, cal tenir en compte que, a causa de la naturalesa dels contextos forenses en els quals pot ser necessari dur a terme anàlisis de mostres lingüístiques, l'existència de mancances, limitacions i biaixos en el coneixement científic disponible sobre les qüestions (imprevisibles i completament dependents de les mostres que calgui analitzar) que resultin rellevants per a aquestes anàlisis poden tenir conseqüències destacades no només en el pes que tingui finalment una anàlisi lingüística en una investigació privada, policial o judicial per a la qual hagi estat sol·licitada, sinó també en el resultat mateix d'aquesta investigació (i, per tant, en la vida de les persones involucrades). L'existència de mancances, limitacions i biaixos en la recerca, però, és inevitable, ja que la ciència és, al capdavall, una empresa humana, subjecta, per tant, a imperfeccions.

En termes generals, no és desgavellat afirmar que la major part dels estudis lingüístics aplicables a l'anàlisi de produccions lingüístiques en contextos forenses, així com els més influents globalment, s'ocupen de parlants monolingües en comunitats de parla anglesa (vegeu, per exemple, Nagy i Meyerhoff, 2008: 7-10; Stanford, 2016), malgrat els esforços acadèmics destinats a diversificar la recerca. Per això, cal prudència a l'hora d'avaluar l'aplicabilitat dels treballs existents a mostres en llengües diferents i provinents de comunitats de parla multilingües i, a més, és necessari considerar la possibilitat que les tendències observades a la recerca no siguin universals o rellevants per explicar els usos lingüístics presents a una mostra determinada.

Un altre aspecte relacionat amb les investigacions científiques existents de les quals es nodreixen les anàlisis lingüístiques en contextos forenses i que pot limitar-ne l'abast o la precisió (amb tot el que això pot comportar en el marc d'un procés investigatiu, com s'ha comentat anteriorment) és el fet que la recerca sincrònica sobre els usos lingüístics difícilment pot acompanyar els canvis que es produeixen a les llengües de manera constant. La llengua evoluciona diàriament, a mesura que els parlants la utilitzen per donar resposta a les seves necessitats

comunicatives. El nivell d'anàlisi en què aquests canvis continuats són més visibles és, segurament, el lèxic i ni tan sols els recursos dedicats específicament a donar compte d'aquesta efervescència en el vocabulari de les llengües (diccionaris, reculls terminològics) aconseguen sempre reflectir fidelment els usos lingüístics dels parlants a cada moment. Així doncs, és molt necessari, per a l'àrea de la lingüística aplicada que tractem en aquesta tesi, garantir que la recerca que descriu les varietats lingüístiques utilitzades actualment a diverses comunitats de parla es mantingui actualitzada, ja que, la majoria de produccions lingüístiques que són rellevants en investigacions forenses contenen varietats contemporànies de la llengua. Degut, però, al biaix que ja s'ha comentat, el volum d'estudis sobre varietats actuals de llengües amb menys parlants que les hegemòniques a l'àmbit acadèmic (com l'anglès, el francès o l'espanyol) és també més reduït, de manera que, sovint, la tasca de trobar estudis recents i directament aplicables a una mostra d'una llengua com el català es complica.

Per això, aquesta tesi pretén contribuir a revertir el biaix monolingüe i anglocèntric de gran part de la recerca lingüística i a ampliar el volum d'estudis descriptius actuals sobre les varietats lingüístiques emprades per parlants adolescents que puguin ser aplicables no sols en àrees amb una innegable tradició acadèmica als territoris de parla catalana, com l'estudi dels trastorns del llenguatge, l'ensenyament i aprenentatge de llengües o la política i la planificació lingüístiques, sinó també a camps encara menys explorats en el nostre àmbit sociolingüístic com el de les anàlisis lingüístiques rellevants per a processos forenses, és a dir, sol·licitades en el marc d'investigacions privades, policials o judicials.

Malauradament, milers d'adolescents entren en contacte cada any amb el sistema judicial a Espanya. Segons l'Institut Nacional de Estadística (2021), per exemple, l'any 2020 un total de 11.238 menors d'edat d'entre 14 i 17 anys van ser condemnats per sentència ferma emesa per jutjats i tribunals de l'ordre jurisdiccional penal a Espanya i, segons l'Institut d'Estadística de Catalunya² (2021), la Direcció General d'Execució Penal a la Comunitat i Justícia Juvenil va

² D'ara endavant, Idescat.

atendre 5.367 menors d'edat i joves. En un nombre desconegut de casos com els recollits en aquestes estadístiques i altres per als quals, fins on arriba el nostre coneixement, no disposem de dades i en què es veuen involucrats menors d'edat i, en especial, adolescents (per exemple, aquells en què actuen com a testimonis, víctimes o familiars implicats) l'anàlisi de les seves produccions lingüístiques podria aportar informació útil per a processos investigatius i/o judicials, no només pel que fa a la possible inculpció de delictes a menors d'edat, sinó també per tal de millorar els processos en sí mateixos. Per exemple, tenir més coneixement sobre les seves pràctiques comunicatives podria contribuir a millorar la comunicació entre participants adults i adolescents en les entrevistes sobre els possibles fets delictius en les quals han de participar els implicats (presumes autors, víctimes o testimonis), de vegades, diverses ocasions com a part d'un mateix procés. D'aquesta manera, estudis, com el present, sobre l'actuació lingüística dels parlants dissenyats especialment amb la finalitat de contribuir a les pràctiques investigatives i judicials que es realitzen a través de la llengua constitueixen un pas endavant en l'impuls de la lingüística aplicada a contextos forenses a l'àmbit sociolingüístic català, en el qual no està tan desenvolupada com en altres països (el Regne Unit, Alemanya, Austràlia o els Estats Units d'Amèrica, per exemple).

Per assolir aquest objectiu, estructurarem el treball de la manera següent. En primer lloc, després de la contextualització àmplia que acabem de presentar, al capítol següent, descrivim el marc teòric en el qual cal situar la nostra recerca, que es fonamenta, d'una banda, en l'àrea de la lingüística aplicada que anomenem lingüística forense i, de l'altra, en la sociolingüística i, més concretament, en l'estudi de la variació. A més de descriure i revisar les ocupacions principals i la història de les dues grans disciplines en què es basa el nostre estudi, en el segon capítol presentem també les variables extralingüístiques que més s'han utilitzat en la lingüística de la variació.

En segon lloc, dediquem el tercer capítol a l'exposició de la pregunta de recerca, els objectius generals i específics i les tres grans hipòtesis que ens plantejem en aquesta tesi. Tot i que ja hem tractat el panorama social en el qual cal emmarcar

la nostra investigació, en aquest capítol tindrem ocasió de presentar més específicament els objectius que perseguim en el nostre treball, així com els aspectes concrets del nostre plantejament que es basen en els fonaments teòrics que tractem al segon capítol, dedicat, com hem vist, al marc teòric.

En tercer lloc, al quart capítol presentem el context sociolingüístic de la nostra investigació, és a dir, el coneixement previ amb què comptem sobre la realitat demolingüística de la qual prové el nostre objecte d'estudi. Així mateix, tractem en aquest capítol els treballs previs que s'han realitzat en el marc del projecte de recerca en què s'insereix aquesta tesi i els que versen sobre el català que utilitzen els parlants joves, ja que és aquest sector de la població el que ens interessa aquí.

A continuació, al cinquè capítol, exposem els aspectes i les decisions metodològiques que hem seguit a l'estudi. Presentem, primer, el projecte de recerca en què s'emmarquen els nostres esforços. Tot seguit, el corpus que analitzem en aquesta investigació. Continuem amb el disseny del nostre estudi, que consta de tres grans parts que tractem detingudament, la preparació del corpus, la de les variables independents (l'edat, el sexe i la llengua inicial) i dependents (aproximadament, una vintena de variables pragmàtiques, morfosintàctiques i lèxiques) que utilitzem a l'estudi i l'anàlisi estadística a la qual sotmetem aquests factors que hem identificat i quantificat al corpus. Després d'explicar en què han consistit aquestes tres etapes en què dividim la nostra recerca, exposem les consideracions específiques que es hem seguit per a la codificació de cadascuna de les variables independents i de les realitzacions presents al corpus de les variables dependents. Després d'això, encara en el capítol dedicat a la metodologia, exposem l'adequació als nostres objectius i les peculiaritats de les tècniques d'anàlisi estadística que hem utilitzat i que agrupem en dos grans grups, les que ens serveixen per explorar les relacions entre les variables (l'Anàlisi de Components Principals) i caracteritzar els nostres informants (els arbres de classificació), d'una banda, i les que ens permeten comprovar si les relacions entre les variables que examinem són estadísticament significatives (la prova U de Mann-Whitney, la prova de Kruskal-Wallis i la prova dels rangs amb signe de Wilcoxon).

Al sisè capítol ens ocupem dels resultats quantitius obtinguts mitjançant l'aplicació de les eines metodològiques presentades anteriorment a les dades que hem analitzat. Tractem primer els resultats de l'estudi exploratori descriptiu que realitzem sobre les dades i que ens permet obtenir una primera impressió de l'ús que fan els informants de cadascuna de les variables lingüístiques que analitzem. A continuació, exposem els resultats de l'exploració de cadascuna de les variables duta a terme mitjançant l'Anàlisi de Components Principals, els resultats de les proves de significació estadística i els arbres de classificació obtinguts per cadascuna de les variables independents de l'estudi.

En el capítol previ a les conclusions de la nostra investigació, el setè, aprofundim en el significat de les dades quantitatives i estadístiques obtingudes sobre les relacions identificades entre les variables lingüístiques i extralingüístiques mitjançant les proves de significació estadística. Per fer-ho, lliguem els nostres resultats amb una revisió de la bibliografia especialitzada més rellevant per a la nostra aproximació a cadascuna de les variables dependents.

Les conclusions de l'estudi, per últim, es presenten al vuitè capítol i giren entorn de les contribucions d'aquesta recerca a la descripció de les varietats contemporànies emprades per adolescents de parla catalana i de les implicacions que poden tenir els resultats obtinguts aquí per a la lingüística forense catalana (concretament, per a la perfilació sociolingüística de mostres produïdes per parlants adolescents). A més, també descrivim possibles línies d'investigació que es desprenen de les limitacions d'aquest treball i de les vies de recerca relacionades amb els nostres objectius que, al nostre parer i com demostra la revisió de la literatura que s'ha dut a terme, encara no s'han impulsat prou a l'àmbit acadèmic català.

2 MARC TEÒRIC

En aquest capítol fem un recorregut pels fonaments teòrics i metodològics de l'estudi que es recull en aquesta tesi. El nostre treball se serveix de la metodologia desenvolupada per la sociolingüística variacionista per aconseguir un objectiu emmarcat en la lingüística forense, obtenir informació útil per a l'elaboració de perfils lingüístics de mostres anònimes en català. Així doncs, comencem explicant com va sorgir la branca de la lingüística aplicada que anomenem lingüística forense, com l'entnem en aquesta tesi i les seves àrees més rellevants per als nostres objectius.

Després d'aquesta introducció, tractem la sociolingüística, la seva història, les seves característiques i els enfocaments que comprèn per tal de comprendre la visió dels sistemes lingüístics que fa possible el nostre treball. A continuació, ens centrem en l'estudi de la variació lingüística, aproximació de la qual parteix aquest estudi.

Finalment, abans de descriure els objectius i les hipòtesis de què partim, fem un repàs del que sabem sobre les variables centrals d'aquest treball. D'una banda, les variables independents, l'edat, el sexe i la llengua inicial dels parlants. També expliquem les particularitats que presenta l'anàlisi de la parla dels adolescents catalans. D'altra banda, tractem les variables dependents i els estudis anteriors en què en basem la descripció.

2.1. La lingüística forense

2.1.1. Àmbits d'actuació de la lingüística forense

La lingüística forense s'ha definit en sentit ampli com la interfície entre la llengua i el dret (Turell, 2005; Cicres i Bosch, 2007: 30; Queralt, 2015: 24; Queralt, Marquina i Giménez, 2018: 1560-1561). Tal com subratllen diverses fonts (per exemple, Coulthard i Johnson, 2007: 5), el creixement actual de la lingüística forense (entesa en sentit ampli) és accelerat, la qual cosa es tradueix en un increment de les línies d'investigació que engloba. Existeixen diverses propostes de divisió de les àrees d'estudi de la lingüística forense, com la distinció entre “the linguistic study of legal texts and processes” i “the provision

of linguistic evidence in criminal and civil cases” (MacLeod, 2013: 1), així com llistes més o menys exhaustives d'exemples de contextos en què lingüistes experimentats han estat d'utilitat en investigacions o en processos judicials (cf. Coulthard i Johnson, 2007; Olsson, 2008; Shuy, 2008). Gibbons i Turell (2008) presenten una distinció de tres àmbits generals: l'estudi del llenguatge jurídic i administratiu, el del llenguatge judicial i, finalment, l'anàlisi del llenguatge probatori o evidencial (*the language of the law, the language of the court i language as evidence*, respectivament), de la qual altres autors com Cicres i Gavaldà (2014) es fan ressò i que resumim a continuació.

El primer àmbit, el del llenguatge jurídic, inclou l'estudi de les estructures i expressions lingüístiques usades en textos normatius i legislatius, de la seva comprensió per diversos col·lectius i del seu significat (Queralt, 2015: 24). En aquest camp, per tant, l'objecte d'estudi són, entre altres, lleis, contractes o pòlisses d'assegurances (Cicres i Gavaldà, 2014: 62).

L'estudi del llenguatge judicial es podria correspondre al que Coulthard, Grant i Kredens (2011: 529) es refereixen com “spoken legal practices” i comprèn l'anàlisi del discurs dels diversos agents que participen en les interaccions que tenen lloc als jutjats i entre els agents de l'ordre i altres ciutadans (Queralt, 2015: 24). Totes aquestes interaccions es caracteritzen per factors socials propis (com posicions desiguals de poder o normes preestablertes sobre la successió dels torns de paraula) que afecten la producció lingüística dels interlocutors.

En tercer lloc, l'àmbit del llenguatge probatori conté l'actuació de lingüistes com a experts en processos judicials o durant investigacions policials. A banda de la seva intervenció en litigis de marques registrades (en què cal comparar elements lingüístics aïllats) o en casos en què es posa en dubte l'autenticitat de certes produccions lingüístiques (com cartes de suïcidi o de rescat), s'hi inclouen l'anàlisi de mostres lingüístiques orals (coneguda sovint com a identificació de parlants) i de mostres escrites. La darrera pot tenir diversos objectius, com l'atribució d'autoria (mitjançant la comparació de corpus lingüístics d'autoria coneguda i desconeguda; Queralt, 2014 en resumeix el protocol d'actuació) o la

detecció de plagi (Marquina i Queralt, 2014, estableixen el nivell llindar per determinar l'existència de dependència textual entre mostres en castellà).

A Catalunya, els dos primers àmbits s'han conreat a través d'estudis sobre la lingüística i la traducció jurídiques (p. ex. González Salgado, 2011; Gelpí, 2013; Domènech Bagaria i Gelpí Arroyo, 2017), així com sobre el dret lingüístic (per exemple, Milian i Massana, 2009; Pons Parera, 2011), de manera que també s'ha entès la lingüística forense com una etiqueta referida exclusivament a l'aplicació de la sociolingüística a processos judicials i investigacions policials i privades (àmbit del llenguatge probatori).

És en aquest sentit que ens centrem en aquest treball. Així doncs, tot seguit, tractem els orígens de la lingüística forense (entesa com l'anàlisi de la llengua aplicat a investigacions judicials, policials i privades) i emmarquem les principals aplicacions de la nostra recerca en els seus àmbits d'actuació i, concretament, en l'elaboració de perfils sociolingüístics i l'anàlisi del discurs de menors aplicat a contextos policials i judicials.

2.1.2. Història de la lingüística forense

Si bé és cert que tant l'interès en les característiques pròpies de la llengua utilitzada en contextos legals i judicials com les disputes sobre l'autoria de certs textos (per exemple, literaris o religiosos) conformen àrees de recerca amb una llarga tradició (Olsson, 2008; Coulthard, Grant i Kredens, 2011), la lingüística forense es considera una disciplina jove dins de la lingüística aplicada (Blackwell, 2013: 1).

El primer ús de l'adjectiu *forense* per descriure aquesta mena d'interessos lingüístics s'atribueix a Philbrick, que va utilitzar l'any 1949, sense gaire èxit, el sintagma *forensic English* al títol de la seva obra sobre el llenguatge legal (Coulthard i Johnson, 2007: 5). Els anys 1960 van dur dos dels esdeveniments més famosos de la disciplina. El primer cas va ser el 1963, quan Ernesto Miranda va ser condemnat a Arizona per robatori i va apel·lar perquè, afirmava, no havia entès els seus drets (Olsson, 2008: 5). Aquest i altres casos semblants a Estats Units van resultar en la implementació en aquest país del que avui sovint es

coneix com *the Miranda warning*, un paràgraf en què els agents expliquen els drets dels arrestats. Amb el temps, molts estudis emmarcats en la lingüística forense han estudiat les implicacions legals lligades a les característiques lingüístiques d'aquest paràgraf, ja que, entre altres fets que conviden a la reflexió, sovint s'enuncia de forma ritualitzada, com una formalitat més, la qual cosa dificulta la seva comprensió per part de l'oient (Ainsworth, 2010).

El segon cas d'ús del terme *lingüística forense* va produir-se l'any 1968, quan el lingüista Svartvik va publicar la seva anàlisi de les declaracions, suposadament literals, que Timothy Evans, un conductor de camions de Gal·les, va fer a la policia britànica en relació amb la seva implicació en la mort de la seva dona i filla l'any 1949. Svartvik demostrava la coexistència de dos estils lingüístics clarament diferents als textos atribuïts només a un autor, Evans. Svartvik va fer servir el terme *lingüística forense* per referir-se a la mena d'anàlisi que havia dut a terme i per això diversos autors veuen la seva publicació com el punt de partida de la branca que ens ocupa (Coulthard i Johnson, 2007: 5; Olsson, 2008: 20; Coulthard, Grant i Kredens, 2011: 530).

A partir de llavors, el nombre de casos en què hi havia lingüistes que estudiaven declaracions escrites en contextos judicials va seguir augmentant paulatinament fins que per primer cop l'any 1989, al Regne Unit, es va sol·licitar un acadèmic com a pèrit per tal que testifiqués sobre un d'aquests textos (Olsson, 2008: 5). A l'inici, doncs, no era habitual que els agents policials o el sistema judicial comptessin amb les opinions de lingüistes i era tan sols de forma esporàdica que es publicaven articles científics sobre confessions sospitoses o inconsistències en textos que havien de ser reproduccions literals de la producció oral espontània (Coulthard i Johnson, 2007: 5). Això no obstant, algunes d'aquestes primeres intervencions de lingüistes en casos judicials van tenir un gran ressò, com en els casos de l'anàlisi de les declaracions atribuïdes a Derek Bentley (Coulthard i Johnson, 2007: 173; Olsson, 2008: 5; Blackwell, 2013: 4) o de la confessió atribuïda a William Power, un dels sis individus acusats de posar bombes a dos bars de la ciutat de Birmingham que van resultar ser innocents i van passar a ser coneguts com *the Birmingham Six* (Coulthard i Johnson, 2007: 130).

Amb el temps, els contactes entre els lingüistes forenses es van anar intensificant. Entre finals dels anys 1980 i inicis dels 1990 van tenir lloc a Europa les primeres trobades entre lingüistes forenses de diversos països (Blackwell, 2013: 1). Mentrestant, a Austràlia també tenien lloc un seguit de reunions en què diversos lingüistes estudiaven la necessitat de millorar les relacions entre aborígens i el sistema legal. L'any 1991 es va establir, arran d'aquests seminaris i reunions, la *International Association of Forensic Phonetics* (IAFPA) i tan sols un any després la *International Association of Forensic Linguists* (IAFL) que, a partir del 25 de setembre del 2021, va passar a anomenar-se *International Association for Forensic and Legal Linguistics* (IAFLL). De llavors ençà, les reunions d'experts van anar-se reproduint en forma de seminaris i congressos, començant a Bonn, Alemanya, l'any 1993.

Actualment, la disciplina s'ha consolidat a diversos països i mostra una tendència expansiva, tal com evidencia el nombre creixent d'articles i altres publicacions, com manuals i obres de referència, l'oferta de cursos d'especialització, el desenvolupament continuat de metodologies o l'actuació de lingüistes forenses com a testimonis experts en un ventall cada cop més ampli de casos judicials (Coulthard i Johnson, 2007: 5; Coulthard, Grant i Kredens, 2011: 530; Blackwell, 2013: 5-6; Heydon, 2019). A més, diverses obres de referència inclouen seccions dedicades a descriure possibles futures línies d'investigació o els debats actius més actuals entre els entesos (vegeu, per exemple, la secció "New directions" a Coulthard, May i Sousa-Silva, 2020), símptoma evident de la vitalitat d'aquesta branca de la lingüística aplicada.

2.1.3. La lingüística forense a Catalunya

Com hem vist al subapartat anterior (2.1.2), la lingüística forense existeix des dels anys seixanta del segle passat. Això no obstant, no va ser fins als inicis del segle XXI que va arribar a Catalunya (i, de fet, a tot l'Estat³) de la mà de la doctora Maria Teresa Turell Julià. Sota el seu lideratge, l'any 2003 va néixer el ForensicLab, el primer laboratori universitari de lingüística forense a Espanya, en el si de la Unitat de Variació Lingüística (UVAL) de la Universitat Pompeu Fabra, que havia establert tan sols dos anys abans i que avui dia s'anomena Morfologia i variació lingüística (Morfoval). Juntament amb el laboratori, va crear també el primer programa de màster en lingüística forense del país, pioner en llengua no anglesa, que esdevindria oficial el 2011. A més, va arribar a un conveni amb el Col·legi Oficial de Doctors i Llicenciats en Filosofia i Lletres i en Ciències de Catalunya per tal que reconegués i inclogués al seu registre de pèrits la figura de pèrit judicial lingüista (UPF-BSM, 2008), la qual cosa va suposar un pas fonamental per a la institucionalització d'aquesta figura professional a Catalunya.

Segons Boix-Fuster i Cicres (2013: 16), “la principal tasca que tenia encomanada [el ForensicLab] era la recerca bàsica i aplicada” i mostra d'aquest fet són, en primer lloc, els tres projectes de recerca competitius que s'hi van dur a terme sobre idiolectometria forense, és a dir, la possibilitat de mesurar objectivament la similitud o distància entre mostres de l'estil idiolectal (Turell, 2010: 240; Turell i Gavaldà, 2013) dels parlants: “Idiolectometría aplicada a la lingüística forense” (EXPLORA HUM2007-29140-E), entre 2007 i 2008; “Idiolectometría forense e Índice de Similitud Idiolectal” (FFI2008-2011-03583/FILO), entre 2008 i 2012, i “Hacia la consolidación de un Índice de Similitud/Distancia

³ Aquí, però, ens centrem en l'evolució de la disciplina a Catalunya i deixem de banda el seu desenvolupament a la resta d'Espanya. Els lectors interessats en aquest darrer poden consultar, per exemple, les publicacions agrupades en el número 9 de *Llengua, Societat i Comunicació* (vegeu Fernández Planas, 2011), Garayzábal, Jiménez Bernal, Reigosa Riveiros (2012), Queralt (2015) o Ramírez Salado (2017: 536-537).

Idiolectal (IS/DI) en Idiolectometría Forense” (FFI2012-34601), que s’havia de dur a terme entre 2013 i 2015 (Forensiclab, 2013; IULA, 2013; vegeu **2.1.4.1.1**).

En segon lloc, les contribucions dels membres del grup de recerca a la comunitat internacional de lingüistes forenses. Una mostra de la seva aportació va ser la celebració del Segon Congrés Europeu de la IAFL sobre Lingüística Forense / Llengua i Dret a l’Institut Universitari de Lingüística Aplicada (IULA) el setembre de 2006 (Spasova, Turell i Cicres, 2007). Una altra mostra de les contribucions científiques del grup són les diverses tesis doctorals que van dur a terme, destacables tant pel seu nombre com pel seu caràcter pioner. Concretament, quatre d’aquestes investigacions van contribuir a l’avenç de l’anàlisi i comparació forense de mostres de parla (Cicres i Bosch, 2007; López Escobedo, 2011; Gavaldà Ferré, 2013; Marquina Zarauza, 2016) i dues al de l’anàlisi i comparació forense de mostres de text (Stefanova Spasova, 2009; Queralt, 2015).

En darrer lloc, cal destacar la contribució que van fer al llarg dels anys els membres del grup de recerca a la lingüística forense de casa nostra. Aquesta contribució es pot dividir, a grans trets, en treballs que exposen les aplicacions forenses de la lingüística (Turell, 2006, 2009; Gavaldà Ferré, 2011), els que tracten l’anàlisi forense de mostres orals (Cicres i Bosch i Turell, 2004; Cicres i Bosch, 2004, 2011a, 2011b), els que exploren la detecció lingüística del plagi (Turell, 2004, 2007, 2008, 2011 i 2012) i, per últim, els que tracten l’anàlisi d’autoria (Turell, 2010, 2011; Queralt, Spasova i Turell, 2011; Queralt, 2013, 2014).

Tant a la recerca acadèmica com a la pràctica professional, les aproximacions metodològiques del ForensicLab es van caracteritzar per la integració capdavantera d’enfocaments qualitatiu i quantitatiu i l’ús d’eines informàtiques per a l’obtenció i la interpretació de les dades (Boix-Fuster i Cicres, 2013: 16, cf. Gibbons, 2013). En aquest sentit, la recerca metodològica d’aquest grup va permetre dissenyar, en primer lloc, un mètode basat en seqüències de categories morfosintàctiques (n-grames) per a la comparació forense de textos escrits (Spasova, 2009; Boix-Fuster i Cicres, 2013: 16), que ha fonamentat diversos estudis posteriors (entre altres, Cicres i Queralt, 2019). En segon, un protocol

per a la mesura quantitativa de la similitud o distància entre idiolectes, l'Índex de Similitud/Distància Idiolectal o IS/DI mencionat anteriorment. I, finalment, una proposta metodològica per a la comparació forense de textos escrits en castellà mitjançant l'aplicació de la raó de versemblança (*likelihood ratio*; Queralt, 2014).

La recerca del ForensicLab va estar sempre lligada a la pràctica professional, ja que durant el temps que va existir s'hi van realitzar més de seixanta peritatges lingüístics a Espanya i Amèrica, “sobretot centrats en el plagi, l'autoria de textos escrits, el litigi de marques i la identificació de parlants” (Boix-Fuster i Cicres, 2013: 16; Casesnoves Ferrer i Gavaldà Ferré, 2014: 712) i va oferir serveis d'assessorament lingüístic gratuït “adreçat a persones immigrades sense recursos” (UPF, s. d.). Mostra d'aquest compromís social del ForensicLab és el reconeixement que va rebre el 2009 de la Universitat Pompeu Fabra en forma de Premi a la Transferència de Coneixement del Consell Social en la modalitat de Ciències Socials i Humanitats (Boix-Fuster i Cicres, 2013: 17) i mostra del reconeixement a la feina de la doctora Turell per part de la comunitat científica és la seva elecció com a presidenta de la *International Association for Forensic Linguists* el 2011.

Després de la defunció de la doctora Turell l'any 2013, el doctor Jordi Cicres Bosch en va prendre el relleu a la direcció del ForensicLab. Malauradament, però, l'any 2017 es va produir la clausura oficial tant de la unitat de recerca com del màster en lingüística forense a la Universitat Pompeu Fabra. El llegat d'aquest període inclou, a més de les fites ja mencionades, tres obres de referència per a la disciplina: *Lingüística forense, lengua y derecho. Conceptos, métodos y aplicaciones* (Turell, ed., 2005), *Dimensions of Forensic Linguistics* (Gibbons i Turell, eds., 2008) i *Ens queda la paraula: estudis de lingüística aplicada en honor a M. Teresa Turell* (Casesnoves-Ferrer, Forcadell i Gavaldà Ferré, eds., 2013).

Alguns dels membres del ForensicLab van reunir-se per reprendre les seves tasques pericials i acadèmiques des de l'àmbit privat. D'aquesta manera, el 2017 es va fundar la primera empresa dedicada exclusivament a la lingüística forense

de l'Estat espanyol, el Laboratorio SQ-Lingüistas Forenses. Les seves línies de treball són els serveis de peritatge i assessorament professional, la docència i la investigació. Ha establert convenis de col·laboració amb diverses associacions (com la Asociación No al Acoso Escolar o Metro A Metro, que posa en contacte familiars de persones desaparegudes amb professionals legals i forenses). Quant a la docència, a més de cursos en línia de formació bàsica en lingüística forense, els membres d'aquesta companyia col·laboren amb universitats espanyoles i llatinoamericanes, tant per impartir diverses assignatures de grau i postgrau relacionades amb la lingüística forense com per oferir cursos d'estiu introductoris.

Finalment, pel que fa a la recerca, cal destacar, d'una banda, els esforços que s'han fet des de l'àmbit privat per continuar investigant en diversos aspectes de la lingüística forense a casa nostra, com els projectes de recerca que s'han dut a terme des del Departament I+D+i del Laboratorio SQ-Lingüistas Forenses des de l'any 2018. Alguns dels temes que s'han abordat fins ara són l'aplicació de la tècnica dels n -grames (conjunts d'un nombre n d'elements que apareixen junts) a mostres relacionades amb casos d'assetjament escolar (Cicres i Queralt, 2019; Queralt i Cicres, en premsa), l'impacte de la pandèmia de la COVID-19 en el benestar emocional de la comunitat internacional de lingüistes forenses (Giménez, Elstein i Queralt, 2020) o l'aplicació de l'anàlisi del discurs a mostres relacionades amb casos d'estafes en el marc d'una relació afectiva (Queralt, 2022) o de terrorisme (Queralt i Giménez, en premsa; Giménez i Queralt, 2021). De l'altra, també ha continuat la recerca des de l'àmbit universitari mitjançant diverses tesis doctorals actualment en marxa; dues centrades en l'anàlisi lingüística en relació amb casos de violència de gènere i una en les diferències individuals en produccions orals en anglès com a llengua estrangera de parlants inicials d'holandès.

Així doncs, en resum, malgrat els diversos obstacles que s'ha trobat la consolidació de la lingüística forense en el nostre context sociolingüístic, aquesta parcel·la de la lingüística aplicada s'obre camí i avança a Catalunya des de fa aproximadament una vintena d'anys. Com demostren, d'una banda, la seva

presència cada cop més freqüent a trobades interdisciplinàries i mitjans de comunicació i, de l'altra, el nombre d'investigadors que s'hi dediquen i l'abast de la recerca actual en lingüística forense a casa nostra, es tracta d'una disciplina fèrtil i amb un marcat creixement potencial.

Després d'aquest repàs al desenvolupament de la lingüística forense a Catalunya, seguim amb la contextualització de la nostra investigació, concretament, en relació al projecte del qual prové el nostre corpus d'anàlisi, un dels escassos projectes longitudinals de la sociolingüística educativa catalana.

2.1.4. L'anàlisi forense del discurs dels menors i l'elaboració de perfils

En aquesta tesi ens interessen especialment dues línies d'investigació. La primera se situa a mig camí entre l'àrea del llenguatge judicial i la del llenguatge probatori, mentre que la segona s'emmarca clarament en l'àrea del llenguatge probatori. La primera línia d'investigació en què ens centrem tot seguit analitza el discurs dels menors d'edat que entren en contacte amb el sistema jurídic i/o les forces de l'ordre. La segona consisteix en l'elaboració de perfils sociolingüístics. Aquesta tasca es pot dur a terme tant a partir de mostres orals (per exemple, fragments de converses telefòniques) com escrites (com correus electrònics).

2.1.4.1. L'anàlisi del discurs en àmbits forenses

L'anàlisi del discurs s'ha utilitzat en un ventall cada cop més ampli de contextos per millorar la nostra comprensió sobre com interactuem i construïm les diverses identitats que exercim. Un dels àmbits en què l'anàlisi del discurs s'ha fet servir sovint és el judicial. Dins d'aquest àmbit, s'esdevenen situacions que són especialment rellevants per als objectius d'aquesta tesi, aquelles en què interactuen, d'una banda, ciutadans (i, en particular, menors d'edat) llecs en temes judicials i, de l'altra, juristes o agents de les forces de l'ordre.

Tot i que és una àrea complexa, s'hi poden distingir, en primer lloc, els casos en què el lingüista estudia mostres produïdes fora d'aquests contextos que resulten rellevants per a diversos tipus de processos judicials o investigacions policials i,

en segon lloc, aquells altres en què l'objecte d'anàlisi són les produccions lingüístiques que fan els parlants en un context policial o judicial.

2.1.4.1.1. L'anàlisi del discurs produït fora de l'àmbit forense

Dins del primer grup hi incloem les anàlisis de mostres lingüístiques de la parla originades fora del context forense que tenen com a objectiu extreure'n informació que pugui ser rellevant per a investigacions policials o processos judicials. Aquestes mostres poden ser de naturalesa molt variada, com gravacions, missatges de correu electrònic, missatges de text o enviats a través de plataformes de missatgeria instantània, documents escrits a mà, etc. En aquests casos, les anàlisis poden tenir com a objectiu examinar el significat d'expressions lingüístiques en el context situacional i comunicatiu en què s'han produït, comparar mostres de diversos parlants per avaluar la possibilitat que siguin els autors de la mostra en qüestió, determinar si el discurs d'un parlant pot haver estat manipulat o elaborar un perfil sociolingüístic del possible autor de la mostra. En relació amb menors d'edat, aquests estudis poden ser crucials en casos de possible abús o assetjament, segrest, desaparició, suïcidi o disputa per la seva custòdia, entre altres (per exemple, IAFL, 2021: 74).

Les característiques pròpies de les situacions en què es requereixen aquesta mena d'anàlisis lingüístiques limiten la publicació de bibliografia sobre els detalls dels casos en què participen els lingüistes forenses. Amb tot, disposem d'un nombre considerable i creixent d'informes de casos reals, inclosos alguns casos i anàlisis de textos escrits que s'han convertit en icònics dins de la disciplina. És el cas, per exemple, de la participació del perfilador James Fitzgerald en el cas del terrorista nord-americà conegut com *Unabomber*, que va ser identificat gràcies al seu manifest (vegeu, per exemple, Coulthard i Johnson, 2007: 162-163), o les anàlisis que el lingüista Malcolm Coulthard va fer dels missatges de text enviats des del mòbil de Jenny Nicholl i que van ser crucials en el judici contra la seva parella (per exemple, Grant, 2010).

Ara bé, a la bibliografia de l'àrea les descripcions de casos en què s'han dut a terme anàlisis de produccions lingüístiques són limitades, especialment en el cas de mostres provinents de parlants menors d'edat. Aquesta situació s'explicaria per l'impacte sobre la publicació de dades relacionades amb casos reals que tenen les restriccions legals i la naturalesa sensible dels casos. Tot i això, comptem amb certes publicacions que llisten algunes de les variables lingüístiques que s'han utilitzat en casos reals. Segons Garayzábal, Queralt i Reigosa (2019), a finals dels anys noranta del segle passat es va publicar una revisió de publicacions de lingüística forense i computacional que va arribar a comptabilitzar més d'un miler de variables lingüístiques (Rudman, 1998), però, malgrat això, “no existe un consenso ni una lista estable para abordar el problema” (Garayzábal, Queralt i Reigosa, 2019: 63).

A aquesta dificultat cal afegir-hi el fet que el decidit caràcter pioner de la recerca que ens ocupa dificulta la identificació de variables lingüístiques rellevants per als nostres propòsits en estudis previs en lingüística forense. D'una banda, trobem estudis sobre mostres escrites, tant qualitatives o combinats (per exemple, Wright, 2014) com computacionals, com els que recull Rudman (1998). La identificació de variables en les anàlisis d'autoria qualitativa depèn de les característiques lingüístiques de les mostres de què es disposa en cada cas, de manera que les utilitzades en un estudi rarament són aplicables a mostres diferents. En aquest sentit, destaquen per la seva ambició els projectes de recerca sobre idiolectometria forense (és a dir, la mesura de l'estil idiolectal) que es van desenvolupar al desaparegut Forensiclub de la Universitat Pompeu Fabra entre el 2007 i el 2013.

Entenent per estil idiolectal el conjunt d'eleccions lingüístiques que fa cada parlant d'entre les possibilitats que li ofereix la llengua (Turell, 2010; Forensiclub, 2013), aquests projectes tenien l'objectiu d'“explorar i desenvolupar la possibilitat de mesurar la diferència lingüística entre diversos idiolectes i la distància idiolectal de cada individu, de manera que al final es pugui aconseguir un Índex de Similitud Idiolectal (ISI), una mesura numèrica que compari diverses mostres lingüístiques i calculi la distància lingüística entre elles”

(Forensiclàb, 2013: 2-3). Es van utilitzar tres corpus (en castellà, català i anglès, respectivament) i es van analitzar des de les perspectives fonològica, morfosintàctica i discursiva o pragmàtica, cadascuna de les quals corresponia a un mòdul del projecte. Pel que fa al català, es van tenir en compte les variables morfosintàctiques i discursivopragmàtiques que reproduïm tot seguit. A les fases inicials del nostre estudi, es va considerar la idoneïtat d'aquestes variables per als nostres objectius (vegeu l'apartat 5.3).

Taula 1. Variables per al mòdul morfosintàctic del català del projecte ISI (2013).

Variable	Variants	Exemples
1 Negació amb <i>pas</i>	1. <i>pas</i>	“Jo que no hi és pas aquest problema de...” (VC) - RT
	2. \emptyset	“sembla que no puguin anar aquí o allà” (DC) - RT
2 Realització del pronom del subjecte	1. subjecte	“ jo ho vai descobrir no fa gaire, eh.” (DC) - RT
	2. \emptyset	“ \emptyset No érem aquí, érem a Vila-seca.” (VC) - RT
3 Concordança nom-verb col·lectius	1. singular	“ No hi ha massa problemes de convivència” (VC) - RT
	2. plural	“però veig que vénen gent de molts de pobles d'aquí al voltant eh” (DC) - RT
4 Caiguda del clític	1. \emptyset	[“Jo penso que no hi ha massa problemes de convivència entre lo que és comunitat catalana-castellana” (VC) - RT
	2. clític	[“... al poble pràcticament hi vaig a dormir.” (VC)] - RT
5 Dislocació a la dreta	1. Accent	“la pobre gent no m'han fet mai re, eh al contrari vull dir si ells s'hi porten bé ” (MG)
	2. Dislocació a la dreta	“a mi me sembla que sí que n'hi han dos de comunitats ” (MG)
6 Dislocació	1. Dislocació a l'esquerra	[“bandera, home, sí que la tenen, no?” (VC) - RT
	2. Dislocació a la dreta	“Últimament n'hem fet moltes de coses eh.” (DC)] - RT
7 Morfologia de clítics	1. <i>me, te, se</i>	“ me sembla a mi” (DC)
	2. <i>em, et, es</i>	“Aquestes, no, em sembla que no” (VC)
8 Relatiu de lloc	1. <i>que</i>	“vam anar l'altre dia aquella botiga que tenen animals” (DC)
	2. <i>on</i>	“cap al Sector Nord en concret que és allà on es feia aquesta festa...” (VC)

9 Article masculí	1. <i>lo, lu</i>	“ni que ho reguis no és lo mateix” (VC).
	2. <i>el</i>	“ el nostre fill per exemple va estar a la Canonja” (VC)
10 Caiguda de preposició	1. \emptyset	“Però fins i tot jo veig companys del meu fill \emptyset que [qui] ell es relaciona” (VC)
	2. preposició	Però fins i tot jo veig companys del meu fill amb qui ell es relaciona
11 Alternança d'adverbi	1. <i>en aquí, en allà</i>	“I durant el matí fins a les tres i llavors en allà en veiem” (VC)
	2. <i>aquí, allà</i>	“ara canvia , ara es posarà allà, es veu que ho farà més gran” (DC)
12 Tria de preposició	1. <i>amb</i>	“Sí i el vai trucar corrents amb ell” (DC)
	2. <i>a, en</i>	“Naltros tenim gent de la que veus en fotos d'aquestes” (VC)
13 “Dequeisme”	1. <i>de que</i>	“contactes que he tingut amb altra gent pos parlen de que hi ha molta gent” (VC)
	2. <i>que</i>	“Sí doncs per allà, i diu que ell és especialista” (DC)
14 Alternança de funció de pronom	1. Datiu	“que jo em sembla de que” (VC)
	2. Subjectiu	“a mi m'és igual” (DC)
15 Nosaltres-naltros	1. <i>naltros</i>	“i de vegades naltros nem a casa seva” (MG)
	2. <i>nosaltres</i>	“la gent quan dius nosaltres baixem cap allà i ells pugem amunt” (E4)
16 Venir-vindre	1. <i>venir</i>	“es veu que ja es va cansar de venir o no sí què” (MG)
	2. <i>vindre</i>	“i bueno vai vindre un any aquí amb les amigues” (MG)
17 Voler-volguer	1. <i>voler</i>	“home voler voler no crec que volgués eh?” (RG)
	2. <i>volguer</i>	“jo em penso que es deuen volguer integrar” (MG)

Font: Forensiclab (2013: 25-26).

Taula 2. Variables per al mòdul pragmàticodiscursiu del català del projecte ISI (2013).

Variable	Variants	Exemples
1 Resposta a pregunta oberta amb marcador discursiu	1. Presència de marcador discursiu	E- Del 89 fins ara? — A- A veure , les, una entitat històrica com el Casino
	2. Una altra estratègia i/o \emptyset	E- I els Paral·lels? — B- No no
2 Resposta a pregunta oberta amb repetició de la pregunta	1. Repetició	E- O sigui que hi ha hagut una única festa al poble — A- Hi ha hagut una única festa al poble... passa que
	2. Una altra estratègia i/o \emptyset	E- La reforma? — B- \emptyset El que no es vegi

3 Resposta a pregunta oberta amb exclamació	1. Exclamació	E- I la Maria què diu? — A- Ui! La Maria
	2. Una altra estratègia i/o o ø	E- sembla que el poble ha anat millor? — B- ø amb aspect de fet doncs
4 Posició del marcador discursiu <i>pues</i>	1. Inicial en resposta a pregunta oberta	E- la teva filla on va estudiar? — A- pues va anar a Barcelona
	2. Una altra: intermèdia o final	E- Tu d'on ets? — B- de Tortosa
5 Ús de <i>eh?</i> i <i>no?</i>	1. <i>eh?</i>	A- de fet perquè sóc bastant perucot, eh? , però vull dir no se m'hauria ocorregut mai en general
	2. <i>no?</i>	B- de fet pràcticament això ja començava a bullir no? però jo recordo les coses més típiques
6 Posició discursiva <i>eh?</i>	1. Final de torn o absoluta (funció apel·lativa, espera resposta)	A- Què vols dir amb això, eh?
	2. Intermèdia (final d'acte) (funció fàtica, no espera resposta)	B- jo aquest poble no l'havia sentit ni anomenar eh? , per començar i bueno vai vindre un any
7 Posició discursiva <i>no?</i>	1. Final de torn o absoluta (funció apel·lativa, espera resposta)	A- Vindreu demà, no?
	2. Intermèdia (final d'acte) (funció fàtica, no espera resposta)	B- jo penso en l'altre extrem, no? saps i doncs jo no sé si les opinions que jo te doni són massa d'allò
8 Intensificació morfològica vs. lèxica de l'enunciat	1. Ús de prefixos i quantificadors	A- la meva filla està súper feliç en aquest nou col·legi.
	2. Ús de la repetició	B- ja està bé, ja està bé.
9 Intensificació del subjecte de primera persona singular	1. 1a persona singular <i>jo</i>	A- Jo era sempre la que ho feia tot.
	2. ø	B- Cada any ø vinc a les festes
10 Atenuació del subjecte per impersonalització del JO	1. <i>jo</i> + ø	A- Jo sempre li dic a la nena que no faci cas.
	2. <i>Es, un/una, 2a p. sing</i>	B- Perquè clar, tu vens i t'agrada, i no saps què dir
11 Dislocació a la	1. Accent	A- jo n'he sentit a parlar del tema perquè ho havien venut o no sé què

dreta	2. Dislocació a la dreta	B- hi ha pobles que m'atrau algo
12 Dequeisme	1. <i>de que</i>	A- també influeix de que vinguin o no ells
	2. <i>ø que</i>	B- Llavons és una manera que molta gent agafi l'autobús
13 Canvi funció sintàctica	1. Datiu	A- Me sembla que vindrà molta gent aquest any.
	2. Subjecte	B- Penso que la veïna era de Portobou.
14 Estil comunicatiu	1. Control monitorització	A- No sé, eh! És que saps? cada vegada que et trobes un
	2. Confirmació	B- Els nens ara van més contents als, oi que sí?

Font: Forensiclab (2013: 29-30).

Quant als estudis estilomètrics que tracta Rudman (1998, 2012), tot i fa una revisió exhaustiva de centenars de treballs, el gran obstacle per a la seva aplicació en el nostre estudi és que la seva llista de variables utilitzades comprèn trets lingüístics identificats a mostres escrites i en llengua anglesa, mentre que el nostre treball es proposa aïllar variables lingüístiques presents al català parlat. Quinze anys després de la seva publicació de finals dels noranta, Joseph Rudman va visitar el seu article l'any 2012. En tots dos articles assenyala clarament moltes de les raons de fons crucials per entendre per què avui dia, més de 55 anys després de la primera aplicació de mètodes computacionals a problemes d'autoria, encara no hi ha acord entre els acadèmics que s'hi dediquen sobre temes fonamentals com quins mètodes són acceptables⁴. Rudman (2012) resumeix així la proliferació d'estudis estilomètrics amb l'augment corresponent de variables utilitzades:

⁴ Una d'aquestes raons és que l'èmfasi massa sovint s'ha posat a dissenyar metodologies noves i no a validar les ja existents, fet que porta Rudman (2012: 269) a plantejar-se: "What is disturbing about this is that most practitioners are not yet using established techniques and methodologies to solve real world problems. They are not verifying, not replicating studies. With all of the innovating and improving, will we ever reach a valid and accepted method?"

“In 1997, I reported an inventory of over one thousand style markers. With Suguru Ishizaki and David Kaufer’s Docuscope inventory of over forty-one million language strings, and other additions, the total number of style markers has become moot—approaching infinity, which Richard Forsyth and others claim. There is no one set of style markers that is correct for every study. If authors have a verifiably unique style, it stands to reason that in most cases this will be evidenced by a set of different style markers.” (Rudman, 2012: 267-8).

Així doncs, seria impossible tenir en compte totes les variables que s’han utilitzat en estudis estilomètrics i, a causa de les discrepàncies entre autors i les mancances de les descripcions metodològiques que apunta Rudman (1998, 2012), un triatge de les més habituals tampoc solucionaria els problemes fonamentals de replicabilitat i capacitat d’explicar perquè unes variables i no unes altres han aportat resultats aparentment fiables en estudis previs.

D’altra banda, trobem els treballs que tracten mostres orals. A més de la llengua d’anàlisi, que, de nou, tot sovint és l’anglès, aquestes publicacions solen basar-se en variables fonètiques (com Nolan i Grigoras, 2005; però cf. Kredens i Góralewska-Łach, 1998), que queden fora del nostre abast.

Malgrat tot, com a mostra de les variables lingüístiques que s’han utilitzat en anàlisis d’autoria i perfilació sociolingüística en altres treballs, recollim i comentem breument a continuació, d’una banda, les que es van utilitzar a l’estudi pilot de Queralt (2015) sobre l’ús de distribucions poblacionals i relacions de verosimilitud en tasques de comparació forense de textos escrits i, de l’altra, les que s’enumeren al manual de Garayzábal, Queralt i Reigosa (2019: 64), ja que *a priori* les dues llistes de variables lingüístiques podrien tenir-se en compte en el nostre estudi. (Les variables que es van tenir en compte a les etapes inicials es tracten al subapartat **5.3.2.3**).

Taula 3. Variables lingüístiques utilitzades a Queralt (2015) i en anàlisi d'autoria segons Garayzábal, Queralt i Reigosa (2019: 64).

Nivell lingüístic	Variables replicables	Comentaris
Variables de complexitat	Longitud de paraula Longitud de frase o longitud mitjana de paraules per oració Nombre de paraules diferents Nombre d'oracions Índex de variació lèxica (<i>type-token ratio</i>) Nombre de paràgrafs Longitud mitjana de paraules per paràgraf	Sembla que “se ven influidas significativamente por la longitud del texto” (Garayzábal, Queralt i Reigosa, 2019: 64).
Variables lèxiques	Freqüència relativa de paraules funcionals Hàpax legòmena i hàpax dislegòmena Errors gramaticals Errors ortotipogràfics Temps verbals de futur Perífrasi verbal <i>ir a</i> + INF Perífrasis verbals d'obligació Expressió de la condició (<i>si / como</i>) Expressió del pretèrit imperfecte del subjuntiu (terminació <i>-ra</i> o <i>-se</i>) Abreujaments (com euros / € / EUR) Nombre de paraules malsonants	Cal recordar que el vocabulari que s'utilitza en un text, a més de respondre a l'estil idiolectal dels autors, està fortament determinat pel tema que s'hi tracta.
Variables sintàctiques	Construccions sintàctiques Seqüències de categories gramaticals (<i>part of speech n-grams</i>) Tipus d'oracions (simples / complexes) Tipus d'oracions complexes (subordinades / coordinades / juxtaposades) Tipus d'oracions subordinades Tipus d'oracions coordinades Tipus d'oracions juxtaposades	“[L]as variables morfosintàcticas resultan más estables que las léxicas y que tienden a caracterizar el estilo del autor de un texto con un grado de fiabilidad mayor” (Garayzábal, Queralt i Reigosa, 2019: 64).

	Tipus de sintagmes i elements constituents	
Variables pragmàtiques	Intensificació del subjecte 1a persona del singular Expressió de l'èmfasi (majúscules / puntuació / repetició) Tractament (<i>tu / vostè</i>) Fórmules de salutació Fórmules de comiat Marcadors discursius Actes de parla Puntuació Nombre de preguntes Nombre d'exclamacions	“Los resultados de dichos estudios destacan el alto poder discriminante gracias a la baja variación intraautor y la alta variación interautor de estos marcadores.” (Garayzábal, Queralt i Reigosa, 2019: 64).

Font: elaboració pròpia a partir de Queralt (2015) i Garayzábal, Queralt i Reigosa (2019: 64).

Finalment, cal destacar la rellevància per als nostres objectius dels estudis d'Illamola i Vila (2015) i Illamola i Queralt (2020), en què s'explora l'ús de pronoms enclítics o proclítics en perífrasis verbals en castellà, utilitzant com a mostra part del corpus compilat al projecte RESOL (vegeu **punt D** al subapartat **5.3.2.3.2**).

2.1.4.1.2. L'anàlisi del discurs produït dins de l'àmbit forense

Dins del segon grup de casos exposats al subapartat **2.1.4.1**, hi trobem l'anàlisi, a partir de gravacions, transcripcions o resums, d'interaccions orals entre ciutadans llecs en temes judicials i representants de les institucions que ens ocupen, com agents policials, advocats, fiscals, jutges o jurats. Aquestes interaccions es poden originar per diversos motius que poden condicionar-ne el contingut i la forma, ja que pot ser que el ciutadà llec hagi estat implicat en activitats delictives com a persona d'interès per a la investigació, víctima, testimoni o autor del delictes. L'anàlisi lingüística d'aquests intercanvis pot també tenir objectius diversos.

Entre altres, el lingüista pot avaluar l'existència d'inconsistències en les intervencions, detectar intervencions coercitives per part dels professionals o contribuir a l'ús d'expressions lingüístiques adequades al context comunicatiu (per exemple, tenint en compte la capacitat comunicativa, el nivell educatiu,

l'origen social o ètnic del parlant, etc.). En casos en què el ciutadà implicat és menor d'edat, aquestes anàlisis lingüístiques han de tenir en compte, entre altres factors, les diferències en el coneixement del món i en el nivell de desenvolupament cognitiu entre els interlocutors.

Un dels contextos comunicatius que ha suscitat l'interès d'acadèmics de diverses disciplines (psicòlegs, educadors socials, lingüistes, etc.) són les entrevistes policials a menors. Aquestes interaccions es regeixen sovint per protocols preestablerts elaborats per experts de diversos àmbits com agents de l'ordre, psicòlegs o treballadors socials. A més, es diferencien d'altres converses entre adults i menors, entre altres factors, pels complexos objectius dels entrevistadors, que han de deixar de banda la manera en què normalment es dirigeixen als infants en contextos diferents. Mentre que en situacions quotidianes els adults poden interrompre'ls, fer-los preguntes concretes, donar-los opcions i guiar-los en la forma d'exposar els esdeveniments que descriuen, en contextos forenses aquest estil és inadequat perquè augmenta “the risk that children will speculate, misinterpret questions, and answer falsely because of memory intrusions, source-monitoring errors, and confabulations” (Poole, 2016: 21).

Recentment, s'han investigat les pràctiques discursives dels menors d'edat interrogats com a testimonis o víctimes durant investigacions policials (Fogarty, 2010; Fogarty, Augoustinos i Kettler, 2013; MacLeod, 2016; Jol i Stommel, 2021). Una de les troballes recurrents d'aquests treballs és que els menors d'edat tracten l'agentivitat de les accions que descriuen i la relacionen amb la culpabilitat, de forma que els objectius dels investigats i dels investigadors no coincideixen: “Consequently, children themselves do not treat the interview as simply retrieving uncontaminated information but as a social event in which issues of blame are at stake” (Jol i Stommel, 2021: 300).

Sembla que existeix cert acord entre els experts en aquest àmbit sobre el fet que els investigadors han d'adaptar-se als entrevistats segons la seva edat (entre altres factors; vegeu Goodman i Melinder, 2007: 9-10, sobre les diferències individuals i la suggestibilitat), de manera que haurien de tractar de forma

diferent infants joves i adolescents, però, segons Poole (2016: 5), “what this entails is still largely based on common sense rather than systematic research”. Així doncs, una de les necessitats que pot cobrir la recerca en lingüística i que es destaca més sovint a la bibliografia especialitzada en aquest tema és la d’ajudar els entrevistadors a entendre les característiques i les limitacions pròpies de les diverses etapes del desenvolupament lingüístic i comunicatiu dels menors amb què interaccionen.

2.1.4.2. Els perfils sociolingüístics

L’elaboració d’un perfil sociolingüístic és la tasca que consisteix a identificar informació sobre els antecedents⁵ de l’autor d’un text⁶ anònim a partir de la llengua del text (Nini, 2014: 13). Així doncs, per a l’elaboració d’un perfil sociolingüístic, l’expert ha d’analitzar els textos d’autoria desconeguda que li són entregats i identificar-hi trets lingüístics que puguin proporcionar informació sobre diverses variables sociolingüístiques de l’autor.

Aquesta tasca és possible gràcies a la disponibilitat d’estudis sociolingüístics i, molt concretament, de treballs variacionistes com els que tracta al subapartat **2.2.4**, que han demostrat reiteradament el lligam existent entre els usos de mecanismes lingüístics concrets i les característiques socials dels individus. Per tant, mitjançant l’elaboració d’un perfil sociolingüístic, poden vincular-se diversos recursos lingüístics identificats en el text a trets del parlant com el seu nivell d’instrucció, la seva possible relació amb una professió o àmbit laboral específics o el seu origen geogràfic o ètnic (vegeu, per exemple, Queralt, 2022).

Això fa que conèixer el màxim possible d’aquestes variables a partir de l’ús de la llengua que es fa en textos vinculats a investigacions delictives sigui potencialment molt útil per als agents policials. Sobretot, pot servir per reduir el grup de possibles sospitosos (Coulthard i Johnson, 2007: 148; Coulthard, Grant i

⁵ Tot i que hem optat per la paraula *antecedents*, Nini (2014) utilitza *background*, un terme possiblement més ampli i amb unes connotacions diferents en anglès.

⁶ En aquesta definició, entenem el terme *text* en sentit ampli com una producció lingüística, independentment de la modalitat utilitzada.

Kredens, 2011: 536), més que no pas per identificar definitivament l'autor de la mostra en qüestió (Shuy, 2008: 686).

Per a l'elaboració de perfils lingüístics de mostres anònimes, els lingüistes poden recórrer, a més dels estudis variacionistes disponibles i del seu coneixement com a parlants experts i amb consciència lingüística, a corpus de referència per als quals disposin de la informació necessària sobre les característiques dels parlants i que els serveixin com a punt de partida per a la identificació dels trets lingüístics marcats de la seva mostra. Malauradament, els recursos d'aquesta mena que es troben a l'abast dels lingüistes forenses sovint són molt limitats. És per això que és imprescindible que puguin comptar amb un ampli ventall d'estudis sociolingüístics que descriguin els usos de la llengua que fan els parlants que pertanyen a diverses categories socials quan es comuniquen en contextos diversos.

Quant a les variables lingüístiques que s'han utilitzat per elaborar perfils sociolingüístics, generalment ens trobem amb les dificultats exposades anteriorment sobre les variables que s'han considerat a estudis sobre atribució d'autoria (vegeu 2.1.3.1.1). Tot i això, recentment s'han dut a terme dos estudis que investiguen la utilitat de les seqüències de categories gramaticals o n-grames (concretament, bi- i trigrames, és a dir, conjunts de dos i tres elements que apareixen seguits) per classificar mostres escrites en català produïdes per menors per edat (alumnes de primer, segon i tercer cicle de primària) i sexe (Cicres i Queralt, 2019 i Queralt i Cicres, en premsa, respectivament). Els resultats d'aquests estudis mostren una eficàcia d'aproximadament un 70 % en la classificació, la qual cosa demostra que aquesta tècnica d'anàlisi pot resultar d'utilitat per determinar l'edat i el sexe més probable de mostres anònimes escrites per menors en català.

2.1.5. La lingüística forense: síntesi

En aquest apartat hem definit la lingüística forense, n'hem resumit la història i hem descrit els seus àmbits d'actuació principals. Tot seguit, hem fet un repàs a la seva arribada i evolució a casa nostra i ens hem centrat en dues de les tasques que desenvolupen aquests lingüistes i per a les quals poden ser especialment rellevants les troballes del nostre estudi: l'anàlisi del discurs de menors d'edat en contextos forenses i l'elaboració de perfils sociolingüístics. Al llarg de l'apartat, hem assenyalat les mancances en la bibliografia quant a aspectes metodològics i la descripció de variables lingüístiques que s'han utilitzat en estudis previs i, en la mesura del possible, hem assenyalat les variables ja utilitzades que resulten rellevants per a aquest treball.

2.2. La sociolingüística

La relació entre els sistemes lingüístics i les societats en què sorgeixen i s'utilitzen s'ha explorat mitjançant moltes tècniques i des de moltes òptiques diverses que, generalment, s'han encabir sota el gran paraigua de la sociolingüística. A continuació, ens ocupem d'aquesta rica i àmplia empresa científica en què se sustenta la nostra recerca. Primer, en resumim la història i descrivim la seva naturalesa interdisciplinària. Després, la diferenciem d'altres perspectives que serveixen per estudiar la llengua i la facultat humana del llenguatge. Finalment, ens aturem en les diverses subdisciplines que la formen, amb un focus especial sobre la sociolingüística de la variació, mare dels constructes teòricometodològics que formen la base d'aquest estudi.

2.2.1. Història de la sociolingüística

2.2.1.1. El sorgiment de la sociolingüística

Com veurem més endavant, en aquest estudi prenem una perspectiva àmplia del concepte *sociolingüística*. Ara bé, des dels seus inicis, acadèmics d'àmbits d'especialització ben diversos han intercanviat opinions sobre la història, els límits i la definició d'aquest camp interdisciplinari.

Segons diverses fonts (per exemple, la publicació anònima *The Origin of 'Sociolinguistics'* i Paulston i Tucker, 2003: 1) i malgrat d'altres que assenyalen com al creador del terme a Currie (Chambers, 2003: 15; Koerner, 1991: 65) o que afirmen que el primer lingüista a emprar-lo va ser Eugene Nida l'any 1949 (Paulston i Tucker, 2003: 1), el primer ús del terme *sociolingüística* sembla que correspon a la publicació de Thomas Hodson (1939) titulada "Sociolinguistics in India". Més enllà del fet anecdòtic i, en certa manera, l'"honor" d'esdevenir la primera persona a utilitzar un substantiu concret per designar un corrent acadèmic, com subratlla Koerner (2002: 274), aquest fet es podria considerar com la marca que assenyalava "the point at which professional identification of a particular enterprise is regarded as desirable by at least some of its practitioners". A més a més, no es pot associar l'ús d'una denominació concreta a l'inici d'una perspectiva social de la llengua, ja que, com subratlla Neubert (1976: 152, citat a Figueroa, 1994: 1):

"[...] practical questions of how language is an integral part of society and, above all, how social or rather extralinguistic factors have shaped and are shaping language... were in fact the concern of linguistic research long before the term sociolinguistics was coined, so that linguistics has always had a social aspect."

Figueroa (1994: 2), en la mateixa línia, afegeix que la història d'aquesta empresa no comença amb el primer ús del terme que la designa i que una sola persona o punt de vista no poden definir la sociolingüística.

Ara bé, molts dels manuals sobre sociolingüística que tracten els orígens d'"estudis pròpiament sociolingüístics", els situen, com Boix i Vila (1998: 12), als anys cinquanta del segle passat, en què destaquen Uriel Weinreich amb

Languages in Contact de 1953, Marcel Cohen amb *Pour une sociologie du langage* de 1956 o Charles A. Ferguson amb *Diglossia* el 1959. Pel que fa als Estats Units, diversos autors nord-americans vinculen el naixement d'aquest camp a les conferències sobre dialectes socials de Bloomington (Shuy, 1965) i sobre sociolingüística de Califòrnia (Bright, 1966) que van tenir lloc l'any 1964 (cf. Shuy, 2003; Koerner, 2002: 258).

El fet és que aquest corrent acadèmic tan bon punt es designava amb aquest nou terme com amb el de *sociologia de la llengua*, considerats sinònims en bona mesura intercanviables durant dècades (Paulston i Tucker, 2003: 1). Els seus antecedents mostren aquest origen mixt. D'una banda, una de les disciplines que sovint es consideren influències directes de la sociolingüística moderna és l'anomenada "dialectologia tradicional" (Tagliamonte, 2012: 1; Chambers, 2003; Figueroa, 1994; Ball, 2010), és a dir, "el estudio sistemático de los dialectos regionales, concretamente de su descripción" (Hernández Campoy, 1993: 151). Les altres arrels de la sociolingüística moderna serien la lingüística històrica i l'estudi de les llengües en contacte (Tagliamonte, 2012: 1).

Dins de la dialectologia tradicional destaquen, per exemple i a banda d'un gran nombre de dialectòlegs europeus en què també s'hi inclouen Ferdinand de Saussure o Antoine Meillet, les idees del nord-americà William Dwight Whitney⁷, el qual, tot i no ser "a profound student of dialects, yet he did share the early dialectologists' main insight: he saw linguistic diversity as something that is always present within a speech community" (Alter, 2005: 260) i també, tot i que adoptant un angle més social, els estudis de Louis Gauchat, qui "broke the conventions of traditional dialectology by correlating linguistic variability with the sex and age of the individual informants, prefiguring at least sociolinguistic network studies" (Chambers, 2003: 15-6).

⁷ Per a un "arbre genealògic" de William D. Whitney (1827-1894) a William D. Labov (1927-), vegeu Koerner (2002: 263-4).

Quant als primers sociolingüistes americans, no eren tan sols lingüistes, sinó que estaven relacionats, més o menys directament, amb camps com l'antropologia, la sociologia, la psicologia social i les ciències polítiques (Paulston i Tucker, 2003: 1-2). Concretament, els “pares fundadors” que llista Spolsky (2011: 3) provenen d'àrees tan variades com la química (William D. Labov i John J. Gumperz), la sociologia i l'antropologia (Basil Bernstein, Allen D. Grimshaw i Dell Hymes), la filosofia (Charles A. Ferguson) o la psicologia social (Joshua A. Fishman i Susan M. Ervin-Tripp), a més de la lingüística o la filologia (els anteriors i també William Bright, Einar Haugen i Uriel Weinreich).

Com avançàvem, cap a la meitat del segle passat es va crear l'ambient idoni per tal que aquests “small number of key individuals”, com els anomenen Paulston i Tucker (2003: 2), compartissin diàlegs que esdevindrien clau per al futur de la sociolingüística. Alguns d'aquests espais d'interacció que van fer créixer l'interès en les possibilitats de col·laboració entre disciplines diferents inclouen la reunió de l'*American Anthropological Association* de 1963, les actes de la qual es van publicar l'any següent (Gumperz i Hymes, 1964), la conferència sobre sociolingüística mencionada més amunt i l'escola d'estiu de la *Linguistic Society of America* (LSA) a Indiana, on Gumperz i Ferguson van impartir els cursos anomenats “Language and Society” i “Sociolinguistics” (Shuy, 2003: 12).

El clima social de l'època també va tenir un paper rellevant per al desenvolupament de l'àrea acadèmica que ens ocupa. En paraules de Paulston i Tucker (2003: 2): “[...] the field appears to have emerged partially in response to a number of well-articulated and compelling social issues”. Per exemple, per a Tucker (1997: 320, segons Spolsky, 2011: 11), dues de les característiques del període que van contribuir al creixement de la sociolingüística van ser “a sense of the social justice aspect, and a view of its relevance to politics and government”.

Des d'una perspectiva històrica més àmplia, Boix i Vila (1998: 12) proporcionen un resum dels diversos canvis socials que van tenir lloc durant els segles XIX i XX i que “expliquen l'afermament de l'estudi dels vincles entre les societats i les varietats lingüístiques que parlen”:

“Al llarg dels segles XIX i XX, la humanitat ha presenciada unes transformacions sense precedents. El creixement demogràfic, la industrialització, la urbanització, les colonitzacions i les migracions, la reorganització política, primer en imperis colonials, després en estats nació i encara després en blocs polítics i comercials, les creixents facilitats de comunicació i transport, han trasbalsat profundament la realitat social, política, econòmica, cultural i, per tant, lingüística de tot el món.”

A més, a diferència d'altres formes d'estudiar les llengües, la sociolingüística es va impregnar ben aviat de preocupacions que anaven més enllà de la mera descripció lingüística i que estaven relacionades amb les jerarquies i les dinàmiques socials. Aquestes preocupacions expliquen, entre altres, les influències marxistes, en forma de teoria crítica i interès pels conceptes de poder i conflicte, que relata Calvet (2003), així com també l'atenció en objectes d'estudi inusuals fins aleshores com el de Putnam i O'Hern (1955), centrat en la relació entre l'estatus social i la parla d'una comunitat d'afroamericans de Washington DC⁸. Shuy (2003: 8) resumeix la relació entre el moment històric i tres de les disciplines que donarien forma a la sociolingüística (sociologia, antropologia i lingüística) de la següent manera:

“In American society, it was the time of increased problems with racial segregation, poverty education and social structures. The problems were clear enough and these three disciplines had some of the tools needed to address them, but not apart from each other.”

A més, els avenços teòrics i tecnològics de l'època, inimaginables poc abans, constitueixen tant per a Shuy (2003: 5) com per a Labov (2006: 12) un tercer factor que va possibilitar el ràpid desenvolupament d'aquestes noves aproximacions a la lingüística. Aquests avenços, a més d'una predisposició teòrica favorable als descobriments de relacions directes entre factors socials i fets lingüístics, van suposar també una revolució metodològica en la recollida i

⁸ De fet, l'interès per les varietats parlades per afroamericans seria creixent en la sociolingüística incipient (p. ex. Labov, 1973; Thomas, 2007) i aquestes rebrien diversos noms, com *Black English*, *Ebonics* o *African American Vernacular English* (Mufwene, Rickford, Bailey i Baugh, 1998; Holmes, 2001: 351).

l'anàlisi de les dades a partir de l'ús generalitzat de gravadores, espectrogrames i ordinadors. De sobte, el volum de dades que es podia gestionar s'havia incrementat exponencialment i el seu processament i mostreig era comparativament fàcil. En paraules de Labov (2006: 12):

“We can now return to this area of work with more adequate equipment than Meillet could have brought to bear upon such difficult problems. Not only do we have a more explicit theory of phonological structure, but we also possess such useful tools as tape recording, spectrograms and methods of sampling and handling large quantities of data.”

2.2.1.2. La interdisciplinarietat de la sociolingüística

Com a camp inherentment interdisciplinari, els estudis sociolingüístics estan marcats per una sèrie d'avantatges i inconvenients resultants de l'intercanvi de conceptes i metodologies originats en diversos corrents de pensament.

D'una banda, els avantatges resultants de la naturalesa interdisciplinària de la sociolingüística es poden exemplificar amb la fructífera dinàmica de col·laboració entre Joshua A. Fishman (representant de la sociologia) i Charles A. Ferguson (de la lingüística), ja que, segons Spolsky (2011: 5), les primeres converses entre ells i la seva associació continuada van influenciar enormement el creixement del camp.

De l'altra, en el cas de la sociolingüística, la col·laboració entre tradicions científiques no va estar exempta de certes tensions. Shuy (2003: 8-9) explica que, a l'inici, els científics socials no volien renunciar de cap manera a la seva perspectiva per comprendre la lingüística i viceversa: “Each wanted to keep its own field, goals and theory-building foremost while enjoying the most minimal fruits of the other”. Així, es van formar “dues cultures” en certa manera separades.

Avui, aquesta frontera correspondria encara segons alguns acadèmics als usos diferenciats dels termes *sociolingüística* (en sentit restrictiu) i *sociologia de la llengua*. Per a Paulston i Tucker (2003: 1), mentre que la primera inclou els esforços de lingüistes i antropòlegs per descriure les llengües des

d'aproximacions cada cop més diverses i profundes, l'última tracta d'explicar i predir "language phenomena in society at the group level" i és principalment una empresa de sociòlegs, si bé els han acompanyat uns quants lingüistes. Alguns afirmen que els interessos de lingüistes i sociòlegs rarament coincideixen (Shuy, 2003: 8). Tot i això, és possible identificar esforços d'alguns acadèmics formats en tots dos àmbits, com Mallinson (2009), que identifica línies d'investigació paral·leles en la sociologia i la lingüística i proposa formes en què aquestes disciplines poden col·laborar.

Sobre la contribució d'experts en lingüística i en sociologia al camp que ens interessa, els antropòlegs han tingut un pes creixent en la disciplina, si bé al principi els lingüistes van fer el gruix d'aportacions teòriques a la sociolingüística (Boix i Vila, 1998: 16). Des d'una perspectiva americana, que no es correspon exactament amb l'europea en aquest cas, Spolsky (2011: 8) explica que, segons Ferguson (1997: 78), els sociòlegs, malgrat les seves importants contribucions al seminari de Bloomington, van delegar en antropòlegs i lingüistes la tasca de desenvolupar aquest camp. A més, sembla que les opinions d'experts que avui considerem fundadors de la sociolingüística més àmplia divergien des dels orígens del camp sobre el camí que havien de seguir les seves investigacions pioneres. Així doncs, per exemple, mentre que Labov volia posar el focus en la millora de la lingüística, Fishman preferia contribuir a la sociologia i "[o]ver the years, most did not change their opinions" (Spolsky, 2011: 8).

En certes qüestions, això no obstant, alguns d'aquests coneguts experts sí que van canviar de parer amb el temps. Possiblement el canvi més conegut és el de l'actitud de Labov vers el terme *sociolingüística*. Al principi, va ser reticent al seu ús perquè s'oposava a la perspectiva d'estudiar els sistemes lingüístics sense tenir en compte la seva materialització en les societats⁹, però finalment la realitat acadèmica va imposar la necessitat de diferenciar-la d'altres aproximacions mitjançant aquesta etiqueta (vegeu, per exemple, Boix i Vila, 1998: 13; Koerner, 2002: 274; Cornips i Gregersen, 2016: 504).

⁹ "[T]he language of individuals cannot be understood without knowledge of the community of which they are members" (Labov, 2006: 5).

2.2.2. Divergències teòriques entre la sociolingüística i la lingüística formal

Al llarg del temps s'han fet diversos intents de definir la sociolingüística, cosa que ha suscitat controvèrsia entre acadèmics, ja que, com recorda Silva-Corvalán (2001: 1), “definir y delimitar esta disciplina de manera precisa es tarea difícil y, en todo caso, poco fructífera”. Tot i això, en aquesta tesi entenem la sociolingüística en sentit ampli com el conjunt d'aproximacions que estudien els usos lingüístics sense deslligar-los de la societat en què ocorren i dels complexos fenòmens que s'hi relacionen i poden servir per entendre'ls, ja siguin lingüístics, socials, psicològics o d'altra naturalesa.

Des del seu establiment, la sociolingüística es va oposar, principalment, a dos corrents imperants fins aleshores entre els científics interessats pels sistemes lingüístics: d'una banda, l'estructuralisme de Ferdinand de Saussure i l'escola de Praga i, de l'altra, el generativisme de Noam Chomsky¹⁰. Tots dos tenen objectius diferents. Mentre que el primer estudia “la llengua com a codi i com a sistema que transmet informació”, l'últim la veu “com a model o mirall de la ment” (Boix i Vila, 1998: 111). A més, els seus mètodes tampoc són estrictament els mateixos, però sí que coincideixen en l'acceptació de la inducció i la intuïció lingüística dels investigadors com a metodologia de recollida de dades (Turell, 2003: 2).

¹⁰ De vegades, força aferrissadament: “[...] aquesta és una proposta teòrica i modèlica, argumentant en contra d'aquells lingüistes (els generativistes) que ens volen fer creure que només ells es dediquen a la lingüística; això sí, aquest model implica uns mètodes prou diferents dels d'aquells, ja que fa ús de l'observació i la descripció de les dades observades, mentre ells utilitzen la introspecció. De fet, aquest model representa un trencament amb l'estructuralisme — que només considera la funció referencial de la llengua— i amb Chomsky, perquè encara que la proposta dels conceptes de competència, actuació i creativitat és prou interessant, aquesta proposta és incompleta (Hymes, 1972)” (Alturo i Turell, 1991: 128).

En canvi, una de les reivindicacions més fonamentals de la sociolingüística va ser la necessitat d'entendre la llengua en la societat, la qual cosa implicava que els estudis havien de partir d'una mostra representativa de la societat, i no de les reflexions dels propis lingüistes¹¹.

Actualment comptem amb diverses obres que tracten les oposicions i els punts de contacte entre la sociolingüística i altres maneres d'estudiar les llengües. Per exemple, Chevrot i Foulkes (2013) plantegen la relació entre la sociolingüística i la psicolingüística cognitiva. A més, gairebé tots els manuals sobre sociolingüística incorporen seccions dedicades a contextualitzar-la dins de la lingüística o en relació a altres perspectives (entre altres, Boix i Vila, 1998: 13-33; Hudson, 2000: 13-15; Chacón Beltrán, 2004: 23-5). En aquest sentit, destaca l'obra de Figueroa (1994), que persegueix precisament establir la relació entre la sociolingüística i altres models teòrics de la lingüística més general.

Segurament, la confrontació més clara entre la sociolingüística i altres perspectives és la premissa que Hymes (2000: 312) resumeix amb l'oposició "+social vs. -social" i que reflecteix el nom mateix de *sociolingüística*. La segona diferència és el rebuig de l'axioma o postulat de la categoricitat, que havia predominat a les empreses lingüístiques des de temps remots i segons el qual el material de l'anàlisi lingüística "must be regularized to eliminate real-world variability" (Chambers, 2003: 12). Aquest postulat descansa en la distinció saussureana entre *langue*, la facultat abstracta del llenguatge, i *parole*, l'ús real que en fan els parlants, que segons Chambers (2003: 26) dona la mà a altres dicotomies com la de *competència* i *actuació* de Chomsky o la d'*energeia* i *ergon* de Humboldt.

Una tercera oposició és l'alternativa, coneguda com "paradigma dinàmic", a la separació metòdica i taxativa entre sincronia i diacronia propagada per l'estructuralisme. La sociolingüística treballa des d'un marc d'anàlisi pancrònica,

¹¹ Segons Chambers (2003: 26), això correspon a "[...] what Labov (1972a: 185-6) calls the "Saussurean Paradox," whereby the data from a single speaker, often linguists relying on their own intuitions, are all that are needed to study the communal *langue*, but in order to study the individualistic *parole* one must study language as it is actually used in the community".

és a dir, que uneix en una sola dimensió analítica el temps històric, el real i l'aparent (Silva-Corvalán, 2001: 2-3; vegeu el subapartat **2.2.4.1.3**). En aquest sentit, per exemple, Labov, principal exponent del variacionisme (el corrent sociolingüístic que més ens interessa en aquest estudi), deia en una ponència de 1977 que “utilitz[a] el present per explicar el passat” (Pradilla, 2011: 129). La distinció entre temps real i temps aparent va obrir un gran ventall de possibilitats per a la recerca lingüística, com s'explica més detingudament al subapartat **2.2.4.1.3**.

La quarta diferència entre la sociolingüística i els seus precedents és el rebuig de la idea de Bloomfield (1933/1984: 347) que el canvi fonològic no es pot observar directament. El gran nombre d'estudis sociolingüístics que s'han centrat precisament en aquest objectiu contradiu per si mateix aquesta posició. El desenvolupament de la metodologia variacionista, que permeté estudiar canvis mentre es produïen, va ser un dels factors clau que va fer que molts lingüistes deixessin de banda la idea bloomfieldiana i s'interessessin pels canvis fonològics.

El penúltim punt de discrepància entre les dues orientacions d'estudis lingüístics és l'expressat per Bloch i Trager (1942; citat a Labov, 2006: 9), que creien que els sentiments dels parlants cap a les llengües són inaccessibles per als investigadors. Això no obstant, actualment diverses línies de recerca sociolingüística estudien les actituds lingüístiques dels parlants, sobretot en l'àmbit de la psicologia social. Existeixen nombroses publicacions centrades en la psicologia social de la llengua, com Meyerhoff (2002), que la tracta en relació amb la variació lingüística, o la revisió bibliogràfica que fan Baker, Watson i Gallois (2020). En català, un exemple d'aquesta mena de recerca és l'estudi de Sinner i Wieland (2015).

L'últim punt que diferencia els enfocaments sociolingüístics d'altres concepcions de la lingüística és una idea especialment estesa després d'André Martinet que diu que els lingüistes no haurien de servir-se de dades no lingüístiques per explicar els canvis lingüístics (Labov, 2006: 10-11). En canvi, l'acceptació i la proliferació d'estudis que se centren efectivament en les relacions entre factors externs i interns a la llengua per explicar la variació demostra la gran utilitat

teòrica (per exemple, per a la descripció dels canvis lingüístics) i pràctica (per a la planificació i política lingüístiques, entre altres) d'aquesta aproximació metodològica.

2.2.3. Els enfocaments sociolingüístics

Així doncs, tenint en compte la seva emergència com a camp interdisciplinari i el seu desenvolupament, és fàcil comprendre per què diversos autors es concentren en diferents aspectes concrets dels usos lingüístics i divergeixen a l'hora de llistar aquells enfocaments de l'estudi de la llengua que consideren sociolingüístics (Tagliamonte, 2012: 1).

A tall d'exemple de les diverses subdisciplines que es poden considerar sociolingüístiques, presentem a continuació com a Taula 4 una llista que considerem no exhaustiva, adaptada de Preston (2004: 141). Tot i que són indiscutiblement complementàries, com demostra l'intercanvi continuat de construccions teòriques entre elles, la diversitat de perspectives genera debats i desacords entre acadèmics que persegueixen objectius específics semblants i, fins i tot, superposats.

Taula 4. Algunes de les primeres línies sociolingüístiques.

Aproximació	Autors de referència
Sociolingüística de la variació	William D. Labov, Walt Wolfram, Sali A. Tagliamonte
Etnografia de la comunicació, interaccionisme	Dell Hymes, John J. Gumperz, Deborah Tannen
Psicologia social de la llengua	Wallace E. Lambert, Jean-Marc Dewaele, Howard Giles
Sociologia de la llengua	Joshua A. Fishman, David Singleton, Patrick Studer
Etnometodologia, anàlisi de la conversa i anàlisi crítica del discurs	Harvey Sacks, Emanuel A. Schegloff, Norman Fairclough
Política i planificació lingüístiques	Einar Haugen, Bernard Spolsky, Robert L. Cooper

A banda de la seva definició, un dels temes més discutits al voltant de la sociolingüística és precisament la seva compartimentació. Per exemple, Coulmas (1997: 2) i altres (com Gimeno Menéndez, 1990 o Pradilla, 2003: 4) presenten una distinció entre *microsociolingüística*, o dimensions socials de la llengua, i *macrosociolingüística*¹², o dimensions lingüístiques de la societat, mentre que altres (com Hudson, 2000 o Chambers, 2003) utilitzen el terme *sociolingüística* per referir-se a la primera i el de *sociologia de la llengua* per a la segona (Boix i Vila, 1998: 14). Per poder comprendre la relació entre diversos corrents sociolingüístics, en aquesta tesi optem per la concepció de la sociolingüística en sentit ampli com un contínuum tridimensional proposada a Boix i Vila (1998). Aquests autors distingeixen tres pols lligats a tres dimensions: el pol social (dimensió social-col·lectiva), el pol psicològic (dimensió individual-mental) i el pol lingüístic (dimensió lingüística-estructural).

Amb una orientació cap al primer pol hi identifiquen tres línies de recerca principals: la sociologia de la llengua (estudis, entre altres, sobre el multilingüisme i la substitució lingüística), la planificació lingüística i estudis antropològics d'interaccions cara a cara. Alguns dels noms més cèlebres d'aquesta dimensió són: Joshua Fishman, Einar Haugen, Nancy D. Dorian, Susan Gal, Erving Goffman, John J. Gumperz o Dell Hymes.

Els estudis de les ideologies, els de les actituds lingüístiques i els que s'ocupen dels vincles entre llengua i ensenyament estan més orientats cap al pol psicològic. Alguns noms destacats d'aquests camps són Louis-Jean Calvet, Howard Giles, Henri Tajfel, Jerzy J. Smolicz, Tove Skutnabb-Kangas o Marilyn Martin-Jones.

¹² Si bé és cert que Coulmas (1997: 3) apunta més endavant que: "Although sociolinguistic research centers about a number of different key issues, any rigid micro-macro compartmentalization seems quite contrived and unnecessary in the present state of knowledge about the complex interrelationships between social and linguistic structures."

L'últim dels pols és el lingüístic, que atrau aproximacions a l'estudi de les llengües com el variacionisme (descriu en més profunditat al subapartat **2.2.4**), el contacte de llengües o la planificació del corpus. Hi destaquen noms com el de William D. Labov, Uriel Weinreich, Einar Haugen o Heinz Kloss.

Malgrat la utilitat d'aquesta concepció tridimensional de la sociolingüística, cal subratllar que, com passa a totes les disciplines científiques, les fronteres entre aproximacions diferents són difuses i (per sort) existeixen estudis que desenvolupen mètodes combinats. Entre molts altres, Snell (2010) constitueix un bon exemple d'aquesta mena d'estudis sociolingüístics, ja que combina:

“[...] quantitative sociolinguistic analysis of a linguistic variable (the first person possessive singular) with interactional analysis of the way one particular variant of that variable (possessive 'me', as in *Me pencil's up me jumper*) is used by speakers in interaction” (Snell, 2010: 630).

Aquesta tesi, com ja s'ha dit anteriorment, també se serveix de diversos camps i aproximacions. L'esbós del marc general de la sociolingüística que s'ha presentat fins ara es complementa tot seguit amb un repàs de la sociolingüística variacionista, una de les fonts teòriques i metodològiques de les quals beu més directament el nostre estudi.

2.2.4. El variacionisme

Com hem vist a la Taula 4 del subapartat **2.2.3**, la sociolingüística variacionista és una de les aproximacions a la sociolingüística. A continuació, es resumeixen els conceptes bàsics d'aquest enfocament, se'n presenta sumàriament l'evolució a nivell internacional i després es contextualitzen les aportacions sobre el català a aquest corrent sociolingüístic.

En aquesta tesi, utilitzem com a sinònims diversos termes que s'han usat al llarg dels anys per denominar la subdisciplina que ens ocupa. Així doncs, no fem cap distinció entre els termes *sociolingüística variacionista* o *de la variació*, *variacionisme*, *sociolingüística laboviana* o *quantitativa*, *sociolingüística*

*correlacional*¹³, *dialectologia urbana* o *estudi de la variació i el canvi lingüístics*, si bé utilitzem de forma majoritària els tres primers.

A continuació, caracteritzem els fonaments i l'evolució del variacionisme i, més endavant, repassem la seva aplicació a l'estudi de la llengua catalana.

2.2.4.1. Les bases teòriques i metodològiques del variacionisme

2.2.4.1.1. L'aplicació del mètode hipoteticodeductiu a l'estudi de la variació

Com apunta Casesnoves (2012: 5), “la sociolingüística de la variació consisteix, com el seu propi nom indica, a afirmar, demostrar i explicar la *variabilitat* de la llengua”. Anteriorment s'ha explicat que la sociolingüística general, i per tant també la variacionista, no veu la llengua com Saussure definia la *langue*, homogènia, sinó com un sistema inherentment variable. Bailey (1973: 9), per exemple, afirma que la distinció saussureana entre *langue* i *parole* “was destined to have ominous consequences”.

Els acadèmics d'altres corrents lingüístics que estudien la llengua com a sistema homogeni prenen com a objecte d'estudi vàlid el seu propi idiolecte. En canvi, el variacionisme defensa la necessitat de dur a terme estudis empírics, basats en l'observació de varietats vernaculars, i, fins i tot, hipoteticodeductius, en què les hipòtesis es corroboren o refuten mitjançant l'observació de la realitat.

Ara bé, quina realitat ha de ser l'observada continua sent un dels debats actius entre sociolingüistes de la variació. Alguns, entre ells Labov (2006), afirmen que “[l]a detecció de fenòmens variables passa indefectiblement per observar-ne la presència en el marc d'una determinada comunitat de parla” (Pradilla, 2001: 1). Altres acadèmics, però, defensen les aproximacions que observen la variació en l'individu, com el paradigma dinàmic de Bailey (1973).

¹³ Aquest terme s'ha usat, sobretot a l'inici de la subdisciplina, amb una certa intenció despectiva i ha estat rebutjat per alguns acadèmics per als quals sembla implicar que “the counting and measuring of linguistic features and the correlating of linguistic variability with different forms of social differentiation was the simple and banal goal of our work”, mentre que aquest és merament el punt de partida de les seves investigacions (Trudgill, 2006: 373).

Malgrat les possibles discrepàncies entre ells, i sigui quina sigui la naturalesa de la seva mostra, tots els variacionistes comparteixen la premissa que, observant els usos que fan els parlants de les varietats vernaculars, s'hi poden identificar regularitats o patrons sociolingüístics (en anglès, *sociolinguistic patterns*). En aquest treball, entenem els patrons sociolingüístics com un seguit de “factors [socials i lingüístics] que determinen i es relacionen de manera regular i sistemàtica amb l'ús lingüístic dels parlants” (Casesnoves, 2012: 14).

Així, els estudis variacionistes han demostrat reiteradament mitjançant l'aplicació del mètode hipoteticodeductiu que la variació no és caòtica, sinó sistemàtica. Aquest fet contradiu una idea que va ser àmpliament acceptada fins a l'establiment de la subdisciplina variacionista. Fins aleshores, un gran nombre de lingüistes defensaven que “algunes diferències lingüístiques no comport[av]en cap diferència i que exist[ia] una equivalència de variants, i per tant, una variació lliure” (Turell, 2003: 2).

A més, també s'ha pogut provar que els canvis diacrònics s'esdevenen a mesura que les comunitats de parla, que són heterogènies, adopten i abandonen certes característiques lingüístiques. Pradilla (2001: 1) resumeix aquesta premissa sota l'etiqueta *hipòtesi de la variabilitat*:

“El mètode hipoteticodeductiu, per definició, parteix de consideracions provisionals que hauran de ser validades (o invalidades). Aquestes suposicions prèvies van des de plantejaments molt generalistes a previsions d'alta concreció. Un exemple de les primeres pot ser la mateixa hipòtesi de la variabilitat, segons la qual la parla dels individus d'una determinada comunitat es caracteritza per la presència de certs trets lingüístics variables la freqüència d'aparició dels quals es trobarà condicionada per factors independents de tipus lingüístic i/o extralingüístic.”

2.2.4.1.2. Les nocions de variable i regla variable

Per dur a terme l'empresa variacionista, va caldre, entre altres, desenvolupar els constructes teòrics bàsics de *variable* i *regla variable*¹⁴. La primera és, per a Fasold (1990: 223-4), un conjunt de formes alternatives de dir el mateix, en què cada alternativa té rellevància social, és a dir, que cada variant (o possible realització d'una variable) “may index a series of stylistic, social and regional characteristics” (Trousdale, 2010: 122).

Hazen (2011: 32) resumeix els beneficis revolucionaris que va comportar el concepte de variable sociolingüística. En primer lloc, va permetre gestionar la variació de forma sistemàtica mitjançant el seguiment quantitatiu de la producció de les variants en diversos contextos socials i lingüístics. En segon, va permetre el desenvolupament i l'aplicació de models quantitius multivariants de la variació lingüística.

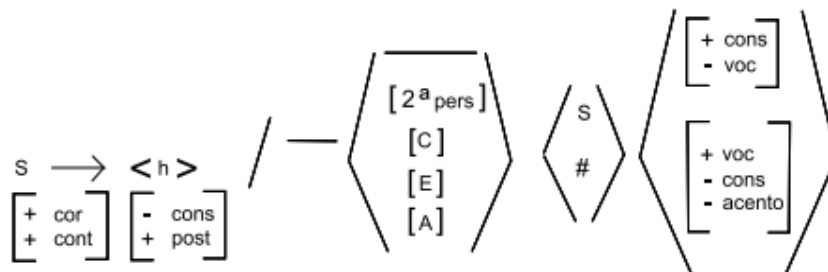
Amb el temps, s'han estudiat diversos tipus de variables: fonològiques, lèxiques, morfosintàctiques o discursives. A causa de les seves característiques, cadascun d'aquests tipus ha tendit a aparèixer en certa mena d'estudis. Cal tenir en compte, entre altres, que les dues últimes són sovint més complexes que les dues primeres, segons Hazen (2011: 33). Autors com Wolfram (2006: 334) indiquen, per exemple, que la variació fonològica i morfosintàctica ha dominat les recerques sincròniques i de temps aparent; la majoria d'estudis diacrònics s'han centrat en la variació sintàctica. En el paradigma quantitatiu, però, independentment del tipus de variable estudiada, com apunten Boix i Vila (1998: 30) “[l]a diferència, per tant, no es basa pas en la categoria que fa servir [el parlant] sinó en els percentatges d'utilització de cada variant en cada estil”.

Pel que fa a la noció de regla variable, Gimeno Menéndez (2003: 76) explica que és una representació de l'efecte que un conjunt de factors lingüístics i socials tenen sobre una variable lingüística. Per exemple, Etxebarria (2013: 218) mostra

¹⁴ Pradilla (2011: 128-9) proporciona una síntesi de les principals aportacions teoricometodològiques de la sociolingüística quantitativa de Labov en només 5 punts.

la regla variable següent sobre els processos d'aspiració i elisió del fonema /s/ en l'espanyol parlat a Las Palmas.

Figura 1. Regla variable de l'aspiració de /s/ a Las Palmas.



Font: Etxebarria (2013: 218).

Aquesta regla variable descriu que la posició interna de /s/ n'afavoreix l'aspiració i que els contextos preconsonàntics i prevocàlics atons l'afavoreixen a final de paraula, però no l'afavoreixen ni la posició prèvia a una pausa ni a una vocal tònica (Etxebarria, 2013: 218). Així doncs, aquesta regla variable descriu factors lingüístics (la posició del fonema) que afecten la variable lingüística observada (l'aspiració del fonema /s/).

Hazen (2011: 33) explica que la regla variable es remunta a Labov (1969), mentre que la variable lingüística es va integrar a les teories per a l'estudi de la llengua a Weinreich, Labov i Herzog (1968: 167). Aquests dos conceptes han conegut un èxit molt diferent, ja que, segons Fasold (1991), "scholars have largely abandoned variable rules as a descriptive mechanism, although the construct of the variable is widely used" (Hazen, 2011: 33). Com es descriu més àmpliament al capítol 3, en aquesta tesi no intentem identificar regles variables, però, en canvi, els conceptes de variable i variant són al centre del nostre estudi.

2.2.4.1.3. La revolució variacionista del temps

Com s'ha avançat anteriorment, una altra de les aportacions més revolucionàries de la sociolingüística de la variació a l'estudi de la llengua es relaciona amb la concepció del temps que es proposa en els estudis dels canvis lingüístics. Des d'una perspectiva sociolingüística, la barrera que distingia fins al seu sorgiment entre les perspectives diacrònica i sincrònica es torna difusa, i més després de la

introducció i acceptació de la proposta de Labov (2006: 8) d'estudiar els canvis lingüístics a partir de la noció de temps aparent:

“By studying the differences between the linguistic behavior of successive age levels in our sample, we can make inferences about linguistic change. This type of approach may be referred to as a *pseudo-trend study*, it is a series of cross-sections in apparent time as opposed to real time. The dimension of apparent time lies along the axis of the age levels of present-day informants, taken as representatives of the native speech pattern of the years in which they grew to maturity”.

Per tant, els estudis que observen canvis lingüístics en temps aparent són aquells que ofereixen “anàlisi[s] de la distribució de les variables lingüístiques en diferents grups d'edat” (Turell, 2003: 7). Aquests estudis no poden demostrar canvis en marxa en la comunitat de parla analitzada, sinó patrons sociolingüístics de gradació per edats (en anglès, *age grading*; trobeu-ne la definició al subapartat **2.3.2.1**).

Les limitacions metodològiques dels estudis a temps aparent es poden superar mitjançant estudis a temps real, en què es recullen dades en dos moments diferents. L'investigador de canvis en temps real té tres opcions: intentar trobar recerques anteriors sobre la comunitat de parla que vol analitzar (“ressenyar el passat”); elaborar un estudi rèplica, en què realitza exactament el mateix procés de selecció i anàlisi d'informants a la mateixa comunitat en dos moments diferents; o elaborar un estudi panel, en què localitza els mateixos informants un temps després del primer estudi i repeteix el procediment original (Turell, 2003: 7-8). Com veurem, el nostre estudi encaixa en l'última opció, ja que s'estudien els mateixos informants en dos moments diferents i en la mateixa situació comunicativa. Val a dir que, com passa amb altres decisions metodològiques (per exemple, les anàlisis quantitatives i qualitatives), és possible combinar corpus a temps real i temps aparent en un sol projecte d'investigació, com exemplifica el projecte d'idiolectometria mencionat anteriorment (al subapartat **2.1.4.1.1**), que, per al català, utilitza el corpus de La Canonja recollit a la dècada de 1980

(temps aparent) i el complementa amb gravacions obtingudes anys més tard dels mateixos informants (temps real) (Forensiclab, 2013: 9).

Així doncs, la perspectiva temporal de l'anàlisi sociolingüística de la variació sovint es descriu amb el terme *pancrònica*, ja que engloba tant la dimensió temporal específica en què es du a terme l'estudi com l'evolució en el temps dels trets lingüístics que observa. A més, segons Turell (2003: 5):

“Labov (1989) demostra que algunes variables lingüístiques, que no estan motivades sincrònicament, reflecteixen continuïtat històrica amb pocs canvis al llarg del temps i que la separació entre lingüística sincrònica i diacrònica no està fonamentada, ja que els infants són historiadors perfectes del llenguatge”.

2.2.4.1.4. L'oposició variacionista entre els canvis des de dalt i des de baix

Una altra de les aportacions més àmpliament reconegudes de l'estudi de la variació i el canvi lingüístics és la distinció laboviana entre els canvis des de dalt i els canvis des de baix. Com veurem al subapartat **2.3.1.2**, aquests conceptes s'han relacionat sobretot amb els efectes que tenen sobre el comportament lingüístic en relació amb el sexe dels parlants.

Els canvis lingüístics des de dalt tenen lloc més freqüentment en estils de parla formals, sovint estan subjectes a fenòmens d'hipercorrecció i, de vegades, en aquests canvis, els parlants mostren estereotips explícits sobre variables sociolingüístiques estables. Els canvis des de baix operen “within the system, below the level of social awareness” (Labov, 1990: 215; citat a Turell, 1995b: 284). La Taula 5 recull a continuació les diferències entre aquests dos tipus de canvi lingüístic.

Taula 5. Diferències entre canvis lingüístics des de dalt i des de baix.

Canvi des de dalt	Canvi des de baix
<ul style="list-style-type: none">• Típicament sorgeixen en les anomenades situacions sociolingüístiques estables.• Comporten alteracions en la distribució social de moltes variables lingüístiques.• Es correlacionen clarament amb variables socials com el sexe i la classe social.• Les variables lingüístiques tenen un prestigi manifestament reconegut per la comunitat de parla.• Els parlants tenen actituds molt positives vers aquestes variables; les valoren com a més correctes.• Les variables presenten estratificació estilística.	<ul style="list-style-type: none">• En principi els parlants no en són conscients i les variables lingüístiques en qüestió no provoquen cap tipus de valoració.• Originàriament, el canvi està motivat per factors lingüístics interns.• En les fases finals del procés de canvi les variables presentaran estratificació estilística.• També en aquestes fases els parlants tindran actituds molt negatives vers les variables.

Font: elaboració pròpia a partir de Turell (1995b: 284) i Casesnoves (2012: 31-32).

L'enfocament i la metodologia que acabem de descriure van influenciar les pràctiques de molts sociolingüistes i van propiciar canvis profunds del paradigma utilitzat fins a l'establiment de l'estudi de la variació i el canvi lingüístics.

2.2.4.2. L'evolució del variacionisme

Des dels seus inicis, el variacionisme s'ha anat transformant, sense deixar de banda els fonaments delineats als apartats anteriors. Eckert (2005, 2012) proposa la classificació de les contribucions variacionistes en tres onades segons els seus mètodes i objectius. Els canvis que han comportat cadascuna de les etapes han desembocat en una inversió de la visió del vincle que uneix les llengües i les societats (Eckert, 2012: 97).

Segons aquesta autora, durant la primera onada d'estudis es va desenvolupar el panorama general (*big picture*) mitjançant estudis a gran escala de comunitats definides geogràficament. Aquestes investigacions usaven les jerarquies

socioeconòmiques com un mapa de l'espai social i concebien les variables lingüístiques com a marcadors de categories socials primàries carregades de prestigi o estigma social. A més, l'estil es definia com “attention paid to speech, and controlled by orientation to prestige/stigma” (Eckert, 2005: 3).

La segona onada, que va ajudar a desenvolupar el panorama local (o *local picture*), es resumeix com el conjunt d'estudis etnogràfics de comunitats, encara definides geogràficament, que veien diverses categories locals com vincles amb etiquetes demogràfiques. Per als investigadors d'aquest període, les variables lingüístiques observades serveixen per indiciar categories definides a nivell local i l'estil es concep com actes d'afiliació a aquests grups.

En canvi, a la tercera onada s'adopta una perspectiva estilística. Aquest èmfasi “places speakers not as passive and stable carriers of dialect, but as stylistic agents, tailoring linguistic styles in ongoing and lifelong projects of self-construction and differentiation” (Eckert, 2012: 97). Les investigacions etnogràfiques d'aquesta etapa analitzen comunitats de pràctica i proposen que les categories locals es construeixen sobre posicions comunes (*common stances*) i que les variables indicien posicions, activitats o característiques.

Així doncs, amb el temps i els canvis metodològics, “[i]t has become clear that patterns of variation do not simply unfold from the speaker's structural position in a system of production, but are part of the active —stylistic— production of social differentiation” (Eckert, 2012: 98). Aquest canvi fa, doncs, que els estudis més recents sobre la variació lingüística concebin els informants com a participants que juguen un rol actiu en la construcció de les seves identitats mitjançant els recursos lingüístics al seu abast.

2.2.4.3. Els estudis variacionistes de la llengua catalana

En el nostre context acadèmic, l'enfocament sociolingüístic variacionista gaudeix de certa tradició. En aquest subapartat es presenta un repàs d'alguns dels estudis de la variació en català més significatius per al nostre treball, partint de les descripcions de l'estat de la qüestió que fan diversos experts (Mas i Montoya, 2004; Pradilla, 2003, 2008, 2011, 2020; Turell, 2013).

El gruix de la sociolingüística de la variació en llengua catalana està format per les investigacions de fenòmens fonològics, ja que aquesta mena de variables:

“[...] present[en] uns avantatges que no trobem en [les] altres, com poden ser la major freqüència d'aparició, la facilitat de comprovació del requisit d'equivalència entre totes les variants d'una variable i la propietat de les variants fonològiques de marcar fàcilment l'estratificació social i estilística” (Mas i Montoya, 2004: 247).

Troblem estudis sobre els sistemes vocàlics de diverses varietats, com el de Plaza (1995) sobre [i], [e] i [ə] a la Conca de Barberà o Carrera (2006) sobre les vocals pretòniques a inici de paraula a Alguaire i Lleida, i també sobre sons consonàntics (com Güell, Buldó, Pradilla, Saura i Verge, 2006 sobre la realització de la fricativa labiodental a Valls)¹⁵.

La sociolingüística variacionista del català ha anat canviant al llarg del temps. Turell (2013) fa un repàs de nombrosos estudis variacionistes sobre la llengua catalana i conclou que:

“[...] d'una banda, s'ha pogut avançar en la comprensió específica de les causes de la variació i el canvi lingüístics del català, i de la forma en què els canvis lingüístics s'estenen a través del temps i de l'espai, i de l'altra, i sense cap mena de dubte, s'han fet aportacions importants a la sociolingüística variacionista general, tot i que en ambdós casos queda molt camí per recórrer” (2013: 49).

¹⁵ Aquí no tractarem amb detall les investigacions de fenòmens fonològics del català, ja que aquest nivell lingüístic es troba fora dels objectius de la nostra tesi. Els lectors interessats en el gran nombre d'investigacions sobre variació fonètica i fonològica de diverses varietats del català poden trobar-ne resums molt aclaridors a les obres en què es basa aquest subapartat: Mas i Montoya (2004), Pradilla (2003, 2008, 2011 i 2020) i Turell (2013).

En una línia semblant, Pradilla (2011: 127) sosté que s'ha produït un canvi dins la sociolingüística catalana des dels anys vuitanta del segle passat, anteriorment enfocada gairebé exclusivament a la sociologia de la llengua, gràcies a “l'aparició d'un bon grapat de recerques de caràcter variacionista i etnogràfic”.

A continuació es presenten de forma resumida i per ordre cronològic d'aparició les aportacions variacionistes rellevants per al nostre estudi. La primera d'elles és l'estudi realitzat per Mier (1986) de la parla de 43 informants del barri de Gràcia que tenien el català com a primera llengua. A banda de tres variables fonològiques, va observar l'ús de *dos* o *dues*, de les preposicions *en* i *a*, així com de les formes verbals *tenir que* o *haver de*. Les variables socials amb les quals les va relacionar són l'edat, el sexe, el nivell d'ocupació i el grau d'alfabetització en català dels participants. S'ha assenyalat que aquest estudi capdavanter presenta mancances perquè “s'obvien factors tan importants com la procedència del mateix informant i dels pares, en el cas que no hagin nascut allí” (Pradilla, 2008: 68).

Alonso i Suils (1993) van estudiar el canvi des de dalt que afectava les terminacions del subjuntiu al Segrià amb una mostra de 87 informants de 7 pobles d'aquesta comarca, alguns dels quals experimentaven aleshores canvis econòmics que podrien reforçar l'assimilació de trets lingüístics de la varietat estàndard (l'impuls del sector comercial en detriment de part de l'agrícola). Els autors van analitzar la producció de formes verbals al nivell formal mitjançant les variables lingüístiques de persona verbal i conjugació i les socials de sexe i grup d'edat i van concloure, en relació als canvis socioeconòmics de la comarca, que “[l]es dones de trenta a cinquanta anys són les més innovadores” (Turell, 2013: 43).

Alturo (1995) va estudiar el canvi de *ser* per *haver* en formes de plusquamperfet al Pont de Suert. Observava les variables d'estil de parla, nivell de coneixement del català escrit, temps de l'acció i temps de referència i va trobar que el procés d'estandardització s'afavoria en parlants de grups intermedis (Turell, 2013: 43). A més, va proposar un sistema en què “les situacions d'aspecte cíclic s'oposarien

d'una banda a les d'aspecte no cíclic/[+anterior], i d'una altra a les d'aspecte no cíclic/[−anterior]” (Alturo, 1995: 248; citat a Turell, 2013: 43).

També de la dècada dels noranta cal destacar, d'una banda, l'estudi de Turell (1995a) en què explorava des d'una perspectiva variacionista el contacte entre el català i el castellà a La Canonja mitjançant l'observació de dos préstecs (o manlleus) lingüístics (*bueno* i *casi*) i de l'alternança de codi (*code-switching*, o l'ús de diverses llengües en un mateix torn de paraula). Quant als primers, va concloure que “són més nombrosos els préstecs distants, de tipus *bueno*, que els préstecs propers, de tipus *casi*” i que “la majoria de préstecs, tant distants com propers, estan integrats a la llengua receptora, és a dir, el català” (Turell, 1995a: 289). Pel que fa a l'alternança de català i castellà en un mateix torn de paraula, va trobar que la llargada de la situació de contacte entre les llengües i les adscripcions dels parlants a “les pràctiques discursives de cada grup” regien el seu comportament durant les interaccions (Turell, 1995a: 288).

De l'altra banda, a finals dels noranta (1998) es va publicar *Oralmunt: Estudis de variació funcional*, editat per Lluís Payrató, que “aplega un conjunt d'estudis sobre la variació funcional del llenguatge manifestada a través del cas concret de la llengua catalana” (Payrató, 1998: 5). Aquests treballs se centren en diversos aspectes d'un mateix corpus, dissenyat expressament per a l'ocasió i que “es compon de 16 mostres, de 2000 paraules aproximadament cada una, corresponents a varietats funcionals diferenciades de català contemporani de procedència dialectal central” (Tomàs i Pallejà, 1998: 35). Com veurem més endavant, diversos treballs d'aquest volum es relacionen directament amb certes variables del nostre estudi.

Ja a la dècada dels 2000, Romero (2001; mencionat a Turell, 2013: 43) analitzava variables morfosintàctiques (adjectius demostratius, l'article definit masculí, formes dels pronoms febles personals i adjectius possessius) de la Conca de Tremp en relació amb l'estil de parla, l'edat, el grau d'instrucció i coneixement del català escrit i factors lingüístics (entre altres, la posició respecte el verb, el gènere o el nombre). Els seus resultats sobre l'ús oral de la llengua en situacions formals mostraven estratificació segons l'edat dels parlants i

“comportaments més conservadors entre els informants de més edat” (Turell, 2013: 43).

Montoya (2002) es va situar en la línia de la sociolingüística històrica. En aquesta obra, l'expert alacantí exemplificava “el comportament de tres tipus de canvis lingüístics a base de la morfologia del català balear” segons la direcció que seguien en relació al català i el castellà. En concret, observava els articles, els pronoms clítics i la flexió verbal a les tres illes. Va poder concloure que, mentre que Menorca havia tendit a incorporar certes innovacions que l'alineaven amb Catalunya, Eivissa s'assemblava més aviat al País Valencià en determinades solucions morfològiques i Mallorca “presentaria solucions intermèdies” (Montoya, 2002: 297).

La tesi doctoral de Busquet (2010) representa l'interès continuat i encara vigent per la variació de l'article salat i el literari. En aquest cas, el territori estudiat comprèn diversos pobles del nord-est de Catalunya. L'autora constata que l'article salat només se segueix utilitzant a Cadaqués, on també se n'ha interromput la transmissió cap a les generacions més joves i conclou que el parlar salat d'aquestes regions s'ha extingit.

Finalment, Valls (2013) investiga aspectes fonològics i morfològics del canvi lingüístic en curs del català nord-occidental mitjançant tècniques d'anàlisi dialectomètrica. D'una banda, troba diferències en el grau d'anivellament de diverses varietats nord-occidentals amb l'estàndard (molt més avançat a Catalunya i Andorra que a La Franja, amb els parlars ribagorçans i tortosins entre els dos pols, actuant de “veritables garants del manteniment del contínuum nord-occidental”). De l'altra, confirma que el canvi lingüístic és més ràpid en els aspectes morfològics que en els fonètics.

Així doncs, i malgrat la indiscutible observació que feia Turell (2013) sobre la necessitat de continuar produint investigacions sociolingüístiques variacionistes sobre el català, hem de coincidir amb Mas i Montoya (2004: 250) quan assenyalen que “[...] la variació morfosintàctica té una presència en la sociolingüística catalana més que considerable”.

2.2.5. La sociolingüística com a camp de recerca: síntesi

Fins ara, hem presentat el marc teòric del nostre treball, que beu de la sociolingüística, entesa en sentit ampli com el conjunt d'aproximacions que estudien fenòmens lingüístics en relació amb les societats en què es produeixen. Hem recorregut la història del seu desenvolupament i hem tractat les diferències d'aquesta perspectiva i d'altres que s'emmarquen en la lingüística formal. També hem observat que hi ha certa controvèrsia entre els acadèmics sobre la classificació de les diverses aproximacions sociolingüístiques i ens hem alineat amb la distinció de tres pols que proposen Boix i Vila (1998).

Hem vist que el variacionisme és el punt de partida metodològic del nostre estudi, i n'hem resumit algunes nocions bàsiques, com l'ús del mètode hipoteticodeductiu, el constructe de variable, els estudis a temps real i temps aparent o els canvis des de dalt i des de baix. Hem recuperat la distinció en tres onades de l'evolució del variacionisme proposada per Eckert (2005, 2012) i hem fet un resum de les aportacions variacionistes sobre la llengua catalana més rellevants per als nostres objectius.

2.3. Les principals variables independents dels estudis sociolingüístics

Els estudis de la variació han tractat diverses variables extralingüístiques. Algunes de les analitzades des de l'inici són el sexe, l'edat, la classe social (tot i que, com indica Ash, 2013: 365, no s'ha desenvolupat una perspectiva consensuada sobre com calcular aquesta variable) o la pertinença ètnica (Trudgill, 2006: 373; Feagin, 2013: 28). D'aquestes, les que atrauen més consens entre els variacionistes quant a la seva interacció amb els usos lingüístics dels parlants són el sexe i l'edat, com mostra el fet que es destaquin com les dues variables extralingüístiques principals a la gran majoria de manuals sobre la sociolingüística de la variació (Blas Arroyo, 1994; Fought, 2004; Tagliamonte, 2006, o Chambers i Schilling, 2013, entre altres). Més concretament, Areiza Londoño, Cisneros Estupiñán i Tabares Idárraga (2012: 50) diuen en el seu manual que la variable edat "es una de las más interesantes en la investigación sociolingüística y siempre está relacionada con las otras variables". Cheshire

(2004: 423) afirma en relació a la variable sexe que, juntament amb la classe social, l'edat i l'etnicitat, és una de les categories demogràfiques més utilitzades, tot i que cap altra ha esdevingut tan problematitzada i carregada de connotacions com el sexe.

Al costat d'aquestes variables clàssiques del variacionisme, a la societats plurilingües com ara Catalunya és fonamental la primera llengua. Com demostren Flors-Mas, Sorolla, Pradilla i Vila (2021: 50), la llengua inicial és un paràmetre fonamental de les dinàmiques demolingüístiques”.

2.3.1. La variable sexe

La primera de les variables independents de la recerca sociolingüística és el sexe dels informants. A continuació, repassem els treballs previs que justifiquen la rellevància d'aquesta variable en estudis que investiguen els usos lingüístics de la població, tant des de la sociolingüística més àmplia com de la variacionista i tractem, a més, treballs previs sobre la variable sexe en contextos interaccionals. Finalment, abans de passar a les altres dues variables independents del nostre estudi, explorem les principals crítiques que s'han fet del tractament d'aquesta variable i en resumim l'evolució.

2.3.1.1. La rellevància de la variable sexe

La variable sexe és fonamental per entendre la variació lingüística perquè, com recorda Blas Arroyo (1994: 162), a partir dels setanta del segle passat, la recerca sociolingüística ha confirmat que en diverses llengües occidentals “ciertas variables lingüísticas covarían significativamente con el sexo de los hablantes”. Per exemple, en castellà contemporani, López García i Morant (1991) identifica diferències quantitatives en l'ús de certs elements lingüístics com substantius i adjectius relacionats amb colors o sentiments, alguns prefixos (com *super-*) o truncaments lèxics (*gordi, ilu, chuli*, etc.).

Tot i que es poden trobar exemples d'estudis previs a l'establiment de la sociolingüística variacionista que demostren diferències entre el comportament lingüístic d'homes i dones (com el de Salvador, 1952 sobre les diferències fonètiques entre la parla d'homes i de dones de Granada), els corrents previs al

variacionisme, com la dialectologia tradicional, no es van ocupar sistemàticament d'aquesta variable. Tuson (1998: 207) resumeix el raonament en què se solen basar els estudis dialectològics:

“Com plantegen J. K. Chambers i P. Trudgill (1980), el prototipus d'informant de la dialectologia ha estat, fins fa relativament molt poc, el que ells anomenen amb el acròstic NORMs: *nonmobile, older, rural males* (mascles rurals, vells i no mòbils). Aquest *NORM* era considerat qui millor *conservava* la parla «autèntica» objectiu de recerca de la dialectologia tradicional; qui viatja es «contagia» d'altres formes de parlar, qui viu en un medi urbà també està exposat a d'altres parles i llengües, els joves o encara «s'estan fent» i tenen un llenguatge «pobre» o per tal de distingir-se dels adults busquen formes de parlar diferents i, finalment, les dones resulten poc fiables perquè o bé són considerades massa «normatives» o massa variables i lingüísticament «volubles».”

En canvi, els variacionistes sí que han estat sistemàticament interessats en la variable independent sexe. Fins ara, els estudis s'han centrat en dues línies de recerca principals (Blas Arroyo, 1994: 162). D'una banda, l'anàlisi de variables sociolingüístiques des d'una perspectiva quantitativa, i, de l'altra, les diferències entre els sexes en contextos interaccionals.

2.3.1.2. L'anàlisi variacionista de les diferències per sexe

Des de la perspectiva quantitativa, els estudis de la variació lingüística sovint han inclòs la variable sexe com un dels factors socials relacionats amb el comportament lingüístic dels parlants. La bibliografia variacionista assenyala diferències entre sexes en diversos nivells lingüístics. Bastida Serra (2017: 93) afirma que:

“S'han observat diferències en com les dones i els homes utilitzen el llenguatge en tots els nivells d'anàlisi, des de qui domina la conversa fins a l'ús de la cortesia o el tipus de lèxic emprat (Coates, 2015; Mulac, Bradac i Gibbons, 2001; Tannen, 1991, 1993).”

També s'han identificat aquesta mena de diferències en comunitats de parla diferents. Malgrat tot, cal recordar un cop més que la majoria d'estudis de la variació s'ha centrat en el món occidental i en el nivell fonològic i fonètic (per

exemple, Trudgill, 1974, Macaulay, 1978, o Newbrook, 1982). Si bé aquest predomini de variables fonètiques no exclou l'existència d'estudis que demostrin diferències entre els sexes pel que fa a variables morfològiques i sintàctiques, algun dels quals publicat fins i tot abans dels anys setanta, com el de Shuy (1969).

En termes generals, hi ha nombrosos estudis que apunten que les dones fan un ús més proper a la norma de la llengua (Serrano, 2008). El perquè d'aquesta tendència és un tema àmpliament debatut (vegeu, per exemple, Meyerhoff, 2014; Milroy i Gordon, 2003: 100-108, o Goldschmidt i Weller, 2000: 120). Aquesta tendència es concreta, segons Labov (2001: 266-292), en tres principis plantejats que han estat corroborats pels comportaments lingüístics analitzats en diversos estudis variacionistes.

El primer principi (Labov, 1991: 205; 2001: 266) estipula que, en condicions estables d'estratificació social, les dones usen formes estàndards amb més freqüència que els homes. Per exemple, segons Trudgill (2000: 70) la recerca sobre les diferències entre l'anglès utilitzat per homes i per dones a Detroit mostra que:

“In Detroit, higher-class speakers use fewer instances of nonstandard multiple negation (e.g. *I don't want none*) than lower-class speakers. Allowing for social class, however, women on average use fewer such forms than men do.”

El segon principi de Labov (2001: 274; Trudgill, 2000: 77-78) diu que en canvis lingüístics *des de dalt* (descrits al subapartat **2.2.4.1.3**), les dones adopten formes de prestigi més ràpidament que els homes. Turell (1995b: 285) usa com a exemple d'aquest principi el canvi observat a Alturo i Turell (1990) de [tʃ] per [ʒ] al català del Pont de Suert, “where women are the real innovators”.

El tercer principi labovià (Labov, 2001: 292) defensa que en canvis lingüístics *des de baix* (subapartat **2.2.4.1.3** també), les dones usen freqüències més elevades de formes innovadores que els homes. Un dels exemples de treballs que donen suport a aquest principi, segons Turell (1995b: 286) és el de Pradilla (1995), en què s'observen dos canvis en marxa a Benicarló que afecten la

realització de [dʒ], [iʒ] i [ʒ]. En aquest cas, la tendència a l'africament és majoritàriament afavorida per dones i correspon a la variant més innovadora (Turell, 1995b: 286).

El conjunt d'aquests tres principis s'ha denominat “el patró de gènere” (*the gender pattern*) i es pot sintetitzar en l'afirmació que les dones lideren els canvis lingüístics mentre que els homes solen usar més formes socialment concebudes com a no estàndards (Queen, 2013: 374).

2.3.1.3. Diferències entre els sexes en contextos interaccionals

Des d'una perspectiva complementària a la quantitativa, els estudis de les diferències entre els sexes en contextos interaccionals permeten observar els comportaments típics d'homes i dones. Tuson (1998: 214-216) fa un recull de les principals diferències que s'han detectat entre parlants dels dos sexes. Distingeix entre elements que engloba sota l'etiqueta de “prosòdia i elements paralingüístics”, “morfosintaxi”, “lèxic”, “organització temàtica textual” i “mecànica conversacional”. Per als objectius del nostre estudi, ens interessen especialment les diferències morfosintàctiques, lèxiques i interaccionals o pragmàtiques que reproduïm tot seguit en forma de Taula 6.

Cal subratllar que el comportament lingüístic descrit en la Taula 6 per a homes i per a dones és la suma de les observacions fetes a diversos estudis que analitzen mostres diferents i, per tant, contextos interaccionals i comunitats de parla també diferents. Així doncs, es tracta de tendències i no s'ha d'entendre aquest resum com a trets que necessàriament siguin presents en la mateixa mesura en tots els homes o totes les dones de qualsevol comunitat de parla analitzada, sinó com a indicadors del que sol distingir els dos grups segons els estudis disponibles.

Taula 6. Alguns dels trets més recurrents i identificadors per a homes i dones.

Nivell lingüístic	Dones	Homes
Morfosintaxi	2a persona i 1a persona del plural per incloure els interlocutors (modalitats elocutiva i apel·lativa)	1a persona del singular, 3a persona i formes impersonals (modalitat més delocutiva)
	Més freqüència d'oracions interrogatives i exclamatives	Més freqüència d'oracions enunciatives
	Més formes indirectes	Més enunciats directes
	Més modalitzadors (adjectius, adverbis, apreciatius, minimitzadors, etc.)	Menys ús de modalitzadors
	Més preguntes «eco»: (<i>no?</i> , <i>oi?</i> , etc.)	Menys preguntes «eco»
	Més oracions inacabades	Menys oracions inacabades
Lèxic	Més paraules que designen matisos	Menys paraules que designen matisos
	En algunes cultures, paraules tabú només per a dones	En les cultures observades, absència de paraules tabú només per a homes
	En algunes cultures existeixen paraules tabú per a homes i dones	
Mecànica conversacional	Més solapaments i interrupcions cooperatius (per manifestar comprensió, completar la intervenció anterior, etc.)	Més solapaments i interrupcions competitius (per aconseguir espai per parlar, manifestar desacord, desautoritzar, etc.)
	Més «feina» per mantenir la conversa (preguntes, exclamacions, «ajudes» temàtiques, etc.)	Menys «feina» per mantenir la conversa

Font: Elaboració pròpia a partir de Tuson (1998).

D'altres estudis assenyalen que els homes centren els seus esforços a transmetre informació mentre que les dones fan un ús de la llengua més aviat orientat a la interacció social i a expressar els seus sentiments (Kring i Gordon, 1998; Goldschmidt i Weller, 2000; Baron i Ling, 2011; Bastida Serra, 2017: 93). Aquestes tendències concorden amb les que recull Tuson, que les resumeix així: “Els comparteixen idees i activitats, elles comparteixen sentiments i intimitat” (1998: 214).

Tot i que aquestes diferències s'han observat en diversos estudis, cal tenir present certes consideracions. En primer lloc, no podem ignorar que la variació individual és present, en un grau més o menys important, dins de cada grup de parlants analitzat. En segon lloc, cal que recordem que no tots els trets que acabem d'enumerar es presenten en la mateixa freqüència en una mostra determinada, sinó que sovint es treballa amb mitjanes per poder arribar a conclusions. En tercer lloc, cal tenir en compte que la variació no és unicausal, sinó que es deu a la interacció entre diverses variables. Així, per exemple, diversos grups d'edat poden mostrar comportaments lingüístics que s'apropin o s'allunyin en un grau diferent a aquesta caracterització estereotípica dels sexes. Tuson (1998: 216) indica, per exemple, que els adolescents poden distanciar-se substantivament d'aquests estils conversacionals estereotípics. També cal tenir en compte els canvis socials que afecten les comunitats de parla i, per tant, les seves eleccions lingüístiques.

2.3.1.4. Explicacions i crítiques a l'anàlisi del sexe i llengua

A més de la detecció de diferències lingüístiques entre homes i dones, bona part dels estudis de la variació han tractat les possibles causes de les troballes variacionistes relacionades amb el sexe dels parlants.

Per exemple, Trudgill (2000: 73-74) proposa diverses explicacions de la troballa recurrent en estudis de la variació que les dones usen en general més formes considerades correctes que els homes a diferents parts del món i apunta a la confluència de diversos factors. Concretament, l'autor suggereix que els rols assignats tradicionalment a homes i dones, així com certs valors socials podrien condicionar l'actuació lingüística d'uns i altres en aquesta direcció:

“[...] it has also been pointed out that many societies seem to expect a higher level of adherence to social norms - better behaviour - from women than they do from men. [...] A woman interviewed in a Norwegian dialect survey said, when asked why she used the prestige pronunciation [ɛg] 'egg' while her brothers said [æg]: 'It isn't *done* for a woman to say [æg].’” (Trudgill, 2000: 73-74).

Una altra noció que s'ha proposat per explicar el comportament lingüístic dels parlants és el de prestigi encobert (*covert prestige*). Aquesta noció explicaria per què els parlants continuarien usant variants que en principi tindrien connotacions socials poc desitjables (per exemple, lligades a la pertinença a classes socials menys afavorides; Trudgill, 2000: 74). El prestigi encobert (és a dir, no explícit) d'aquestes variants serviria per expressar solidaritat amb grups socials i en certa manera, per tant, per contribuir a la construcció de la identitat dels parlants mitjançant els seus usos lingüístics.

La recerca contemporània sovint ha criticat un gran nombre d'investigacions per la seva concepció massa simplificada d'aquesta variable i dels usos lingüístics dels participants. Podem distingir dues menes de crítiques en aquests treballs més recents.

D'una banda, hi ha la diferència entre sexe i gènere. Bona part de la recerca contemporània estableix la distinció entre la variable sociolingüística sexe, com a categoria biològica, i la variable gènere, com a constructe social (Milroy i Gordon, 2003: 100; Coates, 2004: 4). Aquestes dues etiquetes han originat i segueixen sent objecte de llargues disputes teòriques en l'àmbit de la sociolingüística (vegeu, per exemple, la crítica de McElhinny, 2008: 23 o el repàs del debat entre acadèmics que fa Serrano, 2008). En aquesta tesi, ens cenyim a la perspectiva de la variable sexe com a variable independent binària, sobretot per qüestions pràctiques com les que descriu Precht (2008: 91):

“Gender research is, by and large, high context. Large-scale corpus research is low context and therefore precludes the study of gender, which requires a researcher to search by hand for masculine and feminine characteristics and to take contextual elements into consideration in making claims.”

D'altra banda, el patró de gènere labovian tractat fins ara no és l'únic que s'ha proposat en relació amb la variable sexe. Un dels grans retrets als primers estudis de la variació que contemplaven el sexe dels parlants és el fet que adoptessin una visió massa superficial d'aquesta variable i deslligada del context d'interacció.

Aquesta visió pot conduir, en paraules de Serrano, “a realizar afirmaciones estereotipadas que correlacionan el habla de las mujeres con unos mayores niveles de corrección y cuidado [que el de los hombres] (Cameron 1993: 3)”, per la qual cosa cal que els investigadors tinguin en compte les complexes interaccions entre diverses variables (vegeu el subapartat **2.3.5**) i el rol del context situacional en la interpretació dels resultats, així com en el disseny de la mostra que utilitzaran (Serrano, 2008: 183).

2.3.1.5. L'evolució de l'estudi de la variable sexe

Durant la tercera onada sociolingüística, com ja hem vist, s'entén que cada parlant construeix la seva identitat en les interaccions en què pren part amb l'ajuda dels seus interlocutors. Així doncs, a diferència dels primers estudis variacionistes, els més recents incorporen clares influències constructivistes en la seva perspectiva de les relacions entre, d'una banda, variables socials (com el sexe) i variables lingüístiques i, de l'altra, les identitats dels parlants.

A aquest canvi de mentalitat de molts sociolingüistes cal afegir-hi, com hem vist al subapartat anterior (**2.3.1.4**), la preferència de la variable gènere per sobre de la variable sexe. La primera s'entén en aquests estudis com una “dimensión sociocultural adquirida” (Serrano, 2008: 177) o “los patrones de socialización y de interacción social diferenciados según sexo” (Silva-Corvalán, 2001: 100). Tot plegat ha contribuït que les explicacions proposades als estudis de la primera onada, que veien els comportaments lingüístics com a conseqüència de característiques estàtiques i predefinides dels parlants, s'hagin anat qüestionant i, fins a cert punt, substituint per explicacions més elaborades.

Ara bé, tot i que la inclusió de la variable gènere pot aportar informació molt rellevant en estudis de caire qualitatiu, per comprendre la utilitat de considerar la variable sexe en estudis quantitius amb grans volums de dades, com en el nostre cas, cal no oblidar que:

“La división simple, dicotómica, de la población en hombres y mujeres ahorra al sociolingüista una operación complicada e ineludible en el resto de las variables: la conversión de modalidades en clases, es decir la organización de los distintos y numerosos intervalos en los que aparece dividida la variable en el censo en

grupos metodológicamente operativos y expresivos.” (Larrosa Barbero, 2003-2004: 149).

En aquest sentit, a més, si bé hi ha un percentatge de la població (encara difícil de precisar per falta de dades) que no es pot classificar com a home ni com a dona o que s'identifica amb un gènere diferent al que se li va assignar al naixement, aquest no és el cas de la majoria de la població, que sí que s'identifica amb una d'aquestes dues categories, que es correspon a la que se li va assignar al naixement (Goodman *et al.*, 2019).

2.3.1.6. La variable sexe: síntesi

Com hem vist, per tant, la bibliografia indica que seria possible trobar diferències en el comportament lingüístic d'homes i dones. A la Taula 4, hem recuperat la llista de troballes que resumeix Tuson (1998: 214-216) i hem avançat que les possibles diferències que es relacionen més directament amb els nostres objectius són les que corresponen als nivells morfosintàctic, lèxic i pragmàtic.

Quant al primer, les tendències observades en altres treballs indiquen que les dones utilitzen més formes gramaticals que inclouen els interlocutors (2a del singular i 1a del plural) en comparació amb els homes, que usen més la 3a persona i formes impersonals, i que les dones usen més modalitzadors, preguntes, exclamacions i oracions inacabades que els homes.

Pel que fa al lèxic, s'ha dit que les dones fan més ús de paraules que designen matisos i s'han observat paraules tabú per a dones i homes i exclusivament per a dones, però no exclusivament per a homes.

Finalment, quant a estratègies pragmàtiques, l'esperable, segons la bibliografia, és que les dones usin més recursos cooperatius i més esforços lingüístics per mantenir la conversa, mentre que els homes farien més ús d'estratègies competitives i menys orientades a continuar la interacció.

En aquest apartat també hem explicat que entenem la variable sexe com a categoria binària i la variable gènere com a constructe social. Com hem descrit, en aquest estudi ens centrem en la variable sexe, i no en el gènere dels parlants,

per motius pràctics lligats al disseny metodològic, però també a les característiques del corpus observat (vegeu apartat 5.3.2.2.1).

Així doncs, caldrà veure si els resultats del nostre estudi confirmen o desmenteixen les troballes d'estudis previs relacionades amb la variable sexe pel que fa a la nostra mostra d'adolescents catalanoparlants. A continuació, fem un repàs d'estudis que analitzen l'ús de la llengua en relació amb l'edat.

2.3.2. La variable edat: la llengua dels adolescents

Després d'ocupar-nos de la rellevància de la variable independent sexe dels parlants per als estudis sociolingüístics i la nostra recerca, ens centrem ara en la segona variable extralingüística més usada habitualment, l'edat dels informants. Primer, introduïm alguns dels conceptes clau en la sociolingüística que ha observat la relació entre llengua i edat, com *gradació per edats* i la diferència entre *estudis rèplica* i *estudis panel*. Tot seguit, ens centrem en els treballs que han investigat el procés d'adquisició lingüística en l'etapa (pre)adolescent i, finalment, resumim algunes de les crítiques més freqüents respecte el tractament d'aquesta variable en estudis sociolingüístics.

2.3.2.1. La importància de la variable edat

La variable edat és una de les que més s'ha estudiat en la sociolingüística i el variacionisme. En paraules de Blas Arroyo (1994: 190):

“Las diferencias generacionales y su impacto en la variación lingüística han sido puestas de relieve desde antiguo en comunidades de habla muy diferentes. Por lo que se refiere a la sociolingüística, una de las ideas más recurrentes en la bibliografía es que la edad representa un factor que puede condicionar la variación en un grado incluso mayor a como lo hacen otros parámetros sociales tan relevantes como el sexo o la clase social.”

A més, els canvis sociolingüístics a l'adolescència són destacables a les societats monolingües, ja que diversos estudis han assenyalat que aquest grup etari hi lidera processos de canvi, però encara més en societats multilingües com la nostra (vegeu 2.3.4), en les quals s'ha identificat els adolescents com “key agents of language shift” (Vila, Ubalde, Bretxa i Comajoan-Colomé, 2020: 1160).

Un dels conceptes més rellevants a través dels estudis de la variació que tenen en compte l'edat és el de gradació per edats (Turell, 2003: 7; en anglès, *age grading*). Si bé a l'inici aquest terme feia referència als usos lingüístics que són propis dels infants però no se solen trobar en altres grups generacionals, a l'actualitat s'utilitza per parlar de trets que caracteritzen els diferents grups. Blas Arroyo (1994: 193) el defineix com “[...] las diferencias en el lenguaje que son específicas de las diversas edades en la vida de los individuos y que contribuyen a singularizarlas desde un punto de vista sociolingüístico (Romaine 1984b: 761)”. Aquestes diferències poden deure's al fet que només un determinat grup d'edat usi una variable lingüística o bé a què aquest cert grup d'edat l'usi de forma diferent a la resta de la població (per exemple, de manera molt més freqüent).

Com explica Eckert (1997: 153), si estudiem mostres de grups generacionals equivalents en dos moments històrics i hi observem diferències, podem concloure que es tracta d'un canvi a temps real (vegeu el subapartat **2.2.4.1.3**), mentre que si observem canvis en el comportament lingüístic d'un grup en dos moments diferents, ens trobem davant de diferències atribuïbles a la gradació per edats, és a dir, a l'evolució que pateixen els parlants al llarg de la seva vida. El primer enfocament és l'adoptat en estudis rèplica (*trend studies*, en anglès) i el segon és el dels estudis panel (Turell, 2003). Com veurem més endavant (vegeu l'apartat **5.3**), el nostre treball s'assembla més al segon tipus d'estudi, ja que observem els mateixos informants en dos moments temporals, de manera que, en cas de detectar diferències en el seu comportament lingüístic, caldrà veure si es poden atribuir a la gradació per edats.

2.3.2.2. L'adquisició lingüística durant l'adolescència

Actualment, sabem que l'adquisició de les llengües inicials s'estén durant la infantesa i a través de l'adolescència i que, de fet, mai acaba, ja que tots els parlants segueixen incorporant vocabulari i noves formes d'usar la llengua al llarg de tota la seva vida (Nippold, 1998: 1; Singleton i Ryan, 2004: 57). Els estudis sobre els processos d'adquisició lingüística que ocorren a l'adolescència ens permeten detectar possibles variables lingüístiques per al nostre estudi que

puguin mostrar diferències en els informants de 12 i 16 anys. És per això que tot seguit repassem els resultats d'aquesta mena d'estudis, dividits per nivells lingüístics. Cal assenyalar que la majoria han observat adolescents de parla anglesa o espanyola.

Quant a les habilitats lèxiques, s'ha trobat que durant l'adolescència es desenvolupen la capacitat de descriure substantius abstractes i complexos, frases fetes i proverbis (Brasseur i Jimenez, 1989; McGhee-Bidlack, 1991; Nippold i Rudzinski, 1993; Nippold *et al.*, 1997, citats a Singleton i Ryan, 2004: 55), la capacitat de recuperar paraules ràpidament i amb precisió, el raonament verbal complex i la comprensió avançada de llenguatge figurat (Nippold, 2004: 2-3).

A més, segons McLaughlin (1978: 55 i s., citat a Singleton i Ryan, 2004: 58), la fluïdesa lèxica arriba al 80% de la competència adulta als 20 anys d'edat. Durant l'etapa adolescent també es desenvolupa el lèxic lletrat (*literate lexicon*), “a mental dictionary of thousands of complex and low frequency words, coexisting in an elaborate semantic network” (Nippold, 2004: 2), caracteritzat, a diferència del lèxic d'infants més joves, per la incorporació de substantius abstractes (Nippold, 1998: 14; 2004: 2), termes polisèmics (Nippold, 1998: 21-23), adverbis de probabilitat i magnitud (Nippold, 1998: 25-27), verbs metalingüístics i metacognitius (com *afirmar* o *concloure*; Nippold, 1998: 27-28), i verbs fàctics (que pressuposen la veritat de la proposició que els segueix, com *saber* o *adonar-se*) i no fàctics (amb els quals “the truth of the complement clause is uncertain”, com amb *pensar* o *creure*; Nippold, 1998: 28-29).

Segons Nippold (1998: 29), part de l'adquisició lèxica de l'adolescència es basa a millorar la comprensió de paraules que no necessàriament es troben per primera vegada:

“Developmental studies have shown that the subtle meanings of literate words are acquired gradually during the school-age and adolescent years. Although preschool children may use many of the words in their spontaneous speech (e.g., *above, below, warm, cold, know, think, believe, etc.*), a full understanding of the words may not be reached until late adolescence or early adulthood and is closely tied to educational experience and performance.”

Pel que fa al desenvolupament sintàctic, les estructures adquirides pels adolescents anglòfons inclouen la veu passiva, les oracions subordinades i de relatiu, el plusquamperfet, els modals o conjuncions adverbials poc freqüents, entre altres (Nippold, 2004: 3). A mesura que els adolescents incorporen cadascun d'aquests recursos a les seves produccions, la llargada d'aquestes augmenta de manera gradual. Paral·lelament, també adquireixen l'habilitat d'usar aquestes estructures de manera flexible i el coneixement necessari sobre quan recórrer a estructures pertanyents a llenguatge més o menys formal segons la situació comunicativa (Nippold, 2004: 3).

Singleton i Ryan (2004: 60) llisten diverses característiques sintàctiques pròpies de la llengua anglesa dels adolescents:

“Scott (1988) and Scott and Stokes (1995) found that as students mature noun phrases are expanded through the use of appositives, elaborated subjects and post-modification via prepositional phrases, non-finite verbs and relative clauses, while verb phrases are expanded through the use of structures such as modal auxiliaries, perfect aspect and the passive voice. [...] Through the use of low-frequency structures, the number of words per clause increases, along with density of information expressed. Sentence length also increases through the use of co-ordinating, subordinating and correlative conjunctions and use of adverbial conjuncts. However, it is worth noting that sentence length can vary greatly in relation to discourse mode and the context of the conversation.”

També s'han estudiat les capacitats pragmàtiques que es desenvolupen en aquesta etapa. Khorounjaia i Tolchinsky (2004), per exemple, exploren l'estructura argumental d'infants, adolescents i adults en castellà. Entre les seves conclusions, hi inclouen que: “Spanish-speaking children and adolescents have a preferred argument structure that is identical to that of Spanish-speaking adults” (2004: 104) i també que “[a] major feature of later language development is an increasing sensitivity to the demands of different communicative circumstances and interlocutor needs” (2004: 108).

Tots aquests estudis mostren que, si bé l'adquisició lingüística dels primers anys pot resultar més espectacular per la seva rapidesa, el procés segueix durant l'adolescència. Així doncs, un estudi longitudinal com el que ens ocupa, en què es comparen produccions dels mateixos parlants en la mateixa situació comunicativa passats uns anys, permetrà analitzar l'adquisició del català a l'adolescència i comparar les nostres troballes amb les que acabem de veure pel que fa a l'adquisició de l'anglès i el castellà.

Els estudis disponibles sobre varietats de llengua catalana utilitzades per parlants adolescents són escassos (com també fan notar Bernal i Sinner, 2009 o Santamaria, 2010). Quant a l'adquisició del català en etapes infantils i adolescents, destaca Tolchinsky, Martí i Llauradó (2010) sobre el creixement del lèxic escrit en infants d'entre 5 i 16 anys. En aquest estudi es troba un increment notable en el vocabulari dels informants al llarg de l'escolarització obligatòria, tant des d'un punt de vista conceptual com d'actuació, i també constaten que "the written lexicon is rather impermeable to multilingual interference" (2010: 226).

Pel que fa a estudis descriptius de les varietats juvenils del català, hem identificat dues grans àrees d'interès. La primera comprèn les característiques del lèxic dels joves. En aquesta àrea hi trobem, entre altres, el treball de Bernal i Sinner (2009) en què descriuen l'ús de diversos recursos de creació de paraules (vegeu l'apartat 4.4).

La segona àrea d'interès detectada és la versemblança de la representació del llenguatge juvenil en produccions audiovisuals, representada per Wieland (2009a i 2009b) o Santamaria (2010). Una de les principals preocupacions d'aquests esforços, però, és com tractar la presència d'elements lèxics clarament influenciats pel contacte amb el castellà i que entren en conflicte amb les normes prescriptives del català, tema que no coincideix amb els nostres objectius, ja que la nostra perspectiva sobre les produccions orals que analitzem és descriptiva. Tot i això, cal dir que ens ocupem de les particularitats del context demolingüístic del territori català i les conseqüències del contacte amb el castellà sobre el lèxic emprat pels joves als apartats 4.1 i 4.4.

2.3.2.3. Explicacions de la variació lingüística relacionada amb l'edat

Així doncs, els estudis sobre l'adquisició, encara que majoritàriament observen parlants d'altres llengües, indiquen que en l'etapa adolescent es produeixen canvis en els nivells lèxic, sintàctic i pragmàtic.

Un altre aspecte que cal tenir en compte és que la variació lingüística observada en parlants de grups d'edats diferents ha suscitat explicacions diferents a la bibliografia consultada. Com assenyalen autors com Eckert (1997: 157) i Blas Arroyo (1994: 191), moltes de les anàlisis de la parla dels infants s'han centrat en processos de socialització, mentre que les observacions de l'ús de la llengua que fan els joves han relacionat el comportament lingüístic observat amb l'aprenentatge de rols típicament desenvolupats per adults. Les anàlisis dels usos dels parlants de més edat han girat al voltant de la seva pèrdua d'habilitats pròpies de l'edat adulta.

Per tant, en la majoria de casos, la perspectiva dels investigadors ha pres com a punt de partida l'actuació lingüística i social dels adults per entendre la d'altres grups generacionals. Eckert (1997: 157), en canvi, proposa adoptar una perspectiva combinada que permeti reconèixer el caràcter continuat del desenvolupament lingüístic al llarg de la vida (com hem destacat a l'inici de l'apartat **2.3.2.2**), així com les propietats específiques de la competència sociolingüística de cada etapa vital o grup d'edat.

2.3.2.4. La variable edat: síntesi

En resum, tot i que en general hi ha pocs estudis sobre l'adquisició lingüística a l'etapa adolescent, comparats amb els treballs sobre etapes anteriors del desenvolupament lingüístic, existeixen estudis que indiquen que durant l'adolescència es produeixen canvis en les capacitats lèxiques, sintàctiques i pragmàtiques dels parlants. A més, hem trobat treballs anteriors que tracten el desenvolupament d'adolescents de parla catalana, sobretot pel que fa al nivell lèxic i a la influència del castellà sobre el vocabulari.

2.3.3. Les variables classe social i nivell educatiu

Una altra de les variables més recurrents en estudis sobre la variació lingüística és la classe social. A final del segle passat, Milroy i Milroy (1997: 55) afirmaven que “[s]ocial class has been by far the most widely used social variable” d’ençà de l’estudi de Labov (1966) a Nova York. A més, altres autors, com Verdoodt (1997: 38), subratllen que és necessari que els estudis sobre la distribució social de varietats i variables lingüístiques tinguin en compte el desenvolupament socioeconòmic, incloent-hi les classes socials i els espais geogràfics a què pertanyen els parlants. Nombrosos estudis sobre diverses comunitats de parla (per exemple, Trudgill, 1974, sobre Norwich; Macaulay, 1978, sobre Glasgow; Gregersen i Pedersen, 1991, sobre Copenhaguen) han identificat variables lingüístiques estratificades de forma estable a través de classes socials i contextos estilístics diferents, la qual cosa demostra que la variació lingüística a grans ciutats no és caòtica, sinó explicable (Labov, 2004: 9).

Malgrat l’ús reiterat i marcadament productiu de la variable classe social en aquests estudis, com explica Ash (2013: 350), no hi ha consens entre els acadèmics sobre exactament quins factors específics i quantificables formen la classe social, sinó que se sol definir de forma més aviat abstracta. De fet, s’hi ha fet referència mitjançant diversos termes. Per exemple, en un mateix manual (*The Routledge Handbook of Sociolinguistics Around the World*, editat per Martin J. Ball) trobem que Collins utilitza *estatus socioeconòmic* (2010: 151), Mugler i Benton (2010: 174) empen *classe social*, mentre que Di Vito parla d'*estatus social* (2010: 313).

A més, per a l’estudi de la variació lingüística, habitualment no s’aprofiten aproximacions sociològiques al constructe que ens ocupa (Ash, 2013: 350). No obstant això, aquesta autora proposa que “it may be appropriate to accept this state of affairs without excessive lamentation”, ja que els marcs teòrics i metodològics desenvolupats en altres àrees científiques per comprendre el funcionament de les classes socials no són necessàriament aplicables a l’estudi de la variació lingüística, ja que “the object of inquiry of these studies is the

dynamics of linguistic variation and change, not the study of social class itself” (Ash, 2013: 364-365).

Ara bé, malgrat que l'aproximació quantitativa a la classe social és productiva i permet obtenir resultats rellevants per explicar la variació lingüística de les societats, Milroy i Milroy (1997: 54) la descriuen com la variable social més controvertida i Wardhaugh (2006: 148) en diu que és “the most complicated factor of all”. Els primers estudis quantitius que consideraven aquesta variable distingien els usos lingüístics dels diversos grups de forma categòrica (per exemple, per la presència o l'absència de certa variant lingüística en els grups d'informants), però, en canvi, com assenyalen Milroy i Milroy (1997: 49), “[i]n practice such usages are seldom categorical for any group of speakers”. En aquest sentit, Labov (2004: 8) explica que, mentre que en un gran nombre d'estudis es representa la classe social com una variable binària, formada per dues categories (per exemple, classe alta / classe treballadora), es pot arribar a un nivell de certesa més elevat sobre la relació entre aquesta variable i la dependent quan es consideren quatre o cinc categories en lloc de només dues.

Amb tot, la divisió de la societat en categories de classes socials no pot sinó ser una decisió metodològica fins a cert punt arbitrària, ja que, en realitat, com passa amb altres variables socials, es tracta d'un contínuum. De la mateixa manera, les varietats dialectals i funcionals també són contínuums segmentats en categories en els estudis sobre la variació. És per això que, segons Durrell (2004: 204):

“Empirical studies support the view that few, if any, linguistic varieties can be unambiguously marked as *regional dialect*, *sociolect* or *register*, since the boundaries between the dimensions of variation are not clear-cut, and, although some variables may be primarily indicative of a particular dimension, it is quite typical for a single variable to correlate with all three in some measure.”

Establir una categorització vàlida de la variable classe social és, doncs, un problema metodològic que va sorgir amb els primers estudis de la variació i segueix sense una solució clara (Durrell, 2004: 201; Ash, 2013: 364). Arran de la crítica del paradigma labovià pel fet que la forma amb què considera la classe

social pot amagar diferències individuals i l'efecte d'altres factors, com el nivell educatiu, Milroy (1987) va desenvolupar la variable de xarxes socials (Milroy i Milroy, 1997: 53-54; Durrell, 2004: 201). Tant aquesta variable com la classe social, però, són variables complexes, que s'han de calcular “by reference to a number of indicators” (Milroy i Milroy, 1997: 54). En el cas de la classe social, a més de la professió (Ash, 2013: 365), se solen tenir en compte els ingressos econòmics o el nivell educatiu. Tradicionalment, els individus amb un poder adquisitiu més elevat (les classes socials més altes) han tingut més possibilitats d'accedir al sistema educatiu i cursar estudis superiors. Sobre el paper que juguen factors com el nivell educatiu en la construcció de la variable classe social, Hudson (2000: 186-187) conclou que:

“Los datos sociolingüísticos, pues parecen sugerir que los factores como profesión y educación deberían ser anotados por separado, aunque dejando la posibilidad de que actúen entre sí, exactamente de la misma forma que pueden afectar a otros factores tales como edad y sexo”.

El nivell educatiu dels parlants és una de les variables utilitzades en la sociolingüística quantitativa (per exemple, López Morales, 1990: 81; Chacón Beltrán, 2004: 33) amb una tradició més establerta, ja que la dialectologia tradicional ja se'n servia (Trudgill, 2000: 27). Com respecte a la classe social, tampoc hi ha consens entre els autors en el nom de la variable nivell educatiu. Hi trobem referències mitjançant denominacions com, entre altres, *nivell d'instrucció* (Calderón Noguera i Durán Mendivelso, 2009), *nivell educacional* (Bobadilla i Bobadilla, 1980-1981), *assoliment educatiu* (Ash, 2013: 353), *escolaritat* o *educació* (López Morales, 1990: 81). A més, com explica Wardhaugh (2006: 148), es pot conceptualitzar d'acord a diverses categories: “graduate or professional education; college or university degree; attendance at college or university but no degree; high school graduation; some high school education; and less than seven years of formal education”.

En qualsevol cas, en un nombre considerable d'estudis, s'ha trobat que el nivell educatiu aporta informació rellevant per comprendre fenòmens de variació lingüística (per exemple, en estudis sobre el català com Romero, 2001 i 2003 o

Suïls, Ariño, Alturo i Turell, 2009). Això es deu al fet que una de les conseqüències de l'escolarització és un augment de la competència comunicativa dels parlants, ja que alguns dels objectius principals del sistema educatiu són precisament assegurar que els alumnes adquireixen coneixements lingüístics de les varietats considerades estàndards i considerades adequades a diverses situacions comunicatives, fomentar la seva consciència metalingüística i, en paraules de Trudgill (2000: 82) “introduce pupils to the registers of particular subjects”, de manera que dominin el vocabulari propi de les diverses matèries que treballen als centres educatius. En definitiva, el que Hudson (2000: 240) resumeix com “ensanchar” la competència comunicativa amb què els parlants accedeixen al sistema educatiu.

En relació amb la competència comunicativa als sistemes educatius, cal assenyalar que un bon nombre d'estudis sobre la variació lingüística han tractat desigualtats en l'assoliment educatiu de l'alumnat d'orígens ètnics i/o classes socials diferents. Aquests treballs han demostrat que els parlants les varietats dels quals són més allunyades de les considerades estàndards i utilitzades en les institucions educatives poden tenir més dificultats que els seus companys per assolir els mateixos nivells educatius i experimentar les conseqüències de diversos prejudicis lingüístics vers les seves varietats vernaculars (vegeu, per exemple, Trudgill, 2000: 82; Hudson, 2000: 240). Altres estudis han demostrat que l'accés a l'alfabetització pot condicionar processos de canvi lingüístic (Bright, 1997: 89).

2.3.4. Llengua inicial i varietat dialectal

Fins ara, hem tractat la rellevància per als estudis sociolingüístics de les variables independents sexe, edat, classe social i nivell educatiu. En aquest apartat, fem una primera aproximació a les particularitats del context sociolingüístic català (vegeu el capítol 4 per una descripció més detallada del context en què es desenvolupa la nostra recerca) relacionades amb la rellevància de les variables llengua inicial i varietat dialectal, així com la bibliografia consultada que s'ocupa dels usos lingüístics.

Anteriorment, hem assenyalat el predomini d'informants de parla anglesa en els estudis de l'adquisició lingüística dels adolescents (subapartat **2.3.2.2**). Aquesta tendència es pot observar en altres parcel·les de la lingüística i de la sociolingüística i suposa una comprensió esbiaixada dels fenòmens analitzats. La situació ideal seria poder comptar amb més estudis d'altres comunitats de parla que han estat comparativament poc estudiades o sobre les quals encara no tenim dades. Un indicador d'aquest desequilibri són les anàlisis quantitatives de les publicacions *Language Variation and Change* (entre 1989 i 2015) i *Journal of Sociolinguistics* (entre 1997-2008), que mostren una clara manca de diversitat lingüística en els objectes d'estudi dels treballs publicats en aquestes revistes de referència en favor de comunitats de parla americanes i europees i de parla anglesa (Nagy i Meyerhoff, 2008: 7-10; Stanford, 2016).

Això no obstant, el desequilibri en les comunitats estudiades no és l'únic problema que presenta el fet que gran part dels sociolingüistes hagin centrat els seus esforços en societats de parla anglesa. Una altra conseqüència que se'n desprèn és el predomini d'una visió monolingüe dels fenòmens sociolingüístics, un perill del qual diversos autors han advertit al llarg dels anys (per exemple, Crystal, 1987: 360; Ellis, 2006: 174; Nagy i Meyerhoff, 2008: 2, o Stanford, 2016). Part de la recerca variacionista, sobretot de tradició anglosaxona, ha tendit a observar els parlants com a representants de comunitats de parla homogènies, malgrat que la major part del món està format per individus i comunitats de parla multilingües (Nagy i Meyerhoff, 2008: 1). El monolingüisme s'ha convertit massa sovint en el cas considerat no marcat de manera inqüestionada. Aquest fet té dues explicacions fonamentals, segons Ellis (2006: 189):

“One explanation is that it is a consequence of the rise of the European nation state, later exported to their colonies, and another is that powerful English-speaking nations are both the producers and beneficiaries of English as a global language, and they tend to be monolingual”.

Aquest biaix monolingüe de la recerca comporta nombrosos desavantatges, dels quals en destaquem tres. El primer és que considerar per defecte que els parlants analitzats són monolingües comporta ignorar el fet que el conjunt de recursos lingüístics que utilitzen en les seves interaccions provenen, en realitat, d'un ventall de llengües (Nagy i Meyerhoff, 2008: 1). El segon és que “[t]o consider single languages in isolation in scholarly work when they don't exist that way in their social lives is to reify idealisations about the discreteness of language systems and the norm of monolingualism” (Nagy i Meyerhoff, 2008: 10). I el darrer és que, en els estats occidentals més poderosos, els discursos monolingües tendeixen a dominar els debats sobre polítiques educatives i socials, que tenen repercussions directes sobre els plans docents de l'escolarització obligatòria o els protocols que se segueixen a l'hora de determinar l'origen dels sol·licitants d'asil, per exemple (Eades *et al.*, 2003; Ellis, 2006: 185).

En contraposició, augmentar el volum de recerca disponible que tingui en compte el multilingüisme dels parlants i de les comunitats de parla proporciona diversos beneficis i oportunitats de fer avançar el coneixement sobre els sistemes lingüístics. En aquest sentit, Nagy i Meyerhoff (2008: 3) indiquen que estudiar comunitats de parla multilingües ens permet comprendre processos socials i lingüístics rellevants per a la sociolingüística més àmplia. A més, les mateixes autores afirmen que “sociolinguistically interesting situations of multilingualism often involve contact and alternation between internationally dominant languages and socially subordinate ones” (2008: 6). Així doncs, en resum, si volem assolir una comprensió plena sobre els processos lingüístics, caldrà que puguem comptar amb estudis de societats complexes des d'un punt de vista lingüístic, ja que el multilingüisme és el més comú en moltes parts del món (Stanford, 2016: 531).

Des d'aquesta perspectiva, la situació sociolingüística de Catalunya es pot considerar un escenari idoni per dur a terme investigacions que contribueixin a revertir el biaix monolingüe de la recerca que acabem de descriure. Des d'una perspectiva demolingüística, les transformacions que han tingut un impacte més destacat en el coneixement i l'ús del català s'han produït en els darrers dos segles

(Vila, 2020). En especial, cal destacar la profunditat dels canvis socials de la segona meitat del segle passat. Després de la Guerra Civil, Catalunya (com la resta de territoris de llengua catalana) va deixar de ser etnolingüísticament homogènia¹⁶ i va passar a estar formada per dos grans col·lectius per l'arribada de milers de ciutadans del sud d'Espanya (Vila, 2013: 11; Larrea Mendizabal, 2016: 9-10).

Després de la tornada de la democràcia, aquestes migracions entre els territoris autonòmics va minvar dràsticament, però, cap als anys noranta del segle passat, es va produir un altre fenomen migratori que va reconfigurar la demografia de Catalunya. Aquesta vegada, els nous catalans provenien de molts països diferents, empesos per causes econòmiques i polítiques diverses. Per a Domingo, Vidal-Coso, i Serra (2012: 152), aquesta onada va representar un “boom migratori”. D'altra banda, Vila (2013: 12) resumeix aquesta nova situació així:

“Vinguts dels indrets més allunyats, al llarg de la primera dècada del segle XXI, els nouvinguts van anar transformant profundament el paisatge lingüístic i cultural dels carrers, les aules, els centres de treball i d'esbarjo i les institucions dels països de llengua catalana d'una manera i en unes proporcions que, en termes generals, ningú no hauria gosat vaticinar tan sols deu anys enrere”.

A més, l'autor explica que aquesta nova situació demolingüística va tenir repercussions en la recerca sociolingüística catalana, tant pel que fa a investigacions empíriques com a reflexions teòriques sobre llengua i societat (Vila, 2013: 14). De fet, fins i tot parla d'una nova etapa en la sociolingüística catalana a partir de la dècada del 2000, impulsada per aquesta nova immigració i la globalització, a la qual proposa anomenar l'etapa del multilingüisme, “en què la sociolingüística catalana s'ha vist obligada a reconcebre's partint de la base de la reflexió sobre la diversitat idiomàtica al món i a casa seva (Marí 2011)” (Vila, 2013: 15).

¹⁶ Entenent la homogeneïtat com un terme relatiu com puntualitza Vila (2020), ja que, si bé “[t]he Catalan Countries have been the natural corridor between Europe and Africa for millennia” (2020: 631), “in the Catalan-speaking territories until the 20th century, immigration never really led to the formation of separate linguistic communities” (2020: 632).

En aquesta etapa de la sociolingüística catalana, hi trobem diversos estudis sobre contextos escolars i la gestió del multilingüisme a les aules, com Vila (2004) o Purí (2006). Durant aquests anys també observem un canvi en la terminologia utilitzada en la legislació lingüística, que passa de tractar la *normalització* i la *planificació lingüística* a la *política lingüística* (Marí, 2011). Aquest és el context en què es concep i desenvolupa el projecte RESOL, iniciat el 2006, del qual prové la nostra mostra d'informants (vegeu el capítol 4) i que va recollir un gran volum d'informació, inclosa la referent a l'espai que ocupaven en la vida dels informants les diverses llengües del seu entorn (llengües inicials, llengües utilitzades amb els amics, llengües de consum cultural, etc.).

Una altra variable que ha resultat productiva en treballs sobre fenòmens de variació lingüística és la varietat dialectal. Com s'ha explicat anteriorment (al subapartat 2.2.1.1), una de les precursoras de la sociolingüística actual és l'anomenada dialectologia tradicional. L'objectiu d'aquesta disciplina era "l'estudi de l'evolució lingüística en el decurs del temps" i per aconseguir-lo, observava les varietats diatòpiques de la llengua: "[es] buscaven les petges i les regles d'aquesta evolució en les formes vives dels dialectes i en les característiques que els diferenciaven, que presentaven associades a la seva distribució geogràfica" (Viaplana, 2002: 149).

Ja s'ha mencionat també que els informants objectiu dels primers treballs dialectològics eren homes adults fortament arrelats a localitats rurals (subapartat 2.3.1.1). Els nivells lingüístics més explorats per la dialectologia i els estudis de variació diatòpica són el foneticofonològic (Trousedale, 2010: 44-45) i el lèxic (Veny i Massanell, 2015: 154), si bé també s'han interessat, en menor mesura, per les diferències regionals quant a aspectes morfològics i sintàctics (vegeu, entre altres, Kortmann i Szmrecsanyi, 2004).

Als territoris de parla catalana, l'anàlisi de la diversitat dialectal ha produït nombrosos i valiosos fruits (per exemple, Viaplana, 2002; Veny i Massanell, 2015) i gaudeix d'una llarga tradició (Rafanell, 1991). A més de treballs centrats en trets propis d'una zona dialectal determinada (com Gómez Duran, 2011, sobre el català rossellonès o Borràs-Comes, 2013, sobre el català nord-

occidental), comptem amb estudis descriptius sobre els parlars de transició (p. ex. Navarro, 1999) i fronterers (com Gargallo Gil, 2002). D'altra banda, cal destacar també el desenvolupament recent de la dialectometria aplicada a les varietats catalanes (Clua Julve, 1999; Valls, 2010, 2013; Valls i Clua, 2013; Perea, 2012; Fernández Planas *et al.*, 2015; Valls i González, 2016).

2.3.5. La interacció entre les variables i els biaixos en la recerca existent Com hem vist anteriorment (subapartat **2.2.4.2**), la sociolingüística de la variació actual es troba lluny de les primeres investigacions que explicaven comportaments lingüístics concrets a partir de factors socials estables i inherents als parlants. Així doncs, un nombre creixent d'estudis observen diverses variables extralingüístiques alhora (és a dir, aposten per enfocaments multivariables) per poder explicar de manera més encertada les realitzacions lingüístiques que fan els parlants a partir de les interaccions entre elles (com Moreno Fernández, 1998: 34-36 o Queen, 2013: 374).

Per exemple, Nichols (1986) troba que les diferències entre els usos lingüístics d'homes i dones a Carolina del Sud es poden explicar per patrons econòmics que estan indirectament relacionats amb el sexe biològic dels parlants (molts homes treballen en el turisme i, per això, mantenen l'ús del gullah, mentre que moltes dones treballen en sectors en què entren en contacte amb l'anglès més sovint i per això l'adopten en la seva parla).

A l'estudi de Paunonen (1994), dut a terme entre els 1970 i els 1990, les dones tendien a fer usos menys normatius a mesura que envellien, mentre que els homes mostraven la tendència contrària. Eckert (1997: 154) troba que aquesta mena d'interaccions entre el sexe i l'edat probablement sigui habitual, particularment, especula, "in times of change in gender norms and practices". En aquest sentit, Milroy i Milroy (1997: 55) expliquen que la variable classe social ha anat guanyant pes en la interpretació de les diferències observades en el comportament lingüístic entre els sexes i que, per exemple, el fet que les dones facin usos més normatius que els homes s'ha interpretat en alguns estudis "as arising from a desire on their part to acquire social prestige through their speech,

as they could not traditionally acquire this through career success – as males could”, si bé aquest argument ha estat objecte de diverses crítiques.

I és que els resultats vinculats a variables extralingüístiques també poden interactuar amb dinàmiques socials entre els parlants. Sobre aquesta observació, és rellevant el concepte d’afiliació a grups socials emprat en estudis de Cheshire (1987) o Eckert (2000), entre altres. Per exemple, la variable sexe pot arribar a ser tan sols intermediària entre aquestes dinàmiques grupals dels parlants i el seu comportament lingüístic, com passa a l’estudi de Cheshire (1987), en què s’observen produccions de grups masculins i femenins d’alumnes en edat escolar. En paraules de Queen (2013: 375):

“Cheshire argues that it is not specifically gender that explains her findings but the structure of the students’ peer groups and individual students’ orientation toward the norms of those groups that were the primary predictors of linguistic behavior.”

Pel que fa a la variable edat, un dels factors que hi poden interaccionar és la pertinença dels parlants a diverses cohorts o grups generacionals (*cohort membership*). A més, cal destacar que no sempre és fàcil distingir aquestes cohorts de l’edat dels parlants i que aquestes dues nocions poden solapar-se. Així, doncs, s’ha demostrat reiteradament que els enfocaments multivariables poden explicar comportaments lingüístics complexos millor que els bivariables. No obstant això, abans de poder fer els primers, va caldre establir una sòlida base empírica d’estudis que demostraven diverses relacions bivariables com les que s’exploren en el nostre estudi (vegeu el capítol següent), per al qual no hi ha precedents en el nostre àmbit lingüístic.

Altres factors que poden afectar els resultats per edat són l’existència de canvis socials en marxa o, fins i tot, la comunitat que s’estudia, ja que les diverses condicions culturals i materials poden afectar les trajectòries vitals dels parlants de manera diferent. És en aquest sentit, per exemple, que Eckert (1997: 167) afirma que “the overstudied English-speaking white middle-class child dominates our understanding of development”. Aquest fet s’ha de tenir present en analitzar

els usos lingüístics espontanis de parlants reals que representen societats cada vegada més diverses culturalment.

D'una banda, com que la majoria d'estudis variacionistes s'han centrat en comunitats de parla britàniques o americanes, sobretot als inicis d'aquest corrent, cal considerar la possibilitat que les tendències observades no siguin universals, malgrat la insistència d'alguns experts, com Turell (1995b: 284), en el fet que els principis labovians es basen en anàlisis de comunitats de parla d'arreu del món (cf. Serrano, 2008: 180-181). De l'altra, com indica Trudgill (2000: 73), l'accés a l'educació o el contacte amb altres comunitats varia entre regions diferents. Per exemple, en comunitats en les quals les dones no tinguin accés a l'educació, podria ser que no utilitzessin, com proposa el segon principi de Labov (subapartat 2.3.1.2), més formes considerades estàndards que els homes.

A més del nivell educatiu, el sexe pot interaccionar també amb la classe social. Feagin (2013: 28) explica que les classes socials altes tendeixen a utilitzar varietats locals de l'estàndard, mentre que les classes treballadores (i, especialment, els parlants de més edat) solen mantenir vives formes considerades no estàndard i que en altres varietats han quedat obsoletes. Labov (2004) exemplifica les possibles relacions entre el sexe i la classe social mitjançant l'estudi de Trudgill (1974), una de les troballes del qual sobre la variació en el morfema *ing* a Norwich va ser que "it is the women in the second highest status group who show an exceptionally steep slope of style shifting, quite different from any other group" (Labov, 2004: 10).

Ara bé, alguns autors apunten que els principis generals sobre variables com el sexe en relació amb la classe socioeconòmica estan basats tan sols en els nivells socioeconòmics estudiats i que, per tant, no es poden concebre com a principis universals perquè no és possible garantir que s'hagin considerat tots els nivells i estatus socioeconòmics existents. És a dir, per a la interpretació de resultats, cal tenir en compte que un comportament lingüístic concret pot tenir implicacions socials diferents segons la classe social a la qual pertanyi el parlant que el realitza (Serrano, 2008: 181).

Una altra variable que també pot interactuar amb la classe social és l'edat dels parlants. Segons Feagin, (2013: 28), els integrants de més edat de totes les classes socials tendeixen a mostrar els trets foneticofonològics més conservadors, mentre que els adolescents de classe treballadora solen liderar les innovacions lingüístiques.

Malgrat tot, alguns acadèmics han fet notar que les jerarquies socials poden funcionar de forma diferent en societats diferents i, per tant, presentar interaccions diverses amb altres variables sociolingüístiques. Per exemple, Durrell (2004: 202) apunta que en països de parla anglesa és molt difícil discernir entre dialectes i sociolectes, ja que “all *dialects* typically correlate with both regional and social factors” (cursiva en l'original). Alhora, però, subratlla que aquestes correlacions poden no ser aplicables a totes les societats altament industrialitzades (per exemple, països europeus de parla no anglesa) i que cal buscar els motius d'aquestes discrepàncies “in the peculiar socio-historical development of individual countries” (Durrell, 2004: 203).

Una altra de les variables que interactua amb la classe social és el nivell educatiu, com hem mencionat anteriorment (al subapartat **2.3.3**). Publicacions recents mostren que tant l'accés al sistema educatiu com el grau de formació assolit varia entre sectors socioeconòmics diferents (p. ex. Cabanellas Torres, 2020). En aquest sentit, la sociòloga educativa Aina Tarabini (2021: 17) aporta dades estadístiques eloqüents sobre les relacions entre classe social i assoliment educatiu la societat espanyola:

“Les dades més recents indiquen que a Espanya un 43,7 % dels nens i nenes de 0 a 2 anys estan escolaritzats a algun servei educatiu de caràcter formal. No obstant això, aquest percentatge es redueix fins al 26,3 % en el cas de nens i nenes que provenen dels quintils inferiors de renda, mentre que augmenta fins al 62,5 % a les llars amb majors recursos econòmics (Save the Children, 2019). [...] Així, al 2019 el percentatge d'AEP [abandonament escolar prematur] dels i les joves les mares dels quals tenen estudis superiors se situa només al 4 %, i puja fins al 41,8 % si aquestes només tenen educació primària o inferior (MEFP, 2020a). Cal tenir en compte, a més, que l'accés a l'educació postobligatòria està

profundament esbiaixat en funció de les seves modalitats formatives. Així, la taxa bruta d'accés als cicles formatius de grau mitjà (CFGM) al curs 2017-18 era del 35,2 %, mentre que al batxillerat era del 64 % del total (MEFP, 2020a). I de nou, el perfil social dels estudiants, tant pel que respecta a la seva classe social com també al seu origen i al seu gènere, determina l'accés a una o altra via formativa.”

A més, el nivell educatiu també interacciona, als territoris de parla catalana, amb diversos trets dialectals, com mostren estudis sobre la influència que han tingut agents clau en el procés de difusió social de les varietats estàndards (com el sistema educatiu i els mitjans de comunicació) sobre les varietats dialectals utilitzades per parlants de generacions diferents (vegeu, per exemple, Segura Llopes, 2003: 216, i Valls, 2013).

L'estudi de Nussbaum i Unamuno (eds., 2006) tracta els usos i les competències multilingües de l'alumnat d'origen immigrant mitjançant una aproximació qualitativa. Entre altres troballes interessants per a la comprensió dels usos lingüístics de l'alumnat “immigrant ubicat socialment en sectors pobres o empobrits, urbans o semiurbans, localitzats geogràficament en espais caracteritzats per una història migratòria prèvia amb forta presència de persones d'origen espanyol” (2006: 167), els autors examinen el poder explicatiu de l'estatus socioeconòmic respecte a les diferències comunicatives entre alumnes d'orígens diversos i conclouen que:

“En definitiva, el resultat del nostre estudi ens porta a afirmar que la categoria immigrant econòmic no és explicativa de diferències entre competències, ni és indicatiu de fracàs o d'èxit escolar. En canvi, sí que ho és el grau d'afiliació de cada noi o noia, amb el seu correlat en la inversió que fan en els processos de socialització lingüística i escolar. Hem pogut observar alts graus d'afiliació i d'inversió en el cas d'escolars d'origen immigrant, i alts graus de reluctància entre l'alumnat autòcton.” (Nussbaum i Unamuno, eds., 2006: 173).

Els resultats d'aquesta investigació pel que fa als àmbits d'ús de les llengües utilitzades mostren, a més, que entre l'alumnat estudiat, tant l'autòcton com l'al·lòcton, el castellà s'utilitza en entorns “més diversificats que els de la resta de

llengües escolars”, mentre que els usos que fan del català són més sovint receptius que interactius (Nussbaum i Unamuno, eds., 2006: 169). Aquestes diferències d’ús, però, no impliquen que l’alumnat no hagi desenvolupat la seva competència comunicativa en català.

Com mostra l’estudi de Nussbaum i Unamuno (eds., 2006), les llengües catalana i castellana tenen una prominència social diferent a diverses àrees del territori catalanòfon, ja que les diverses onades immigratòries que hem tractat al subapartat **2.3.4** no han tingut una influència homogènia a totes les regions, sinó que diversos nuclis poblacionals (sobretot, de l’àrea metropolitana de Barcelona) destaquen per concentrar un percentatge més elevat de població d’origen al·lòcton que altres. Per això, les variables de llengua inicial i varietat dialectal poden mostrar correlacions en estudis sobre els territoris de parla catalana.

2.3.6. Les principals variables independents dels estudis sociolingüístics: síntesi

En els subapartats precedents hem fet un repàs d’estudis que han tractat les variables sexe, edat, classe social, nivell educatiu, llengua inicial i varietat dialectal des de diverses aproximacions. Hem ressaltat que s’haurà de considerar les múltiples possibilitats d’interacció entre variables. Quant a la primera variable, el sexe dels informants, hem observat que segons la bibliografia disponible hauria de ser possible identificar diferències en els usos dels recursos morfosintàctics, lèxics i pragmàtics que fan servir homes i dones. També hem distingit entre sexe i gènere.

Pel que fa a l’edat, hem comprovat que els estudis sobre l’adquisició lingüística adolescent són menys nombrosos que els que estudien l’adquisició d’infants, malgrat que les diverses capacitats lingüístiques també evolucionen durant l’adolescència. Els treballs sobre el català en aquesta etapa se centren sobretot en el lèxic i la influència del castellà.

Quant a la classe social i al nivell educatiu, hem vist que totes dues s'han utilitzat en un bon nombre d'investigacions sobre la variació lingüística i que el nivell educatiu ja s'utilitzava en treballs dialectològics, malgrat que no podem afirmar que existeixi un consens entre els acadèmics sobre com definir clarament i mesurar cap de les dues variables. En aquest sentit, hem observat també que el nivell educatiu és un dels indicadors que s'ha emprat més sovint per classificar els parlants en classes socials. Per últim, s'ha explicat que la variable nivell educatiu ha permès a un bon grapat de treballs variacionistes descobrir desigualtats en els sistemes educatius, sustentades en prejudicis envers certes varietats lingüístiques.

Hem tractat igualment la necessitat de tenir en compte la llengua inicial dels informants a l'hora d'explicar els seus usos de diverses variables lingüístiques en relació al biaix monolingüe de la recerca a nivell internacional i a la rellevància d'aquesta variable independent en el context sociolingüístic català (qüestió que es tracta amb més deteniment al capítol 4).

L'última de les variables més utilitzades en treballs sobre la variació lingüística de la qual ens hem ocupat és la diversitat dialectal. Com s'ha exposat, comptem amb nombrosos treballs sobre les varietats dialectals catalanes, originats tant des de la dialectologia com des de la sociolingüística de la variació.

Finalment, hem pogut observar que les variables independents actuen sobre el comportament lingüístic dels parlants de manera interrelacionada. Entre altres, hem trobat que els estudis disponibles han permès identificar relacions entre la classe social i el sexe, el sexe i l'edat, l'edat i la pertinença a cohorts, el nivell educatiu i la varietat dialectal o la varietat dialectal i la llengua inicial. Amb el temps, els treballs sobre la variació lingüística (especialment sobre llengües com l'anglès o el castellà) han evolucionat cap a aproximacions multivariables. Això no obstant, aquesta mena d'estudis requereixen una base prèvia de treballs sobre relacions bivariables. Per últim, hem vist que diversos autors subratllen la necessitat d'ampliar el ventall de societats i comunitats de parla analitzades per evitar (o pal·liar) els biaixos en la recerca sobre la variació lingüística.

3 OBJECTIUS I HIPÒTESIS

La revisió bibliogràfica exposada al capítol anterior mostra que existeixen nombrosos estudis previs que permeten comprendre millor la llengua produïda per infants o proporcionen dades que podrien ser útils per a la perfilació lingüística. La revisió ha permès, també, identificar un buit en els estudis de sociolingüística catalana, ja que no comptem amb precedents que se serveixin de la metodologia variacionista per investigar el català oral utilitzat per informants adolescents amb la finalitat d'informar les pràctiques de perfilació sociolingüística de mostres d'autoria desconeguda. Aquest és, precisament el nostre objectiu: analitzar un seguit de variables que permetin la perfilació sociolingüística de mostres d'autoria desconeguda produïdes per parlants d'edat preadolescent i adolescent en llengua catalana. Per assolir-lo l'estudi es basarà en la concepció variacionista de la variable lingüística formada per variants.

Més concretament, en aquesta tesi ens proposem identificar trets lingüístics que mostrin una correlació estadísticament significativa amb les variables edat, sexe i llengua inicial dels parlants que formen la nostra mostra i que, per tant, es puguin explotar en l'elaboració de perfils sociolingüístics forenses. Per tal de complir aquest objectiu, utilitzem eines metodològiques de la sociolingüística de la variació. De fet, podem dividir la nostra empresa en quatre objectius específics. En primer lloc, identificar les variables susceptibles de ser realitzades pels informants. En segon lloc, identificar les variants realitzades pels informants de les diverses variables que analitzem. En tercer lloc, quantificar el total de realitzacions de cada variant present en les transcripcions del corpus d'anàlisi i, en quart lloc, aplicar tècniques estadístiques al conjunt de dades obtingudes per comprovar si les variables lingüístiques que analitzem mostren diferències estadísticament significatives entre informants d'edat, sexe i llengua inicial diferents.

La pregunta de recerca que ens formulem, així doncs, és la següent:

→ Hi ha variables lingüístiques que permeten identificar l'edat, el sexe i la llengua inicial més probable en el discurs d'un parlant adolescent en català?

Aquesta pregunta es concreta en tres hipòtesis generals que caldrà corroborar o refutar a partir dels resultats. Les nostres hipòtesis sorgeixen dels estudis previs que, com hem vist al capítol anterior, han detectat o apunten al fet que hi pugui haver diferències en el comportament lingüístic dels parlants en relació amb les tres variables independents que ens ocupen.

Hipòtesi 1 – Les variables lingüístiques seleccionades com a diferenciadores per grups d'edat mostren diferències estadísticament significatives.

Hipòtesi 2 – Les variables lingüístiques seleccionades com a diferenciadores per grups de sexe mostren diferències estadísticament significatives.

Hipòtesi 3 – Les variables lingüístiques seleccionades com a diferenciadores per grups de llengua inicial mostren diferències estadísticament significatives.

També es parteix de les subhipòtesis sobre quines variables independents és més probable, segons la revisió bibliogràfica realitzada, que es correlacionin amb cadascuna de les variables dependents que s'exposen al capítol 5 (subapartat **5.3.2.3**). Si es corroboren aquestes hipòtesis, les diferències que identifiquem entre els grups d'informants diferents són les que contribuiran a l'elaboració de perfils lingüístics forenses (tractats al subapartat **2.1.4.2**) a partir de mostres dubitades de casos relacionats amb adolescents en què apareguin les variables analitzades en aquest treball. Per aconseguir els objectius i posar a prova les hipòtesis que acabem de descriure utilitzem part d'un corpus recollit en quatre onades entre 2006 i 2015 en un seguit de projectes de recerca duts a terme al Centre de Recerca en Sociolingüística i Comunicació de la Universitat de Barcelona (CUSC-UB). Concretament, utilitzem dades obtingudes a les poblacions catalanes de Manlleu, a Osona, i Mataró, al Maresme (vegeu els apartats **5.1** i **5.2**).

4 CONTEXT DE LA RECERCA

Un cop descrit el marc teòric en el qual cal encabir aquesta tesi i exposat els objectius que perseguim, cal ara descriure el context sociolingüístic en què es va dur a terme la recollida de les dades que analitzem en aquest treball. És per això que, a continuació, tractem, primerament, diversos estudis que ens ajuden a comprendre la situació sociolingüística de Catalunya i de les poblacions de les quals prové el corpus de l'estudi en el període en què es van dur a terme les onades de recollida de dades. En segon lloc, completem aquesta contextualització de la nostra investigació amb un repàs de la bibliografia disponible sobre la varietat oral col·loquial de la llengua utilitzada per la població jove, caracteritzada, entre altres trets, per la influència del castellà.

4.1. La realitat demolingüística dels joves catalans

D'una banda l'Idescat i la Secretaria General de Política Lingüística (fins el 2021, Direcció General de Política Lingüística) de la Generalitat de Catalunya han dut a terme enquestes d'usos lingüístics (*Enquesta d'usos lingüístics de la població*, a partir d'ara, EULP) cada cinc anys des de 2003 per monitoritzar els usos de diverses llengües que fan els ciutadans més grans de 15 anys, amb especial interès pel català, “i com hi incideixen els canvis sociodemogràfics dels darrers anys” (Idescat i DGPL, 2015: 13).

De l'altra, el Consell Superior d'Avaluació del Sistema Educatiu (d'ara endavant, CSASE), òrgan de consulta i d'assessorament del Departament d'Educació de la Generalitat de Catalunya, realitza, cada set anys i des de 2006, un estudi sobre usos lingüístics dins i fora del context educatiu i dades sociolingüístiques de l'alumnat a l'últim curs de l'escolarització obligatòria (*Estudi sociodemogràfic i lingüístic de l'alumnat de quart d'ESO a Catalunya*) i emet un informe que té l'objectiu de donar a conèixer “el perfil sociodemogràfic i lingüístic de l'alumnat” (CSASE, 2014: 9). Així doncs, els resultats sobre els 2.552 alumnes de 4t d'ESO d'arreu de Catalunya enquestats pel CSASE el 2012 ens aporten informació encara més directament relacionada amb una meitat del nostre corpus (descriu a l'apartat 5.2) que l'obtinguda a l'EULP. Per aquesta raó, a continuació complementem la perspectiva àmplia que proporciona l'EULP sobre els usos

lingüístics dels joves catalans amb la més específica de l'estudi del CSASE, centrada en l'alumnat de 4t d'ESO.

Les estadístiques rellevants per al nostre estudi són les compreses entre 2008 i 2013, ja que les entrevistes orals que conformen el nostre corpus es van realitzar entre 2007 i 2012. Per comprendre les dades que tractem a continuació, cal tenir presents els fenòmens migratoris que hem vist a l'apartat anterior (2.3.4). Sobretot, cal recordar que entre 2000 i 2013 la població estrangera a Catalunya va passar d'un 3% a un 15%, aproximadament (Ortega-Rivera i Solana, 2015: 53). Segons Flors-Mas, Sorolla, Pradilla i Vila (2021: 43), la immigració va ser el factor principal dels canvis demolingüístics produïts a Catalunya entre el 2003 i el 2018.

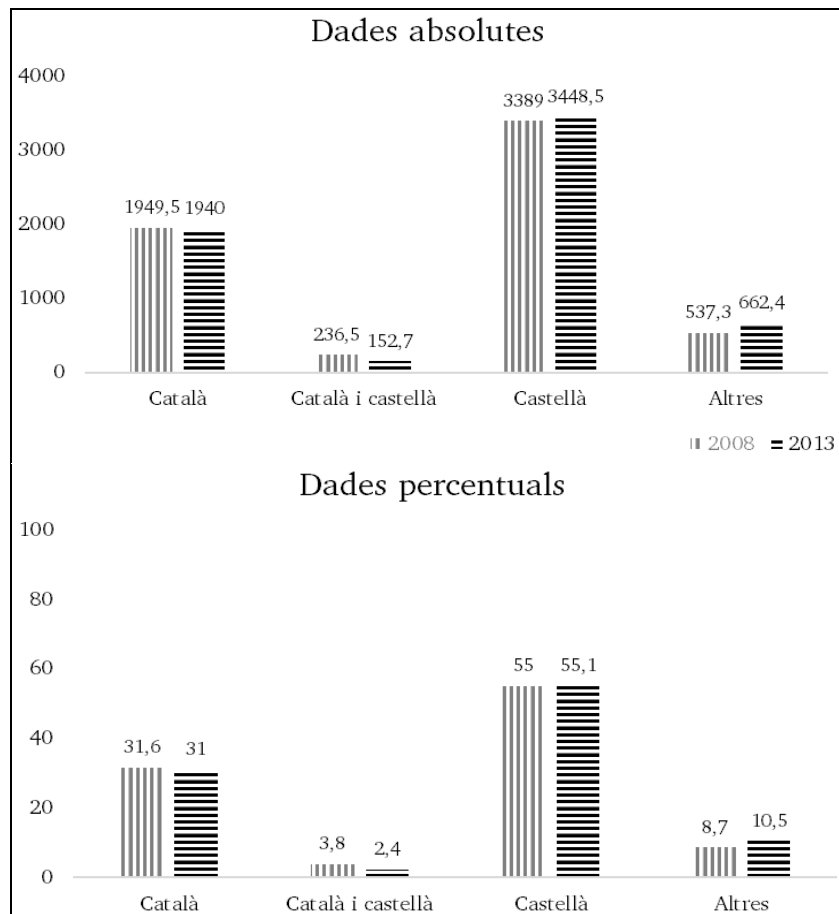
Per entendre les característiques dels nostres informants i el seu comportament lingüístic en la nostra mostra, en aquest apartat ens centrem en la llengua inicial, la llengua d'identificació, el coneixement de català i castellà (en les habilitats orals) i la llengua habitual dels sectors més joves de la població descrits a l'EULP de 2008 i 2013, així com les dades sobre llengua inicial, llengua d'identificació i ús global català i de castellà que proporciona l'estudi del CSASE de 2013, ja que aporta informació sobre l'alumnat de 4t d'ESO de l'any 2012, el mateix en què els nostres informants es trobaven en aquest curs.

En primer lloc, quant a la llengua inicial (Figura 2), l'EULP ens mostra que entre 2008 i 2013 el percentatge d'al·loglots inicials va augmentar lleugerament (d'un 9,3% a un 10,7% i de 537,3 a 662,4), la qual cosa és congruent amb el fenomen migratori que hem descrit. Quant a la resta d'informants, les dades absolutes mostren que es produeix un petit augment en el nombre de parlants inicials del castellà (de 3.389 a 3.448,5) mentre que els parlants inicials de català i de català i castellà disminueixen (de 1.949,5 a 1.940 i de 236,5 a 152,7, respectivament). En percentatges, els canvis són mínims per al català i al castellà com a llengua inicial. El primer en una tendència no gaire pronunciada a la baixa (de 31,6% a 31%) i el segon, a l'alça (de 55% a 55,1%). El grup de bilingües inicials mostra una baixada més clara (passa del 3,8% al 2,4%). Tot i aquests petits canvis, l'ordre dels grups és manté intacte entre 2008 i 2013: el grup més nombrós, amb

una diferència notable, és el de parlants amb el castellà com a llengua inicial, seguits pels parlants inicials del català, els al·loglots i, finalment, els bilingües en català i castellà.

Tot seguit, a la Taula 7, comparem les dades sobre la llengua inicial de l'EULP de 2013 amb les obtingudes pel CSASE el 2012 entre l'alumnat de 4t d'ESO. Com s'hi pot observar, segons les darreres l'alumnat de l'últim curs de l'educació obligatòria està format per uns percentatges més elevats de catalanoparlants i de bilingües inicials que el conjunt de la població més gran de 15 anys de Catalunya. Alhora, presenta un percentatge menor de castellanoparlants i al·loglots inicials.

Figura 2. Llengua inicial a Catalunya segons l'EULP 2008-2013.



Font: elaboració pròpia a partir de l'EULP de 2008 (Idescat i SPL, 2009: 43) i l'EULP de 2013 (Idescat i DGPI, 2015: 43).

Taula 7. Comparació dels resultats percentuals sobre la llengua inicial a l'EULP (2013) i l'informe del CSASE (2014).

	EULP (2013) (n=7.500, >15 anys, Catalunya)	CSASE (2014) (n=2.552, 15-16 anys, Catalunya)
Català	31 %	34,9 %
Castellà	55,1 %	40,8 %
Català i castellà	2,4 %	15 %
Altres	10,5 %	9,3 %

Font: Elaboració pròpia a partir de l'EULP de 2013 (Idescat i DGPL, 2015 : 43) i CSASE (2014: 45).

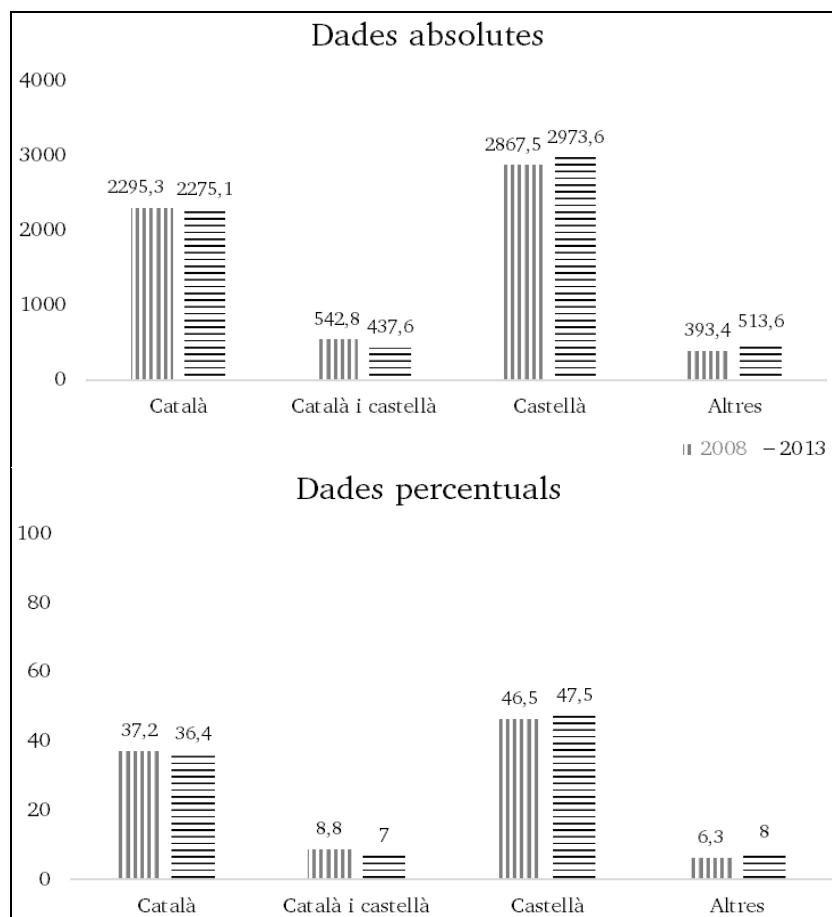
Com hem vist al capítol anterior, en el nostre estudi utilitzem la variable llengua inicial, ja que podria explicar diferències en el comportament lingüístic dels individus de la nostra mostra i, a diferència de la llengua d'identificació o de la llengua habitual, que tractem tot seguit, es manté estable al llarg del temps. A més, en el context català, per la situació històrica de contacte lingüístic amb el castellà (cf. Argenter, 2020), l'evolució de la llengua inicial en el temps esdevé una variable clau per comprendre el panorama sociolingüístic (Flors-Mas, Sorolla, Pradilla i Vila, 2021: 33).

En segon lloc, quant a la llengua d'identificació, diversos estudis han subratllat la seva rellevància a l'hora d'explicar els usos lingüístics dels parlants (per exemple, DGPL, 2015 o Fabà Prats i Torres-Pla, 2020). En relació als nostres objectius, destaca l'estudi de Sorolla i Flors-Mas (2020), ja que demostra que la llengua d'identificació dels parlants pertanyents a la generació Y, és a dir, els nascuts entre el 1980 i el 1999 (que inclou els informants del nostre corpus, nascuts entre el 1994 i el 1996), és la variable sociolingüística que millor explica les seves tries lingüístiques.

Així doncs, si ens fixem en la llengua d'identificació dels parlants enquestats a l'EULP (Figura 3), veiem que es produeix un canvi en l'ordre dels quatre grups d'identificació entre 2008 i 2013. En tots dos anys els dos primers grups són el castellà i el català, però, mentre que el 2008 hi havia més parlants que s'identificaven com a bilingües en català i castellà (542,8 o un 8,8%) que no com a al·loglots (393,4 o un 6,3%), el 2013 el nombre de parlants que

s'identificaven com a al·loglots (513,6 o un 8%) va superar el de bilingües en català i castellà (437,6 o un 7%). Aquest canvi es deu, alhora, a l'increment que es reflecteix tant en dades absolutes com percentuals del grup d'al·loglots i a la disminució que va patir el grup de bilingües entre les dues enquestes. Pel que fa als informants que s'identificaven amb el castellà i el català, les dades absolutes i les percentuals mostren que el grup de castellà va augmentar (de 2.867,5 a 2.973,6, és a dir, d'un 46,5% a un 47,5%) mentre que el de català va disminuir lleugerament (de 2.295,3 a 2.275,1 o d'un 37,2% a un 36,4%).

Figura 3. Llengua d'identificació a Catalunya segons l'EULP 2008-2013.



Font: elaboració pròpia a partir de l'EULP de 2008 (Idescat i SPL, 2009 : 43) i de 2013 (Idescat i DGPL, 2015 : 46).

Com mostra la Taula 8, les dades sobre la llengua habitual dels enquestats a l'estudi sobre l'alumnat de 4t d'ESO realitzat l'any 2012 pel CSASE mostren, en relació amb les de l'EULP de 2013, uns percentatges similars (entorn el 8 %) d'al·loglots inicials, unes proporcions menors d'individus catalanoparlants i

castellanoparlants inicials (de vora el 30 % en l'alumnat de 4t d'ESO) i força més bilingües inicials respecte la població catalana de més de 15 anys.

Taula 8. Comparació dels resultats percentuals sobre la llengua habitual a l'EULP (2013) i l'informe del CSASE (2014).

	EULP (2013) (n=7.500, >15 anys, Catalunya)	CSASE (2014) (n=2.552, 15-16 anys, Catalunya)
Català	36,4 %	31,5 %
Castellà	47,5 %	32,8 %
Català i castellà	7 %	27,9 %
Altres	8 %	7,8 %

Font: Elaboració pròpia a partir de l'EULP de 2013 (Idescat i DGPI, 2015: 46) i CSASE (2014: 55).

Pel que fa a les dades sobre el coneixement declarat de català, la Taula 9 mostra els percentatges dels informants de l'EULP que van declarar entendre i parlar el català el 2008 i el 2013 dividits per franges d'edat. S'hi pot observar que els percentatges de parlants que declaren poder entendre el català són més elevats que els que declaren poder parlar-lo. Pel que fa a les diferències per edats, hi veiem que les més destacables a l'EULP de 2008 pel que fa a la capacitat d'entendre el català són entre els parlants de 45 a 64 anys (96%) i els de 65 anys i més (93,1%), separats per un 2,9% i, pel que fa a la capacitat de parlar-lo, entre els parlants més joves (84,9%) i més grans (72,3%), distanciat per un 12,6%. Aquestes diferències augmenten en les dades de 2013, en què la distància més gran (un 6,3%) quant a comprensió oral es troba de nou entre els parlants de 15 a 29 anys (96,4%) i els de 65 anys o més (90,1%) i també quant a producció oral, capacitat que un 92% dels joves declaren tenir respecte un 66,9% dels més grans (la qual cosa significa una diferència d'un 25,1%).

Si ens centrem en la franja d'edat que més s'acosta a l'edat dels nostres informants, hi observem un augment en el percentatge de parlants que declaren posseir les dues habilitats entre 2008 i 2013 (en un 1,7% i un 7,1%, respectivament). Tot i que més tímid, aquest augment també es pot constatar en les xifres totals (un 0,1% i un 2,1%), de manera que la franja més jove i el conjunt de la població enquestada mostren una tendència a l'alça.

Aquestes dades mostren, doncs, que tant en general com pel que fa a la població més jove, la gran majoria d'enquestats declaren entendre i parlar el català i, a més, es produeix un augment en aquestes xifres entre 2008 i 2013. A més, tot i que es basen en les creences dels parlants, les dades sobre comprensió oral són millors que les dades sobre producció oral, la qual cosa està relacionada amb la confiança lingüística dels parlants i les seves actituds vers el català, temes en els quals no aprofundirem en aquesta tesi, però que val la pena tenir presents.

Taula 9. Població segons coneixement del català i grups d'edat 2008-2013.

	Entendre'l		Parlar-lo	
	2008	2013	2008	2013
De 15 a 29 anys	94,7	96,4	84,9	92,0
De 30 a 44 anys	94,3	95,3	78,2	82,6
De 45 a 64 anys	96,0	95,2	77,2	80,9
De 65 anys i més	93,1	90,1	72,3	66,9
Total	94,5	94,6	78,3	80,4

Font: Idescat i DGPL (2014a).

La Taula 10 mostra els percentatges corresponents als enquestats a l'EULP que van declarar entendre i saber parlar el castellà el 2008 i el 2013. Tot i que amb una diferència menor que en el cas del català, hi observem que també en el cas del castellà, les dades de comprensió oral superen les de producció oral.

Quant a les diferències per edats, podem observar a simple vista que són bastant reduïdes. El 2008, la distància més gran en la capacitat d'entendre el català es dona entre els parlants de 45 anys o més i la resta, però és tan sols d'un 0,1%, mentre que la distància més destacable en la capacitat de parlar-lo separa els parlants de 45 a 64 anys i els de 65 anys i més per un 0,5%. El 2013, les distàncies més grans entre franges d'edat augmenten lleugerament. El percentatge més elevat de parlants que diuen entendre'l correspon a la franja d'entre 45 i 64 anys (99,85%) i el més reduït al de parlants de 65 anys o més (99,25%), distanciats per un 0,6%. Pel que fa a capacitat de parlar-lo, la distància més destacada (malgrat ser només del 0,13%) es troba entre la franja de 15 a 29 anys (99,83%) i la de 30 a 44 anys (99,7%).

La franja d'informants entre 15 i 29 anys mostra una lleugera davallada en el percentatge d'enquestats que van declarar entendre el castellà entre 2008 i 2013 (d'un 0,07%) i un augment, també lleuger, entre els que van declarar saber parlar-lo entre aquests dos anys (d'un 0,13%). La primera tendència també es pot observar en les dades totals de la capacitat de comprensió oral, que disminueixen en un 0,1%, però no en les de producció oral, que es mantenen.

Així doncs, les dades de comprensió i capacitat de producció oral declarades pel castellà, són encara més elevades que les del català. L'augment que constatarem anteriorment en el cas del català entre 2008 i 2013, però, en general no s'observa en el castellà. El que sí que coincideix i, per tant, sembla un fenomen comú, és que, en general, les dades de comprensió superen les de producció.

Finalment, cal destacar que la comparació de les dades de català i de castellà per a la franja que ens interessa en relació a la nostra mostra, és a dir, la més jove, mostra que mentre que el percentatge d'enquestats que diuen saber parlar el català oscil·la entre el 84,9% i el 92% entre 2008 i 2013, en el cas del castellà, no baixa del 99,7%. Aquestes dades semblen indicar que una part dels nostres informants poden sentir-se capaços de parlar el castellà però no el català, fet que podria reflectir-se en el seu comportament lingüístic.

Taula 10. Població segons coneixement del castellà i grups d'edat 2008-2013.

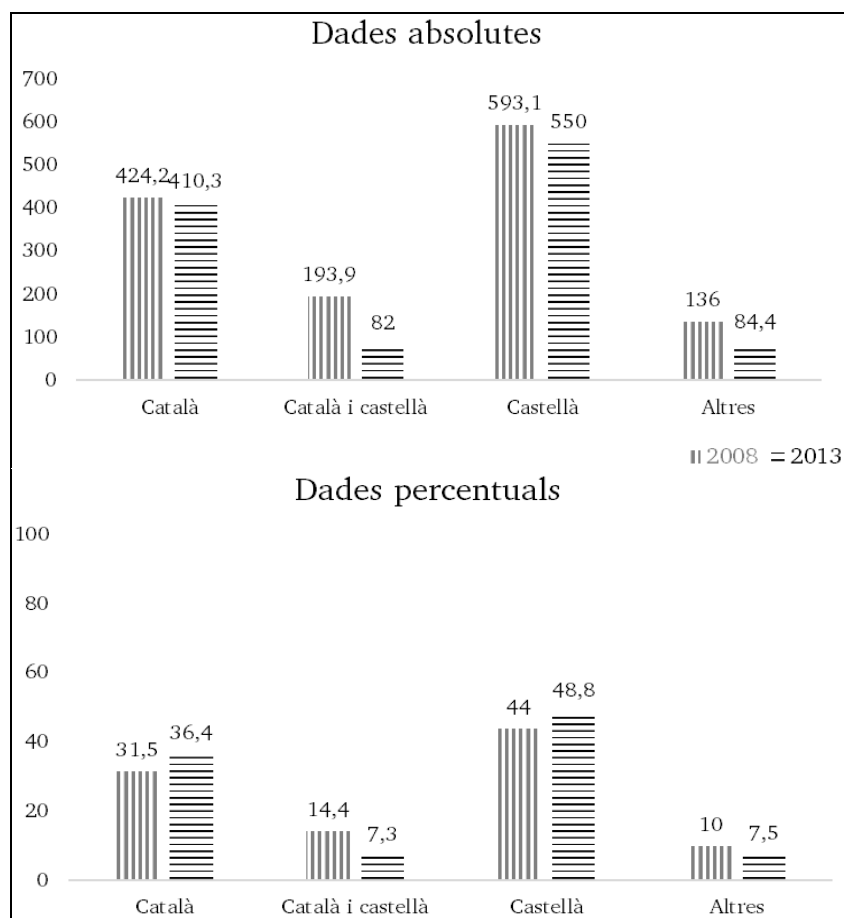
	Entendre'l		Parlar-lo	
	2008	2013	2008	2013
De 15 a 29 anys	99,9	99,83	99,7	99,83
De 30 a 44 anys	99,9	99,73	99,8	99,7
De 45 a 64 anys	100	99,85	99,9	99,8
De 65 anys i més	100	99,25	99,4	98,73
Total	99,9	99,8	99,7	99,7

Font: Idescat i DGPL (2014) i EULP (Idescat i SPL, 2009: 143).

A continuació, la Figura 4 recull l'evolució de les dades corresponents a la llengua habitual dels enquestats d'entre 15 i 29 anys entre l'EULP de 2008 i de 2013. Si ens fixem en dades percentuals, hi veiem que la proporció de joves que utilitzen habitualment només el castellà (un 44% el 2008 i un 48,8% el 2013) o

només el català (31,5% i 36,4%) augmenta entre 2008 i 2013 (tot i una lleugera caiguda en les dades absolutes), mentre que tant el percentatge de parlants joves que utilitzen habitualment tant el castellà com el català i el de parlants habituals d'altres llengües disminueixen (passen d'un 14,4% a un 7,3% i d'un 10% a un 7,5%, respectivament). Així doncs, el grup més nombrós, tant el 2008 com el 2013, és el de parlants habituals del castellà, seguit del grup de parlants habituals del català, mentre que els altres dos grups mostren percentatges similars, si bé inverteixen l'ordre d'un any a l'altre.

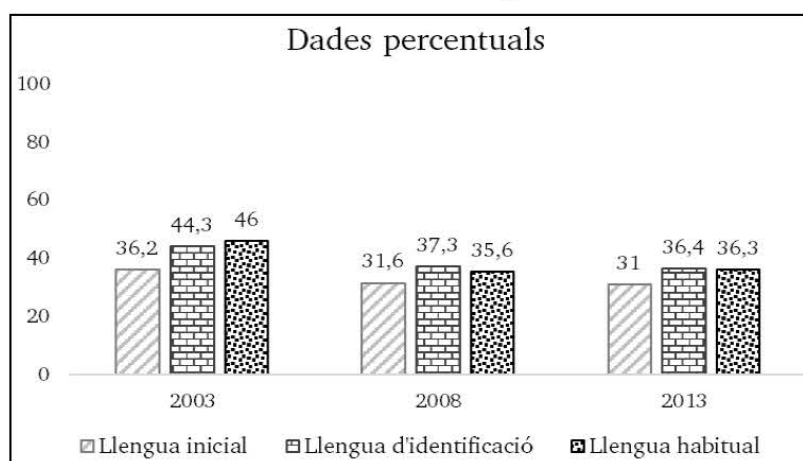
Figura 4. Llengua habitual dels parlants entre 15 i 29 anys a Catalunya.



Font: elaboració pròpia a partir de l'EULP de 2013 (Idescat i DGPL, 2015: 193).

Com mostra la Figura 5, el percentatge de parlants de 15 anys o més que tenen el català com a llengua inicial és inferior al d'aquells que s'hi identifiquen o que l'utilitzen de manera habitual. Així doncs, és possible que existeixin diferències en l'ús del català entre parlants que el tenen com a llengua inicial i parlants inicials d'altres llengües.

Figura 5. Població de 15 anys o més amb el català com a llengua inicial, d'identificació i habitual a Catalunya entre 2003 i 2013.



Font: elaboració pròpia a partir d'Idescat i DGPL (2018a).

L'enquesta distribuïda l'any 2012 pel CSASE entre l'alumnat català de 4t d'ESO no incloïa una pregunta sobre la llengua habitual, sinó que preguntava als informants en quina mesura utilitzaven el català i el castellà i l'informe (CSASE, 2014) agrupa les respostes en tres categories: sovint, sempre o gairebé sempre; mitjanament, i poques vegades, mai o gairebé mai. Els resultats mostren que un 51,5 % de l'alumnat declara utilitzar el català sempre (sovint o gairebé sempre), mentre que un 66,1 % afirma emprar el castellà amb aquesta freqüència. A més, “[e]l baix percentatge d'alumnat que utilitza poques vegades, mai o gairebé mai tant el català com el castellà [el 23,9 % i el 16,4 %, respectivament] demostra que la majoria de l'alumnat té contacte amb les dues llengües en les seves interaccions quotidianes. En canvi, són pocs els alumnes majoritàriament monolingües” (CSASE, 2014: 115).

Per tant, fins ara hem vist una forta presència del castellà en les llengües inicials, d'identificació, en el coneixement i les llengües d'ús habitual dels joves catalans, tant a través de les dades sobre el conjunt de la població (EULP) com de les que específicament provenen de l'alumnat de 4t d'ESO (CSASE). Com hem vist, en tots els casos és el castellà la llengua que compta amb percentatges més elevats, seguida del català. També hem observat a les dades de l'EULP que el grup d'altres llengües va superar el de bilingües en català i castellà tant en llengua d'identificació com en llengua habitual el 2013, fet que sembla correspondre a l'augment de població d'origen estranger. Quant al coneixement declarat de

català i castellà, hem observat, d'una banda, que es produeix un cert increment en el percentatge d'informants que diuen saber parlar cada llengua entre 2008 i 2013 i, de l'altra, que, en totes dues enquestes, els percentatges d'enquestats que es consideren capaços de parlar el castellà són superiors als que diuen saber parlar el català (els primers no són inferiors al 99,7%, mentre que el percentatge més alt per al català és del 92%). Finalment, també hem vist que entre 2008 i 2013 augmenta el percentatge de parlants que habitualment fan servir només castellà o català i disminueix el de parlants habituals d'altres llengües o tant de català com de castellà.

Damunt d'aquestes dades estructurals, cal tenir en compte factors com, en primer lloc, el consum cultural que fan els joves. Segons Bretxa (2014), que analitza part dels individus del mateix corpus observat en aquesta tesi¹⁷, i segons l'informe del CSASE (2014: 108), que se centra en l'alumnat que cursava 4t d'ESO al mateix any que els nostres informants, la llengua en què es produeix el consum cultural per part dels joves és el castellà o una combinació del castellà i el català. En segon lloc, cal tenir present les llengües que empren els joves amb les seves xarxes socials¹⁸, és a dir, els vincles que mantenen amb els membres de la seva comunitat que els són més propers: família, amics, coneguts. En tercer lloc, cal considerar també la llengua vehicular en què són escolaritzats (en la gran majoria de casos, el català) i els usos lingüístics que fan en els centres educatius tant l'alumnat com el professorat (sobre aquest tema, vegeu, entre altres, Bretxa i Vila, 2014, que tractem a l'apartat 4.3, i Martínez Pons *et al.*, 2021: 58-71). A més, pel que fa a la tria de llengua dels adolescents, Flors (2015) fa un repàs de la bibliografia que s'ha ocupat dels parlants joves i la situació sociolingüística a Catalunya i explica que aquests parlants justifiquen les

¹⁷ “Desde el mismo proyecto RESOL, se ha llevado a cabo un exhaustivo trabajo sobre el consumo cultural de los adolescentes catalanes, concretamente de la población de Mataró. [...] La autora, posteriormente, crea un índice para el análisis estadístico que permite afirmar que el consumo cultural de los jóvenes es principalmente en castellano (Bretxa, 2014: 194-197)” Illamola (2015: 143).

¹⁸ La relació entre les xarxes socials dels parlants i els seus usos lingüístics s'ha estudiat, per exemple, a Blanco Canales (2000) o a Sorolla Vidal (2016).

seves tries en base a característiques individuals, com la facilitat que tenen per una llengua o una altra, més que no pas en base a la seva afiliació lingüística (Flors, 2015: 33) i que majoritàriament s'acomoden a la llengua triada pel seu interlocutor (Flors, 2015: 34).

Segons les dades recopilades a l'EULP i l'estudi del CSASE, una mostra de la població més jove de Catalunya inclouria parlants amb el català, el castellà, totes dues o altres llengües com a llengua inicial. Com veurem, aquest és el cas del corpus RESOL, tot i que en aquesta tesi només estudiem adolescents dels tres primers grups de llengua inicial. Com també hem pogut observar, la llengua inicial no sempre coincidirà en la nostra mostra amb la llengua d'identificació o la llengua d'ús habitual dels informants, ja que totes aquestes variables es veuen influenciades per factors molt diversos i preferències personals. Amb tot, entenem que la llengua inicial és una variable estable que pot explicar part del comportament lingüístic dels individus i que és crucial en el context sociolingüístic català (Flors-Mas, Sorolla, Pradilla i Vila, 2021). Estudis com l'informe *Joventut i ús del català a Catalunya* (Sorolla i Flors-Mas, 2020: 55) subratllen que, juntament amb la llengua d'identificació i el nivell de competència oral, és una de les variables sociolingüístiques que millor pot explicar els usos lingüístics dels joves a Catalunya. Per tant, serà una de les tres variables que incloem en el nostre estudi.

Val la pena tenir en compte la hipòtesi que formula Tenorio Segarra (2013: 411), en el seu estudi sobre la relació entre llengua d'ús amb els progenitors i llengua inicial, que “la llengua inicial estaria influïda pel grup de pertinença on l'individu se sent identificat”, no tant “pels usos lingüístics reals (expressats per la llengua amb progenitors) sinó per un seguit de factors ideològics o sentimentals que deriven en una identificació amb el grup de referència”. Sorolla i Flors-Mas (2020) també troben una forta relació entre la llengua inicial i la d'identificació mitjançant un estudi multivariant a partir de les dades de l'EULP de 2013.

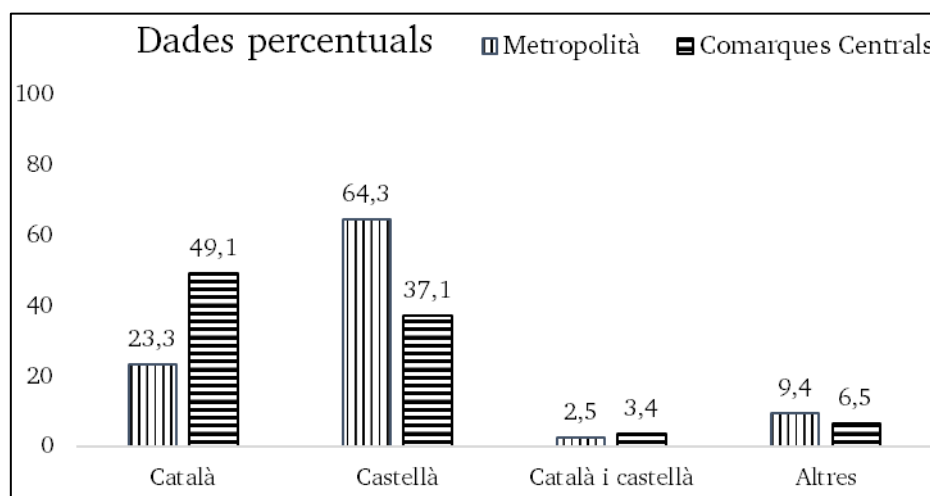
A més, caldrà tenir en compte que el nivell de bilingüització dels informants serà elevat però heterogeni, la qual cosa es podrà observar en les variables lingüístiques que mostrin el grau de contacte entre el català i el castellà. En relació amb aquesta observació, com hem vist, les dades de l'EULP mostren percentatges majors d'enquestats que declaren poder parlar el castellà que el català.

4.2. La realitat demolingüística de Manlleu i Mataró

La nostra mostra es va recollir en centres educatius dels municipis de Manlleu i Mataró (vegeu el capítol següent). A l'EULP de 2013 (Idescat i DGPL, 2015), aquestes poblacions pertanyen a dos àmbits territorials diferents: Comarques Centrals i Metropolità. Per això, després del repàs a les dades generals que hem dut a terme a l'apartat anterior, cal filar una mica més prim i veure les dades de cada àmbit territorial pel que fa a la llengua inicial i la llengua habitual de la població. Això ens permetrà entendre millor la realitat sociolingüística que representa la mostra d'informants del nostre estudi, ja que, com han demostrat diversos estudis, la composició demolingüística de l'entorn afecta l'adquisició i els usos lingüístics dels alumnes (especialment dels al·loglots, vegeu Flors, 2015: 38).

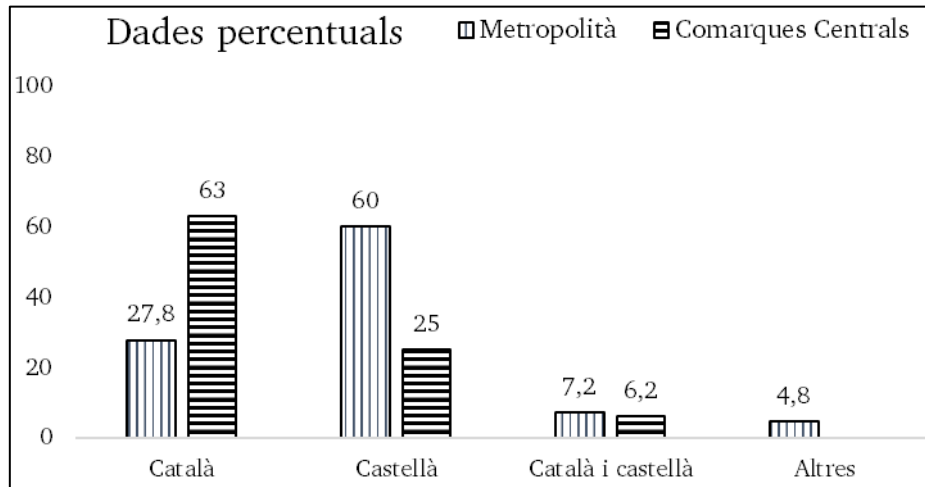
Primer, pel que fa a la llengua inicial, la Figura 6 mostra que a la regió metropolitana la llengua inicial majoritària és el castellà (64,3%), seguida a gran distància del català (23,3%). A les comarques centrals, en canvi, l'ordre és l'invers i la distància entre el català com a llengua inicial (49,1%) i el castellà (37,1 %) és considerablement més reduïda. Quant als altres dos grups de població, el que representa parlants inicials bilingües en català i castellà i els parlants inicials d'altres llengües o combinacions de llengües, s'observa que el darrer és més nombrós en els dos territoris. Tot i això, es pot observar una diferència entre la distància que els separa, ja que mentre que a l'àrea a la qual pertany Mataró hi ha un 9,4% de parlants amb llengües inicials diferents al català i al castellà, a l'àrea en què es troba Manlleu aquest grup de població representa un 6,5%, i mentre que a la primera hi trobem un 2,5% de bilingües inicials, a la segona, representen el 3,4% del total poblacional.

Figura 6. Llengua inicial als àmbits territorials Metropolità i Comarques Centrals segons l'EULP 2013.



Font: elaboració pròpia a partir d'Idescat i DGPL (2014c).

En segon lloc, la Figura 7 comprèn les dades sobre la llengua habitual de les dues regions. Hi veiem, de nou, l'ordre invers dels dos primers grups, en el mateix sentit descrit anteriorment per a la llengua inicial. Mentre l'àmbit metropolità mostra un predomini del castellà com a llengua habitual (amb un 60% de la població), a les zones centrals és el català la llengua més habitual (amb un 63% dels parlants enquestats). La distància amb el segon grup és semblant a les dues regions: el català és la segona llengua més habitual a la regió metropolitana per a un 27,8% i el castellà ho és al centre per a un 25% dels parlants. El primer dels dos grups restants mostra comportaments similars a les dues regions. A l'àrea metropolitana hi ha un 7,2% d'enquestats que diuen utilitzar tant el català com el castellà com a llengües habituals, mentre que a les comarques centrals aquest grup d'enquestats és d'un 6,2% del total. Finalment, pel que fa al grup que emprava habitualment altres llengües, a la regió metropolitana representen un 4,8%, però a les comarques centrals no disposem de dades per a aquesta categoria.

Figura 7. Llengua habitual als àmbits territorials Metropolità i Comarques Centrals segons l'EULP 2013.

Font: elaboració pròpia a partir d'Idescat i DGPL (2014c).

A continuació, complementem aquestes dades provinents de l'EULP de 2013 amb dades més específiques de cada municipi pel que fa al coneixement declarat de català, recollides al cens de 2011. Com es pot observar a la Taula 11, el cens mostra que la diferència més notòria pel que fa a les competències orals en català entre les dues poblacions es troba en la capacitat de parlar-lo. Mentre que a Manlleu el percentatge de població que diu poder parlar el català (gairebé un 80%) és superior al del conjunt de Catalunya (al voltant d'un 73%), a Mataró aquest percentatge és inferior al del conjunt (gairebé un 70%). Les dades de comprensió són semblants per als tres territoris (que ronden el 95%), tot i que la més baixa torna a ser la de Mataró (un 94%).

Taula 11. Coneixement del català a Mataró i Manlleu segons cens de 2011.

	Entendre'l	Parlar-lo
Manlleu	95,55 %	79,98 %
Mataró	94,06 %	69,75 %
Catalunya	95,12 %	73,17 %

Font: Idescat (2019a i 2019b).

Així doncs, les dades regionals i municipals de llengua inicial, llengua habitual i coneixement del català reflecteixen diferències entre les dues poblacions de les quals prové la nostra mostra. D'una banda, la majoria de població de Manlleu i de les comarques centrals té el català com a llengua inicial i més habitual,

mentre que, de l'altra, a Mataró i l'àrea metropolitana, aquesta posició l'ocupa el castellà.

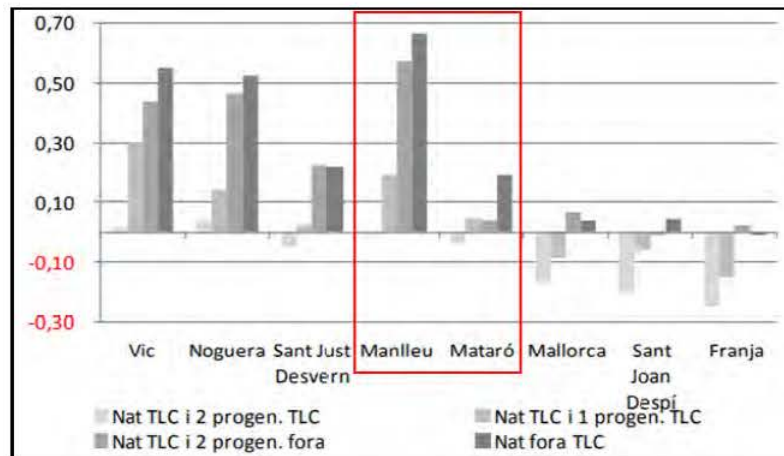
4.3. Estudis sociolingüístics previs sobre els adolescents de Manlleu i Mataró provinents de RESOL

Com hem avançat anteriorment, aquest estudi s'emmarca en un seguit de projectes de recerca (habitualment englobats sota la denominació "projecte RESOL¹⁹") duts a terme entre 2006 i 2015, mitjançant els quals es van obtenir i analitzar dades sociolingüístiques de més de 2.500 informants escolaritzats a poblacions d'arreu del territori català, amb característiques poblacionals i demolingüístiques diverses. Aquests projectes tenien l'objectiu d'investigar l'evolució de les competències, els usos i les actituds lingüístics d'alumnes de diverses localitats catalanòfones. Es van realitzar quatre onades de recollida de dades dels mateixos informants, corresponents als cursos de 6è de primària, 1r i 4t d'Educació Secundària Obligatòria i 2n de batxillerat o cicles formatius (per a més informació, vegeu l'apartat 5.1). En aquest apartat tractem altres estudis desenvolupats en el marc d'aquests projectes que aporten informació complementària a la que tractem en el nostre treball.

L'estudi de Comajoan, Vila, Bretxa, Sorolla, Tenorio i Melià (2013) observa dades provinents del projecte RESOL i investiga l'ús del català que fa l'alumnat amb els amics en relació amb l'ús que en fa amb els progenitors (Figura 8). Els resultats mostren que a Manlleu és més alt l'ús del català amb iguals que amb els progenitors, ja que el català és la llengua de comunicació entre amics, amb indiferència del lloc de naixement tant dels joves com dels progenitors. En canvi, a Mataró, "el nivell d'ús del català amb els progenitors és similar al d'ús amb els amics, la qual cosa amaga cert progrés entre els (pocs) al·lòctons i estancament, i fins i tot retrocés, entre els autòctons" (Comajoan *et al.*, 2013: 59).

¹⁹ Acrònim del nom del primer projecte, *Resocialització i llengües: els efectes lingüístics del pas de l'educació primària a secundària en contextos plurilingües*, finançat pel Ministeri de Ciència i Innovació i realitzat entre 2006 i 2009.

Figura 8. Diferència entre l'ús del català amb els amics i l'ús amb els progenitors (punts de diferència entre mitjanes) segons l'àrea i l'origen geogràfic familiar d'alumnes de sisè de primària.



Font: Comajoan *et al.* (2013: 59). TLC = territoris de llengua catalana.

Com mostren les dades disponibles, per tant, el context sociolingüístic de les dues poblacions de les quals prové el corpus que analitzem a l'estudi és força diferent. Resumidament, podem afirmar que, mentre que a Manlleu la llengua dominant és el català, a Mataró aquesta posició l'ocupa el castellà. D'aquesta manera, la nostra mostra representa dues realitats sociolingüístiques catalanes.

Tot seguit, complementem la descripció del context del qual provenen les nostres dades a partir d'estudis previs que investiguen les tries i els usos lingüístics dels joves que conformen la mostra recollida al projecte RESOL. Concretament, ens ocupem de quatre dels estudis sobre el català o les tries lingüístiques dels joves que han utilitzat dades del projecte RESOL i que resulten més rellevants per arribar a una comprensió holística dels nostres resultats, en el sentit que, a més dels valors estadístics obtinguts i de la seva relació amb la bibliografia ressenyada fins ara també tingui en compte el que ja sabem dels usos lingüístics dels informants del projecte.

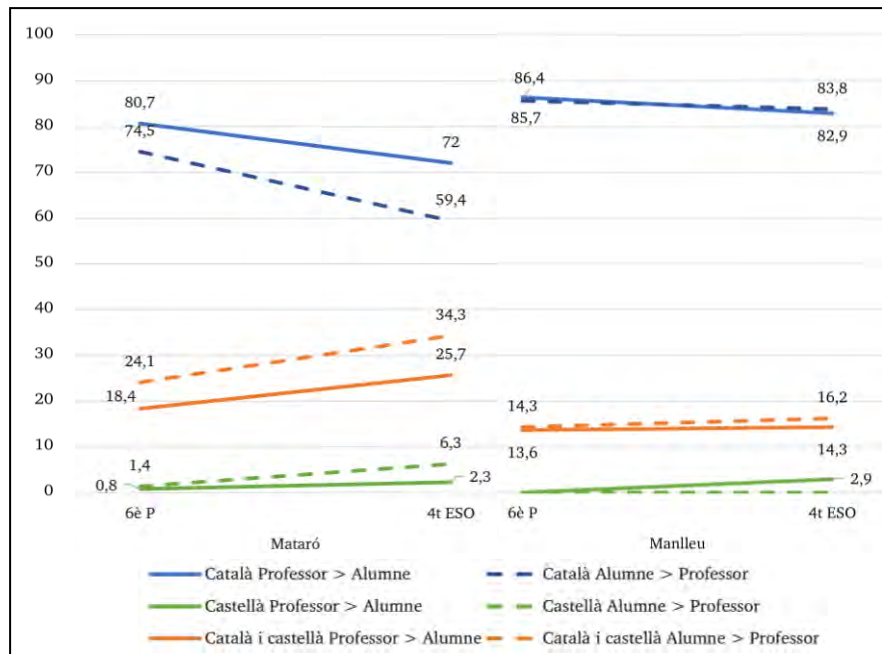
Aquests quatre estudis tenen diverses característiques comunes. Comparteixen el fet que exploten dades provinents, com a mínim, de les mateixes ciutats que ens interessin aquí, Mataró i Manlleu; que utilitzin dades recollides mitjançant el qüestionari sociolingüístic del projecte (Flors-Mas, 2018, les complementa amb entrevistes grupals), i que estudien els usos lingüístics declarats pels informants o la seva confiança lingüística, per la qual cosa els seus resultats són rellevants

per al nostre estudi, ja que ens poden ajudar a comprendre els usos que fan els informants de les diverses variables que analitzem.

En primer lloc, Bretxa i Vila (2014) estudien els usos lingüístics entre alumnes i professors a les aules de sisè de primària, primer i quart de secundària i troben diferències significatives en relació al municipi i a la llengua inicial dels alumnes. Utilitzen dades recollides mitjançant dues de les preguntes del qüestionari corresponents a tots els informants enllaçats de la primera a la tercera onada dels projectes, de manera que la seva mostra prové de cinc municipis, tres de l'àmbit metropolità de Barcelona (AMB) i dos de la resta de Catalunya.

Si ens centrem en les dades de Manlleu i Mataró per any i municipi (que recull la Figura 9), cal destacar que els resultats d'aquest estudi mostren que, en els tres anys analitzats, l'ús predominant de català tant per part de professors com d'alumnes és superior a Manlleu (entre un 86 % i un 83 %) que a Mataró (entre un 80 % i un 60 %) i, de fet, a la resta de municipis de fora de l'AMB observats (on oscil·la entre un 91 % i un 83 %) en comparació amb els de dins de l'AMB (on els percentatges es mouen entre el 80 % i el 48 %, Bretxa i Vila, 2014). En aquest sentit, també observen que només a l'AMB es perceben diferències destacables entre alumnat i professorat: el darrer utilitza més el català, mentre que els alumnes fan servir més tant el català com el castellà a les aules. Altres observacions destacables que fan els autors a partir d'aquestes dades són, d'una banda, la diferència, a les de Mataró, entre el percentatge d'informants que utilitzen de forma predominant el castellà per dirigir-se als docents (un 6,3 %) i el d'informants que afirmen que els professors se'ls adrecen en aquesta llengua (un 2 %). De l'altra, a Manlleu es dibuixa un descens de l'ús predominant de català per part dels alumnes que afavoreix l'ús del castellà entre sisè de primària i quart de secundària i, a més, en aquest darrer curs hi veiem que gairebé un 3 % de l'alumnat afirma que els docents els parlen en castellà malgrat que "aquest percentatge d'informants a qui tots els professors s'adrecen en castellà no diu pas parlar en castellà amb els seus companys" (Bretxa i Vila, 2014: 119).

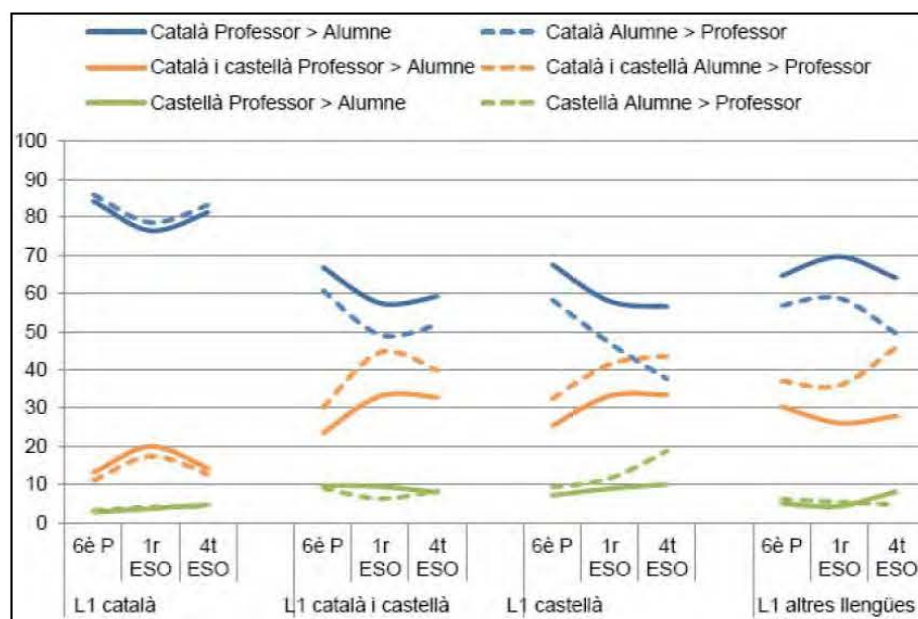
Figura 9. Llengües dins les aules per municipi. Dades percentuals de Bretxa i Vila (2014).



Font: adaptació de Bretxa i Vila (2014).

Pel que fa als usos lingüístics de tots els informants de Bretxa i Vila (2014) per llengua inicial (Figura 10), els autors observen, en primer lloc, que els catalanoparlants inicials, tant professors com alumnes, disminueixen el seu ús del català (representat per les parelles de línies blaves, a la part superior del gràfic) i augmenten els usos bilingües (les parelles de línies de la meitat del gràfic, en taronja) entre sisè i primer i que entre primer i quart gairebé retornen als percentatges de primària (Bretxa i Vila, 2014: 115). En segon lloc, quant als bilingües inicials, l'alternança entre les dues llengües és més elevada per part dels alumnes que no pas dels docents i, en el pas de sisè de primària a primer d'ESO es produeix un increment notable de l'ús combinat de les dues llengües. En darrer lloc, pel que fa als castellanoparlants inicials, cal destacar la davallada en l'ús del català entre sisè i primer i l'augment de l'ús predominant del castellà entre primer i quart d'ESO per part de l'alumnat.

Figura 10. Llengües dins les aules segons la llengua inicial de l'alumnat (6è primària-1r-4t ESO). Dades percentuals.



Font: Bretxa i Vila (2014: 117).

Així doncs, els autors conclouen que, en el pas de primària a secundària, disminueix l'ús a les aules del català i que la llengua inicial dels parlants està estretament relacionada amb els usos lingüístics analitzats, ja que els catalanoparlants inicials fan servir i són interpel·lats majoritàriament en aquesta llengua, mentre que els alumnes que tenen una altra llengua o una combinació de llengües com a inicial fan servir i són interpel·lats més en castellà.

“Això fa que sigui a les localitats amb fort influx immigratori del tercer quart del segle XX on més castellà es detecti dins l'aula, fins al punt que, en algun cas, el nombre d'alumnes que parlen en català i castellà amb els docents pràcticament s'equipara al dels que hi parlen en català. Amb tot, convé remarcar que els usos combinats de català i castellà també es detecten en percentatges significatius a les localitats amb percentatges més elevats de catalanoparlants.” (Bretxa i Vila, 2014: 120).

Per tant, és força probable que els usos lingüístics entre alumnat i professorat als centres recollits al nostre corpus siguin diferents segons la localitat (Manlleu vs. Mataró), la qual cosa pot influir en els usos d'algunes de les variables del nostre estudi com, per exemple, l'ús de fenòmens de contacte entre llengües (V38). Ara bé, també cal tenir en compte que les interaccions que analitzem no són entre els

alumnes i els seus professors, sinó amb les entrevistadores. Tot i això, donades les dinàmiques establertes normalment a les aules entre docents i alumnes i, com que el context de les interaccions que observem s'assembla al d'una prova de competència, caldrà tenir en compte la influència de les tries lingüístiques observades en estudis com el de Bretxa i Vila (2014) a l'hora d'interpretar els nostres resultats.

En segon lloc, Bretxa, Comajoan, Ubalde i Vila (2016) exploten les dades sobre llengua inicial, confiança lingüística i competència declarada de 941 informants de les mateixes ciutats estudiades a Bretxa i Vila (2014) recollides mitjançant el qüestionari sociolingüístic del projecte. Investiguen l'evolució de la confiança lingüística i la seva interacció amb la llengua inicial dels alumnes en el complex context sociolingüístic català. Els resultats mostren, primerament, uns nivells de confiança lingüística elevats tant per al català com per al castellà. Segonament, que la confiança lingüística no és un tret estable dels informants, sinó una construcció dinàmica que fluctua en interacció amb diverses variables depenent del moment vital en què es troben els parlants (per exemple, la transició entre l'educació primària i la secundària). En tercer lloc, es va poder establir una correlació estadísticament significativa entre els grups de llengua inicial i el nivell de confiança lingüística dels participants i, a més, es van observar quatre tendències diferenciades en l'evolució de la confiança lingüística segons la llengua inicial: períodes de més canvis que altres, grups de llengua inicial que pateixen més canvis en la confiança lingüística que la resta, una progressió generalment a l'alça (amb algunes excepcions) en la confiança lingüística amb el temps i diferències en la confiança lingüística en català i en castellà. Finalment, es va comprovar que existeixen certes diferències significatives quant al nivell de confiança lingüística segons la llengua predominant a l'entorn, de manera que la confiança lingüística en català és més elevada fora de l'AMB que dins.

Per tant, de forma semblant als resultats de l'estudi anterior, caldrà tenir en compte el paper que els canvis detectats per Bretxa, Comajoan, Ubalde i Vila (2016) en la confiança lingüística dels informants entre els dos moments temporals del nostre estudi puguin jugar a l'hora d'explicar els nostres resultats.

En tercer lloc, Flors-Mas (2018) també parteix de les dades recollides a Mataró i Manlleu mitjançant el qüestionari sociolingüístic (en aquest cas, quan els informants cursaven 4t d'ESO), però les complementa amb dades de Castelló de la Plana i entrevistes grupals semiestructurades a una submostra de 207 alumnes per observar fins a quin punt el comportament lingüístic declarat s'alinea amb les polítiques lingüístiques educatives de Catalunya o del País Valencià i com afecta la institucionalització del català a les aules. Combina, per tant, una aproximació quantitativa a les dades del qüestionari amb una de qualitativa sobre l'ús d'una llengua o una altra per part dels docents i els discents, que són vistos com agents capaços de decidir quins usos lingüístics fan i d'intervenir en els dels seus interlocutors. Els resultats d'aquest estudi assenyalen que, a Manlleu i Mataró, el català està força institucionalitzat (és la llengua en què s'espera que es duguin a terme les explicacions per part del professorat, en què els docents solen interactuar entre ells, etc.) i és la llengua en què s'espera que es comuniqui la majoria dels alumnes, si bé s'observen usos combinats de català i castellà a les aules per part d'alguns professors. El castellà s'utilitza de vegades per a la comunicació interpersonal entre alumnes i una part dels informants que no tenen el català com a llengua inicial prefereixen expressar-se en castellà a classe (Flors-Mas, 2018: 113).

A Castelló, en canvi, el català no està institucionalitzat, com es pot observar en l'aparició continuada al corpus analitzat de referències a docents que imparteixen en castellà assignatures que, segons les polítiques del territori, haurien de tenir el català com a llengua vehicular. Segons les dades recollides, el català pràcticament només l'utilitzen els alumnes que el tenen com a llengua inicial, de forma que es reproduïx a la classe l'estructura sociolingüística de fora de l'àmbit escolar (Flors-Mas, 2018: 113). Finalment, cal assenyalar que als dos territoris els informants observen que els docents de secundària intervenen menys que els de primària en els usos lingüístics a les aules i que, generalment, “they do not question students speaking in Spanish in subjects where, in theory, Catalan is the medium of instruction” (Flors-Mas, 2018: 112).

Així doncs, tenint en compte els resultats de Flors-Mas (2018), serà interessant veure si els alumnes del nostre estudi mostren diferències en l'ús de les variables analitzades per llengua inicial, possible indicador de la llengua en què s'expressen amb més comoditat, malgrat la forta institucionalització del català tant als centres educatius de Manlleu com de Mataró en comparació amb els de Castelló de la Plana.

Finalment, Vila, Ubalde, Bretxa i Comajoan-Colomé (2020) fan una anàlisi minuciosa dels canvis en els usos lingüístics entre sisè de primària, primer i quart d'ESO de 1.046 informants amb el català, el castellà, ambdues o altres llengües com a inicials de Mataró, Sant Joan Despí, Sant Just Desvern, Balaguer i Manlleu. Per fer-ho, utilitzen les respostes del qüestionari sociolingüístic sobre les llengües en què els informants parlen amb els seus amics i companys, la seva xarxa social, per calcular un índex d'ús per al català, el castellà i altres llengües per a cada informant en cada moment en què va ser enquestat.

En general, els autors observen canvis poc significatius en els usos lingüístics dels alumnes, si bé l'ús del català disminueix lleugerament i el castellà augmenta, també lleugerament, mentre que l'ús d'altres llengües es manté força estable al llarg del temps. Els resultats d'aquesta investigació permeten classificar els informants en cinc grups segons les seves trajectòries lingüístiques i caracteritzar aquests grups per llengua inicial, nivell d'escolarització de la mare i àrea de residència. En aquest estudi no es troben correlacions significatives entre els usos lingüístics i el sexe dels alumnes.

El primer grup (un 40 % dels informants) inclou majoritàriament catalanoparlants inicials fills de mares amb estudis universitaris que viuen fora de l'àrea metropolitana de Barcelona, fan servir fonamentalment el català i no experimenten canvis significatius entre els tres moments temporals. El segon grup (32,9 % de la mostra) està format bàsicament per castellanoparlants inicials fills de mares amb estudis mitjans o bàsics que viuen a l'àrea metropolitana, fan servir sobretot el castellà i tampoc experimenten canvis significatius amb el temps. El tercer grup conté un 13,8 % dels alumnes que redueix substancialment el seu ús del català i, en canvi, augmenta el del castellà

de manera significativa amb els anys, especialment en el pas de primària a secundària i engloba majoritàriament castellanoparlants i bilingües inicials les mares dels quals sovint tenen estudis bàsics i viuen a l'àrea metropolitana. El penúltim grup d'informants (un 11,1 %) disminueixen el seu ús de castellà amb el temps i incrementen el de català i el formen sobretot bilingües inicials que viuen a l'AMB. L'últim grup només comprèn un 2,2 % de la mostra analitzada i inclou al·loglots inicials que fan un ús majoritari de llengües diferents al català i al castellà, sense canvis significatius entre els moments temporals estudiats, i també viuen a la regió metropolitana.

Així doncs, sembla que la llengua inicial dels informants, que també incloem com a variable en el nostre estudi, es relaciona directament amb els usos lingüístics que fan, mentre que el sexe no hi tindria gaire relació. Una vegada més, la regió de residència interactua amb les dades lingüístiques, de manera que caldrà tenir present aquesta possibilitat a l'hora d'interpretar els resultats, ja que no és una de les variables que mesurem i els nostres informants provenen tant de dins com de fora de l'àrea metropolitana de Barcelona.

També cal destacar que Vila, Ubalde, Bretxa i Comajoan-Colomé (2020) només troben un 22 % d'informants que declaren no haver experimentat cap canvi en el transcurs de temps que analitzen (quatre anys), si bé, en la majoria d'ocasions, els canvis que s'identifiquen són poc significatius. En aquest sentit, si les tries i els usos lingüístics es relacionen amb les eleccions lingüístiques a nivell micro, podria ser que els nostres informants mostrin canvis significatius en l'ús de les variables lingüístiques que observem segons la seva edat.

Així doncs, els estudis previs de RESOL que acabem de veure s'ocupen de les relacions entre variables independents com l'edat (en termes de moment temporal) o la llengua inicial i els usos i la confiança lingüística dels informants a partir de les dades declarades recollides mitjançant el qüestionari sociolingüístic que van completar tots els participants del projecte en les diverses onades. És per això que els seus resultats complementen, des d'una perspectiva macrolingüística, els objectius que perseguim en el nostre estudi, que se centra en els usos de variables lingüístiques concretes, és a dir, el nivell microlingüístic.

4.4. El català dels joves

El volum d'estudis que han intentat descriure el català col·loquial utilitzat pels joves no és tan elevat com ens agradaria d'acord amb els objectius d'aquesta tesi²⁰. De la mateixa manera, existeixen estudis que, tot i observar la llengua utilitzada pels parlants (pre)adolescents, adopten perspectives i objectius que difereixen dels del nostre treball. Per exemple, Bastardas (2012) tracta el comportament lingüístic d'aquest grup d'edat amb l'objectiu de fer propostes de polítiques lingüístiques per fomentar l'ús del català entre els joves.

Malgrat tot, l'interès acadèmic pel català oral col·loquial té certa tradició (vegeu, com a exemples de dues aproximacions molt diferents, Payrató, 1996 i Casanova, 1998) i hi ha estudis que ens ajuden a entendre quines són (algunes de) les característiques del català que parlen els joves. Una de les característiques per excel·lència d'aquesta modalitat és el contacte amb el castellà i comptem amb nombrosos estudis que analitzen les pràctiques d'infants i adolescents quant a alternances de codi (per exemple, Galindo Solé, 2006), les mudes lingüístiques (com González Balletbò, Pujolar Cos, Font Tanyà i Martínez Sanmartí, 2014), la freqüència d'ús d'una o altra llengua i els factors que hi influeixen (vegeu, com a mostra, Vila, 2012).

Ara bé, certament hi ha una manca aparent d'esforços centrats a descriure la llengua emprada en interaccions reals en català col·loquial entre joves o de joves amb parlants d'altres generacions. La seva naturalesa efímera i inestable expliquen en part aquesta manca d'estudis. Tot i això, l'explicació que fan Bernal i Sinner (2009: 9-10) de la perpetuació de la parla juvenil justifica l'interès acadèmic i social d'intentar descriure-la. Segons aquests autors, la llengua

²⁰ En paraules de Bernal i Sinner (2009: 8): "S'ha lamentat repetidament la falta de descripcions completes de la llengua oral contemporània catalana, i particularment de la llengua col·loquial (que s'ha explicat amb la preocupació fàcilment comprensible per la formulació d'una norma escrita de la llengua catalana), fet que, per cert, també dificulta l'estudi del contacte entre català i castellà i constitueix un problema particularment greu per a la distinció de les dues llengües en els registres més allunyats de la llengua de distància (vegeu Vila, 1996: 275; Sinner, 2004: 27; Sinner i Wieland, 2008; Sinner, 2008; Wieland, 2008a)".

juvenil no es renova completament amb cada nova incorporació d'un col·lectiu concret de parlants als grups d'edat que la utilitzen, sinó que “[e]ls grups etaris que fan ús de varietats identificables com a parla juvenil se solapen; els joves entren en els grups que les usen o surten d'ells a mesura que es fan més grans, però en diferents moments, és a dir, els grups es renoven a poc a poc”, de manera que la seva renovació gradual i perpetuació a través de grups generacionals no són mútuament excloents. Aquest procés explica per què certes expressions es vinculen a generacions concretes (com adjectius que expressen avaluació positiva que es posen de moda i són ràpidament reemplaçats per homòlegs quan es queden antiquats), mentre que altres estructures poden trobar-se en la parla juvenil de diverses generacions (com el fet d'utilitzar paraulotes, per exemple).

A més, aquest procés complementa l'explicació de per què és tan difícil estudiar la llengua juvenil. Els autors defensen que:

“Aquestes modificacions constants fan encara més difícil la determinació d'allò que es pot considerar propi de la parla juvenil i allò que és, simplement, llengua col·loquial, ja que part dels usos juvenils entren en els registres col·loquials generals a mesura que els parlants es fan grans i “exporten” elements del llenguatge juvenil, és a dir, els continuen usant fora dels registres exclusivament juvenils.” (Bernal i Sinner, 2009: 9).

Una altra de les característiques del llenguatge juvenil és la creativitat i les “transgressions lingüístiques” en forma de neologismes, manlleus d'altres llengües, calcs semàntics, etc. Bernal i Sinner (2009) analitzen el lèxic i els mecanismes de creació d'argot juvenil català recollits en dos llibres que compilen intervencions dels oients d'un programa de ràdio adreçat a adolescents emès entre el 2003 i el 2006. En total, recopilen 175 paraules i expressions de diverses categories gramaticals formades per sufixació, prefixació, conversió sintàctica, siglació, abreviació o processos semàntics, entre altres. La seva anàlisi els permet concloure que “el català juvenil de cap manera depèn únicament del castellà com a font enriquidora i que els joves fan usos de paraules clarament recognoscibles com a catalanes i òbviament formades seguint les pautes de la

gramàtica catalana”, si bé no hi ha dubte que “el castellà hi juga un paper molt important, ja que sembla servir de model que se segueix en les formacions catalanes” (Bernal i Sinner, 2009: 31).

Una iniciativa que sembla corroborar aquesta conclusió és el portal *Com ho diria?* (creat per Soler, Mas i Viladomat, s. d.), que recull unes 130 expressions pròpies del lèxic juvenil en català. Molts d'aquests ítems provenen del castellà (com *vacil·lon* o *huevo*), malgrat que també inclou paraules d'origen català (com *ou* o *hòstia*) o d'altres llengües (com *hype*).

Com s'ha indicat més amunt, el consum cultural dels joves és un altre dels factors que influencia la seva forma d'utilitzar la llengua. En aquest sentit, és especialment rellevant l'estudi de Wieland (2009a i 2009b), que analitza dos corpus de la primera dècada dels 2000: un de format per 40 converses entre grups reduïts d'adolescents d'entre 12 i 19 anys de Barcelona i un altre per programes de televisió i capítols d'una sèrie adreçats a joves i adolescents. Tot el material és de 2003, menys la sèrie de televisió, que és de 2001. El seu objectiu és comparar la parla dels joves i la utilitzada als mitjans de comunicació per explorar “quina quantitat de llenguatge juvenil contenen els programes destinats als adolescents i als joves adults de la televisió catalana” (Wieland, 2009a: 4).

Quant a la influència del castellà en la parla dels joves, l'autora troba que la parla dels adolescents reflecteix un grau més alt de contacte amb el castellà que la dels mitjans destinats a ells. Concretament, un dels trets que destaca és l'absència en aquests últims de marcadors discursius com *vale* i *bueno*, que explica per les connotacions negatives que reben de la gramàtica normativa i la posició de “garants de la llengua” dels mitjans de comunicació.

A més, pel que fa al contacte amb l'anglès, Wieland (2009a: 5) troba que “els mitjans de comunicació de masses, com a força innovadora, utilitzen molts anglicismes i, en canvi, els adolescents es mostren més aviat reservats envers els nous termes procedents de l'anglès”. Així doncs, arriba a la conclusió que, mentre que els joves veuen en el bilingüisme una “font d'ampliació de les possibilitats lingüístiques”, els mitjans de comunicació es troben en la situació

d'haver de mantenir un ús normativament acceptable de la llengua i, alhora, apropiat-se a com parlen els joves en el seu dia a dia.

Així doncs, les anàlisis de l'ús del català que fan els adolescents mostren que els fenòmens de contacte amb el castellà en són característiques intrínseques, si bé les expressions de nova creació que utilitzen també segueixen patrons propis del català. A més, els seus usos lingüístics estan clarament influïts per les seves relacions interpersonals però també per la seva exposició als mitjans de comunicació.

4.5. Context de la recerca: síntesi

En aquest capítol hem descrit el context sociolingüístic que emmarca les dades del nostre treball. Així, respecte la situació demolingüística dels adolescents a Catalunya, hem observat que el castellà té una influència clara sobre aquest sector de la població, ja sigui com a llengua inicial, com a llengua habitual o com a llengua de consum cultural. A més, caldrà tenir en compte que els individus mostren nivells de bilingüïtzació diferents. També hem vist que en els anys en què es va recollir el nostre corpus es van produir fenòmens migratoris que van afectar el paisatge lingüístic, que la frontera entre parla col·loquial i juvenil no sempre és fàcil d'establir i que els joves, a més de mecanismes que evidencien el contacte amb el castellà, utilitzen recursos de creació lèxica propis del català.

5 METODOLOGIA

5.1. El projecte RESOL

El nostre estudi analitza dades obtingudes en el projecte RESOL (*Resocialització i llengües: els efectes lingüístics del pas de l'educació primària a secundària en contextos plurilingües*). Aquest projecte es va dur a terme entre 2006 i 2009 amb el finançament del Ministeri de Ciència i Innovació. Es va continuar en els projectes RESOLSEC (*Resocialització lingüística a secundària*), que es va desenvolupar entre 2010 i 2012, i RESOL3 (*Plurilingüisme social i educació secundària (no obligatòria)*), entre 2012 i 2015. L'objectiu d'aquests projectes era estudiar l'evolució de les competències, els usos i les actituds lingüístics d'alumnes de diverses zones de parla catalana. Per assolir-lo, es van recollir dades dels mateixos informants dels municipis seleccionats en quatre onades: quan cursaven 6è de primària (és a dir, quan tenien aproximadament 11-12 anys), 1r d'ESO (12-13 anys), 4t d'ESO (15-16 anys) i 2n de batxillerat o cicles formatius (17-18 anys).

Així doncs, es tracta d'un disseny d'estudi panel o panel d'un total de més de 2.500 informants de localitats amb característiques poblacionals i demolingüístiques diferents, seleccionades perquè representen la diversitat sociolingüística del territori catalanòfon. En les quatre onades, es va aplicar un qüestionari sociolingüístic a tots els informants dels diversos municipis. Una submostra dels informants (els informants clau), de Manlleu, Mataró i Sabadell també van participar en dues proves de competència oral que es van enregistrar en vídeo. Malgrat això, a Sabadell només es van seleccionar informants d'origen amazic, que no tindrem en compte en aquest estudi.

A la primera prova de competència oral (que anomenem *prova acadèmica*; Illamola, 2015: 407), es demanava als informants que triessin una de les tres làmines que se'ls oferien, sense veure'n la cara amb els dibuixos que haurien de descriure i, a continuació, se'ls demanava una descripció el més detallada possible de les figures geomètriques que hi apareixien. Després de descriure-les, també es demanava a cada informant que expliqués els canvis necessaris per convertir les figures de la part inferior de la cartolina en les de la part superior.

En aquesta prova, que solia durar entre tres i cinc minuts, els informants van mostrar la seva competència acadèmica en una situació comunicativa més aviat formal i descontextualitzada.

A la segona prova (que anomenem *prova interactiva*; Illamola, 2015: 407), en canvi, els informants havien d'adoptar el paper de venedors d'una botiga d'animals i aconsellar i convèncer la clienta (l'entrevistadora) per tal que comprés l'animal més adequat a les necessitats i característiques del destinatari de l'animal, un familiar de la clienta que feia anys. Es dotava els informants de cinc cartolines amb informació dels animals que podia oferir: una imatge de l'animal, del seu menjar i hàbitat, i indicacions sobre les mides i el preu. Aquesta prova durava normalment entre deu i quinze minuts i permetia als informants mostrar les seves capacitats comunicatives en un context interaccional informal i espontani. Les intervencions de les entrevistadores s'havien de cenyir a un protocol establert prèviament per fomentar l'aparició dels mateixos temes (de manera que es pogués garantir la comparabilitat de les converses) i assegurar que influïa el mínim possible la producció de l'informant, tenint en compte, entre altres aspectes, que alguns dels criteris que s'avaluarien més endavant serien la fluïdesa i la riquesa lèxica o la capacitat d'argumentar o de posar-se en el paper de venedor o venedora (Illamola, 2015: 411-413).

En total, es van seleccionar 120 informants clau de les poblacions de Manlleu (60) i Mataró (60) per completar les proves acadèmica i interactiva. Les variables de selecció (llengua ambiental, llengua inicial i tipus de centre) i la distribució d'aquests informants s'especifiquen a Illamola (2015: 400-401). Tot seguit descrivim les dades que analitzem en aquest estudi, així com els passos i les tècniques que hem seguit per al tractament i l'anàlisi de les dades.

5.2. El corpus

En aquesta tesi, utilitzem una part de les dades recopilades durant la segona i la tercera onada del projecte RESOL als municipis catalans de Manlleu i Mataró, on es van enregistrar els informants clau durant les proves orals descrites anteriorment. Més concretament, tenim en compte les intervencions dels informants amb el català, el castellà o totes dues llengües com a llengua inicial (és a dir, no al·loglots) en la prova interactiva. Quant a les dades provinents dels qüestionaris sociolingüístics, només n'utilitzem les referents a la llengua inicial, el sexe i l'edat dels informants.

Com a part del projecte, les entrevistes amb els informants clau es van digitalitzar, codificar i transcriure. Les pautes de transcripció utilitzades (vegeu **Annex 2**) es basen en Payrató (1996 i 2010; cf. Illamola, 2015: 418; Flors-Mas, 2017: 182). Durant la preparació del corpus per a aquest estudi, es van comprovar les transcripcions i els enregistraments. En aquesta fase, es van detectar petites incorreccions en les transcripcions, que van ser modificades per reflectir més fidelment el contingut dels vídeos. D'altra banda, es van descartar els informants per als quals no es disposava de tots els enregistraments i les transcripcions corresponents als dos moments temporals analitzats (1r i 4t d'ESO). Diversos motius expliquen la manca de part d'aquestes dades. De vegades, un informant no va assistir a classe algun dels dies en què es van realitzar les entrevistes; d'altres, els investigadors van trobar que algun informant havia repetit curs o s'havia canviat de centre educatiu entre les dues entrevistes. En altres ocasions, no és possible enllaçar el qüestionari i l'enregistrament perquè les dades són incompletes o no coincideixen (a causa d'errors dels informants a l'hora d'emplenar el qüestionari, per exemple). Encara altres vegades, l'explicació rau en què es va produir algun problema tècnic amb els arxius (a causa de cintes defectuoses o de la pèrdua d'arxius). En total, el corpus analitzat en aquest estudi consta de 69 informants: 33 de Manlleu (18 nois i 15 noies) i 36 de Mataró (14 nois i 22 noies), distribuïts per llengua inicial i sexe com mostra la taula següent.

Taula 12. Informants de l'estudi per població, llengua inicial i sexe.

Població	Llengua inicial	Sexe	Informants
Manlleu	Català	Femení	L0004, L0037, L0050, L0052, L0060
		Masculí	L0020, L0021, L0025, L0027, L0035, L0038, L0043, L0051, L0057, L0058
	Català i castellà	Femení	L0005, L0032, L0046, L0048, L0049
		Masculí	L0001
	Castellà	Femení	L0002, L0007, L0008, L0018, L0031
		Masculí	L0003, L0006, L0009, L0012, L0016, L0019, L0047
Mataró	Català	Femení	M0054, M0094, M0462, M0469, M0914, M0917
		Masculí	M0045, M0048, M0051, M0453, M0470, M0918
	Català i castellà	Femení	M0297, M0723, M0737
		Masculí	M0925
	Castellà	Femení	M0043, M0049, M0096, M0105, M0289, M0295, M0471, M0472, M0719, M0725, M0728, M0732, M0734
		Masculí	M0089, M0090, M0290, M0293, M0456, M0718, M0722

En resum, al corpus d'estudi hi ha un total de 27 informants amb el català com a llengua inicial (11 noies i 16 nois), 10 informants bilingües inicials (8 noies i 2 nois) i 32 informants amb el castellà com a llengua inicial (18 noies i 14 nois). En termes percentuals, el corpus d'anàlisi el formen un 39,13 % d'informants amb el català, un 14,49 % amb totes dues llengües i un 46,38 % amb el castellà com a llengües inicials i aquest mateix percentatge de nois, un 46,38 %, respecte un 53,62 % de noies. Quant a la representativitat d'aquesta mostra, caldrà tenir present les diferències en la distribució dels informants entre el corpus d'anàlisi i les poblacions de Manlleu i Mataró que es mostren a les taules següents. Com es pot observar, les més importants són en la proporció de nois i noies de Mataró i en la distribució per llengua inicial als dos municipis, si bé l'ordenació dels grups és comparable a les dades de l'EULP 2008 i del corpus.

Taula 13. Distribució dels informants segons sexe al corpus d'anàlisi i a les poblacions de Manlleu i Mataró.

Sexe	Corpus d'anàlisi Manlleu	Manlleu 10 a 14 anys (2007)	Corpus d'anàlisi Mataró	Mataró 10 a 14 anys (2007)
M	54,55 %	52,01 %	38,89 %	51,79 %
F	45,45 %	47,99 %	61,11 %	48,21 %

Font: elaboració pròpia a partir d'Idescat (2018a i 2018b).

Taula 14. Distribució dels informants segons llengua inicial al corpus d'anàlisi i als àmbits territorials de Manlleu i Mataró.

Llengua inicial	Corpus d'anàlisi Manlleu	Comarques Centrals (2008)	Corpus d'anàlisi Mataró	Maresme (2008)
CA	45,45 %	50,9 %	33,33 %	37,2 %
ES	36,36 %	34,8 %	55,55 %	46,8 %
BI	18,18 %	No disp.	11,11 %	No disp.

Font: elaboració pròpia a partir d'Idescat (2019a i 2019b).

Pel que fa al material, com recull la taula següent, el corpus analitzat conté, d'una banda, un total de 138 enregistraments audiovisuals d'una duració mitjana de 13:28 minuts (amb una desviació estàndard menor a 0,01) i que sumen 30:59:10 hores i, de l'altra, les 138 transcripcions corresponents, que formen un total de 333.612 paraules i contenen una mitjana de 2.417 mots cadascuna (desviació estàndard de 679,77).

Taula 15. Dades sobre el corpus de l'estudi per edat, sexe i llengua inicial.

Variables		Nombre d'enregistraments	Durada mitjana	Hores totals	Mitjana de mots transcrits	Mots transcrits totals
Edat	12	138	00:13:19	15:18:49	2.128	146.884
	16	138	00:13:38	15:40:21	2.706	186.728
Sexe	M	64 (32+32)	00:14:05	15:01:30	2.572	164.621
	F	74 (37+37)	00:12:56	15:57:40	2.283	168.991
L1	CA	54 (27+27)	00:13:30	12:09:19	2.498	134.914
	ES	64 (32+32)	00:13:35	14:29:21	2.376	152.089
	BI	20 (10+10)	00:13:01	04:20:30	2.330	46.609

D'aquesta mostra, cal assenyalar que compleix amb els requisits generals que permeten assegurar resultats fiables, ja que superen el mínim representatiu de 30 informants de les lleis estadístiques generals (Guy, 1980, 1993) per als dos sexes. A més, es van prendre altres mesures per aconseguir prou dades lingüístiques, com establir un nombre mínim d'ocurrències per a les variables lingüístiques i les seves variants i seguir el principi de responsabilitat (*principle of accountability*, en anglès) defensat per Labov (1982: 30), que s'expliquen a l'apartat 5.5.

Quan tractàvem algunes de les crítiques habituals que han rebut alguns estudis sobre la variable sexe (apartat 2.3.1.4), s'ha fet notar la reflexió d'autors com Serrano (2008: 183) sobre la necessitat de tenir en compte el rol del context comunicatiu tant en el disseny de la mostra com en la interpretació dels resultats. Al nostre estudi, el rol del context comunicatiu s'ha tingut en compte, sobretot, en la inclusió de certes variables pragmàtiques i en la recopilació de les mostres, ja que es va intentar garantir al màxim possible la comparabilitat de les diverses gravacions controlant els diversos factors a l'abast de les entrevistadores (per exemple, mitjançant el seguiment d'un protocol preestablert, com s'ha explicat a l'apartat 5.1).

Finalment, pel que fa a la varietat de llengua que utilitzen els informants durant la prova interactiva, es tracta de varietats pròpies del català central no planificat, però no és fàcil determinar en quina posició exacta del contínuum formal-informal caldria situar les mostres que formen el corpus. D'una banda, provenen de situacions de parla dialogals (interaccions), no preparades i espontànies per part dels informants, en tant que tenien llibertat per controlar el tema de la conversa, inserir-hi elements fantàstics (com vacunes màgiques contra la caiguda de pèl, per exemple), mostrar-se més o menys empàtics i col·laboratius amb l'entrevistadora, etc. Ara bé, per part de les entrevistadores, com hem vist, havien de cenyir-se al temari i protocol preestablerts. A més, l'objectiu del joc de rol era produir una conversa que podrien mantenir un/a botiguer/a i una clienta, amb la qual cosa, la conversa objectiu podria veure's com una situació comunicativa que requereix un ús de la llengua a mig camí entre formal (entre

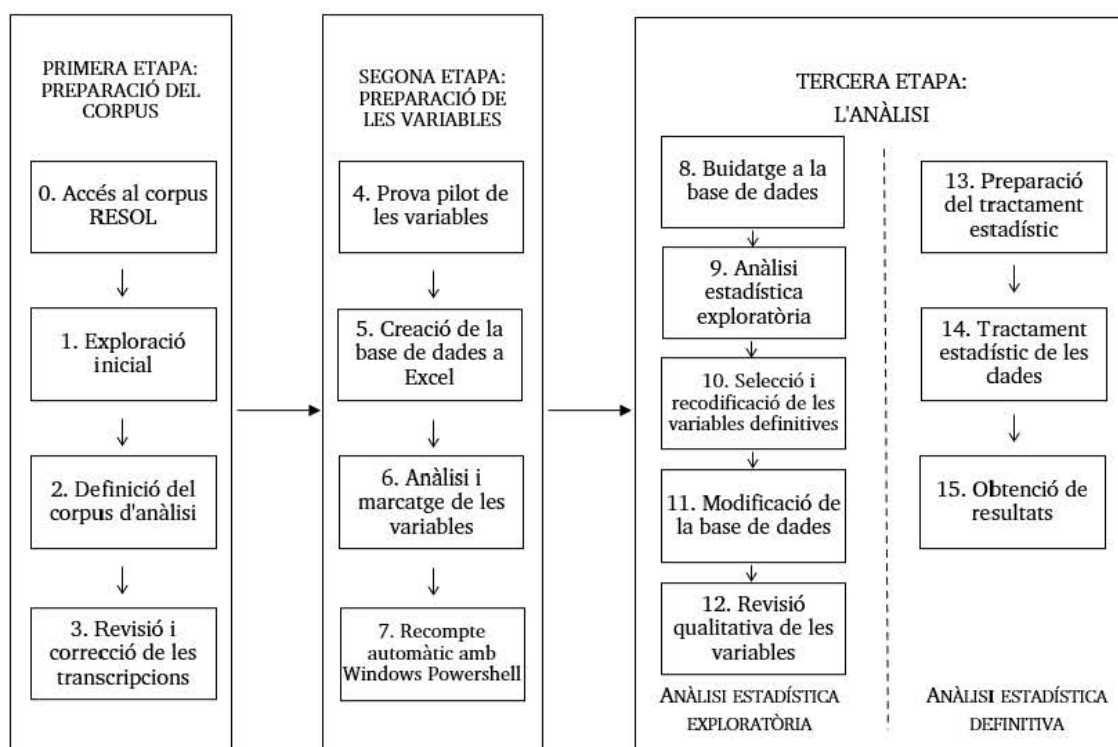
dos interlocutors desconeguts) i no formal (només un dels interlocutors parla des d'una identitat professional). D'altra banda, el context real (fora del joc de rol) on té lloc la interacció entre els informants i les entrevistadores és, al capdavant, el d'una prova de competència oral que es realitza en un centre educatiu. En qualsevol cas, sembla clar que, pel que fa al registre de la llengua oral utilitzada pels adolescents durant la prova interactiva, se situaria lluny dels extrems del contínuum formal-informal.

5.3. Disseny de l'estudi

Com s'ha avançat anteriorment, aquest és un estudi panel, ja que observa el comportament lingüístic de 69 informants (32 nois i 37 noies) en la mateixa situació comunicativa en dos moments temporals diferents (quan els informants cursaven 1r i 4t de l'ESO). Bretxa (2014: 123) recull els principals avantatges i desavantatges d'aquests estudis. En primer lloc, tot i que requereixen una gran inversió de temps i recursos, permeten diferenciar les característiques transitòries i permanents dels fenòmens que s'estudien. En segon lloc, si bé sotmetre els mateixos informants a les proves més d'una vegada pot fer que modifiquin el seu comportament o les seves respostes, aquest procés permet descriure l'evolució entre informants diferents (interindividuals) com també entre les diverses respostes de cada individu (intraindividuals) i identificar els patrons que expliquin aquests canvis. En tercer lloc, cal tenir en compte el possible impacte negatiu en la validesa de l'estudi que pugui tenir el desgast de la mostra amb el temps, és a dir, el fet que certs informants de la primera onada no participin en onades següents. Però, alhora, “[l]es dades longitudinals també contribueixen a la identificació de les causes dels fenòmens socials, o almenys ajuden a permetre que es puguin identificar els antecedents i les conseqüències” (Bretxa, 2014: 123). El balanç d'aquests avantatges i inconvenients és que el disseny d'aquest estudi és el més adequat per analitzar l'evolució de les variables lingüístiques en relació amb les variables independents que ens ocupen, especialment, l'edat dels informants.

El nostre estudi es basa sobretot en mètodes quantitatius, si bé per a l'etiquetatge de les realitzacions de certes variables, com veurem en aquest capítol, va caldre adoptar decisions metodològiques des d'una perspectiva més aviat qualitativa. Tot i això, cal subratllar que la combinació de mètodes qualitatius i quantitatius permet als investigadors beneficiar-se dels punts forts de cada aproximació (Angouri, 2010: 30-31). D'una banda, entenem la metodologia qualitativa com la que se centra en aspectes que no es poden quantificar per tal d'adquirir una comprensió global del fenomen que s'estudia mitjançant la inducció. De l'altra, definim l'anàlisi quantitativa com la que se centra en aspectes observables susceptibles de quantificació (variables), que l'investigador mesura i controla, que utilitza l'estadística per a l'anàlisi de les dades i que es fonamenta en el raonament deductiu (D. A., 2008a i 2008b; Reguera, 2010: 75-76). En el nostre cas, però, es va dividir l'estudi en una etapa inicial, en què es va explorar i preparar el corpus, una segona etapa, en què es van seleccionar i preparar les variables lingüístiques i una etapa final, en què es va dur a terme l'anàlisi estadística de les variables.

Figura 11. Procés de preparació, anàlisi i explotació del corpus d'anàlisi.



5.3.1. Primera etapa: preparació del corpus

Com mostra la Figura 11, la primera etapa del projecte va consistir a accedir i explorar el material disponible per delimitar el corpus d'anàlisi i garantir-ne la qualitat. Així doncs, a les primeres fases es va accedir i explorar el corpus RESOL que, recordem, conté diverses bases de dades dels informants, qüestionaris sociolingüístics de tots ells, gravacions de les proves acadèmica i interactiva realitzades pels informants clau de tres poblacions catalanes (Manlleu, Mataró i Sabadell) i les transcripcions d'aquestes gravacions. A la fase següent, com s'ha explicat anteriorment, es va limitar el corpus d'anàlisi a les gravacions i transcripcions de la prova interactiva dels informants de Manlleu i Mataró per als quals es disposava del material corresponent als dos temps de l'estudi (1r i 4t d'ESO). D'aquesta manera, es van descartar tant els qüestionaris sociolingüístics²¹ com les gravacions i/o transcripcions d'informants al·loglots i d'aquells per als quals no es disposava de material provinent de les dues onades del projecte que ens proposàvem analitzar. A continuació, es van realitzar diverses revisions exhaustives del material audiovisual i de les transcripcions, tal com s'ha explicat més amunt, per garantir que reflectien fidelment el contingut de les gravacions.

²¹ Amb excepció de la informació relativa a la llengua inicial i el sexe dels informants (que, a la cinquena fase, es va introduir a la base de dades de l'estudi).

5.3.2. Segona etapa: les variables

5.3.2.1. Les variables independents

A l'apartat **2.3** ens hem ocupat de les variables independents més habituals de la sociolingüística de la variació. Com s'ha exposat al capítol 3, el nostre estudi utilitza tres d'aquestes variables, l'edat, el sexe i la llengua inicial dels informants. La selecció de variables independents s'ha fet tenint en compte les consideracions sobre cadascuna de les variables que exposem tot seguit.

5.3.2.1.1. Edat

Com hem vist al subapartat **2.3.2**, els estudis disponibles sobre l'adquisició lingüística demostren que aquest procés s'estén més enllà de les primeres etapes de la infantesa i que durant l'adolescència es produeixen desenvolupaments importants en els nivells lèxic, sintàctic i pragmàtic. Tot i això, els treballs centrats en l'etapa adolescents no són tan nombrosos com els disponibles sobre etapes més prèvies. A més a més, estudis com el de Vila, Ubalde, Bretxa i Comajoan-Colomé (2020: 1160) indiquen que els adolescents juguen un paper destacat en processos sociolingüístics com el canvi i la substitució.

Hem observat, igualment, que el volum d'estudis sobre les varietats catalanes emprades per parlants adolescents són escassos i que, en general, a la sociolingüística variacionista són més freqüents els estudis rèplica que els panel, ja que aquests darrers requereixen més temps i recursos. Cal fer notar també que molts estudis sobre la variació lingüística se centren en informants adults, de manera que sovint els parlants adolescents s'obvien o s'agrupen sota categories que no permeten observar l'evolució en les habilitats lingüístiques que té lloc a la darrera etapa de la infantesa. Per tot això, la comparació de mostres de parla dels mateixos informants en dues edats diferents (12-13 anys i 15-16 anys) obtingudes en el marc del projecte RESOL presenta l'oportunitat d'observar diferències atribuïbles a la gradació per edats i d'ampliar la literatura especialitzada existent sobre llenguatge adolescent oral no planificat i català.

5.3.2.1.2. Sexe

Les dades que formen el corpus, com s'ha explicat a l'apartat **5.2**, provenen d'un volum elevat d'interaccions orals entre els informants i les entrevistadores en una situació comunicativa controlada, però no planificada. Aquest corpus, per tant, permet comprovar les afirmacions quant a les diferències per sexe realitzades, d'una banda, pels estudis quantitius (el patró de gènere labovità segons el qual les dones lideren els canvis lingüístics i l'ús de formes considerades estàndard) i, de l'altra, pels treballs sobre els estils interaccionals d'homes i dones (subapartat **2.3.1.3**). A més, permet també posar a prova la idea que proposava Tuson (1998: 216) en relació al fet que, possiblement, els parlants adolescents mostrin estils conversacionals diferents als dels parlants adults.

S'ha observat també que la variable sexe ha resultat rellevant en l'estudi de tots els nivells lingüístics (com apunta Bastida Serra, 2017: 93) i que en la nostra investigació contemplem variables lèxiques, morfosintàctiques i pragmàtiques. Així doncs, la variable sexe podria, en principi, mostrar correlacions amb totes les variables lingüístiques, independentment del nivell al qual pertanyin.

De la mateixa manera, cal tenir en compte que el gruix d'estudis variacionistes sobre la llengua catalana, com hem vist al subapartat **2.2.4.3**, es va realitzar a final del segle passat (si bé, com hem destacat, aquesta disciplina segueix produint recerca a casa nostra). En aquest sentit, el corpus del nostre estudi permet comprovar, des d'una perspectiva quantitativa, si una mostra obtinguda més recentment també conté diferències quant al comportament lingüístic de parlants de sexe diferent.

5.3.2.1.3. Llengua inicial

Al subapartat **2.3.4** hem subratllat la necessitat de comptar amb estudis sobre llengües menys estudiades que l'anglesa i comunitats de parla diverses i lingüísticament complexes (no monolingües) per tal de revertir biaixos existents en la sociolingüística actual i d'ampliar l'abast del nostre coneixement sobre els usos lingüístics dels parlants. La societat catalana constitueix, com hem vist, un escenari adient per dur a terme estudis que contribueixin a aquests objectius. En aquest context sociolingüístic, com s'ha exposat al capítol 4, l'anàlisi de la llengua inicial dels parlants resulta una variable clau (Sorolla i Flors-Mas, 2020: 55; Flors-Mas, Sorolla, Pradilla i Vila, 2021: 33). Per això, la nostra investigació considera la variable llengua inicial com a factor inherent als adolescents analitzats que podria condicionar-ne l'actuació lingüística. Cal tenir en compte, a més, que la llengua inicial presenta un avantatge clar sobre altres variables rellevants en el context català, com la llengua habitual o la llengua d'identificació, pel que fa als objectius del nostre estudi (comparar els mateixos informants en dos moments temporals diferents): que es manté estable al llarg del temps.

5.3.2.2. Altres variables independents

Tot i que els estudis variacionistes han utilitzat, com hem vist, un gran nombre de variables independents, a continuació presentem de manera succinta alguns dels aspectes que es van tenir en compte a l'hora de decidir que les següents no eren tan adients per al nostre estudi com les tres que acabem de tractar. Ara bé, aquestes variables (i altres) es podrien incloure en estudis futurs comparables al que presentem aquí (com s'explica al capítol 8).

5.3.2.2.1. Gènere

Al capítol 2 hem vist que part de l'evolució que s'ha produït en l'estudi de la variació lingüística des dels seus inicis ha girat entorn de la conceptualització de les característiques dels parlants, entre elles, com hem vist, s'ha tendit a qüestionar l'ús de la variable sexe i a afavorir aproximacions a la variació que, en canvi, tenen en compte el gènere dels parlants (subapartat **2.3.1.4**). El gènere és un constructe social i el pes d'aquesta variable en els estudis sobre la variació, sobretot de la tercera onada (subapartat **2.2.4.2**), respon al canvi de perspectiva des de la qual un bon nombre d'acadèmics s'aproximen a les relacions entre els fenòmens lingüístics i les identitats dels parlants. Com hem vist, aquest canvi ha afavorit visions construccionistes que tenen en compte l'agentivitat dels parlants i el rol de patrons culturals amb els quals els individus poden interactuar de formes diferents mitjançant els seus usos lingüístics. És en aquest context acadèmic que la variable gènere s'entén com una "dimensió sociocultural adquirida" (Serrano, 2008: 177) i pren un paper destacat, especialment, en treballs qualitius.

Com hem vist, però, aquesta variable exigeix informació que sobrepassa la disponible sobre la nostra mostra, ja que no es va recollir informació sobre la identitat de gènere dels informants, és a dir, el seu posicionament sobre el grau amb què s'identifiquen, o no, amb els rols socials atribuïts generalment a homes i dones. Som conscients que aquesta informació de la qual no disposem pot coincidir (o no) amb el sexe dels informants (el conjunt de característiques biològiques a partir de les quals s'estableix la distinció entre homes i dones²²), la qual cosa constitueix la raó principal per la qual ens abstenim d'incloure la variable gènere a l'estudi.

Podria argumentar-se que una anàlisi minuciosa dels usos lingüístics que fan els informants podria servir per dividir la mostra en relació amb la variable gènere. No obstant això, la validesa d'aquesta categorització podria no ser suficient, per estar basada en observacions subjectives de la investigadora i, a més, la mida del corpus d'anàlisi comprometia la viabilitat d'aquesta aproximació qualitativa que,

²² Basem les definicions dels conceptes *identitat de gènere* i *sexe* en les recollides a Breu (2013).

de no ser prou meticulosa, perdria rigurositat i fiabilitat. Com explica Precht (2008: 91), l'estudi de la variable gènere exigeix la identificació “manual” de característiques masculines i femenines (i no binàries, hi afegiríem) i prendre en consideració el context en què apareixen per tal de fer afirmacions sobre el gènere dels informants, un procediment costós que difícilment es pot seguir en corpus de grans dimensions.

5.3.2.2.2. Classe social

La variable classe social és, com hem observat (subapartat **2.3.3**), una de les més productives a la literatura sobre la variació sociolingüística, però, alhora, també una de les menys compreses pels lingüistes²³ i una de les més complexes (Wardhaugh, 2006: 148). Alguns dels factors que se solen tenir en compte en estudis variacionistes per categoritzar els parlants en grups de classe social són la seva professió, els ingressos econòmics i el nivell educatiu. En canvi, aquests factors no són directament aplicables a la nostra mostra, ja que els parlants que analitzem són menors d'edat i, per tant, en comptes de considerar la seva professió o el seu nivell d'ingressos, hauríem de tenir dades corresponents a aquests factors relatius al seu entorn familiar, circumstància que no és la nostra. Quant al nivell educatiu, vegeu el subapartat següent (**5.3.2.2.3**).

A més, tal com hem observat al subapartat **2.3.5**, la literatura prèvia mostra que la variable classe social pot presentar interaccions complexes amb un nombre considerable de variables independents, com una de les que, per les característiques del context sociolingüístic català, resulta més rellevant per al nostre estudi, la llengua inicial. Ja hem comprovat que estudis com el de nussbaum mostren que, a Catalunya, la classe social, la zona geogràfica i la llengua inicial estan fortament relacionades, a causa de la influència desigual que han tingut sobre el territori les diverses onades migratòries que s'han produït. A més, Stuart-Smith i Haddican (2010: 301) expliquen que la classe social té més rellevància per explicar els usos lingüístics no estàndards d'algunes societats que d'altres. En el cas que ens ocupa, tenint en compte els resultats de

²³ “Thus, this variable is universally used and extremely productive, although linguists can lay little claim to understanding it” (Ash, 2013: 350).

Flors-Mas, Sorolla, Pradilla i Vila (2021), podria ser que la llengua inicial mostrés més capacitat per explicar els usos lingüístics a la societat catalana que altres variables, inclosa la classe social. Malgrat tot, seria interessant veure si aquestes possibilitats es confirmen en futurs estudis que considerin totes dues variables.

5.3.2.2.3. Nivell educatiu

Quant al nivell educatiu, a més de mostrar-se, a diversos estudis sobre la variació, estretament correlacionada amb la classe social o la varietat dialectal, en el nostre cas no tenia sentit incloure-la a l'estudi, ja que investiguem els mateixos informants en dos moments temporals diferents, de manera que tots es trobaven, en cadascun dels dos moments en el mateix nivell educatiu. D'altra banda, comparar els dos temps de recollida de dades entre ells tenia més sentit, al nostre parer, des de la perspectiva de la variable edat que no pas de la variable nivell educatiu, per les raons que s'han exposat anteriorment (subapartat 5.3.2.1.1).

5.3.2.2.4. Varietat dialectal

Mentre que el coneixement sobre la relació entre l'edat, el sexe i la llengua inicial i les variables dependents que tractem en aquesta tesi i les varietats catalanes és escàs, els estudis de la variació en català han generat un volum comparativament destacable de dades sobre la variació diatòpica. Així doncs, malgrat que els informants del nostre estudi provenen de dues poblacions diferents, no explorem variables dialectals, ja que en totes dues s'hi parla el mateix dialecte (el central) i, si bé l'exploració inicial del corpus ha revelat l'existència d'alguns trets possiblement subdialectals o distintius de les dues zones geogràfiques on es van recollir les dades (per exemple, la negació amb *pas*), considerem que els lectors meta d'aquest treball d'investigació (els lingüistes forenses interessats en trets lingüístics aplicables a l'elaboració de perfils sociolingüístics de parlants adolescents) poden recórrer a altres obres que es dediquen àmpliament a la descripció de trets dialectals. Per exemple, poden consultar el projecte *Atlas Lingüístic del Domini Català* de Joan Veny i Lídia Pons Griera, Veny i Massanell (2015) o Massanell (2020).

En definitiva, les tres variables independents que sí que incloem a l'estudi ens permeten contribuir al coneixement actualment disponible sobre la relació entre, d'una banda, l'edat, el sexe i la llengua inicial i, de l'altra, la vintena de trets lingüístics que tractem a continuació i que apareixen a la nostra mostra de català oral no planificat utilitzat per informants adolescents.

5.3.2.3. Les variables dependents

Com s'ha descrit al subapartat **2.2.4.1.2**, la noció de variable és central a la sociolingüística variacionista i aquest estudi analitza diverses variables lingüístiques (és a dir, dependents) en relació amb tres variables extralingüístiques (o independents): l'edat, el sexe i la llengua inicial dels informants.

Les variables lingüístiques observades compleixen els requisits o característiques fonamentals llistats en estudis previs amb objectius semblants, com Queralt (2015). Aquestes característiques inclouen el fet que es tracti de variables freqüents i estratificades (cf. Labov, 1972), que siguin relativament fàcils d'extreure i de calcular (Nolan, 1983: 11), que les variants siguin intercanviables en certs contextos (Tagliamonte, 2006: 73), que siguin el més independents possible d'altres variables (Rose, 2002: 52), que els parlants en siguin el menys conscients possible (Labov, 1963) i que estiguin definides en termes lingüístics de manera clara i precisa (Britain, 1991: 56-63). A més, també s'ha de tenir en compte que la variabilitat en el seu ús per part de parlants diferents sigui més gran que la variabilitat entre les produccions d'un mateix parlant (Nolan, 1983: 11) i que com més variants presentin més probable serà que el seu ús es correspongui a preferències individuals dels parlants (Koppel, Akiva i Dagan, 2006).

La selecció de les variables lingüístiques s'ha dut a terme a partir de la conjunció de dos factors. En primer lloc, la majoria de variables dependents del nostre estudi provenen de les utilitzades en treballs previs ressenyats a la bibliografia exposada en els subapartats anteriors. Només dues de les tretze variables han estat dissenyades amb independència de la bibliografia consultada i provenen de l'observació inicial del corpus. Per a la resta, ens nodrim de les variables

utilitzades al projecte ISI, Queralt (2015), Garayzábal, Queralt i Reigosa (2019: 64) i dels estudis sobre la variació del català, les diferències lingüístiques entre homes i dones, i l'adquisició lingüística a l'adolescència (vegeu els subapartats següents per a una descripció detallada dels estudis previs rellevants per a cada variable). En segon lloc, com també s'explica més detalladament al subapartat **5.3.2.4**, s'ha realitzat una anàlisi preliminar del corpus d'anàlisi per tal de testar les possibles variables identificades a la bibliografia i comprovar-ne l'aplicabilitat i l'aparició al corpus en una freqüència suficient com per garantir que el tractament estadístic de les dades seria possible. Així doncs, a més de les variables no rellevants per al nostre estudi, les que van presentar menys de deu ocurrències al recompte final van ser descartades.

Malauradament, no ha estat possible utilitzar en el nostre estudi moltes de les variables enumerades a Queralt (2015) i en anàlisis d'autoria segons Garayzábal, Queralt i Reigosa (2019: 64) que hem vist al subapartat **2.1.4.1.1**. En primer lloc, pel que fa a les variables de complexitat, com hem observat, Garayzábal, Queralt i Reigosa (2019: 64) adverteixen que depenen de manera significativa de la longitud dels textos. Així doncs, donada la variabilitat entre les transcripcions del corpus, no s'inclouen a l'estudi per evitar resultats enganyosos o poc fiables. A més, algunes no són aplicables a produccions orals (com el nombre i la longitud dels paràgrafs).

En segon lloc, quant a les variables lèxiques, de nou, moltes de les enumerades no són aplicables al nostre corpus, tant a causa de diferències en la modalitat (oral, i no escrita) com per diferències en la llengua d'anàlisi (per exemple, en català no es pot observar la terminació en *-ra* o *-se* del pretèrit imperfecte de subjuntiu com castellà) com, encara, a causa del pes del tema de la tasca interactiva (la venda d'animals de companyia) sobre les eleccions lèxiques (no apareixen, per exemple, paraules malsonants de forma comparable a com ho poden fer en textos anònims rellevants per a investigacions delictives).

El tercer bloc de variables enumerades per les autores és el de variables sintàctiques. Com s'ha assenyalat, aquestes solen mostrar resultats més fiables que les lèxiques. Amb tot, si bé mesures com el tipus d'oracions (simples i

complexes, subordinades, coordinades i juxtaposades) o el tipus de sintagmes són relativament fàcils d'obtenir en textos anònims habituals en investigacions de possibles delictes (que no solen superar les 1.000 paraules), aquesta mena de variables presenten moltes més dificultats de computació en un corpus com el del nostre estudi. Les dades que analitzem, com hem vist a l'apartat anterior (5.3.1), provenen de produccions orals no planificades i, a més, l'extensió total del corpus d'anàlisi és molt més elevada que l'habitual en tasques d'atribució d'autoria.

Finalment, en relació a les variables pragmàtiques, algunes, com l'ús dels signes de puntuació o l'expressió de l'èmfasi mitjançant l'ús de majúscules o de la puntuació, estan dissenyades per a mostres escrites i no són aplicables a mostres orals no planificades. Altres, com el tractament de *tu* o *vostè*, no apareixen un nombre suficient de vegades al corpus d'anàlisi. Encara d'altres, com els actes de parla, requereixen, com afirmen Garayzábal, Queralt i Reigosa (2019: 64) d'anàlisis totalment manuals i, per tant, resulten molt costoses en corpus tan extensos com el que utilitzem a l'estudi.

Tot i això, seguint Garayzábal, Queralt i Reigosa (2019), les variables que finalment sí que han estat incloses a l'estudi s'han agrupat segons el nivell lingüístic al qual corresponen. Així doncs, s'han observat variables pragmàtiques, morfosintàctiques i lèxiques. A continuació, s'inclouen a la Taula 16 les variables dependents finals per nivell lingüístic, les variants que les formen i les variables independents que s'espera que s'hi correlacionin a partir de la revisió de la bibliografia rellevant per a l'estudi de cada fenomen lingüístic i a partir de l'anàlisi qualitativa preliminar de cinc transcripcions de cada grup del corpus que s'explica a l'apartat 5.3 (sexe masculí o femení; 12-13 o 15-16 anys; català, castellà o totes dues llengües com a llengua inicial). L'estudi de les variables per a les quals no es mostren variants consisteix en el recompte del total de casos que apareixen a cada gravació. Com s'ha avançat al capítol 3, les correlacions que, segons la bibliografia consultada, seria esperable identificar entre les variables independents i cadascuna de les variables dependents de l'estudi

constitueixen les nostres subhipòtesis (trobeu-ne un resum al subapartat 5.3.2.3.4).

Les referències bibliogràfiques indiquen els treballs previs que ja s'han ocupat de les variables o que indiquen que podria haver-hi correlacions entre les variables independents i dependents. Algunes d'aquestes obres es comenten a continuació i altres, a la interpretació dels resultats (capítol 7). En el cas de V17, s'ha inclòs a l'estudi tot i l'absència de bibliografia prèvia rellevant i, en els de V16 i V41, es considera la possibilitat d'identificar correlacions estadísticament significatives amb la llengua inicial dels parlants, malgrat que no hem trobat estudis previs sobre aquesta possible relació, perquè el corpus d'anàlisi proporciona l'oportunitat de comprovar si els catalanoparlants inicials, pel fet de tenir com a primera llengua la que s'utilitza a la prova, produeixen més preguntes i més referències a colors durant la interacció que els seus companys bilingües i castellanoparlants inicials.

Taula 16. Variables dependents amb les seves variants i les variables independents que s'espera que s'hi correlacionin.

Variables pragmàtiques		Variants	S'espera que correlacioni amb...
V3: Doblets de partícules discursives Vila (1998: 271)	V3b: <i>a ver / a veure</i>	V3b1: <i>a ver</i> V3b2: <i>a veure</i> V3b3: <i>a ve(u)re</i> V3b4: <i>aviam</i>	Sexe (Tuson, 1998; Kring i Gordon, 1998; Baron i Ling, 2011; Bastida Serra, 2017) Llengua inicial (Vila, 1998: 271)
	V3c: <i>bueno / bé</i>	V3c1: <i>bueno</i> V3c2: <i>bé</i>	
	V3f: <i>después / després</i>	V3f1: <i>después</i> V3f2: <i>després</i>	
	V3h: <i>hombre / home</i>	V3h1: <i>hombre</i> V3h2: <i>home</i>	
	V3i: <i>o sea / o sigui</i>	V3i1: <i>o sea</i> V3i2: <i>o sigui</i>	
	V3j: <i>pero / però</i>	V3j1: <i>pero</i> V3j2: <i>però</i>	

		V3j3: <i>p(e)rò</i> V3j4: <i>p(er)ò</i>	
	V3k: <i>pues / doncs</i>	V3k1: <i>pues</i> V3k2: <i>pos</i> V3k3: <i>doncs</i>	
	V3l: <i>vale / d'acord</i>	V3l1: <i>vale</i> V3l2: <i>val</i> V3l3: <i>d'acord</i>	
V5: Obertura de la interacció (Queralt, 2015; Garayzábal, Queralt i Reigosa, 2019: 64)		V5a: <i>hola</i> V5b: <i>bon dia</i> V5c: <i>bona tarda</i> V5d: <i>hola bon dia</i>	Edat (Clark, 2003: 282) Llengua inicial (Arenas Benyto, 2005: 88)
V6: Heterorepetició			Edat (Rydland i Aukrust, 2005; Čurković Kalebić, 2016)
Variables morfosintàctiques		Variants	S'espera que correlacioni amb...
V14: Posició dels pronoms clítics (PC) en construccions perifràstiques	V14a: Proclisi	V14a1: PC + <i>tenir que / de</i> + INF V14a2: PC + <i>haver de</i> + INF V14a3: PC + <i>poder</i> + INF V14a4: PC + <i>anar a</i> + INF	Edat (Illamola i Queralt, 2020) Sexe (Illamola i Vila, 2015; Illamola i Queralt, 2020)
	V14b: Enclisi	V14b1: <i>tenir que / de</i> + INF + PC V14b2: <i>haver de</i> + INF + PC V14b3: <i>poder</i> + INF + PC	Llengua inicial (Illamola i Vila, 2015; Illamola i Queralt, 2020)
	V14c: Doble	PC + perífrasi verbal + PC	
V16: Nombre de preguntes no retòriques (Queralt, 2015)			Edat (Larson i McKinley, 1998) Sexe (Tuson, 1998: 216; Pearson, Turner i Todd-Mancillas, 1991: 115) Llengua inicial
V17: Tipus de preguntes	V17a: Oberta V17b: Tancada		Edat

Metodologia

		<p>Sexe</p> <p>Llengua inicial</p>
<p>V18: Construccions condicionals (Queralt, 2015)</p>	<p>V18b: <i>si</i></p> <p>V18c: <i>mentre(s)</i></p> <p>V18f: <i>a no ser que</i></p> <p>V18g: <i>en tot cas</i></p>	<p>Edat (Aparici Aznar i Noguera Llopart, 2012; Nippold, Frantz-Kaspar i Vigeland, 2017)</p>
<p>V22: Perífrasis d'obligació (Queralt, 2015)</p>	<p>V22a: <i>tenir que</i> + INF</p> <p>V22b: <i>haver de</i> + INF</p> <p>V22c: <i>tenir de</i> + INF</p> <p>V22e: <i>caldre</i> + INF</p>	<p>Llengua inicial (Mier, 1986; Stokes, 2015)</p>
<p>V23: Determinants demostratius</p>	<p>V23a: <i>aquest</i> + N</p> <p>V23b: <i>aquell</i> + N</p> <p>V23c: <i>el</i> + N + <i>aquest</i></p> <p>V23d: <i>el</i> + N + <i>aquell</i></p> <p>V23e: <i>aquest</i></p> <p>V23f: <i>aquell</i></p> <p>V23g: <i>(un +) N + d'aquests</i></p> <p>V23h: <i>(un +) N + d'aquells</i></p>	<p>Edat (Diessel, 2006: 469)</p> <p>Sexe (Fitó, 2009: 55)</p> <p>Llengua inicial (Fitó, 2009: 55)</p>
<p>V25: Pronoms clítics de datiu plural</p>	<p>V25a: <i>els hi</i> + V / V + <i>-els-hi</i> / V + <i>'ls-hi</i> / V + <i>-los-hi</i></p> <p>V25b: <i>els</i> + V / V + <i>-los</i> / V + <i>'ls</i></p> <p>V25c: <i>li</i> + V / V + <i>-li</i></p> <p>V25d: <i>lis</i> + V / V + <i>-lis</i></p>	<p>Llengua inicial (Pineda, 2018: 257)</p>
<p>Variables lèxiques</p>	<p>Variants</p>	<p>S'espera que correlacioni amb...</p>
<p>V38: Fenòmens de contacte de llengües (Vila, 1998)</p>	<p>V38a: Importacions</p> <p>V38b: Híbrids</p> <p>V38c: Calcs</p> <p>V38d: Altres</p>	<p>Edat (Nippold, 2004)</p> <p>Sexe (Labov, 2001: 266)</p> <p>Llengua inicial (Baldaquí Escandell, 2002: 47)</p>
<p>V41: Colors</p>	<p>V41a: <i>color [blau]</i></p> <p>V41b: <i>[blau]</i></p>	<p>Edat (Al-Jehani, 1990)</p> <p>Sexe (Al-Jehani, 1990; Paredes García,</p>

		2006) Llengua inicial
V45: Nombre d'adverbis acabats en <i>-ment</i>		Edat (Barriga Villanueva, 1994; Nippold, 1998) Sexe (Tuson, 1998)

Tot seguit es tracta amb més deteniment cadascuna de les variables dependents en relació a la bibliografia disponible i la selecció de les variants d'acord a les característiques del corpus d'anàlisi.

5.3.2.3.1. Les variables pragmàtiques

Com mostra la Taula 16, s'han inclòs tres variables pragmàtiques a l'estudi. Concretament, la variable V3 Doblets de partícules discursives, V5 Obertura de la interacció i V6 Heterorepetició.

A. V3 Doblets de partícules discursives

La primera d'aquestes variables recull el nombre de realitzacions de diversos elements lingüístics (adverbis, interjeccions, formes verbals, etc.) que agrupem sota l'etiqueta de *partícules discursives* (seguint la terminologia proposada per Briz, Pons i Portolés, 2008). S'ha optat per una classificació de les partícules merament formal per tal de garantir que el recompte es pogués dur a terme de la forma més sistemàtica possible, ja que el gran nombre de treballs que proposen classificacions funcionals diferents (p. ex. Martín Zorraquino i Portolés Lázaro, 1999; Pons Bordería, 2006, o López i Borreguero, 2010) complicaria aquesta tasca en un corpus de les dimensions del nostre i comprometria la validesa de l'anàlisi quantitativa.

A més, les classificacions funcionals de les partícules discursives presenten certs desavantatges que no són presents en les classificacions formals, com el fet que els límits entre una funció i una altra no sempre són nítids i que sovint una sola partícula pot realitzar més d'una funció alhora en un context determinat (López i Borreguero, 2010: 442-3).

Les partícules discursives (o marcadors discursius²⁴) han atret l'atenció de sociolingüistes de diversos corrents en moltes ocasions. Compareu, per exemple, les aportacions a Loureda Lamas i Acín-Villa (2010), centrades en aspectes morfològics (Martín Zorraquino, 2010), sintàctics (Llamas Saiz, 2010) o semàntics (Murillo Ornat, 2010) d'aquestes unitats o que les estudien des de l'estudi de la variació (Carbonero Cano i Santana Marrero, 2010) o de l'evolució diacrònica de la llengua (Pons-Rodríguez, 2010). Comptem també amb nombrosos estudis que en proposen classificacions (p. ex. Halliday i Hasan, 1976; Martin, 1992; Martín Zorraquino i Portolés Lázaro, 1999) i d'altres que estudien les diverses funcions i usos de marcadors discursius concrets (per nombrar-ne només alguns, Cuenca i Marín, 2012 sobre *(és) clar*; Marín i Cuenca, 2012 sobre *és que*, o Guevara, 2015 sobre *pues*), així com treballs que els analitzen tant en produccions orals (p. ex. González, 2001, 2004) com en escrites (Hyland, 2000, per exemple). No obstant això, un repàs bibliogràfic exhaustiu de l'àmplia bibliografia disponible sobre els marcadors discursius del català (i del castellà, que hi tenen una gran influència, especialment en la parla oral espontània) queda lluny de l'abast d'aquesta tesi. És per això i per la seva fàcil operacionalització que ens cenyim a una classificació formal d'aquestes partícules per al disseny d'aquesta variable.

A més, recollim la proposta de Vila (1998: 271), segons la qual l'estudi de "la probable repartició funcional de les parelles de mots formades per les formes no normatives i les normatives" (com *bueno* i *bé*, per exemple) podria aportar informació sobre la distribució sociolingüística d'aquestes formes. En el nostre cas, ens plantejem la possibilitat que la llengua inicial dels informants interaccionin amb l'ús d'una o altra forma, ja que, segons Vila (1998: 271) "[é]s ben probable que algunes d'elles, fàcilment identificables pels parlants, funcionin com a marcadors de formalitat i/o puresa lingüística, mentre que la resta els

²⁴ Deixem per una altra ocasió la controvèrsia que ha suscitat l'ús d'una o altra etiqueta per a referir-se als elements lingüístics que ens ocupen i, en aquesta tesi, les utilitzem com a formes sinònimes. Els lectors interessats en aquest debat terminològic poden consultar Fischer (2006) o Degand, Cornillie i Pietrandrea (2013), entre altres.

passin inadvertides”. Així doncs, ens centrem en parelles de partícules que ocorren més de deu vegades tant en la seva forma catalana com castellana. Per tant, no observem l'ús de *d'acord*, per exemple, de forma aïllada, sinó en comparació als seus equivalents *vale* i *val*.

La selecció de les partícules discursives incloses a la variable V3 s'ha fet, en primer lloc, a partir del *Diccionario de partículas discursivas del español* (DPDE; Briz, Pons i Portolés, 2008) i, en segon, de l'anàlisi qualitativa preliminar del corpus, que va permetre identificar cinc dobles de partícules presents al corpus que no es recullen al DPDE però que, pel seu ús per part dels informants, podrien encabir-se a la categoria de partícules discursives. No obstant això, només tres d'aquestes parelles addicionals van superar, a la fase de l'anàlisi quantitativa, el mínim de deu realitzacions tant per a les formes catalana com castellana: V3b *a ver / a ve(u)re*, V3f *después / después* i V3j *pero / p(e)rò*. Per tal d'assegurar-nos que aquestes partícules identificades al corpus poden classificar-se com a partícules discursives, a més de comprovar que reuneixen els criteris exposats a Briz, Pons i Portolés (2008), es va recórrer a la bibliografia especialitzada en partícules i marcadors discursius de l'espanyol. D'aquesta manera, es va comprovar que diversos autors han investigat el funcionament i els usos d'aquestes partícules (entre altres, Pereira, 2011; Garachana Camarero, 2012; Porroche Ballesteros, 2014; González-Sanz, 2017; González Marfud i Perdomo Carmona, 2021).

A més de la llengua inicial, és possible que aquesta variable interaccionï amb el sexe dels informants. Com recull el subapartat **2.3.1.3**, diversos autors indiquen que, prototípicament, en contextos interaccionals, els homes tendeixen a orientar els seus esforços cap a la transmissió d'informació, mentre que les dones tendeixen a centrar-se en la relació entre els participants i en l'expressió dels seus estats anímics. Per això, creiem que l'anàlisi quantitativa de les diverses partícules per a les quals hem detectat més de deu ocurrences al corpus podria revelar patrons d'ús relacionats amb el sexe dels informants. Concretament, es pretén observar si la nostra anàlisi quantitativa podria corroborar la caracterització que fan Tuson (1998), Baron i Ling (2011), Goldschmidt i Weller

(2000) i Eckert i McConnell-Ginet (2003) del comportament interaccional dels homes i les dones. Eckert i McConnell-Ginet (2003: 122) resumeixen la diferència entre els estils interaccionals d'homes i dones que han trobat la majoria d'estudis de què disposem de la manera següent:

“Research on gender and conversation among English speakers has tended to emphasize women’s strategies for cooperative, supportive talk. Women pick up and build on each other’s themes (Kalčik 1975, Coates 1996), they engage in supportive overlap (Eckert 1990, Coates 1993), they provide plentiful backchanneling (Biliouss and Krauss 1988, Roger and Nesshoever 1987, Edelsky and Adams 1990). This style has been contrasted with a male style that is said to be competitive rather than cooperative.”

B. V5 Obertura de la interacció

La segona variable pragmàtica comprèn les fórmules d'obertura de la interacció i respon a la bibliografia que les ha estudiat des de diverses perspectives teòriques, tant en corpus d'interacció oral com en textos escrits. Com apunten Ninio i Snow (1999: 20-21), Brown i Levinson (1987) descriuen la salutació com un acte amenaçador d'imatge²⁵ (*face-threatening act*). A partir de la seva proposta, molts lingüistes l'han estudiat des de la perspectiva de la seva teoria de la cortesia (per exemple, Rash, 2004, sobre a l'alemany de Suïssa o Li, 2009, sobre l'anglès i el xinès). A la lingüística interaccional, també s'ha estudiat com a part de les anomenades *rutines conversacionals*, que serveixen per afirmar l'autoritat dels parlants i proporcionar-los la confiança necessària en el desenvolupament de la interacció (Blas Arroyo, 1998: 547) o de la Teoria de la Rellevància (entre altres, Padilla, 2003). Altres lingüistes s'han centrat en la seva evolució diacrònica (p. ex. Baran, 2017).

El corpus d'anàlisi d'aquest estudi ofereix la possibilitat de contribuir a aquesta extensa bibliografia sobre les fórmules d'obertura de la interacció. En el cas del català, s'han estudiat les expressions fixades més sovint utilitzades per a saludar (Arenas Beneyto, 2005; Bladas, 2012) i també en relació a la variació textual (Arenas Beneyto i Cuenca, 2007). Arenas Beneyto (2005: 88) conclou que

²⁵ Empro la traducció proposada per Marín Jordà (2005: 53).

“analitzar la forma utilitzada per a saludar i la funció que té en una situació determinada ens pot aportar informació sobre el nivell de formalitat de la situació i la relació interpersonal entre els participants o la competència comunicativa del parlant”. En relació a aquesta darrera, podria ser que la llengua inicial dels informants interactués amb la variable V5.

D'altra banda, com assenyala Clark (2003: 303), una part de l'adquisició lingüística que té lloc a la infància passa per aprendre a contribuir adequadament en converses i a respectar els torns de paraula. A mesura que adquireixen aquestes habilitats comunicatives, els infants descobreixen que s'espera que cada participant ajudi a construir l'intercanvi aportant alguna cosa al que ja s'ha dit, a més de patrons generals que serveixen per a diversos tipus de converses, com el fet que les salutacions s'han de retornar (és a dir, formen parells adjacents).

Per tant, serà especialment interessant observar si hi ha diferències per edat, ja que, en cas afirmatiu, podria ser un indicador que l'habilitat de retornar la salutació encara es troba en procés d'adquisició a l'adolescència. A més, cal incidir en el fet que la llengua analitzada en aquest estudi, el català, encara no s'ha estudiat tan sovint ni des de tantes perspectives complementàries com altres llengües.

C. V6 Heterorepetició

La tercera variable pragmàtica és l'ús d'heterorepeticions, enteses com la repetició de les paraules d'un parlant per part del seu interlocutor, normalment, en torns consecutius. Aquest fenomen ha rebut diversos noms a la literatura. Garcés Gómez (2002-2004: 439-440) enumera els d'*alopetición*, *repetición dialógica* i *repetición de otro* en castellà, i *allo-repetition*, *other-repetition*, *secondspeaker-repetition* i *dialogic repetition* en anglès. Es contraposa a les repeticions de paraules o segments que produeix un mateix parlant en el seu torn de paraula, que també han rebut diversos noms (*autorepeticions*, *repeticions pròpies* o *repeticions monològiques*, entre altres, segons Garcés Gómez, 2002-2004: 439-440).

Malgrat que en aquest estudi ens cenyim a la denominació d'heterorepetició, cal no confondre-la amb les citacions heterofòniques (*heterophonic quotations*) a les quals es refereixen estudis com el de Perrin, Deshaies i Paradis (2003: 1845), en què el que es repeteix és un enunciat pronunciat per primera vegada per un parlant que no participa a la interacció.

La repetició oral ha rebut atenció d'un gran nombre d'acadèmics i s'ha observat des de diverses perspectives i tradicions. Segons Freixa (1998: 196), han estat quatre els punts de vista adoptats per estudiar-la: "la repetició com a mecanisme cohesionador del discurs, la repetició com a mecanisme retòric, els efectes semàntics de la repetició i la repetició en l'aprenentatge de la llengua". Des de la sociolingüística, com apunten Van Lacker Sidtis i Wolf (2015: 265), ha estat objecte de comentaris descriptius i se n'ha proposat classificacions formals i funcionals amb objectius diversos (cf. Norrick, 1987; Tannen, 1989, o Johnstone, 1994). Des d'una altra perspectiva, els estudis de retòrica han proposat també diverses taxonomies de figures poètiques relacionades amb la repetició (com l'anadiplosi, l'epanalepsi o l'epífora, per nombrar-ne només algunes). Això no obstant, Arias Flores (2012: 19) assanya que "[...] las denominaciones se han acumulado sin ofrecer precisión ni en la descripción ni en los criterios de clasificación, lo que se traduce en una limitación para trabajar las unidades".

La repetició és un fenomen molt freqüent i característic de les produccions orals, com han subratllat diversos autors (per exemple, Freixa, 1998: 195; Van Lacker Sidtis i Wolf, 2015: 266, o Zuljan Kumar, 2019: 126-127). Com recull Freixa (1998: 198-199), hi ha diversos criteris segons els quals es poden classificar les formes en què els parlants repeteixen elements discursius, segons la significació, la fixació, el temps o la funció, per exemple. En el nostre estudi, utilitzem el criteri del subjecte de la repetició (distingim, per tant, entre autorepeticions i heterorepeticions, com s'ha explicat més amunt) i ens basem, per identificar realitzacions de V6, en els tipus d'heterorepeticions enumerats a Freixa (1998: 204-206), com recull la Taula 17.

Taula 17. Tipus d'heterorepetició segons Freixa (1998) i exemples del corpus d'anàlisi.

Context d'aparició	Funció	Exemples d'heterorepetició		
Conjunts tancats	H1 Obertures i tancaments d'interacció	Freixa (1998)	L: bona tarda\ TOTS: hola\ bona tarda\ 	
		Corpus d'anàlisi	ENT: bona tarda_ L0001QCI: bona tarda_ 	
Preguntes i respostes	H2 Pregunta d'una afirmació anterior	Freixa (1998)	A: ens van tirar les cartes\ M: les cartes/ A: (0) sí\ @	
		Corpus d'anàlisi	ENT: si en té algun altre/ no:_ M0051QCI: no/ vale_ 	
	H3 Afirmació d'una pregunta anterior	Freixa (1998)	L: avui en dia portar un abric de pells significa ser_ (.) a::_ més o menys respectuós amb la natura/ (...) senyor Prenafeta\ P: (..) home\ (.) portar un abric de pell avui en dia pues és ser_ (..) absolutament respectuós amb la-- amb la-- amb la pròpia natura\ 	
		Corpus d'anàlisi	ENT: i per dormir on m'has dit que dormia/ M0471SCI: el gat dormia com en una espècie de_ de sofà petitó_ 	
	Afirmació d'una declaració	H4 Admetre	Freixa (1998)	C: no\ amb això amb això no hi ha problema\ B: no hi ha problema\
			Corpus d'anàlisi	ENT: no passa re(s) amb el gat/ al ratolí/ tampoc suposo\ M049QCI: no no passa re:(s)_
H5 Assentir		Freixa (1998)	C: (..) i s'ha de fer una inversió\ D: (..) hi tens de fotre una inversió\ 	
		Corpus d'anàlisi	ENT: es pot tocar suposo:_ L0009QCI: sí tocar i_ bueno i ja està\ 	
H6 Destacar		Freixa (1998)	J: (..) gent que necessiten personatges per anuncis_ llavors fan proves_ 	

Metodologia

			(...) però no l'agafen mai\ A: moltes proves n'ha fet\ però a l'últim moment no l'han agafat\ ENT: tu que em recomanaries que agafés\ L0009QCI: jo et recomanaria el gat/
Desacord amb una declaració	H7 Expressar sorpresa o incredulitat	Freixa (1998)	B: (..) imagina't lo bo que és tenir una hipoteca\ A: bo::/ C: home bo::_ home sí\ (.) desgrava\ ENT: home sí:: és atrevit és atrevit_ L0012QCI: atrevit/ ENT: és valent_ sí sí_
		Corpus d'anàlisi	
	H8 Contra-declaració contradictòria	Freixa (1998)	C: encara que suposo que vostès ens demo- -- intentaran demostrar que no_ (.) [estan--] P: [no\ (.) no\ (.) nosaltres no hem d'intentar demostrar res\ ENT: i p(e)rò tu dius que és molt petit i llavors què no el pot tocar o no:_ L0043SCI: sí que el pot tocar p(e)rò hi perilla que li trenqui la cua_
		Corpus d'anàlisi	
	H9 Contra-declaració de correcció	Freixa (1998)	C: és petit\ D: és superpetit\ ENT: llavors aquesta no es pot escapar_ aquesta està allà tancada_ L0019SCI: sí_ home està allà tancat_ no ser-- a no ser que el vulguis treure_ perquè un animal allà dintre deu ser molt avorrit_
		Corpus d'anàlisi	

Això no obstant, la dificultat de proposar un model de classificació de la repetició oral satisfactori i àmpliament aplicable es deu, en gran part, al fet que aquest recurs realitza un gran nombre de funcions (p. ex. Freixa, 1998: 199-200) que fins i tot poden ser contradictòries (Zuljan Kumar, 2019: 127). Per a Arias

Flores (2012: 30), per exemple, la repetició pot tenir dues orientacions: “una hacia el texto mismo, al intervenir en su entramado y crear efectos de sentido; y otra hacia el intercambio comunicativo, estableciendo lazos entre los participantes para crear el texto conversacional”. L’heterorepetició pren més aviat la segona orientació i funciona com a enllaç interlocutiu entre els participants de la interacció, com fan notar diversos estudis. Van Lacker Sidtis i Wolf (2015: 274) diuen que els parlants se serveixen de l’heterorepetició per donar-se suport i validar-se els uns als altres, cosa que concorda amb les posicions de Tannen (1989) i Briz (2001). En aquest sentit, Zuljan Kumar (2019: 134) argumenta que “While short responses convey the message: *I am listening and agree with you*, repeated lexemes carry much more powerful message [sic]: *I fully support and agree with you*”.

Aquesta funció interpersonal de la repetició, però, s’ha de complementar amb la observació que la repetició d’informació ja coneguda, des del punt de vista de la organització de la informació textual, té l’efecte d’aturar la progressió temàtica que es produeix a la interacció, per la qual cosa es pot considerar una estructura marcada (Spang, 1979: 145; Arias Flores, 2012: 20). Per a Arias Flores (2012: 29), aquesta repercussió en l’eficàcia comunicativa converteix l’ús d’heterorepeticions en una elecció pragmàtica i la creativitat en l’ús de la llengua que requereix dels parlants la fa una elecció estilística.

Així, doncs, podria ser que l’ús d’aquest recurs marcat augmenti a mesura que els informants del nostre estudi adquireixen habilitats comunicatives. D’aquesta manera, *a priori* esperem que les mostres recollides durant la tercera onada continguin més heterorepeticions que les de la segona.

5.3.2.3.2. Les variables morfosintàctiques

Les variables morfosintàctiques inclouen la posició dels pronoms clítics en construccions perifràstiques (V14), el número (V16) i el tipus (V17) de preguntes no retòriques formulades en cada entrevista per part dels informants, el nombre de construccions condicionals (V18), les perífrasis d'obligació (V22), els determinants demostratius (V23) i els pronoms clítics de datiu plural (V25).

D. V14 Posició dels pronoms clítics en construccions perifràstiques

Com recullen diverses gramàtiques del català actual (Bonet, 2002: 938; GIEC, 2016a: 200; GEIEC, 2021b), en construccions verbals formades per una forma finita i un infinitiu o un gerundi, els pronoms clítics poden aparèixer com a proclític sobre la forma finita (*El pots treure d'aquest lloc*) o com a enclítics sobre la no finita (*Pots treure'l d'aquest lloc*). Així doncs, la posició d'aquests elements no altera el seu significat en català, com tampoc no ho fa en castellà (però cf. Gauthier, 2016 sobre *tener que* + INF i la posició del clíctic *lo*).

Les formes verbals perifràstiques amb pronoms clítics, per tant, funcionen de la mateixa manera en totes dues llengües, amb l'única diferència que el sistema pronominal català inclou més elements que el castellà (Illamola i Vila, 2015: 39). No obstant això, mentre que el sistema pronominal català ha rebut una considerable atenció acadèmica (p. ex. Vila i Moreno, 1996; Bonet, 2002 o Pusch, 2006), l'estudi de la distribució dels pronoms clítics en construccions perifràstiques només s'ha realitzat des de la perspectiva diacrònica (Todolí i Cervera, 1998; Fischer, 2003), a diferència de la llengua castellana, per a la qual sí que disposem d'estudis previs sobre la distribució dels clítics en aquestes construccions (entre altres, Torres Cacoullous, 1999; Silva-Corvalán, 2001: 184-186; Zabalegui, 2008; Troya Déniz i Pérez Martín, 2011; Aijón Oliva i Borrego Nieto, 2013; Gómez Lázaro, 2018).

En castellà, en general, s'ha trobat que la posició més habitual és la proclítica (Silva-Corvalán, 2001; Zabalegui, 2008; Troya Déniz i Pérez Martín, 2011; Aijón Oliva i Borrego Nieto, 2013; Illamola i Queralt, 2020; però cf. Gómez Lázaro, 2018). La distribució dels clítics s'ha relacionat amb factors molt diversos, com el tipus de verb auxiliar (Silva-Corvalán, 2001; Troya Déniz i Pérez Martín, 2011;

Gómez Lázaro, 2018), el tipus i la funció dels pronoms, el temps verbal o la persona gramatical de l'auxiliar (Troya Déniz i Pérez Martín, 2011), la prominència discursivocognitiva de les entitats a què fan referència els clítics (Aijón Oliva i Borrego Nieto, 2013), l'animicitat del referent (Zabalegui, 2008) o el significat del verb auxiliar (Silva-Corvalán, 2001). Així doncs, com recullen Illamola i Vila (2015: 38):

“En resumen, los estudios disponibles hasta ahora sugieren que la variabilidad en la distribución de formas enclíticas y proclíticas debe atribuirse a factores estructurales, pragmáticodiscursivos y funcionales, aunque el fenómeno también presenta cierta variación de naturaleza dialectal.”

Quant a les dades estudiades per observar aquesta variable, cal destacar que la posició dels pronoms en construccions verbals perifràstiques ja ha estat objecte d'estudi a partir de material recollit en el projecte RESOL. En un primer treball, Illamola i Vila (2015) analitzen la distribució enclítica o proclítica dels pronoms clítics en castellà en un total de 38 entrevistes amb informants d'entre 15 i 16 anys amb el català o el castellà com a llengua inicial. Estudien construccions amb les perífrasis d'infinitiu amb els verbs *poder*, *tener que* i *ir a*. Els seus resultats no mostren diferències significatives ni per sexe ni per llengua inicial.

Amb la intenció de comparar aquests resultats amb els nostres, tenint en compte les diferències entre les mostres analitzades, per al recompte d'aquesta variable s'han considerat els pronoms clítics que apareixen a les construccions perifràstiques d'infinitiu amb els verbs catalans *poder*, *tenir de*, *tenir que*, *haver de* i *anar a*.

En un segon treball sobre la posició dels clítics castellans en perífrasis verbals, aparegut després que l'anàlisi quantitativa de la variable V14 hagués conclòs, Illamola i Queralt (2020) investiguen una nova mostra provinent del projecte RESOL. Aquesta vegada, es compon de 155 transcripcions, de nou de la prova interactiva, i l'exploren en relació a les variables independents edat (12-13 o 15-16 anys), sexe, llengua inicial (català, castellà o ambdues llengües) i municipi (Manlleu o Mataró). Els objectius d'aquest estudi i els del nostre estan encara més directament relacionats que els del treball precedent, ja que es tracta

d"“observar si la variación en la posición de los pronombres con perífrasis verbales puede arrojar información sobre el perfil sociolingüístico del individuo” (Illamola i Queralt, 2020: 2). Els resultats mostren que, en general, els informants utilitzen més freqüentment els pronoms clítics en posició proclítica que enclítica, a excepció dels bilingües inicials de 12 anys de Manlleu. A més, de les variables independents que es consideren, el municipi i la llengua inicial mostren diferències estadísticament significatives entre els grups estudiats.

Així doncs, tenint en compte els resultats d'aquests treballs previs sobre mostres del mateix projecte que utilitzem aquí, resultarà especialment interessant observar si, en el corpus analitzat en aquesta ocasió i per al català, es poden identificar o no diferències en la posició de les formes clítics en construccions perifràstiques pel que fa a parlants d'edats, sexes i llengües inicials diferents.

E. V16 Nombre de preguntes no retòriques

La segona variable morfosintàctica és el número de preguntes no retòriques formulades per part dels informants. Com s'ha explicat al subapartat **2.3.1.3**, alguns autors afirmen que, en contextos interaccionals, les dones tendeixen a fer més preguntes que els homes (p. ex. Fishman, 1983; Tuson, 1998: 216). No obstant això, també hi ha autors que contradiuen aquesta mena d'afirmacions relacionades amb les diferències entre homes i dones (Hawkins i Power, 1999, per exemple, no van trobar diferències significatives en el nombre de preguntes formulades per homes i dones en grups de presa de decisions reduïts²⁶) o que les consideren massa simplificades (com s'ha explicat al subapartat **2.3.1.4**). A més, reculls com el de Pearson, Turner i Todd-Mancillas (1991: 115) conclouen que no hi ha consens sobre quin dels dos sexes sol formular més preguntes.

Malgrat que els estudis longitudinals d'adolescents com el nostre no abunden a la bibliografia consultada, n'hi ha que han tractat aquesta variable i no han trobat diferències significatives en el nombre total de preguntes formulades per nois i noies. Per exemple, Larson i McKinley (1998) analitzen diverses variables lingüístiques i comportaments conversacionals de 4 nois i 4 noies en sis ocasions

²⁶ *Small decision-making groups*, en anglès.

al llarg del temps (des que tenien 12 o 13 anys fins als 16 o 17) en dos contextos diferents (amb un adult desconegut i un amic de la seva edat seleccionat per l'informant) i no troben diferències en el nombre total de preguntes formulades per sexe ni per edat. Així doncs, el nostre estudi pot aportar informació rellevant per poder confirmar o rebutjar amb més certesa afirmacions com les de Tuson (1998) pel que fa a parlants adolescents, tenint en compte que la nostra mostra és més gran que la d'altres estudis longitudinals de què disposem fins ara amb parlants d'aquestes edats, com el de Larson i McKinley (1998).

F. V17 Tipus de preguntes

La tercera variable morfosintàctica del nostre estudi és el tipus de preguntes formulades pels informants, que poden ser obertes o tancades. Com indica Freed (1994: 621), “[q]uestions (or interrogatives) have been investigated at every level of linguistic analysis for decades”. En efecte, les perspectives d'anàlisi des de les quals s'ha estudiat les preguntes són molt variades, així com les definicions i les tipologies que se n'ha proposat amb el temps. Entre altres, se n'han ocupat lingüistes interessats en els actes de parla (començant amb Austin, 1962), l'anàlisi de la conversa (per exemple, Stivers, 2010) o entrevistes policials (com Waterman, Blades i Spencer, 2001 o Oxburgh, Myklebust i Grant, 2010), així com especialistes d'altres àmbits, com la sociologia (per exemple, Raymond, 2003) i la psicologia educativa (com León, Peñalba i Escudero, 2010). En general, hi ha consens sobre certs aspectes que caracteritzen les preguntes, com la possibilitat d'identificar-les per la seva forma sintàctica i naturalesa prosòdica o que el seu contingut proposicional és diferent de la seva funció pragmàtica i interaccional, però els acadèmics que se n'han ocupat no han arribat a un consens sobre la definició de pregunta i la manera més adequada d'analitzar-les (Freed, 1994: 622).

Entre les tipologies de preguntes disponibles a l'extensa bibliografia consultada, es poden distingir tres grans grups: el de les que es basen en la forma de les preguntes (com Peterson, Dowden i Tobin, 1999, i Waterman, Blades i Spencer, 2001), les classificacions que tenen en compte tant la forma sintàctica com la funció que duen a terme (p. ex. Freed, 1994) i les que es basen tan sols en les

molt diverses funcions que poden desenvolupar aquestes construccions lingüístiques (com Hawkins i Power, 1999, o Stivers, 2010). Per això, trobar consens sobre quina classificació és més aplicable a diverses produccions lingüístiques és difícil, ja que cada tipologia respon als objectius concrets de la investigació per a la qual es va dissenyar. Aquesta és la raó principal per la qual, en el nostre estudi, distingim només entre preguntes obertes i tancades, una classificació aplicable tant a les transcripcions que analitzem com a qualsevol altra mostra lingüística. Això darrer és important perquè, com s'ha explicat anteriorment, es preveu que els resultats del nostre estudi tinguin aplicacions en la disciplina de la lingüística forense, en què qualsevol producció lingüística podria esdevenir objecte d'anàlisi (Perkins, 2015: 216).

En resum, doncs, la revisió bibliogràfica sobre els estudis previs dels diversos tipus de preguntes que formulen els parlants durant les converses no ens ha permès identificar cap possible relació entre les variables independents del nostre estudi i les preguntes obertes o tancades. Malgrat això, hem optat per incloure aquesta variable a l'estudi ja que reuneix els requisits que enumeràvem a l'inici d'aquest apartat i permetrà contribuir així a la bibliografia existent tant si els resultats ens permeten concloure que existeix una relació entre l'edat, el sexe i/o la llengua inicial dels informants i el tipus de preguntes que formulen durant la prova com si no és així.

G. V18 Construccions condicionals

La següent variable morfosintàctica comptabilitza el nombre de construccions condicionals i els nexes que fan servir els informants. Aquestes estructures han despertat un interès notable entre acadèmics de diversos àmbits, com la filosofia, la psicologia i la lingüística. A continuació, ens centrem en treballs rellevants de les dues darreres.

La capacitat de conceptualitzar escenes que no han tingut lloc però que podrien haver passat (com el fet que una gerra hagi estat a punt de caure) es desenvolupa molt aviat, entre els 2 i 3 anys (Badger i Mellanby, 2018: 25). Això no obstant, actualment disposem de nombrosos estudis que demostren que les habilitats de comprensió i producció de condicionals més complexos (com *Si*

hagués fet els deures cada dia, no hauria suspès) s'adquireixen més tard, durant la preadolescència (Jiménez, 2010; Aparici Aznar i Noguera Llopart, 2012) o durant l'adolescència (Mariscal i Auza, 2017; Nippold, Frantz-Kaspar i Vigeland, 2017; Badger i Mellanby, 2018). Badger i Mellanby (2018: 22) assenyalen fins i tot que alguns adults no adquireixen mai els condicionals complexos: “Yet acquisition of the more complex forms of the conditional continues to be difficult for some older children (Amidon, 1976), and indeed, some adults never reach full acquisition (Evans, Handley, Neilens & Over 2008)”.

També sabem que el raonament associat a oracions condicionals és diferent en infants, adolescents i adults, de manera que les habilitats cognitives necessàries per processar aquestes construccions com un adult i poder extreure'n les inferències correctes es desenvolupa gradualment durant la infantesa i l'adolescència (Taplin, Staudenmayer i Taddonio, 1974; Müller, Overton i Reene, 2001; Moreno-Ríos i García-Madruga, 2002). Tot plegat, sumat al fet que el domini d'aspectes comunicatius determinats pel context és limitat en parlants adolescents, fa possible que la proposta d'Andersen (2001: 12) que parlants de 12 o 13 anys encara no hagin adquirit completament les oracions condicionals sembli encertada:

“Furthermore, it may be that, say, a 13-year-old masters the use of conditional clauses only at the very concrete content level (e.g. *If it rains, I'll stay in tonight*) without having acquired the conversationally based, and less concrete, uses of conditionals at the epistemic level (e.g. *If that's a bear's footprint, than I'm King Kong!*) or the speech act level (e.g. *If you don't mind my asking, when did you come home last night?*); cf. Sweetser (1990). That this may be the case, and that conditionals generally cause serious comprehension problems even to adolescents, is evidenced by studies mentioned by Perera (1984).”

D'altra banda, pel que fa a la producció d'aquestes estructures, Morales (1999: 48) subratlla el fet que el mode subjuntiu, necessari per a la producció de les oracions condicionals més elaborades, és, possiblement, un dels aspectes més complexos de la llengua espanyola (cosa que també es pot aplicar a la catalana)

i, com mostra el repàs que en fa Floyd (1990), diversos estudis en aquest idioma han trobat que parlants adolescents encara es troben en ple procés d'adquirir-lo:

“Certainly there is evidence in the literature that among bilingual and even monolingual Spanish-speaking children, subjunctive is not fully acquired in all its forms and all its uses even at early adolescence. Findings of past studies show that older children use a larger inventory of subjunctive forms than do younger children, with simple subjunctive forms appearing earlier than complex subjunctive forms”. (Floyd, 1990: 101-102).

Així doncs, el nostre estudi podria aportar dades que confirmin la relació entre l'edat i la producció de construccions condicionals, ja que observem els informants quan, possiblement, encara estarien adquirint aquestes estructures (als 12 o 13 anys) i més endavant, quan podria ser que ja les haguessin adquirit (als 15 o 16 anys).

H. V22 Perífrasis d'obligació

Aquesta variable morfosintàctica són les perífrasis d'obligació. Tot i que la normativa no contempla la perífrasi *tenir que* com a alternativa acceptable d'*haver de* per expressar modalitat deòntica (GIEC, 2016b: 951; GEIEC, 2021d), totes dues construccions són utilitzades pels parlants, com indica Sinner (2008: 534), i *tenir que* és freqüent en situacions informals, si bé més en uns dialectes i zones geogràfiques que en uns altres; especialment, a l'àrea metropolitana de Barcelona (Mier, 1986) i al País Valencià (Stokes, 2015). L'ús d'aquesta variant s'ha documentat en diverses ocasions al llarg del temps (per exemple, Mier, 1982, 1986; Hualde, 1992: 325; Rodríguez-Vida, 1997; Martínez Díaz, 2002: 87; Paloma i Rico, 2008). En efecte, no es tracta d'un fenomen d'aparició recent, sinó que l'ús de *tenir que* es remunta, com a mínim, al segle XV (Cabanes Fitor, 1996). Tampoc és exclusiu de la llengua oral, malgrat que és menys comú a l'escrita, ja que, per exemple, Stokes (2015) analitza la presència de *tenir que* i *haver de* a Twitter, on la llengua escrita sovint recull usos orals. Cal afegir que el fenomen contrari a l'ús en català de *tenir que*, la influència del català en les perífrasis d'obligatorietat del castellà, també s'ha tractat de forma reiterada a la

bibliografia (per exemple, Martínez Díaz, 2002 i 2003; Sinner, 2003; Wesch, 2008: 58-59).

En el nostre estudi, a més de les variants *haver de* i *tenir que*, tenim en compte les variants genuïnes *caldre* i *tenir de* seguides d'infinitiu. L'anàlisi qualitativa preliminar també va incloure la variant *haver que* seguit d'infinitiu, igualment possible a partir del contacte entre el castellà i el català, si bé normalment vinculada a varietats geolèctals o sociolectals concretes, com recullen alguns autors (com Rigau, 1998: 80) i com passa també amb *tenir de* (Pàmies, 2012: 112). Això no obstant, en tot el corpus d'anàlisi només es va poder identificar una realització d'*haver que* + infinitiu, per la qual cosa no es va tenir en compte per a l'anàlisi estadística definitiva.

Caldrà observar si els resultats d'aquesta variable mostren, com seria esperable segons la bibliografia consultada, una correlació estadísticament significativa amb la llengua inicial dels informants.

I. V23 Determinants demostratius

La penúltima variable morfosintàctica contempla l'ús dels determinants demostratius de primer grau (*aquest, aquesta, aquests i aquestes*) i de segon grau (*aquell, aquella, aquells i aquelles*). Seguint les categories que proposa Martín Miñana (2019), aquesta variable conté les realitzacions d'aquests mots que compleixen les funcions de determinant (quan apareixen abans del substantiu, com en *aquest animal*), d'adjectiu (quan apareixen després del substantiu, com en *l'animal aquest*) i de pronom substantivat (quan no apareixen acompanyant un substantiu, sinó que fan la seva funció, com en *No el venc, aquest*). Per raons pràctiques, queden fora d'aquesta variable, per tant, altres demostratius que sí s'estudien a Martín Miñana (2019): els pronoms neutres (*açò, això i allò*) i els adverbis (*ací, aquí i allà*). A més de les estructures que acabem d'enumerar, a l'anàlisi qualitativa prèvia del corpus es van identificar les construccions del tipus *un + N + d'aquests* i *un + N + d'aquells*, que s'han incorporat a la variable V23 com a variants V23g i V23h, respectivament.

L'ús d'una classificació basada en funcions sintàctiques i no pragmàtiques respon a motius pràctics. D'una banda, diversos estudis dels determinants demostratius han proposat classificacions diferents de les funcions que duen a terme en produccions orals. Per exemple, mentre que Albaladejo (1998) estructura una classificació a partir de la distinció entre dixi exofòrica i dixi endofòrica, Cuenca (2010) distingeix sis categories (dixi situacional, dixi textual, dixi *am Phantasma*, dixi presentacional, ús indefinit i expletiu, ús gramaticalitzat), que coincideixen a grans trets amb les que enumera Ribera (2019), però que no són totes les que fa servir Martín Miñana (2019), que distingeix nou usos (dixi situacional, discurs díctic, *Deixis am Phantasma*, ús recognoscible, anàfora, catàfora, ús gramaticalitzat, ús orientador i ús escenificador). Si bé la majoria dels usos que descriuen diversos autors són semblants, cada proposta difereix en l'agrupació que en fan i la inclusió o exclusió d'alguns valors, cosa que dificultaria l'estructuració d'una variable pragmàtica amb variants clarament definides.

D'altra banda, les funcions pragmàtiques són més difícils d'identificar que les sintàctiques, ja que depenen de factors abstractes com la subjectivitat de l'emissor o el coneixement compartit entre emissor i receptor (Martínez, 2000: 246, citat a Cuenca, 2010: 108), entre altres (vegeu els factors subjectius i intersubjectius que llista Ribera, 2019: 2), i no de normes gramaticals com en el cas de les funcions sintàctiques. A més, diverses funcions pragmàtiques poden coexistir en una sola ocurrència, com, per exemple, la dixi textual catafòrica i la dixi *am Phantasma* en el demostratiu *aquesta al fragment i va ser angoixant la sensació aquesta a càmera lenta* (Cuenca, 2010: 113). Per aquestes raons, ens cenyim a les estructures identificades al corpus, distingint-ne el rol sintàctic a partir del grau de proximitat dels demostratius i de les categories que proposa Martín Miñana (2019).

Quant a la possible interacció d'aquesta variable amb les variables extralingüístiques que observem en el nostre estudi, Diessel (2006: 469) argumenta l'ús dels demostratius per establir l'atenció conjunta (*joint attention*)

en la parla dels infants durant les etapes més primerenques de l'adquisició lingüística:

“In addition to gesture and gaze, language can now be used to create a joint focus of attention. While there are many linguistic means that speakers can use to coordinate a joint attentional focus, there is no other linguistic device that is so closely tied to this function than demonstratives (cf. Clark 1996: 168)”.

Per tant, sembla que a les edats que comprèn el nostre estudi (aproximadament 12 i 16 anys) no s'hauria d'observar cap indicador de l'adquisició dels demostratius, ja que s'haurien adquirit molt abans. Malgrat tot, el nostre estudi ofereix l'oportunitat de corroborar, quantitativament, aquesta idea prèvia que es desprèn de la bibliografia especialitzada en l'adquisició lingüística.

A més, ens interessa observar la freqüència d'ús de les diverses construccions sintàctiques possibles amb demostratius. Alguns autors, com Fitó (2009: 55), que recull la opinió de Kerbrat-Orecchioni (1980), defensen que la freqüència d'ús d'elements dítctics (entre ells, els demostratius) “varia en funció de l'emissor –l'estil personal de cadascú és important– i també del tipus de discurs, particularment de si es tracta d'un discurs oral o escrit”. Així doncs, caldrà veure si els nostres resultats quantitius donen suport a aquesta afirmació o si, per contra, mostren algun patró que pugui explicar la variació en els usos dels demostratius en relació amb les variables sexe i llengua inicial, que presumiblement podrien estar vinculades a aquest “estil personal de cadascú”.

J. V25 Pronoms clítics de datiu plural

La darrera variable morfosintàctica comptabilitza les formes de pronoms clítics de datiu plural, que segons la normativa és (*els*) però que en català central es realitza *els hi* (/lzi/, segons Bonet, 1992: 102 o *elsi*, segons Pineda, 2018). Aquesta variació es recull a diverses gramàtiques de la llengua (Badia, 1994: 465; Colomina, 2002: 575; GIEC, 2016a: 213-4, GEIEC, 2021b) i s'ha relacionat amb diverses explicacions lingüístiques (per exemple, Todolí, 1999; Pineda, 2018). Todolí (1999: 263) contrasta els dialectes barceloní i valencià i explica que, mentre al primer, la combinació del datiu plural i de l'acusatiu plural de tercera persona resulta en *els hi* en varietats no estàndards perquè “els morfemes

de nombre i de persona s'apliquen sobre el conjunt, i no sobre cada clíctic per separat", en valencià i a l'estàndard, "cada clíctic té els seus propis morfemes, de manera que constitueixen dominis independents per a la representació d'aquests morfemes" i per això aquesta combinació es realitza com *els els*.

Com recull la Taula 16 a l'inici d'aquest apartat, en el corpus d'anàlisi s'observen també altres variants (*li* o *lis*), que podrien trobar-se en les produccions d'informants amb el castellà com a llengua inicial, ja que són les formes més semblants a les d'aquesta llengua (*les*). En aquest sentit, Pineda (2018) explica aquest fenomen, que anomena *li-per-els[i]*, com un pas més de la duplicació de l'objecte indirecte (en construccions com *Els[i] vaig donar el paquet als treballadors*) i apunta que "s'hi pot veure segurament el reflex de la influència de l'espanyol", com suggereix Todolí (2002), si bé també s'ha de tenir present que es tracta d'un fenomen compartit per altres llengües romàniques i que, per tant, podria veure's com un procés d'evolució interna de la llengua (Pineda, 2018: 257). Així doncs, caldrà veure si els resultats mostren una relació entre la llengua inicial i els clíctics de datiu plural al corpus per confirmar la influència del castellà en aquest fenomen.

5.3.2.3.3. Les variables lèxiques

Les variables lèxiques inclouen els fenòmens de contacte de llengües (V38), l'expressió dels colors (V41) i el nombre d'adverbis acabats en *-ment* (V45).

K. V38 Fenòmens de contacte de llengües

La denominació d'aquesta variable es basa en la utilitzada a Vila i Moreno (1996: 49) i Boix i Vila (1998). Ara bé, els treballs sociolingüístics sobre les marques lèxiques de contacte entre llengües tenen una llarga tradició que es remunta als anys cinquanta del segle passat (amb obres com Haugen, 1950; Weinreich, 1953) i han produït un gran nombre de denominacions i classificacions d'aquests fenòmens al llarg dels anys (com *interferència*, cf. Vila i Moreno, 1996: 47 i s.; Boix i Vila, 1998: 228 i s.; Espuny, 1998). Per exemple, una altra denominació possible hauria estat la de *marques transcòdiques*, introduïda per Lüdi (1987), per referir-se igualment a elements discursius "que remeten d'una manera o d'una altra a la trobada de dos o més sistemes

lingüístics (calcs, manlleus, transferències lèxiques, alternances de codi, etc.)” (Boix i Vila, 1998: 245; cf. Vila, 1998; Winford, 2003; Grant, 2020).

No obstant això, per mantenir la coherència al llarg del treball, només utilitzem *fenòmens de contacte de llengües* per referir-nos a les realitzacions d'elements lingüístics atribuïbles a la influència d'altres llengües i no recollits a les obres de referència normativa (el *Diccionari de la llengua catalana*, DIEC), paranormativa (el *Diccionari.cat* del Grup Enciclopèdia Catalana), ni auxiliars (la versió en línia del *Diccionari català-valencià-balear*, DCVB, el web de consultes lingüístiques de la Secretaria de Política Lingüística, Optimot, i el portal lingüístic de la Corporació Catalana de Mitjans Audiovisuals, ésAdir). Aquestes obres auxiliars s'han utilitzat per comprovar si certs elements que no apareixen a les obres de referència estan en canvi recollits en aquests cercadors i, per tant, podrien considerar-se importacions esteses en les varietats col·loquials de la llengua (vegeu el subapartat **5.5.11**). Els elements que hi apareixen com a castellanismes s'han considerat fenòmens de contacte de llengües.

Aquesta variable, com es reflecteix anteriorment a la Taula 16 i de nou a continuació a la Taula 18, està formada per les variants V38a Importacions, V38b Híbrids, V38c Calcs i V38d Altres. Quant a la primera, que es basa en la concepció d'un contínuum entre manlleus i alternances de codi (Vila i Moreno, 1998: 414), comprèn tant manlleus (és a dir, mots importats d'altres sistemes lingüístics que no apareixen a les obres lexicogràfiques de referència o de consulta) com alternances de codi mono- i plurilèxiques, enteses com juxtaposicions de fragments que pertanyen a diferents sistemes gramaticals, normalment amb un retorn a la llengua base, en un mateix fragment de parla (Galindo Solé, 2006: 12; de Rosselló i Peralta, 2014: 27).

Les propostes de classificació d'aquests fenòmens han proliferat a la bibliografia sobre el contacte lingüístic (vegeu, entre altres, Poplack i Sankoff, 1988; Myers-Scotton, 1992; Poplack, 2012), com també les formes d'operacionalització de la problemàtica distinció entre manlleus ocasionals i alternances de codi monolèxiques (Boix i Vila, 1998: 246). Per exemple, alguns autors han utilitzat un criteri basat en la freqüència d'ús (Vila i Moreno, 1996: 412-414; Galindo

Solé, 2005: 258). La variant V38a Importacions, doncs, ens permet operacionalitzar la noció del complex contínuum entre els manlleus adaptats (fonològicament, morfològicament, semànticament, etc.), els manlleus no adaptats, les alternances de codi monolèxiques i les alternances de codi plurilèxiques.

Pel que fa a la variant V38b Híbrids, inclou aquells elements formats per diversos morfemes o lexemes, almenys un dels quals, però no tots, provenen d'un sistema lingüístic diferent del català (Boix i Vila, 1998: 229; Lorente Casafont, 2016: 113). La variant V38c Calcs recull diversos tipus de calcs (semàntics, morfològics, etc.), és a dir, elements formats per la reproducció d'elements de la llengua donant mitjançant elements propis del català, la llengua receptora (Boix i Vila, 1998: 228; Lorente Casafont, 2016: 113).

Per últim, la variant V38d Altres agrupa els casos que presenten ambigüïtat perquè, depenent de l'enfocament teòric adoptat, podrien classificar-se com més d'una de les altres variants (per exemple, *què va* podria ser una importació o un calc), les unitats que són resultat d'un procés d'ultracorrecció (entès com a una altra conseqüència del contacte de llengües) i les unitats que no s'expliquen directament pel contacte de llengües sinó que són probablement creacions noves a partir de components del català, però que constitueixen elements idiosincràtics d'alguns informants que calia incloure a l'anàlisi.

Taula 18. Variants i exemples de la variable V38 Fenòmens de contacte de llengües.

V38 Fenòmens de contacte de llengües	
V38a Importacions	<i>saltamontes, sardinita, vale, okay, pipes, bueno, apoiar, a mi prima le pasó, no sé cómo se llama</i>
V38b Híbrids	<i>cuarenta-dos, metge de capcera</i>
V38c Calcs	<i>tendries, cànvien, va a la seva bola, sempre i quan</i>
V38d Altres	<i>què /qué va, saltamuntanyes, expositori</i>

Quant a les variables independents amb les quals creiem que podria interactuar aquesta variable, com s'ha explicat al subapartat **2.3.1.2**, el primer principi de Labov, part del que s'ha anomenat el patró de gènere, estipula que, en condicions estables, les dones fan servir menys formes no estàndards que els

homes. Si apliquem aquest principi a la nostra mostra, podem esperar que els informants nois facin servir més fenòmens de contacte de llengües que les noies. A més, com mostren els exemples, la gran majoria de mots que recull aquesta variable s'expliquen pel contacte amb el castellà, per la qual cosa també és possible que la variable independent llengua inicial es correlacioni amb els resultats d'aquesta variable.

Finalment, cal tenir en compte que, com s'ha explicat a l'apartat 4.4, les transgressions lingüístiques (sobre tot les importacions i els calcs) són, a més de l'ús d'altres recursos com els neologismes, molt habituals en les varietats emprades pels joves. D'aquesta manera, caldrà observar si aquesta variable també interacciona amb l'edat dels informants a mesura que s'allunyen del "llenguatge infantil" i fan més seu el "llenguatge juvenil".

L. V41 Colors

La següent variable és l'expressió dels colors (V41), que sovint apareix en relació als animals i als accessoris que els informants, en el seu paper de venedors, han de vendre a l'entrevistadora, que fa de clienta. A continuació, la Taula 19 mostra exemples de les dues variants d'aquesta variable.

Taula 19. Exemples de la variable dependent V41 Colors.

V41 Colors	
V41a Color [blau]	ENT: i: i com és exactament/ M0094SCI: és de color negre _ amb unes ra(t)lles horitzontals de colo:r--no\ verticals de color groc \
V41b [blau]	ENT: i com és aquest gat/ M0043SCI: pues: és marró :_ (a)m::(b) unes taques negres_

El context interaccional tan específic que conformen les entrevistes (la tasca de joc de rol en què els informants fan de venedors d'una botiga d'animals) propicia l'aparició dels sintagmes que comptabilitzen aquesta variable. És per això que, malgrat que no s'ha trobat bibliografia prèvia rellevant per a l'estudi d'aquesta variable, l'incloem a l'estudi, ja que creiem que el contingut de les gravacions conforma una oportunitat d'observar si aquesta mena d'eleccions lèxiques es

correlacionen amb alguna de les variables independents que analitzem (l'edat, el sexe o la llengua inicial dels informants).

M. V45 Nombre d'adverbis acabats en *-ment*

L'última variable és l'ús d'adverbis acabats en *-ment*. En català, la bibliografia que ha tractat aquestes unitats a corpus lingüístics és més escassa que en altres llengües (per exemple, per al castellà, vegeu de Marlangeon, 2002; Torner, 2007; per a un estudi sincrònic i diacrònic de les llengües romàniques, Hummel, 2013). Malgrat això, comptem, per exemple, amb estudis precedents com el de Gelpí (1998), que analitza la freqüència d'aquests adverbis (a més de substantius, verbs i adjectius) en el corpus descrit al volum editat per Payrató mencionat anteriorment (al subapartat **2.2.4.3**). Concretament, investiga la hipòtesi que en els textos amb to formal, la freqüència d'aquests adverbis serà més elevada que en els textos amb to informal i conclou que “les dades obtingudes no confirmen totalment el supòsit inicial, però ofereixen una clara tendència a la validació” (Gelpí, 1998: 136) perquè només tres dels setze textos analitzats tenen un to diferent al que caldria esperar pel nombre d'adverbis acabats en *-ment* que contenen.

Pel que fa a la relació d'aquesta variable amb les variables independents del nostre estudi, d'una banda, com recull la Taula 6 al subapartat **2.3.1.3**, un dels trets que, segons Tuson (1998), caracteritza la parla de les dones en comparació amb la dels homes és un ús més freqüent del que l'autora anomena *modalitzadors*, una categoria que, entre altres elements lingüístics, inclou els adverbis. Per tant, si bé només comptabilitza alguns adverbis (els que acaben en *-ment*), la variable V45 podria aportar dades que corroborin o refutin aquesta afirmació pel que fa a la nostra mostra. D'altra banda, aquests resultats també podrien corroborar o refutar l'afirmació que recollíem al subapartat **2.3.3.2** i que prové de Nippold (1998: 25-27), és a dir, que durant l'adolescència s'incorporen al lèxic lletrat dels parlants adverbis de probabilitat i magnitud (molts dels quals acaben en *-ment*, com *probablement* i *definitivament* o *extremament* i *lleugerament*). Caldrà, doncs, parar atenció a un possible augment d'aquests adverbis entre la primera i la segona entrevista de cada informant.

5.3.2.3.4. Les variables dependents: síntesi

En els subapartats anteriors hem fet un repàs de les variables lingüístiques d'aquest estudi i les seves variants. Hem tractat la seva relació amb estudis previs dels fenòmens lingüístics que comprenen i les interaccions que poden mantenir amb les variables independents. La Taula 20 mostra una síntesi de les correlacions esperables entre les variables dependents i independents segons la revisió bibliogràfica que s'ha dut a terme (les subhipòtesis de l'estudi). Com veurem al capítol següent (apartat 5.6), la naturalesa de les variables i les categories que les formen s'han de tenir en compte per seleccionar les tècniques d'anàlisi estadística més adequades als objectius de l'estudi (Illamola, 2015: 433).

Taula 20. Síntesi de les correlacions esperades entre les variables dependents i les variables independents segons la revisió bibliogràfica.

Variable	Correlació esperada	Bibliografia rellevant per a la subhipòtesi
V3 Doblets de partícules discursives	Sexe	Diferents estils interaccionals d'homes i dones (<i>i. a.</i> Tuson, 1998; Kring i Gordon, 1998; Baron i Ling, 2011; Bastida Serra, 2017).
	Llengua inicial	Vila (1998: 271): L'estudi de "la probable repartició funcional de les parelles de mots formades per les formes no normatives i les normatives" podria informar sobre la seva distribució sociolingüística.
V5 Obertura de la interacció	Edat	Clark (2003: 282): part de l'adquisició lingüística que dels infants consisteix a aprendre a contribuir adequadament a converses i respectar els torns de paraula (per exemple, retornar la salutació).
	Llengua inicial	Arenas Beneyto (2005: 88): "analitzar la forma utilitzada per a saludar [...] pot aportar informació sobre [...] la competència comunicativa del parlant".
V6 Heterorepetició	Edat	L'heterorepetició contribueix a l'adquisició de segones llengües i llengües estrangeres per part d'infants (Rydland i Aukrust, 2005; Čurković Kalebić, 2016).
V14 Posició dels pronoms clítics en construccions perifràstiques	Edat	Illamola i Queralt (2020): en una mostra diferent del projecte RESOL, els bilingües inicials de 12 anys de Manlleu mostren un ús dels pronoms clítics en construccions perifràstiques diferent de la resta de grups.

Metodologia

	Sexe	Illamola i Vila (2015); Illamola i Queralt (2020): no troben diferències significatives quant al sexe en una mostra provinent del projecte RESOL.
	Llengua inicial	Illamola i Vila (2015): no troben diferències significatives quant a la llengua inicial en una mostra provinent del projecte RESOL. Illamola i Queralt (2020): la llengua inicial mostra diferències estadísticament significatives entre els grups d'informants del projecte RESOL.
V16 Nombre de preguntes no retòriques	Edat	Larson i McKinley (1998): en un estudi panel de parlants d'entre 12 i 17 anys mitjançant converses en dos contextos diferents, no troben diferències per edat en el nombre total de preguntes formulades.
	Sexe	Larson i McKinley (1998): en un estudi panel de parlants d'entre 12 i 17 anys mitjançant converses en dos contextos diferents, no troben diferències per sexe en el nombre total de preguntes formulades.
	Llengua inicial	El corpus permet comprovar si els catalanoparlants inicials, pel fet de tenir com a primera llengua la que s'utilitza a la prova, formulen més preguntes que els bilingües i castellanoparlants inicials.
V17 Tipus de preguntes	Edat	El corpus permet comprovar si els informants formulen més preguntes obertes o tancades quan tenen 16 anys que quan en tenien 12.
	Sexe	El corpus permet comprovar si les noies formulen més preguntes obertes (que contribuïrien més al flux interaccional, cf. Tuson, 1998: 216) que els nois.
	Llengua inicial	El corpus permet comprovar si els catalanoparlants inicials, pel fet de tenir com a primera llengua la que s'utilitza a la prova, formulen algun dels tipus de pregunta més que els bilingües i castellanoparlants inicials.
V18 Construccions condicionals	Edat	La comprensió i producció de condicionals més complexos s'adquireixen durant la (pre)adolescència (p. ex. Aparici Aznar i Noguera Llopart, 2012; Nippold, Frantz-Kaspar i Vigeland, 2017; Badger i Mellanby, 2018).
V22 Perífrasis d'obligació	Llengua inicial	Els resultats de Mier (1986) i Stokes (2015) indiquen que algunes de les variants són més utilitzades per parlants que tenen més contacte amb el castellà.

V23 Determinants demostratius	Edat	El corpus permet comprovar si els demostratius ja s'haurien adquirit a la (pre)adolescència (cf. Diessel, 2006: 469).
	Sexe	El corpus permet comprovar si el sexe podria ser un dels factors que contribueixen a l' "estil personal de cadascú" que, segons Fitó (2009: 55), interacciona amb la freqüència de díctics.
	Llengua inicial	El corpus permet comprovar si la llengua inicial podria ser un dels factors que contribueixen a l' "estil personal de cadascú" que, segons Fitó (2009: 55), interacciona amb la freqüència de díctics.
V25 Pronoms clítics de datiu plural	Llengua inicial	L'ús d'algunes de les variants podria respondre, segons Todolí (2002) i Pineda (2018), a la influència del castellà o bé, segons Pineda (2018) ser conseqüència de l'evolució pròpia del català.
V38 Fenòmens de contacte de llengües	Edat	Les habilitats lèxiques es continuen desenvolupant durant l'adolescència, així com el coneixement sobre quan utilitzar estructures més o menys formals segons la situació comunicativa (Nippold, 2004).
	Sexe	En condicions estables d'estratificació social, les dones usen formes estàndards amb més freqüència que els homes (Labov, 2001: 266).
	Llengua inicial	Vila (1998: 271): "[é]s ben probable que algunes d'ell[s], fàcilment identificables pels parlants, funcionin com a marcadors de formalitat i/o puresa lingüística, mentre que la resta els passin inadverti[ts]".
V41 Colors	Edat	Al-Jehani (1990): en una prova d'elicitació i una mostra que inclou adolescents, identifica diferències entre els grups d'edat en el nombre de termes referits a colors que produeixen.
	Sexe	Al-Jehani (1990): en una prova d'elicitació i una mostra que inclou adolescents, identifica diferències entre homes i dones en el nombre de termes referits a colors que produeixen. Paredes García (2006): en una mostra de PRESEEA de Madrid i enquestes de disponibilitat lèxica a alumnes de 3r d'ESO a 2n de batxillerat, troba diferències qualitatives i quantitatives per sexe en la producció de termes referits a colors.

	Llengua inicial	El corpus permet comprovar si els catalanoparlants inicials, pel fet de tenir com a primera llengua la que s'utilitza a la prova, produeixen més referències a colors que els bilingües i castellanoparlants inicials.
V45 Nombre d'adverbis acabats en <i>-ment</i>	Edat	Barriga Villanueva (1994): en un estudi de converses amb 24 infants de 6 i 12 anys, troba més realitzacions d'adverbis en <i>-ment</i> a les dades dels de més edat.
	Sexe	Tuson (1998): les dones fan més ús de modalitzadors (inclosos adverbis) que els homes. Nippold (1998: 25-27): durant l'adolescència es desenvolupa el lèxic lletrat, que inclou els adverbis de probabilitat i magnitud.

5.3.2.4. Prova pilot de les variables

Després d'introduir-hi els canvis necessaris per aconseguir aquest objectiu, a la quarta fase, es va analitzar detingudament una mostra formada per cinc transcripcions escollides a l'atzar corresponents a informants representatius de cada grup en què es divideix el corpus d'anàlisi d'acord a les variables independents (dos grups d'edat, dos grups de sexe i tres grups de llengua inicial). L'objectiu d'aquesta anàlisi preliminar era posar a prova l'aplicabilitat, pel que fa al nostre corpus i objectius, de les variables utilitzades en els estudis previs ressenyats al segon capítol. Per això, es van incloure a la prova pilot les variables utilitzades al projecte ISI, Queralt (2015), Garayzábal, Queralt i Reigosa (2019: 64) i als estudis ressenyats als apartats sobre la sociolingüística de la variació de la llengua catalana i l'adquisició lingüística durant l'adolescència (2.2.4.3 i 2.3.2.2, respectivament) més rellevants per als nostres objectius i les característiques del nostre corpus (mostres de parla dialogal no planificada). Els resultats d'aquesta prova d'anàlisi amb cinc transcripcions escollides a l'atzar de cada grup d'edat, sexe i llengua inicial del corpus ens van conduir a les conclusions següents.

a) Diverses de les variables detectades a la bibliografia no eren aplicables al corpus d'anàlisi per la modalitat (oral, i no escrita) o la varietat de llengua (català central no planificat, com s'ha explicat al final de l'apartat 5.2) que recull i els objectius de l'estudi (identificar variables útils per a la perfilació lingüística,

i no realitzar una comparació forense de textos per a l'anàlisi d'autoria). Aquest grup de variables inclou, per exemple, els errors gramaticals i ortotipogràfics, la puntuació, els abreujaments, els temps de futur, el pretèrit imperfecte de subjuntiu, la freqüència relativa de paraules funcionals, hàpax legòmena i dislegòmena²⁷ o la morfologia dels clítics (*me / em*, etc.).

b) Diverses de les variables testades presentaven dificultats pràctiques per a l'anàlisi i el recompte en un corpus de les característiques i la mida del d'aquest estudi, fins al punt que podien comprometre la viabilitat del projecte en el període de temps previst. Aquestes variables eren les mesures de complexitat, els tipus d'oracions i d'oracions complexes, els actes de parla, el nombre d'oracions inacabades o els solapaments cooperatius i competitius.

c) Diverses variables no es realitzaven en la submostra del corpus amb una freqüència suficient com per concloure que podrien resultar útils per als objectius de l'estudi, com les variables *comiat*, *en aquí / aquí*, *venir / vindre*, *volguer / voler*, oracions passives o "dequeisme".

d) La variable negació amb *pas* va mostrar resultats molt condicionats per la població (Manlleu o Mataró), de manera que sembla un tret dialectal. Per això, es va procedir a l'anàlisi de tot el corpus per observar si, efectivament, aquesta variable que no s'inclouria a l'estudi (la població) podia incidir en els resultats globals i es va trobar que, mentre que a les transcripcions corresponents a Manlleu es realitzava la variant negació amb *pas* en un total de 52 ocasions, a Mataró aquesta variant només sumava 5 realitzacions. Així doncs, finalment es va decidir no incloure aquesta variable a l'estudi per evitar que una variable independent no controlada influís directament en els resultats, ja que possiblement es tracta d'un tret subdialectal (comarques centrals vs. àrea metropolitana).

²⁷ Formes que apareixen únicament una o dues vegades en un text, respectivament.

D'aquesta manera, es va arribar a una proposta inicial de variables dependents²⁸ per a l'estudi a partir de la bibliografia de referència consultada que va permetre crear, a la cinquena fase, una primera versió de la base de dades, en forma de full de càlcul d'Excel (Figura 12). Aquesta base de dades recull set ítems de metainformació per a cada transcripció del corpus, a més de les dades lingüístiques sobre les variables analitzades. Identifica cada transcripció amb el seu codi únic; informa sobre la llengua inicial, l'edat i el sexe de l'informant; i mostra la duració de la gravació, el total de paraules transcrites i el de paraules de la transcripció sense les intervencions de l'entrevistadora.

Figura 12. Fragment de la base de dades utilitzada per a l'estudi.

	A	B	C	D	E	F	G	L	M	N	O
1	VARIAB							V3 Doblet de partícules discursives			
2	TIPUS							V3B			
3	VARIA- BLE										
4	Descripció variant							a ver	a veure	a ve ure	aviam
5	ID Informant	L1	nivell	sexe	Duració gravació	Paraules (total)	Paraules (només informants)				
6	M0043SCI	2	1	2	0:11:20	1616	873	1	0	0	0
7	M0045SCI	1	1	1	0:14:04	2603	1896	0	0	0	0
8	M0048SCI	1	1	1	0:11:53	2183	1603	0	1	2	0
9	M0049SCI	2	1	2	0:14:42	2804	1855	0	0	2	0
10	M0051SCI	1	1	1	0:19:02	2864	2236	0	0	0	0
11	M0054SCI	1	1	2	0:13:54	2318	1405	0	0	0	0
12	M0089SCI	2	1	1	0:08:09	1563	1000	0	0	0	0
13	M0090SCI	2	1	1	0:12:51	1832	952	0	0	0	0
14	M0094SCI	1	1	2	0:13:31	2124	1278	0	0	0	0
15	M0096SCI	2	1	2	0:13:50	2640	1757	0	0	1	0
16	M0105SCI	2	1	2	0:11:00	2017	1161	0	0	0	0
17	M0289SCI	2	1	2	0:14:14	2437	1956	0	0	3	0
18	M0290SCI	2	1	1	0:15:52	2423	1351	0	0	0	0
19	M0293SCI	2	1	1	0:15:10	2574	1512	1	0	0	0
20	M0295SCI	2	1	2	0:16:54	2838	2076	0	2	0	0
21	M0297SCI	3	1	2	0:16:18	2685	1866	0	0	3	0
22	M0453SCI	1	1	1	0:17:51	2314	1630	0	0	0	0

La sisena fase del procés va consistir en l'anàlisi de les 138 transcripcions i el marcatge de les realitzacions de cada variant amb el codi corresponent, tal com es mostra a l'exemple següent.

²⁸ Degut als nombrosos canvis i ampliacions introduïts a aquesta primera proposta incompleta, estalviem als lectors la seva lectura i incloem en aquest treball final tan sols la llista definitiva de variables utilitzades, juntament amb les explicacions pertinents de variables i variants que van ser descartades.

Figura 13. Exemple del procés de codificació de les variants analitzades.

ENT:	bona tarda_
L0001QCI:	bona tarda_[V5c]
ENT:	mira la setmana que ve_

Un cop marcades totes les realitzacions identificades al corpus de cadascuna de les variants d'una variable, es procedia al recompte automàtic mitjançant la interfície de consola Windows PowerShell. Aquest programa permet executar línies de comandaments (o *scripts*) que poden agilitzar el procés de recompte. Per als nostres propòsits, es va dissenyar un conjunt de comandaments que mostren el número total d'ocurrències de les unitats que ens interessin (en aquest cas, el codi d'una variant) en diversos arxius de forma automàtica, de manera que no cal anar arxiu per arxiu per finalitzar el recompte de cada element (paraula o codi alfanumèric) a tot el corpus. Aquest conjunt de comandaments es pot utilitzar, però, per a estudis futurs en què també calgui comptar el nombre de realitzacions de determinades unitats en un conjunt d'arxius concret.

Les figures següents mostren el codi informàtic utilitzat i la visualització del resultat d'executar-lo a Windows Powershell, és a dir, el recompte del total de realitzacions de V22b, corresponent a la variant *haver de* de la variable V22 Perífrasis d'obligació. Com es pot observar a la Figura 14, el retorn està format per una llista, encapçalada per la unitat cercada (en aquest exemple, V22b), que mostra el recompte de total de realitzacions de l'element cercat a cada arxiu de la carpeta utilitzada seguit de punt i coma i el nom de l'arxiu. Els arxius que no contenen cap realització no apareixen a la llista.

Figura 14. Script o conjunt de comandaments executats amb Windows PowerShell per al recompte automàtic.

```

$files = Get-ChildItem [ruta als arxius del corpus d'anàlisi .txt]

$words1 = Get-Content [ruta a l'arxiu .txt que conté les unitats de cerca .txt]

$words = $words1.Split(";")

Write-Output $words1

foreach ($file in $files) {
    $FileContent = Get-Content $file
    $test = ""
    foreach($Word in $words){
        $Matches = Select-String -InputObject $FileContent -Pattern $word -AllMatches
        $test2 = $Matches.Matches.Count
        $test = $test + $test2 + ";"
    }
    $outputFile = Split-Path $file -leaf
    $test = $test + $outputFile
    Write-Output $test
}

```

Figura 15. Exemple del procés de recompte automàtic de les variants analitzades.

```

Windows PowerShell
Copyright (C) Microsoft Corporation. Todos los derechos reservados.

Prueba la nueva tecnología PowerShell multiplataforma https://aka.ms/pscore6

PS C:\Users\Roser> cd C:\Users\Roser\Desktop\V22
PS C:\Users\Roser\Desktop\V22> .\test.ps1
V22b
4;L0001QCI.txt
1;L00015CI.txt
3;L0002QCI.txt
3;L00025CI.txt
10;L0003QCI.txt
4;L00035CI.txt
0;L0004QCI.txt
0;L00045CI.txt
14;L0005QCI.txt
8;L00055CI.txt
8;L0006QCI.txt

```

Com s'ha explicat anteriorment (apartat 5.3.2.3), només es van mantenir a l'anàlisi les variants per a les quals el nombre total d'ocurrències era igual o superior a deu, ja que aquest nombre de dades lingüístiques permet assegurar la fiabilitat dels resultats segons diversos autors (Labov, 1966; Etxeberria, Joaristi i Lizasoain, 1990: 278; Queralt, 2015: 61). Així doncs, algunes variants i variables identificades a la bibliografia, malgrat haver estat identificades a la prova pilot de possibles variables com a potencialment rellevants, finalment no van passar el filtre de realitzacions mínimes i, per tant, no s'inclouen a l'anàlisi definitiu.

Concretament, la Taula 21 mostra les variants de les variables dependents de l'estudi que van ser excloses de l'anàlisi per falta de realitzacions.

Taula 21. Variants excloses de l'anàlisi per falta de realitzacions al corpus.

Variable	Variant	Total realitzacions (<10)
V3a: <i>además / a més</i>	V3a1: <i>además</i>	1
	V3a2: <i>ademés</i>	3
V3d: <i>casi / quasi</i>	V3d2: <i>quasi</i>	8
V3e: <i>claro / (és) clar</i>	V3e1: <i>claro</i>	1
V3g: <i>hasta / fins (i tot)</i>	V3g1: <i>hasta</i>	5
V3m: <i>vaya / vaja</i>	V3m1: <i>vaya</i>	7
	V3m2: <i>vaja</i>	2
V3n: <i>y eso / i (tot) això</i>	V3n1: <i>y eso</i>	9
V5 Obertura de la interacció	V5e: <i>hola bona tarda</i>	0
	V5f: [∅]	1
	V5g: <i>hola bones</i>	1
V14b Enclisi	V14b3: <i>anar a + INF + PC</i>	5
V14c Doble	V14c1: <i>mateix pronom</i>	6
	V14c2: <i>pronoms diferents</i>	6
V22 Perífrasis d'obligació	V22d: <i>haver que + INF</i>	1
V18 Construccions condicionals	V18d: <i>amb / en que</i>	6
	V18e: <i>con que</i>	6
	V18h: <i>només que</i>	1

5.3.3. Tercera etapa: l'anàlisi

Un cop finalitzat el recompte de cada variant i introduïdes les dades quantitatives a la matriu d'Excel (fase 8), es va realitzar una anàlisi estadística exploratòria, un resum de la qual es pot trobar a Queralt i Giménez (2019), que tracta alguns dels resultats preliminars d'aquest projecte i de les nombroses possibilitats d'explotació del corpus d'anàlisi complementàries a les presentades aquí en relació a la perfilació lingüística de mostres anònimes en català produïdes per parlants menors d'edat. Aquesta primera anàlisi estadística va permetre refinar la selecció i recodificació de les variables d'anàlisi (per exemple, es va reconceptualitzar la variable que considera elements com *bueno* o *doncs* i es va operacionalitzar com a V3 Doblets de partícules discursives). Conseqüentment, a l'onzena fase, es va modificar la base de dades per reflectir les variables dependents definitives que encara es van sotmetre a la revisió esmentada anteriorment abans de ser validades per tal de garantir que la correcció del recompte no s'havia vist afectada pels canvis en la conceptualització de les variables i poder adreçar qualsevol conseqüència que se'n derivés. Un cop finalitzada aquesta revisió minuciosa, es va procedir al tractament estadístic de les dades, per al qual es va utilitzar el paquet estadístic SPSS (versió 27). Com s'ha explicat al tercer capítol, l'objectiu d'aquesta etapa central de l'estudi era explorar les relacions entre les variables i determinar si existeix una relació significativa entre les variables dependents i independents. El plantejament de l'anàlisi estadística i les tècniques estadístiques utilitzades s'expliquen a l'apartat 5.6 i els resultats que se'n van extreure es detallen al proper capítol.

5.4. La codificació de les variables independents

Les dades que conformen les variables independents del nostre estudi es van extreure de les matrius d'Excel corresponents al buidatge dels qüestionaris omplerts de forma anònima pels informants. El qüestionari utilitzat a les diverses onades de treball de camp (que es pot consultar a Illamola, 2015: 583-595) va ser confeccionat per un equip de filòlegs, sociòlegs i lingüistes i es va distribuir entre els informants en el marc del projecte RESOL. Consta de set parts i recull informació sociodemogràfica i sociolingüística, com l'ús que fan els informants del seu temps lliure, informació sobre la seva xarxa social o els seus coneixements, actituds i identitats lingüístics (Illamola, 2015: 405).

Per possibilitzar el tractament posterior amb SPSS, les variables independents es van codificar a la matriu d'Excel amb els codis numèrics que mostra la Figura 12. Just a continuació del codi de cada informant, a la columna B, s'hi va introduir un 1 (equivalent a català), un 2 (castellà) o un 3 (ambdues) segons la llengua inicial declarada per l'informant. A la columna C, es va codificar el nivell educatiu o, el que és el mateix per als nostres objectius, el grup d'edat dels informants (1 en aquest camp equival a 12-13 anys i 2 a 15-16). Finalment, la columna D de la matriu recull la informació relativa al sexe dels participants de manera que els nois hi tenen un 1 i les noies un 2.

A continuació, exposem la metodologia seguida per a la codificació de cadascuna de les variables lingüístiques. Tot seguit, acabem aquest capítol amb els passos seguits per al tractament estadístic de les dades i l'explicació de les proves utilitzades per a l'obtenció dels resultats.

5.5. La codificació de les variables dependents

Tant per a la codificació com per al recompte de les ocurrencies de les variables dependents de l'estudi, es va seguir el principi de responsabilitat de Labov (1982: 30), segons el qual s'han d'incloure totes les dades lingüístiques a l'anàlisi, independentment de si confirmen o contradiuen les hipòtesis de partida (Queralt, 2015: 61). A continuació, s'expliquen els criteris específics que s'han seguit per a la codificació de cada variable.

5.5.1. Doblets de partícules discursives (V3)

L'objectiu d'aquesta variable és comptabilitzar partícules discursives que en registres informals se solen produir tant en les seves variants catalanes (per exemple, *a més a més*) com castellanen (*además*). Considerem formes sorgides del contacte lingüístic en forma de manlleus o calcs (com *ademés*) com a formes catalanes. Com s'ha explicat anteriorment (subapartat **2.2.4.3**, nota 15), el nivell fonològic-fonètic queda fora de l'abast d'aquesta tesi, per la qual cosa, en la majoria de partícules analitzades (a excepció de V3j *pero* / *però*), no s'han tingut en compte aspectes de realització fonètica de les variants (per exemple, si un informant pronuncia *bueno* com [βwenu] o com [βweno]), sinó que s'ha seguit criteris morfològics per a la seva classificació com a formes catalanes o castellanen (seguint amb l'exemple de V3c, *bueno* es considera castellà tot i la seva gran extensió en registres informals del català).

Pel que fa a la noció controvertida de partícula discursiva (cf. Aschenberg i Loureda Lamas, 2011; Garachana Camarero, 2012; Degand, Cornillie i Pietrandrea, 2013; Fischer i Alm, 2013), en aquesta ocasió seguim la proposta de Briz, Pons i Portolés (2008). Aquesta obra la defineix com el conjunt de «elementos lingüísticos que guían la interpretación del discurso», que compleixen els requisits següents:

- a) tenen un valor interaccional (serveixen per justificar una opinió o conclusió, canviar el tema de la conversa, reforçar algun element, interpel·lar l'interlocutor, etc.) o duen a terme una funció de connexió argumentativa, modalització, focalització o control de contacte;

- b) «no se corresponden con ninguna de las categorías gramaticales establecidas», i
- c) han patit un procés de gramaticalització.

Per tant, s'han tingut en compte per a aquesta variable aquelles partícules discursives (1) llistades al *Diccionario de partículas discursivas del español* (DPDE) o que compleixen amb els requisits i la definició de Briz, Pons i Portolés (2008) i (2) que tenen formes equivalents en català. La Taula 22 recull els doblats analitzats i el seus codis i variants corresponents. Així, s'han inclòs en aquesta variable partícules que no apareixen al DPDE, però que considerem que, en els usos que se'n fa al corpus, encaixen amb els requisits que acabem de descriure (marcades amb un asterisc a la Taula 22). Només aquelles variants que apareixen amb una freqüència superior a deu van ser incloses a l'anàlisi estadística posterior. En els casos (sombrejats en gris a la taula) en què alguna de les variants en castellà o en català d'alguna de les partícules analitzades no van superar aquest requisit i, per tant no es podia fer la comparació entre la forma castellana i almenys una forma catalana, es va excloure la partícula de l'estudi, ja que la variable V3 contempla tan sols els doblats de partícules en les dues llengües.

Taula 22. Partícules discursives analitzades en la variable V3.

Doblets de partícules discursives (V3)					
V3a	V3a1	<i>además</i>	V3h	V3h1	<i>hombre</i>
	V3a2	<i>ademés</i>		V3h2	<i>home</i>
	V3a3	<i>a més</i>	V3i	V3i1	<i>o sea</i>
	V3a4	<i>a més a més</i>		V3i2	<i>o sigui</i>
V3b*	V3b1	<i>a ver</i>	V3j*	V3j1	<i>pero</i>
	V3b2	<i>a veure</i>		V3j2	<i>però</i>
	V3b3	<i>a ve(u)re</i>		V3j3	<i>p(e)rò</i>
	V3b4	<i>aviam</i>		V3j4	<i>p(er)ò</i>
V3c	V3c1	<i>bueno</i>	V3k	V3k1	<i>pues</i>
	V3c2	<i>bé</i>		V3k2	<i>pos</i>
V3d	V3d1	<i>casi / q(u)asi</i>		V3k3	<i>doncs</i>
	V3d2	<i>quasi</i>	V3l	V3l1	<i>vale</i>
V3e	V3e1	<i>claro</i>		V3l2	<i>val</i>
	V3e2	<i>clar</i>		V3l3	<i>d'acord</i>
	V3e3	<i>esclar / és clar</i>	V3m*	V3m1	<i>vaya</i>
V3f*	V3f1	<i>después</i>		V3m2	<i>vaja</i>
	V3f2	<i>després</i>	V3n*	V3n1	<i>y eso</i>
V3g	V3g1	<i>hasta</i>		V3n2	<i>i això</i>
	V3g2	<i>fins</i>		V3n3	<i>i tot això</i>
	V3g3	<i>fins i tot</i>		V3n4	<i>i tot</i>

5.5.2. Obertura de la interacció (V5)

Aquesta variable contempla les diverses estructures amb què els informants inicien la seva interacció amb l'entrevistadora en la tasca del joc de rol (*bon dia, hola, etc.*). Així doncs, s'adopta una perspectiva formal i no de l'anàlisi de la conversa, és a dir, no es té en compte, per exemple, si l'opció que escull un informant és o no la repetició de la forma que pronuncia l'entrevistadora en el torn immediatament precedent, ja que aquest és el fenomen que s'observa en la variable V6 Heterorepeticions.

S'han codificat totes les ocurrències de cadascuna de les variants de V5. En la majoria de transcripcions, només s'inicia la interacció del joc de rol una vegada. Ara bé, en algunes, els informants produeixen més d'una variant, ja sigui de forma espontània, com en el primer dels dos exemples següents, o perquè l'informant intenta iniciar la tasca de joc de rol, però no aconsegueix que l'entrevistadora hi col·labori. Això succeeix, per exemple, quan l'informant pretén iniciar la tasca abans que l'entrevistadora hagi acabat la introducció o abans d'haver-se familiaritzat amb les làmines de què disposa, com s'observa en el segon dels exemples que es reproduïxen tot seguit.

S'ha diferenciat entre realitzacions de la interjecció *hola* seguides de la fórmula de salutació *bon dia* o *bona tarda* en un sol grup tonal o en dos. Així, mentre que en el primer exemple de la Figura 16 es considera que l'informant produeix dues variants diferents, en el segon es marca *hola bon dia* com una única variant. Això es deu al fet que en el primer cas les unitats *hola* i *bon dia* es realitzen en dos grups tonals, transcrits en línies diferents, però en el segon cas es pronuncia la interjecció i la fórmula de salutació en el mateix grup.

Figura 16. Exemples de múltiples ocurrencies de V5.

<i>Exemple 1. L'informant produeix dues estructures espontàniament.</i>	
ENT:	bon dia_ ((tus))
L0006SCI:	hola \ [V5a] bon dia \ [V5d] voldria_ estaria interessada en comprar un animal/
<i>Exemple 2. L'informant s'anticipa a l'inici de la tasca.</i>	
L0060SCI:	val_ e:m hola _ [V5a]
ENT:	mi- mira-t'ho tot_
L0060SCI:	[ah val_]
ENT:	[mira-te'ls primer_ perquè tinguis una mica clar què és el que tens i llavors comencem_ ((pausa breu))
L0060SCI:	ja està_
ENT:	sí/ vinga bon dia_
L0060SCI:	hola bon dia _ [V5d]

5.5.3. Heterorepeticions (V6)

Per a aquesta variable s'han tingut en compte els fragments en què els informants incorporen en els seus torns elements lèxics dels torns precedents de l'entrevistadora, ja siguin de forma literal (com en el primer exemple reproduït a continuació) com adaptada (com en el segon). Tot i que, en la majoria d'ocasions, els elements lèxics repetits pels informants i els produïts en primer lloc per l'entrevistadora ocorren en torns consecutius (com al primer exemple), aquest no sempre és el cas (com en el segon).

Figura 17. Exemples de V6.

<i>Exemple 1. Ocurrencia d'heterorepetició literal.</i>	
ENT:	i per netejar-lo/
L0006SCI:	per netejar-lo [V6] és fàcil_
<i>Exemple 2. Ocurrencia d'heterorepetició adaptada.</i>	
ENT:	i tot el dia sol dins la gàbia no s'avorrirà/ potser que n'hi posi un altre o què/
L0047QCI:	sí\ pots posar-n'hi un altre [V6] tranquil·lament_

5.5.4. Posició dels pronoms clítics en construccions perifràstiques (V14)

Aquesta variable considera la variació que permeten les estructures perifràstiques amb pronoms clítics, que poden aparèixer en posició de proclisi, abans del verb auxiliar (V'), o d'enclisi, després del verb auxiliat (V''). A continuació, es mostren exemples de les ocurrencies de cada variant marcades al corpus d'anàlisi.

Taula 23. Exemples de les variants de V14.

Perífrasi verbal	Proclisi	Enclisi
<i>poder</i> + INF	M0723QCI: lo que li podies comprar :_[V14a3]	M0723QCI: pots posar-la [V14b3] tancada_
<i>haver de</i> + INF	M0728SCI: i després tu ho has de recollir _[V14a2]	M0728SCI: has de cuidar-lo [V14b2] bé mimar-lo molt e:m_
<i>tenir de</i> + INF	M0105SCI: [jo et tinc de vendre [V14a1] a tu/	M0105QCI: cada tres mesos tens d'anar a fer-li [V14b1] una revisió_
<i>tenir que</i> + INF	M0723QCI: a ve(u)re li tens que comprar [V14a1] la seva gàbia com tots_	M0723QCI: pues llavors tenen que comprar-li [V14b1] un gàbia:_
<i>anar a</i> + INF	M0723SCI: [...] algun familiar ve i li va a posar [V14a4] menjar i aigua_	M0472SCI: ell an- anirà corrents a buscar-la _[V14b4]

Per al marcatge d'ocurrencies, es van prendre diverses decisions metodològiques prèvies. En primer lloc, només es van tenir en compte construccions en què apareixen tots els elements de la perífrasi (pronomen, verb auxiliar, verb auxiliat), la qual cosa no succeeix sempre que es produeix algun d'aquests elements, ja que la producció oral espontània fomenta l'ús d'omissions i reformulacions, com mostren els exemples següents.

Figura 18. Exemples d'ocurrencies descartades per a la codificació de V14.

<i>Exemple 1. No-realització de part dels elements perifràstics.</i>	
ENT:	ja però és que ell ha d'estar a l'escola estudiant i això\ no l'enviaré al camp_
L0001SCI:	es pot _ [absència de V'] pot comprar-ho_[V14b3]
<i>Exemple 2. Elisió de part dels elements perifràstics.</i>	
ENT:	no sé i si li regalo el gat i no li agrada què_
L0002SCI:	doncs pots venir i tornar-lo _ [elisió de V']

Exemple 2'. Elisió de part dels elements perifràstics.

ENT: i què es pot fer llavors/
 si el treus_
 L0020QCI: no no sé_
deixar-lo córrer [elisió de V'] per (a)llà_

La decisió de descartar ocurrencies com la del segon exemple, com expliquen Illamola i Vila (2015: 44) per al cas del castellà, es deu al fet que, en fragments en què s'elideix V', "la variabilidad es nula; es decir, la posibilidad de anteponer el pronombre no es posible, puesto que al explicitar solamente el infinitivo, esta forma no verbal del verbo solo admite el empleo del pronombre enclíticamente (**le dar de comer*, **lo sacar a pasear*)". Aquesta limitació gramatical també és aplicable al cas del català (per exemple, **doncs pots venir i el tornar*).

En segon lloc, durant el procés de codificació, es va crear una nova variant, V14c: Doble, per recollir les ocurrencies en què els informants produïen pronoms tant en posició proclítica com enclítica. Al seu torn, V14c es va dividir en V14c1, per a ocurrencies en què els pronoms preposats i els posposats són els mateixos (pleonasmes), i en V14c2, per a aquelles en què els pronoms preposats són diferents als posposats (la Figura 19 en mostra alguns exemples). Ambdós casos s'entenen com productes no normatius de la parla espontània. Això no obstant, aquestes variants de V14 no van superar el mínim de deu ocurrencies en tot el corpus d'anàlisi (vegeu Taula 21 al subapartat 5.3.2.4) i, per tant, no es tindran en compte per a l'anàlisi estadística de la variable. Sí que s'hi inclou, però, el total de V14c.

Figura 19. Exemples de la variant V14c.

Exemple 1. Producció del mateix pronom en posició proclítica i enclítica (V14c1).

M0722QCI: no **el necessites banyar-lo** [V14c1] molt_

Exemple 2. Producció de pronoms diferents en posició proclítica i enclítica (V14c2).

M0045QCI: i també **s'ha de ficar-li** [V14c2] sorra: /

En tercer lloc, no s'han considerat les ocurrencies del clíctic *es* amb la perífrasi *poder* + INF en què la variació en la posició proclítica o enclítica afecta el significat de la construcció. Com argumenten Illamola i Vila (2015: 45), és possible distingir aquestes ocurrencies de l'ús impersonal del pronom *es* observant els arguments regits per *poder*. Així, no s'han tingut en compte

fragments com *es pot passejar*, en què la posició alternativa del clític (*pot passejar-se*) resultaria en un canvi semàntic per influència del castellà en la varietat vernacle del català utilitzada pels informants.

En darrer lloc, també s'han descartat fragments en què els informants repeteixen algun dels elements que formen la construcció perifràstica (com a *i es pot pot estar per casa*) per la impossibilitat de saber si, mitjançant la repetició, el parlant pretén reformular tot el fragment, inclòs el pronom ja produït, només una part dels elements realitzats abans de la repetició o la repetició no comporta cap intent de reformular el que ja ha pronunciat. Aquest tipus de produccions també es deuen al mode de producció, ja que són resultat de la manca de planificació de la conversa oral.

5.5.5. Nombre de preguntes no retòriques (V16)

S'han analitzat com instàncies de V16 les preguntes formulades pels informants. Ara bé, per a la codificació d'aquesta variable s'han descartat intervencions dels informants formulades amb entonació interrogativa però que no constitueixen preguntes, peticions d'informació, a la seva interlocutora. És a dir, com mostra la Figura 20, no s'han comptabilitzat les preguntes retòriques, en què s'han inclòs ocurrencies de marcadors discursius i heterorepeticions que, tot i ser produïts amb entonació ascendent, no requereixen la intervenció de l'entrevistadora.

Figura 20. Exemples d'ocurrencies descartades per a la codificació de V16.

<i>Exemple 1. Preguntes retòriques.</i>	
L0021QCI:	m:: inconvenient del gat/ que deixa pèl_
<i>Exemple 2. Marcadors discursius que funcionen com preguntes retòriques.</i>	
ENT:	i per rentar com es fa/
L0003SCI:	pues_ obres això no/ i::_
<i>Exemple 3. Heterorepeticions que funcionen com preguntes retòriques.</i>	
ENT:	i què fa aquest_
L0012QCI:	aquest/ doncs igual que el hà- -- que el hámster_ no no pots fer gaire re:s_

5.5.6. Tipus de preguntes (V17)

En estudis lingüístics, se sol distingir entre interrogatives absolutes, aquelles en què es pregunta sobre el contingut del conjunt de l'enunciat, que moltes vegades es poden respondre afirmativament o negativament, i interrogatives parcials, en les quals es pregunta sobre un dels constituents (Payrató, 2002: 1202; Poblete, 2002; Font Rotchés, 2008: 299; Vanrell, Mascaró, Prieto i Torres-Tamarit, 2010). No obstant això, en aquesta variable, per motius de claredat, parlem de preguntes obertes i tancades, seguint la distinció de Busquet, Medina i Sort (2006: 164):

“Podem parlar de *preguntes obertes* o *tancades* segons el grau de llibertat en la resposta. Una pregunta oberta vol dir que l'individu pot respondre lliurement allò que li sembla i com li sembla, mentre que una pregunta tancada significa que l'individu ha de triar entre un ventall limitat d'opcions de resposta. Per exemple ha de triar entre un *sí* i un *no*.”

Així doncs, com es mostra a la Figura 21, aquelles preguntes formulades pels informants que l'entrevistadora pot respondre com cregui més convenient (encara que formalment de vegades siguin de sí o no, com a l'exemple 1b) s'han codificat com ocurrencies de la variant V17a i les que en limiten la resposta s'han codificat com V17b. La limitació de les respostes possibles pot donar-se mitjançant preguntes de sí o no (exemple 2a), la inclusió de les alternatives en la formulació de la pregunta (exemple 2b) o la invitació a escollir entre un grup limitat de respostes possibles a través del pronom *quin* (exemple 2c) o del tipus d'informació que es demana (a l'exemple 2d, quan es pregunta per l'edat del destinatari, segurament no s'espera obtenir com a resposta un nombre molt superior o molt inferior a la dels participants o l'entrevistadora).

Figura 21. Exemples de la codificació de les variants de V17.

<p>1. <i>Ocurrències de la variant V17a Pregunta oberta.</i></p> <p>a. M0295QCI: bon dia què desitja/ [V17a]</p> <p>b. L0019QCI: tens alguna idea/ algun animal que li pugui agradar/ [V17a]</p> <p>2. <i>Ocurrències de la variant V17b Pregunta tancada.</i></p> <p>a. M0045SCI: hi ha algun altre animal a casa/ [V17b] ENT: no no n'ha tingut mai cap\</p> <p>b. ENT: i quant val això/ L0001QCI: e:: el què/ el terrari o la granota/ [V17b]</p> <p>c. M0737SCI: són a- p(e)rò quants anys té el teu cosí_ [V17b] ENT: tretze farà_</p> <p>d. L0051SCI: de moment t'agrada la ide(i)a del gat o del hámster/[V17b] quina és la que t'agrada més/ [V17b]</p>
--

5.5.7. Construccions condicionals (V18)

El propòsit d'aquesta variable és observar el nombre de construccions condicionals produïdes pels informants. Per això, les variants que la conformen recullen els diversos nexes condicionals que poden introduir aquestes construccions i és l'aparició d'aquests elements, els nexes, que s'ha tingut en compte per a la codificació i no pas criteris sintàctics, com el fet que la construcció produïda pels informants sigui o no gramatical o contingui tots els elements d'una oració condicional. Aquesta decisió metodològica respon a les característiques del corpus que, recordem una vegada més, està format per mostres de llengua oral espontània, en què sovint les intervencions són incompletes i, fins i tot, agramaticals, però compleixen satisfactòriament les seves funcions comunicatives.

Pel que fa a decisions específiques sobre cadascuna de les variants en què es divideix la variable V18, cal destacar, primer, que s'ha diferenciat entre els usos condicionals de la partícula *si* i els que, en construccions amb verbs com *saber* o *demanar*, introdueixen proposicions interrogatives (per exemple, *pot agafar-lo p(e)rò no sé si mossega*). Aquests darrers no s'han considerat ocurrències de V18b. Segon, que s'ha considerat que les proposicions introduïdes per *si* en construccions comparatives com *és com si et fan així* o *com si fos la muntanya sí*

que conformen ocurrencies de V18b, ja que aquestes proposicions conserven el seu significat condicional, encara que es troben dins de construccions comparatives.

Tercer, cal tenir en compte que, des d'un punt de vista formal, la variant V18g agrupa diversos nexes condicionals (*si acaso, si a cas, en tot cas i en cas de que*). Aquesta decisió, de nou, respon a la naturalesa de la llengua emprada pels informants. Sovint, en varietats col·loquials, s'utilitzen formes no recollides a la normativa (com *si a cas*, que en canvi sí que apareix al DCVB, o *si acaso*, del castellà). Tot i això, en aquest cas, hem decidit que, per al propòsit de la variable V18 (possibilitar el recompte del total de construccions produïdes pels informants i, potser, observar preferències per l'ús d'un o altre nexa), les diferències entre les formes dels nexes amb *cas* utilitzades al corpus no eren prou accentuades com per considerar-les ocurrencies de tipus de nexa diferents i, per tant, les hem agrupat en una sola variant de V18 (V18g).

5.5.8. Perífrasis d'obligació (V22)

Com s'ha explicat anteriorment, aquesta variable recull totes les ocurrencies de les perífrasis *tenir que* (V22a), *haver de* (V22b), *tenir de* (V22c), *haver que* (V22d) i *caldre* (V22e) seguits d'infinitiu. Això no obstant, el recompte ha revelat que en tot el corpus només apareix una ocurrencia d'*haver que* (*hi ha que*), per la qual cosa aquesta variant no es tindrà en compte a l'anàlisi. Com que l'objectiu de la variable era permetre la comparació entre les diverses variants produïdes pels informants, s'han marcat com a ocurrencies tant aquelles amb verb auxiliar, preposició o conjunció i verb auxiliat com aquelles altres en què només hi apareix el primer (seguit de *que* o *de* en el cas de V22a i V22c).

Figura 22. Exemples de V22.

<i>Exemple 1. Ocurrencia de la perífrasi verbal completa.</i>	
L0004QCI:	i no es té de fer gairebé res_ [V22c]
	només:_
	posar el menjar [tres cops] al di:a_
<i>Exemple 2. Ocurrencia tan sols del verb auxiliar i preposició.</i>	
L0004QCI:	doncs el tens de: _ [V22c]
	bueno_

5.5.9. Determinants demostratius (V23)

Per a la codificació d'aquesta variable, va ser necessari prendre tres decisions metodològiques principals. En primer lloc, com en l'etiquetatge d'altres variables, algunes de les ocurrències de demostratius identificades al corpus constitueixen produccions no admissibles segons la gramàtica normativa, però que apareixen en mostres espontànies de varietats col·loquials de la llengua. Aquest és el cas de les ocurrències agrupades en les variants V23c i V23d (*els gats aquests* o *les rodes aquelles*, per exemple) o també de dues estructures que es reproduïen a continuació i que no s'han inclòs al procés de codificació, atribuïdes a errors relacionats amb la falta de planificació de les mostres o a dificultats en la transcripció.

Figura 23. Ocurrències de demostratius descartades per a la codificació de V23.

M0472SCI:	rates mortes\ de peluche aquests _
M0734SCI:	suposo que: a les tendes_ le:s botigues que hi tenen mussols aquests _

En segon lloc, també cal assenyalar que es consideren ocurrències de les variants V23g i V23h tant contruccions amb l'article indefinit (com *en un coixí especial d'aquells* o *un pal d'aquests*) com aquelles en què apareix el substantiu en plural sense l'article (per exemple, *insectes d'aquests així*).

En darrer lloc, durant el procés de codificació, es va decidir identificar com a ocurrències de V23c i V23d dues construccions amb la preposició *de* abans del demostratiu, ja que també contenen l'article definit (*les joguines d'aquestes* i *les plantes d'aquelles*). Donat el nombre anecdòtic d'aquestes construccions que formalment podrien situar-se a cavall entre V23c/V23d i V23g/V23h, es va decidir no crear una nova variant, sinó marcar-les com a ocurrències de les ja mencionades.

5.5.10. Pronoms clítics de datiu plural (V25)

Aquesta variable recull les instàncies en què els pronoms clítics de datiu, tant preposats com posposats a les formes verbals, es refereixen a un sintagma que té de nucli un nom plural (com *gats*) o un nom col·lectiu (com *gent*). S'han considerat noms col·lectius els recollits a l'Annex 17 de Solé (2002). El contacte entre el català i el castellà, així com el registre i l'espontaneïtat del corpus propicien l'ús del pronom *li* en referència a entitats plurals, de manera que, a més dels pronoms descrits a les obres normatives com plurals (*els* i *els hi*), s'han tingut en compte aquests usos de *li* i també les ocurrències del pronom *lis* (en dues ocasions, realitzat *les*²⁹), que entenem que es deu al contacte lingüístic. Les instàncies de *li* el referent de les quals no es podia establir clarament pel context no s'han inclòs al recompte.

Figura 24. Exemples de V25.

Exemple 1. Ocurrència de V25a.

L0001SCI: do:ncs: una mica gran pe:r_
perquè jugui_
perquè **els hi** agrada molt jugar_[V25a]

Exemple 2. Ocurrència de V25b.

M049QCI: i:: **els** agrada molt estar per casa:_[V25b]

Exemple 3. Ocurrència de V25c.

L0031QCI per la gent que **li** agrada tenir-ho_[V25c]

Exemple 4. Ocurrència de V25d.

M0295SCI: i per lo tant **lis** fa falta calor_[V25d]

²⁹ A causa del limitat nombre d'aquestes realitzacions i a la similitud formal i d'origen entre *lis* i *les*, en comptes de crear una nova variable tan sols per a dues instàncies, s'han incorporat al recompte de V25d.

5.5.11. Fenòmens de contacte de llengües (V38)

Com s'ha explicat al subapartat sobre la definició i la bibliografia prèvia en què es basa aquesta variable (**punt K** al subapartat **5.3.2.3.3**), es divideix en quatre variants i recull aquelles unitats lèxiques (mots, com *cuidado*, i expressions, com *a la seva bola* o *como que no*) que no es troben a les obres lexicogràfiques de referència (DIEC, DCVB i *diccionari.cat*). També s'han utilitzat dues obres de consulta complementàries (els portals Optimot i ésAdir) per comprovar que certes unitats, tot i que força esteses en certes varietats vernaculars, no estan recollides en obres normatives i tenen una alternativa genuïna.

Així, s'ha aplicat per a la codificació d'aquesta variable un criteri normatiu que no té en compte el nombre de realitzacions de les unitats al corpus perquè, possiblement a causa de les seves característiques (tema, situació comunicativa semblant a un examen, informants joves, etc.), algunes unitats que en altres mostres de la llengua o en l'observació de les comunitats de parla semblen força esteses aquí apareixen un nombre reduït de vegades o al revés. Per exemple, mentre que Boix i Vila (1998: 245) utilitzen com a exemple de “manlleu molt integrat” el mot *alfombra*, en canvi, al nostre corpus d'anàlisi només ocorre en una ocasió o, tot i que el verb *aranyar* en comptes d'*esgarrapar* no apareix als cercadors emprats per consultar possibles importacions esteses en varietats col·loquials, al corpus hi apareix fins a 25 vegades (vegeu **Annex 3**).

El procés de codificació d'aquesta variable ha consistit en dues fases. En la primera, s'ha utilitzat el cercador incorporat al processador de text per marcar de forma automàtica les realitzacions de partícules discursives i unitats no normatives molt freqüents (com *vale*, *bueno* o *algo*), tenint en compte que les pautes de transcripció fan possible l'aparició de signes de puntuació entre les lletres que formen els mots, per exemple, per a la representació d'elisions o d'allargaments vocàlics (com en *val(e)* o *algo:*). En la segona fase, s'ha procedit a la codificació “manual”, és a dir, a la revisió transcripció per transcripció, la identificació i codificació de realitzacions de V38. D'aquesta manera, s'han pogut incloure a l'anàlisi unitats poc freqüents i expressions dividides entre línies successives a la transcripció.

5.5.12. Colors (V41)

Aquesta variable inclou totes les mencions a colors que fan els informants al corpus i distingeix si utilitzen o no el substantiu *color* seguit de *blau*, *verd*, *vermell*, etc.

Figura 25. Exemples de V41.

1. Exemple de V41a.	
M0290SCI:	i té ratlles de color marró _ [V41a]
2. Exemples de V41b.	
M0470QCI:	sí hi ha una que és verinosa_ per això no: no la solem vendre_ que és groga [V41b] i negra _ [V41b]

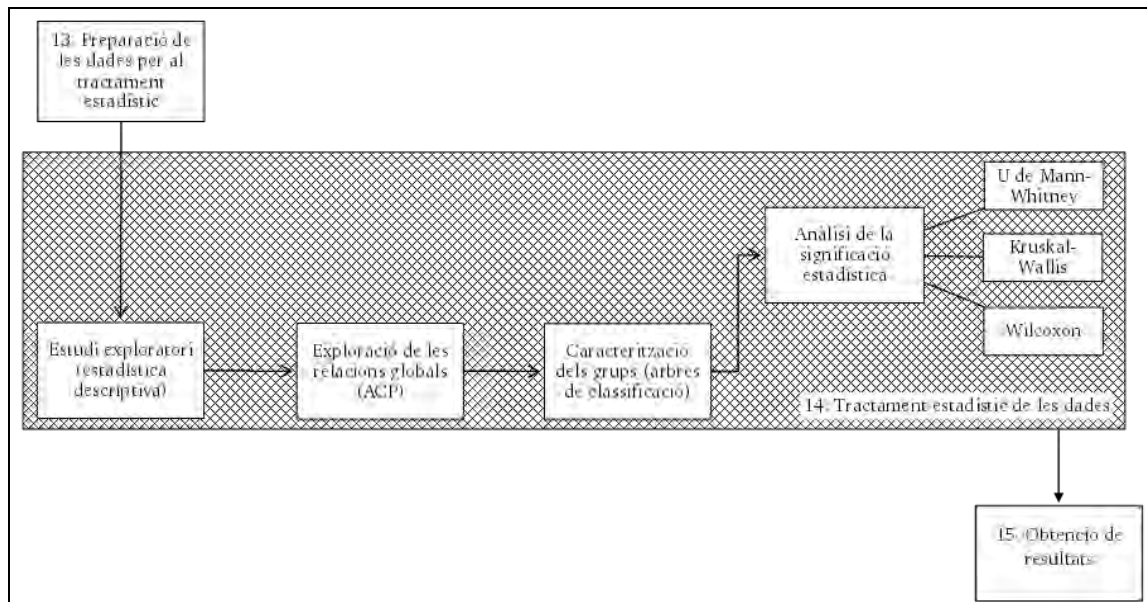
5.5.13. Nombre d'adverbis acabats en *-ment* (V45)

La variable V45 proporciona el recompte total d'adverbis acabats en *-ment* que produeixen els informants en cadascuna de les transcripcions. Cal assenyalar, però, que s'han codificat com a ocurrencies d'aquesta variable tots els adverbis acabats en *-ment* produïts pels informants a excepció de *malament*. S'ha considerat que aquest adverbi d'ús molt comú és la forma més habitual en el dialecte central per expressar el significat '[n]o gens bé, tot altre que bé' (DIEC2009b), de manera que el seu ús desvirtuaria un possible patró d'adquisició de la resta d'adverbis acabats en *-ment*, ja que, tot i que podria incloure's a la categoria de modalitzadors de què parla Tuson (1998), no seria un dels adverbis acabats en *-ment* que, segons Nippold (1998: 25-27) s'incorporen al lèxic lletrat durant l'etapa adolescent (vegeu el subapartat **2.3.2.2**), sinó un dels que s'adquiririen en les primeres etapes del desenvolupament.

5.6. Les tècniques d'anàlisi estadística

Com s'ha explicat a l'apartat 5.3, després del recompte de les realitzacions de cada variant al corpus, es va procedir a preparar les dades compilades a la base de dades d'Excel per al seu tractament amb el paquet estadístic SPSS d'IBM (versió 27). Durant la fase de preparació de les dades per al tractament estadístic, es va observar que, un cop comptabilitzades les realitzacions de les diverses variants de les variables dependents era comparativament senzill crear variables que mesuressin el nombre total de realitzacions de cada fenomen lingüístic a les transcripcions. Per exemple, un cop obtingut el recompte de les variants de V3C (*bueno* i *bé*) al corpus, era fàcil des d'un punt de vista operatiu crear una variable *germana* de V3C que mesurés el nombre total d'aquestes dues variants a cada transcripció. Es va decidir, per tant, crear una variable d'aquesta mena per a la resta de variables que ho permetien i afegir-les a l'estudi univariant, com expliquem més endavant. A continuació es mostra el procés seguit per al tractament estadístic del corpus d'anàlisi³⁰.

Figura 26. Procés d'anàlisi estadística del corpus.



³⁰ En aquesta etapa del projecte vam comptar amb el suport de la cap de l'Àrea d'Informàtica Científica de la Secretaria General Adjunta d'Informàtica del Consell Superior d'Investigacions Científiques (SGAI-CSIC), Laura M. Barrios Álvarez, a qui agraeixo les seves paciència, propostes i explicacions.

El primer pas de l'anàlisi estadística va consistir en un estudi exploratori per descriure l'ús de cada variable dependent al corpus (vegeu l'apartat 6.1). Segons les variables independents, edat, sexe i llengua inicial, el corpus es divideix en els grups d'informants que es mostren a la taula següent, en què la darrera fila (*n*) mostra el nombre total d'informants de cada grup. Per a cada variant, es va calcular el valor mínim i el valor màxim en què apareix al corpus, la mitjana, la mediana i els intervals de confiança inferior i superior de la mediana en cadascun d'aquests grups (per motius d'espai, les taules resultants es poden consultar a l'Annex 4, taules 4.1 a 4.88).

Taula 24. Informants segons les variables independents de l'estudi.

Edat	12 anys						16 anys					
Sexe	Masculí			Femení			Masculí			Femení		
L1	C	E	B	C	E	B	C	E	B	C	E	B
<i>n</i>	16	14	2	11	18	8	16	14	2	11	18	8

C: catalanoparlants inicials; E: castellanoparlants inicials; B: bilingües inicials.

Com s'ha avançat al final del subapartat 5.3.2.3.4, va ser fonamental tenir en compte la naturalesa de cadascuna de les variables a l'hora d'identificar les tècniques d'anàlisi estadística més adequades per tractar-les. Per això, des de l'inici d'aquesta fase, es van dividir les variables segons el tipus d'estudi que requerien (univariant o multivariant). Així doncs, es va crear un full de dades a SPSS per a les divuit variables que requerien una anàlisi univariant, ja que no inclouen diverses variants, sinó que mesuren el nombre total d'aparicions d'un cert fenomen lingüístic a les transcripcions (Taula 25 a continuació) i un altre full de dades per a les dinou variables per a les quals calia dur a terme una anàlisi multivariant perquè sí que contenen diverses variants, és a dir, que es realitzen de formes diferents al corpus (Taula 26).

Taula 25. Freqüència d'aparició total de les variables de l'estudi univariant.

Variable	Descripció	Freqüència total al corpus
Total V3b	Partícules discursives <i>a ver / a veure / a ve(u)re / aviam</i>	250
Total V3c	Partícules discursives <i>bueno / bé</i>	2.125
Total V3f	Partícules discursives <i>después / després</i>	502
Total V3h	Partícules discursives <i>hombre / home</i>	229
Total V3i	Partícules discursives <i>o sea / o sigui</i>	439
Total V3j	Partícules discursives <i>pero / però / p(e)rò / p(er)ò</i>	1.791
Total V3k	Partícules discursives <i>pues / pos / doncs</i>	1.536
Total V3l	Partícules discursives <i>vale / val / d'acord</i>	850
Total V5	Interjeccions <i>hola / bon dia / bona tarda / hola bon dia</i>	144
Total V6	Heterorepeticions	712
Total V16	Preguntes no retòriques	368
Total V18	Construccions condicionals	2.062
Total V22	Perífrasis d'obligació	1.339
Total V23	Determinants demostratius	894
Total V25	Pronoms clítics de datiu plural	302
Total V38	Fenòmens de contacte de llengües	5.598
Total V41	Expressió de colors	115
Total V45	Adverbis acabats en <i>-ment</i>	391

Taula 26. Variables de l'estudi multivariant i freqüència de les seves variants.

Variable	Variant	Freqüència total al corpus
V3b	V3b1: <i>a ver</i>	12
	V3b2: <i>a veure</i>	45
	V3b3: <i>a ve(u)re</i>	182
	V3b4: <i>aviam</i>	11
V3c	V3c1: <i>bueno</i>	2.106
	V3c2: <i>bé</i>	19
V3f	V3f1: <i>después</i>	29
	V3f2: <i>després</i>	473
V3h	V3h1: <i>hombre</i>	11
	V3h2: <i>home</i>	218
V3i	V3i1: <i>o sea</i>	46
	V3i2: <i>o sigui</i>	393
V3j	V3j1: <i>pero</i>	86

Metodologia

	V3j2: <i>però</i>	13
	V3j3: <i>p(e)rò</i>	1.551
	V3j4: <i>p(er)ò</i>	141
V3k	V3k1: <i>pues</i>	477
	V3k2: <i>pos</i>	126
	V3k3: <i>doncs</i>	933
V3l	V3l1: <i>vale</i>	635
	V3l2: <i>val</i>	172
	V3l3: <i>d'acord</i>	43
V5	V5a: <i>hola</i>	27
	V5b: <i>bon dia</i>	94
	V5c: <i>bona tarda</i>	13
	V5d: <i>hola bon dia</i>	10
V14a	V14a1: PC + <i>tenir que / de</i> + INF	168
	V14a2: PC + <i>haver de</i> + INF	439
	V14a3: PC + <i>poder</i> + INF	1.247
	V14a4: PC+ <i>anar a</i> + INF	10
V14b	V14b1: <i>tenir que / de</i> + INF + PC	49
	V14b2: <i>haver de</i> + INF + PC	11
	V14b3: <i>poder</i> + INF + PC	150
V14	V14a: Proclisi	1.864
	V14b: Enclisi	215
	V14c: Doble	12
V17	V17a: Oberta	77
	V17b: Tancada	292
V18	V18b: <i>si</i>	2.009
	V18c: <i>mentre(s)</i>	27
	V18f: <i>a no ser que</i>	15
	V18g: <i>en tot cas</i>	11
V22	V22a: <i>tenir que</i> + INF	360
	V22b: <i>haver de</i> + INF	860
	V22c: <i>tenir de</i> + INF	58
	V22e: <i>caldre</i> + INF	61
V23	V23a: <i>aquest</i> + N	231
	V23b: <i>aquell</i> + N	23

	V23c: <i>el + N + aquest</i>	56
	V23d: <i>el + N + aquell</i>	17
	V23e: <i>aquest</i>	426
	V23f: <i>aquell</i>	11
	V23g: <i>(un +) N + d'aquests</i>	116
	V23h: <i>(un +) N + d'aquells</i>	14
V25	V25a: <i>els hi + V / V + -els-hi / V + 'ls-hi / V + -los-hi</i>	150
	V25b: <i>els + V / V + -los / V + 'ls</i>	37
	V25c: <i>li + V / V + -li</i>	99
	V25d: <i>lis + V / V + -lis</i>	16
V38	V38a Importacions	5.501
	V38b Híbrids	11
	V38c Calcs	67
	V38d Altres	19
V41	V41a: <i>color [blau]</i>	38
	V41b: <i>[blau]</i>	77

5.6.1. Exploració de les relacions globals i caracterització dels informants

En estudis com el que ens ocupa, en què es mesuren un gran nombre de variables, pot resultar molt útil dur a terme una prova estadística multivariant que redueixi el nombre de variables utilitzades tot preservant al màxim la seva capacitat d'explicar les dades recollides per a l'estudi. Aquesta mena de proves, a més, ajuden a comprendre les correlacions que operen entre les variables dependents, entre les independents i entre unes i altres. Com tractem a continuació, en aquest estudi ens servim de l'Anàlisi de Components Principals (ACP). A més, també explorem el corpus d'anàlisi mitjançant una tècnica de mineria de dades, els arbres de classificació, amb l'objectiu d'avaluar la utilitat de les variables de l'estudi per caracteritzar els informants.

5.6.1.1. Anàlisi de Components Principals

Una de les proves multivariants més utilitzades que permet comprendre les relacions entre les variables en estudis que en presenten un gran nombre és l'Anàlisi de Components Principals (ACP). Aquesta tècnica transforma el conjunt de variables original, format per variables correlacionades entre elles, en un segon conjunt de variables (anomenades dimensions, components o factors) que no mantenen correlacions entre elles (Díaz Monroy, 2007: 198). Per fer-ho, combina les variables originals de manera lineal segons l'ordre en què recullen la variància de la mostra (escalament òptim o *optimal scaling*). D'aquesta forma, selecciona aquelles variables que aporten més informació sobre les dades i les combina per crear dues dimensions (components principals).

En el nostre estudi, apliquem aquesta prova als dos fulls de dades per explorar les relacions entre les variables. Un dels resultats que genera aquesta prova és la taula resum del model, que mostra el percentatge de variància del corpus que el model estadístic aconsegueix explicar amb les dimensions obtingudes. Com a exemple, a continuació reproduïm la taula corresponent a l'estudi multivariant de V3C (*bueno / bé*), en què s'observa que el model és capaç d'explicar gairebé el 50 % de la variància de la mostra analitzada.

Taula 27. Resum del model generat per l'Anàlisi de Components Principals.

Resum model multivariant rotat ^a			
Dimensió	Alfa de Cronbach	Variància explicada	
		Total (valors propis)	% Variància
1	,280	1,288	25,759
2	,220	1,212	24,238
Total	,750 ^b	2,500	49,997

^a Mètode de rotació: Varimax amb Normalització Kaiser.

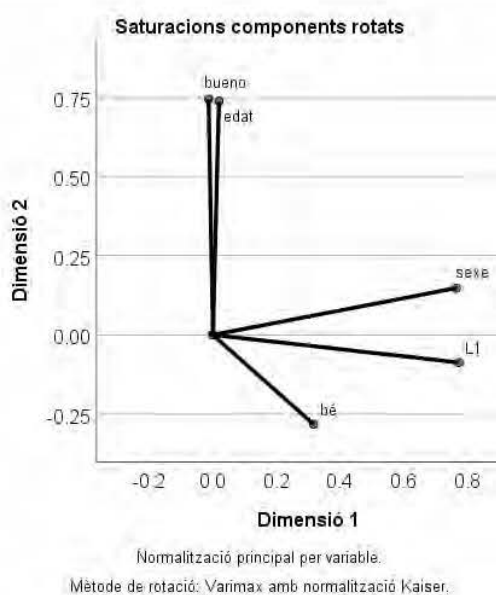
^b Alfa de Cronbach total basat en el total de valors propis.

A més d'aquest resum, els resultats de la prova ens proporcionen la matriu de correlacions de les variables transformades, que mostra les relacions entre les variables transformades mitjançant el procés d'escalament òptim; les matrius de

saturacions als components (*component loadings*) i de saturacions als components rotats (*rotated component loadings*), que ens indiquen el pes de cada variable, és a dir, la contribució que fa a la capacitat de les dimensions (components) per explicar la variància i, per tant, com s'agrupen les variables transformades segons a quin component fan una contribució més destacada (representada per un valor de saturació més pròxim a -1 o 1); els gràfics de transformació, que mostren la reordenació de les categories que formen les variables independents introduïdes al programa i les representacions gràfiques de les matrius de saturacions, que mostren el pes de les variables i les relacions entre elles de forma visual.

En aquests gràfics, hi trobem un punt, situat a la cruïlla entre els zeros de cada eix, del qual parteixen diverses línies que acaben en un cercle. Cada línia representa una variable. La distància entre les línies, representada pels angles que les separen, indica el grau de correlació entre les variables. Així, per exemple, al gràfic següent, extret dels resultats de la prova multivariant, hi podem observar l'obertura dels angles que separen les línies i veure que la variant *bueno* es correlaciona sobretot amb l'edat, mentre que la variant *bé* està més correlacionada amb la llengua inicial dels informants.

Figura 27. Exemple de gràfic de saturacions als components rotats.



A més, la Taula 28 recull els valors de saturació obtinguts abans i després de la rotació. Aquest procés permet mostrar més clarament a quin dels dos components principals (o dimensions) està aportant més informació cada variable. Per exemple, podem veure que, com a resultat del procés de rotació, és més evident en els valors de la segona taula que el sexe dels informants contribueix més al primer component principal (0,766) que no pas al segon (0,147), perquè la diferència entre els valors per als dos components és encara més acusada (concretament, de 0,619 punts) que a la taula prèvia a la transformació, en què es mostra que té una càrrega de 0,647 per a la primera dimensió i de 0,436 per a la segona (és a dir, en què la diferència entre els valors és de 0,211 punts).

Taula 28. Exemples de matrius de saturacions als components i saturacions als components rotats.

Dimensió	Saturacions als components		Saturacions als components rotats	
	1	2	1	2
L1	,747	,223	,775	-,087
Edat	-,272	,688	,019	,739
Sexe	,647	,436	,766	,147
<i>bueno</i>	-,305	,681	-,014	,746
<i>bé</i>	,402	-,135	,317	-,281

Mètode de rotació: Varimax amb Normalització de Kaiser. La rotació no ha convergit en 3 iteracions. (Convergència = ,000).

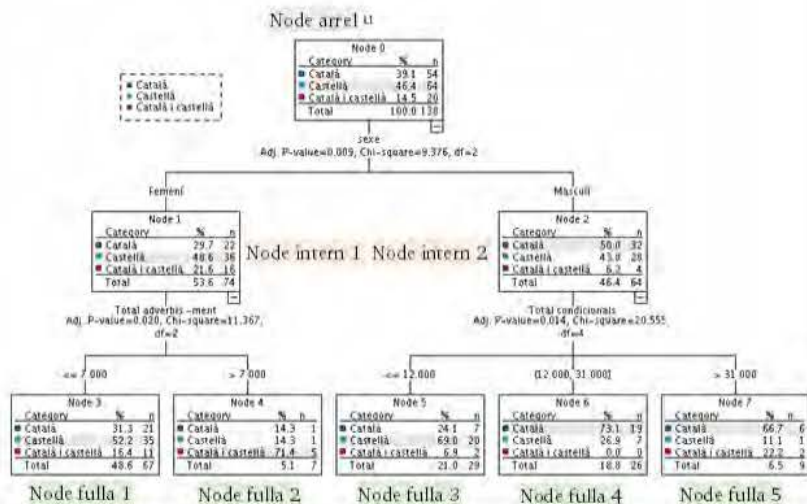
Així doncs, per la seva claredat i per evitar duplicar la informació, al capítol següent ens centrarem en els valors obtinguts després del procés de rotació quan tractem els resultats de l'Anàlisi de Components Principals.

5.6.1.2. Arbres de classificació

Després d'explorar les relacions entre les variables de l'estudi mitjançant l'Anàlisi de Components Principals, el següent pas que vam dur a terme va ser la caracterització dels grups d'edat, sexe i llengua inicial. Per a aquest pas vam utilitzar la tècnica de mineria de dades anomenada arbres de classificació. Es tracta d'una tècnica d'anàlisi discriminant no paramètrica de segmentació binària i se sol classificar com una tècnica predictiva de classificació *ad hoc*, juntament amb l'anàlisi discriminant o les xarxes neuronals (Berlanga Silvente, Rubio Hurtado i Vilà Baños, 2013: 66). Els arbres de classificació generen representacions gràfiques en forma de diagrames de flux jeràrquics que classifiquen els casos d'una variable dependent en funció dels valors de les variables independents.

Per la forma dels ordinogrames resultants d'aplicar els algorismes dels arbres de classificació, és fàcil interpretar-los de forma intuïtiva. A més, permeten identificar i caracteritzar diversos grups de manera visual. Com mostra l'exemple següent, estan formats per un node arrel, situat a la part superior de l'ordinograma, del qual en parteixen com a mínim dues branques, una cap a l'esquerra i una cap a la dreta, que desemboquen en els nodes interns. Els nodes interns reben i també generen branques i poden precedir altres nodes interns o nodes fulla (terminals), que estan situats a la part inferior del diagrama i dels quals no sorgeix cap branca.

Figura 28. Exemple d'arbre de classificació.



Aquest arbre permet caracteritzar els informants per llengua inicial, sexe i les variables lingüístiques de l'estudi. S'hi pot observar que les noies es caracteritzen per l'ús d'adverbis acabats en *-ment*. El primer node fulla mostra que la majoria (un 52,2 %) de les noies que produeixen set o menys adverbis són castellanoparlants inicials. En canvi, segons el segon node fulla, la majoria (un 71,4 %) de les que fan més de set adverbis acabats en *-ment* són bilingües inicials. Pel que fa als nois, la variable dependent que millor els caracteritza és el nombre de construccions condicionals i, segons el tercer node fulla, un 69 % dels que n'utilitzen dotze o menys són castellanoparlants inicials. Els darrers nodes ens informen que, en cas que trobem un noi que produeix més de dotze construccions condicionals, probablement serà catalanoparlant inicial, ja que la majoria dels nois que en produeixen entre dotze i trenta-un (un 73,1 %) o més de trenta-un (un 66,7 %) ho són. Per tant, com veiem mitjançant aquest exemple, els arbres de classificació ens permeten caracteritzar els informants que s'inclouen als nodes terminals de l'ordinograma en funció dels nodes que els precedeixen i de la informació percentual que contenen.

5.6.2. Significació estadística de les dades

En l'últim pas de l'anàlisi estadística, es van posar a prova les hipòtesis de l'estudi i, per tant, es va observar si les relacions que s'havien detectat en els passos previs entre les variables dependents i les independents eren estadísticament significatives. Aquest objectiu es va complir mitjançant l'aplicació de tests no paramètrics, ja que no requereixen una distribució normal de les dades. Concretament, per tal de comprovar la significació de les relacions entre el sexe dels informants i les variables dependents, es va utilitzar la prova U de Mann-Whitney perquè calia comparar dades de mostres independents i, per la seva mida relativament reduïda, no es podia garantir que tinguessin una distribució paramètrica. Per comprovar la significació de les relacions entre la llengua inicial i les variables dependents, es va aplicar el test de Kruskal-Wallis pels mateixos motius que en el cas anterior, però no era possible aplicar la mateixa prova (U de Mann-Whitney) perquè la variable llengua inicial té més de dos grups. Finalment, per comprovar la significació de les relacions entre l'edat i

les variables dependents es va dur a terme un estudi apariat de les dades mitjançant la prova dels rangs amb signe de Wilcoxon perquè es tracta de dades dependents, ja que provenen dels mateixos informants mesurats en dos moments temporals diferents.

5.6.2.1. U de Mann-Whitney

En primer lloc, la prova U de Mann-Whitney es pot veure com l'alternativa no paramètrica a la prova t de Student. Planteja la hipòtesi nul·la que les dues mostres que es comparen, que són de mides diferents (en el nostre cas, hi ha més noies que nois), procedeixen de poblacions contínues idèntiques. La hipòtesi alternativa, en canvi, suposa que les tendències centrals de les poblacions són diferents (Amat Rodrigo, 2017). Com explica, Amat Rodrigo (2017), per tant, el que es comprova mitjançant aquest test és que la probabilitat que una observació d'A, una de les poblacions que es comparen, sigui més alta que una observació de B, l'altra població ($P(A>B)$) és igual a la probabilitat que una observació de la població B sigui més alta que una de la població A ($P(B>A)$), la qual cosa es pot representar de la següent manera: $H_0: P(A>B) = P(B>A)$. Com que aquesta és la hipòtesi nul·la, la podem confirmar quan $P(A>B)$ sigui igual o inferior a 0,5 ($H_0: P(X>Y) \leq 0.5$). Per aplicar aquesta prova, és necessari que les variàncies de les dues poblacions siguin iguals. Per calcular-la, s'ordenen totes les observacions de les dues mostres en ordre descendent i s'assigna un rang a cada observació (el valor més petit rep un 1, el següent un 2, etc.), de manera que sumant els rangs de cada mostra s'obtenen els valors R_1 i R_2 de la fórmula següent.

Figura 29. Fórmula de la prova U de Mann-Whitney.

Valor estadístic U de Mann-Whitney:

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2$$

n_1 = mida de la mostra 1.

n_2 = mida de la mostra 2.

R_1 = sumatori dels rangs de la mostra 1.

R_2 = sumatori dels rangs de la mostra 2.

Després d'obtenir el valor estadístic, es pot calcular el p valor, és a dir, el nivell de significació dels resultats comparant el resultat de Z amb α (0,05). Si Z és més gran que α , es pot descartar la hipòtesi nul·la (H_0).

Figura 30. Fórmula de p valor per a la prova U de Mann-Whitney.

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} \sim N(0,1)$$

5.6.2.2. Kruskal-Wallis

En segon lloc, la prova de Kruskal-Wallis és una extensió de l'anterior per a tres o més grups. Per tant, la hipòtesi nul·la que planteja és, de nou, que les dades de les dues mostres tenen la mateixa distribució, és a dir, provenen de la mateixa població. És l'alternativa per a dades no paramètriques de la prova ANOVA d'una via. Com amb l'anterior, per aplicar aquesta prova, cal que les variàncies de les dues poblacions siguin iguals. Es calcula de forma semblant a la prova U de Mann-Whitney, ja que també es basa en l'ordenació dels valors de les observacions i l'assignació de rangs per comparar-ne la mitjana. Si la H_0 es confirma i les mides de les mostres són més grans que cinc, el valor estadístic de la prova Kruskal Wallis presenta una distribució aproximada de khi-quadrat (χ^2) amb un nombre de graus de llibertat igual al nombre de mostres utilitzades menys 1 ($k-1$) (Alea Riera *et al.*, 2005).

Figura 31. Fórmula de la prova de Kruskal-Wallis.

Valor estadístic Kruskal-Wallis (H):

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1)$$

N = nombre total d'observacions entre totes les mostres.
 n_i = nombre d'observacions a la mostra.
 R_i = sumatori dels rangs de cada mostra.

5.6.2.3. Wilcoxon

En tercer lloc, per a la prova dels rangs amb signe de Wilcoxon no és necessari assumir que les poblacions es distribueixen de forma normal però sí que requereix que es distribueixin de forma simètrica. A més, és necessari que les dades siguin ordinals i dependents (com en el nostre estudi, en què s'aplica a dades provinents dels mateixos informants en dos moments temporals). Parteix de la hipòtesi nul·la que la mediana de les diferències de cada parell de dades és igual a zero ($H_0: M_e(\text{dif}) = 0$). Per testar aquesta hipòtesi, la prova calcula la diferència de cada parell d'observacions, ordena els resultats i els assigna un rang, començant pel valor més petit. Aleshores, aplica la fórmula següent.

Figura 32. Fórmula de la prova dels rangs amb signe de Wilcoxon.

Valor estadístic dels rangs amb signe de Wilcoxon (W):

$$W = \min(W+, W-)$$

$W+$ = suma dels rangs amb signe positiu.
 $W-$ = suma dels rangs amb signe negatiu.

Una vegada més, un cop obtingut el valor estadístic d'aquesta prova, se'n pot calcular el nivell de significació. Si la mida de la mostra és més gran de 25 ($n > 25$) i, per tant, es pot assumir que W té una distribució normal, es pot rebutjar H_0 quan el resultat de Z (Figura 33) sigui menor que α (0,05) (Amat Rodrigo, 2016).

Figura 33. Fórmula de p valor per a la prova dels rangs amb signe de Wilcoxon.

$$Z = \frac{W - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{12}}} \approx N(0,1)$$

Així doncs, l'aplicació de cadascuna d'aquestes tres proves estadístiques per als conjunts de dades corresponents a les variables univariants i multivariants va permetre obtenir els resultats de l'estudi que responen als objectius que ens proposàvem (explicats al capítol 3) i que es detallen al capítol següent.

6 RESULTATS

Aquest capítol recull els resultats de les diferents etapes d'anàlisi a què es va sotmetre el conjunt de dades. Primer, com a mostra de l'estudi exploratori descriptiu, repassem les medianes de realització de cada variable (i variant). Després, ens ocupem dels productes més rellevants per als nostres objectius de l'Anàlisi de Components Principals. Al tercer apartat d'aquest capítol (6.3), observem els resultats de les proves de significació estadística que s'han descrit a l'apartat precedent (5.6.2). Finalment, complementem tot l'anterior amb l'aplicació d'un model de classificació no lineal que selecciona les variables més rellevants per a la caracterització de cada grup d'edat, sexe i llengua inicial del corpus. Tanca el capítol una síntesi dels resultats més destacats de cada fase.

6.1. Estudi exploratori descriptiu

Com avançàvem a l'apartat 5.6, la primera fase del tractament estadístic de les dades va consistir a extreure mesures estadístiques bàsiques per a cadascuna de les variables i les seves variants. Concretament, es van obtenir els valors mínim i màxim, la mitjana i la mediana d'aparició, així com els intervals de confiança inferior i superior de la mediana. A continuació, observem la mediana de realitzacions de cada variable, és a dir, el valor que es troba a l'equador de la sèrie de dades si les ordenem de menor a major. (La resta de mesures estadístiques exploratòries s'inclouen als **Annexos**, taules 4.1 a 4.88). Aquesta mesura, per tant, mostra la tendència central de realització al corpus de cada variant i, a diferència de la mitjana, no es veu influenciada pels valors extrems (per exemple, parlants que produeixen una variant determinada en una freqüència molt més elevada que la resta per qüestions estilístiques o preferències idiolectals).

6.1.1. Estudi estadístic exploratori del doblet *a ver* / *a veure* (V3b)

Tractem, en primer lloc, les variables pragmàtiques (V3, V5 i V6), seguides de les morfosintàctiques (V14, V16, V17, V18, V22, V23 i V25) i, finalment, les lèxiques (V38, V41 i V45). Quant al primer conjunt de partícules discursives (V3), format per les variants *a ver*, *a veure*, *a ve(u)re* i *aviam* (V3b), la Taula 29 mostra que les medianes més elevades corresponen a *a veure* i *a ve(u)re*, el valor central de les quals és més que zero, pel que fa a *a ve(u)re*, en el cas de les noies bilingües inicials als 12 anys, els nois catalanoparlants inicials i les noies bilingües inicials als 16 anys i, pel que fa a *a veure*, en el cas dels nois bilingües inicials als 16 anys. Les dades de l'estudi univariant mostren medianes iguals o superiors a una realització d'aquestes partícules discursives en 6 dels 12 grups de parlants que conformen el corpus, els corresponents al de noies castellano-parlants inicials als 12 anys i a tots els grups de llengua inicial als 16 anys a excepció de les noies catalanoparlants inicials. A més, les noies bilingües inicials als 12 anys mostren una mediana de 0,5 realitzacions d'aquesta variable.

Taula 29. Medianes de realització de la variable V3b.

Estudi multivariant		12 anys						16 anys					
		Masculí			Femení			Masculí			Femení		
Variable	Variante	C	B	E	C	B	E	C	B	E	C	B	E
V3b	<i>a ver</i>	0	0	0	0	0	0	0	0	0	0	0	0
	<i>a veure</i>	0	0	0	0	0	0	0	1	0	0	0	0
	<i>a ve(u)re</i>	0	0	0	0	1	0	1	0	0	0	1	0
	<i>aviam</i>	0	0	0	0	0	0	0	0	0	0	0	0
Estudi univariant													
Total V3b		0	0	0	0	,5	1	1	1	1	0	1,5	1,5

C: catalanoparlants inicials; B: bilingües inicials; E: castellanoparlants inicials.

6.1.2. Estudi estadístic exploratori del doblet *bueno* / *bé* (V3c)

El segon doblet de partícules discursives és el format per *bueno* i *bé* (V3c). Com recull la Taula 30, la primera, a diferència de *bé*, mostra medianes força elevades a tots els grups de parlants, que superen la mediana de 8 realitzacions en tots els casos menys en el grup de noies bilingües inicials als 12 anys. Si ens fixem en les medianes de l'estudi univariant, hi observem alguns valors més elevats que els de la variant *bueno* a l'estudi multivariant, corresponents a l'addició de les realitzacions de la variant *bé* a aquestes dades (concretament, en el grup de noies catalanoparlants i bilingües inicials als 12 anys).

Taula 30. Medianes de realització de la variable V3c.

Estudi multivariant		12 anys						16 anys					
		Masculí			Femení			Masculí			Femení		
Variable	Variant	C	B	E	C	B	E	C	B	E	C	B	E
V3c	<i>bueno</i>	13	14	9	11	3	8	10	23	9	25	11	24
	<i>bé</i>	0	0	0	0	0	0	0	0	0	0	0	0
Estudi univariant													
Total V3c		12,5	14	9	13	4,5	8	10	22,5	9	25	11	24

C: catalanoparlants inicials; B: bilingües inicials; E: castellanoparlants inicials.

6.1.3. Estudi estadístic exploratori del doblet *después / después* (V3f)

El tercer doblet de partícules discursives és el de *después* i *després* (V3f), la més freqüent de les quals és, en tots els grups de parlants, aquesta darrera, com mostra la Taula 31. En aquest cas, són els nois catalanoparlants i bilingües inicials als 12 anys els que mostren medianes més elevades. Pel que fa a l'estudi univariant, reflecteix medianes equivalents a les de *després* en tots els grups excepte el de noies catalanoparlants inicials als 12 anys i el de noies bilingües inicials als 16 anys.

Taula 31. Medianes de realització de la variable V3f.

Estudi multivariant		12 anys						16 anys					
		Masculí			Femení			Masculí			Femení		
Variable	Variant	C	B	E	C	B	E	C	B	E	C	B	E
V3f	<i>después</i>	0	0	0	0	0	0	0	0	0	0	0	0
	<i>després</i>	5	4	2	0	1	3	3	0	2	2	2	2
Estudi univariant													
Total V3f		5	4	1,5	1	1,5	3	3	0	2	2	3	2

C: catalanoparlants inicials; B: bilingües inicials; E: castellanoparlants inicials.

6.1.4. Estudi estadístic exploratori del doblet *hombre / home* (V3h)

La següent parella de partícules la conformen *hombre* i *home*. Una vegada més, la que presenta valors centrals més alts a la majoria de grups de parlants és la variant catalana, com mostra la Taula 32. La mediana més elevada es troba al grup de nois catalanoparlants inicials als 16 anys, tot i que les diferències entre els valors centrals dels diversos grups no excedeix el punt. En aquest cas, els resultats de l'estudi univariant i les medianes de la variant *home* coincideixen.

Taula 32. Medianes de realització de la variable V3h.

Estudi multivariant		12 anys						16 anys					
		Masculí			Femení			Masculí			Femení		
Variable	Variant	C	B	E	C	B	E	C	B	E	C	B	E
V3h	<i>hombre</i>	0	0	0	0	0	0	0	0	0	0	0	0
	<i>home</i>	0	2	1	1	0	0	2	0	1	1	1	1
Estudi univariant													
Total V3h		0	2	1	1	0	0	2	0	1	1	1	1

C: catalanoparlants inicials; B: bilingües inicials; E: castellanoparlants inicials.

6.1.5. Estudi estadístic exploratori del doblet *o sea* / *o sigui* (V3i)

A continuació, observem a la Taula 33 que, d'entre el doblet de partícules discursives *o sea* i *o sigui*, de nou la que presenta medianes més elevades a la majoria de grups de parlants del corpus d'anàlisi (7/12) és la variant catalana. Els nois bilingües inicials als 12 anys són, en aquest cas, els que mostren un valor central més alt (de sis punts), seguits a una distància de dos punts dels catalanoparlants inicials als 16 (tant nois com noies). Les dades de l'estudi univariant confirmen les dades del multivariant i només divergeixen de les corresponents a la variant *o sigui* pel que fa als nois bilingües i castellanoparlants inicials als 16 anys.

Taula 33. Medianes de realització de la variable V3i.

Estudi multivariant		12 anys						16 anys					
		Masculí			Femení			Masculí			Femení		
Variable	Variant	C	B	E	C	B	E	C	B	E	C	B	E
V3i	<i>o sea</i>	0	0	0	0	0	0	0	1	0	0	0	0
	<i>o sigui</i>	0	6	0	0	1	0	4	3	0	4	3	2
Estudi univariant													
Total V3i		0	6	0	0	1	0	4	3	0	4	3	2

C: catalanoparlants inicials; B: bilingües inicials; E: castellanoparlants inicials.

6.1.6. Estudi estadístic exploratori del doblet *pero* / *però* (V3j)

El següent conjunt de partícules discursives és el del doblet *pero* i *però*, que també inclou les variants *p(e)rò* i *p(er)ò*. La Taula 34 mostra que les més habituals als corpus són les dues darreres (i especialment *p(e)rò*), si bé els nois bilingües inicials als 12 anys també presenten una mediana de realitzacions superior a zero. Els valors més elevats de la mediana d'aquesta variable els trobem als tres grups de noies als 16 anys i al de noies catalanoparlants inicials als 12 anys. Les dades de l'estudi univariant confirmen aquesta tendència i, a més, mostren que, tenint en compte totes les variants en conjunt, el grup de nois

catalanoparlants inicials als 16 anys presenta, com els grups de noies d'aquesta edat, un valor central superior a les 15 realitzacions de V3j.

Taula 34. Medianes de realització de la variable V3j.

Estudi multivariant		12 anys						16 anys					
		Masculí			Femení			Masculí			Femení		
Variable	Variant	C	B	E	C	B	E	C	B	E	C	B	E
V3j	<i>pero</i>	0	0	0	0	0	0	0	0	0	0	0	0
	<i>però</i>	0	1	0	0	0	0	0	0	0	0	0	0
	<i>p(er)ò</i>	7	10	7	13	7	11	12	11	11	13	14	15
	<i>p(er)ò</i>	1	2	0	0	0	0	1	1	0	1	1	0
Estudi univariant													
Total V3j		13,5	11,5	8	13	8	13	15,5	11	12	16	15	16,5

C: catalanoparlants inicials; B: bilingües inicials; E: castellanoparlants inicials.

6.1.7. Estudi estadístic exploratori del doblet *pues* / *doncs* (V3k)

La penúltima variable de partícules discursives és el conjunt de *pues*, *pos* i *doncs* (V3k). En aquest cas, tant *pues* com *doncs* mostren medianes de realització superiors a zero en la majoria de grups de parlants. Com s'observa a la Taula 35, els valors més elevats els trobem als nois bilingües inicials als 12 (en el cas de *pues*) i en els de nois catalanoparlants i bilingües inicials als 16 (en el de *doncs*). En tots els grups excepte el de nois bilingües inicials als 12 anys, *doncs* és la variant amb una mediana més alta. L'estudi univariant mostra medianes superiors a les 15 realitzacions d'aquesta variable per als grups de noies bilingües i castellanoparlants inicials als 12 anys i inferiors a set realitzacions per als de nois catalanoparlants inicials a la mateixa edat i castellanoparlants inicials als 16 anys.

Taula 35. Medianes de realització de la variable V3k.

Estudi multivariant		12 anys						16 anys					
		Masculí			Femení			Masculí			Femení		
Variable	Variant	C	B	E	C	B	E	C	B	E	C	B	E
V3k	<i>pues</i>	1	7	1	2	2	1	1	1	1	0	3	1
	<i>pos</i>	0	0	1	0	0	0	0	1	0	0	1	0
	<i>doncs</i>	4	4	2	2	2	2	7	9	2	3	6	5
Estudi univariant													
Total V3k		6,5	10	7,5	15	14,5	11,5	9,5	10	5,5	7	10	11,5

C: catalanoparlants inicials; B: bilingües inicials; E: castellanoparlants inicials.

6.1.8. Estudi estadístic exploratori del doblet *vale* / *d'acord* (V3l)

Les medianes de l'últim conjunt de partícules discursives, formada per les variants *vale*, *val* i *d'acord*, mostren que, mentre que *vale* i, en menor mesura, *val*, es distribueixen de forma regular entre els grups de parlants, *d'acord* només presenta una mediana superior a zero en el cas dels nois bilingües inicials als 16 anys. Les medianes més elevades de *vale* i *val* les trobem, respectivament, en les noies castellanoparlants inicials als 16 anys i en els nois bilingües inicials als 12 anys, els dos grups amb una mediana conjunta més alta també a les dades de l'estudi univariant.

Taula 36. Medianes de realització de la variable V3l.

Estudi multivariant		12 anys						16 anys					
		Masculí			Femení			Masculí			Femení		
Variable	Variant	C	B	E	C	B	E	C	B	E	C	B	E
V3l	<i>vale</i>	3	4	3	2	4	5	5	2	3	3	4	6
	<i>val</i>	1	5	1	2	1	1	0	1	0	0	2	0
	<i>d'acord</i>	0	0	0	0	0	0	0	1	0	0	0	0
Estudi univariant													
Total V3l		4	9	5,5	5	5	6	5	3	5,5	4	6,5	7

C: catalanoparlants inicials; B: bilingües inicials; E: castellanoparlants inicials.

6.1.9. Estudi estadístic exploratori de les fórmules d'obertura de la interacció (V5)

Quant a la següent variable pragmàtica, les fórmules d'obertura de la interacció (V5), els resultats de l'estudi multivariant són clars pel que fa a la variant més habitual al llarg del corpus, *bon dia*. L'únic grup de parlants que presenta una mediana superior a zero per a altres variants és el de nois bilingües inicials als 16 anys, que, igual que de *bon dia*, té una mediana d'una realització d'*hola* i de *bona tarda* d'una realització. En canvi, el valor central de la variant *hola bon dia* és zero en tots els grups de parlants. L'estudi univariant també mostra aquesta diferència en la diversitat de formes utilitzades entre els nois bilingües inicials als 16 anys i la resta de parlants mitjançant una mediana més elevada de realitzacions totals de V5.

Taula 37. Mediane de realització de la variable V5.

Estudi multivariant		12 anys						16 anys					
		Masculí			Femení			Masculí			Femení		
Variable	Variant	C	B	E	C	B	E	C	B	E	C	B	E
V5	<i>hola</i>	0	0	0	0	0	0	0	1	0	0	0	0
	<i>bon dia</i>	1	1	1	1	1	1	1	1	1	1	1	1
	<i>bona tarda</i>	0	0	0	0	0	0	0	1	0	0	0	0
	<i>hola bon dia</i>	0	0	0	0	0	0	0	0	0	0	0	0
Estudi univariant													
Total V5		1	1	1	1	1	1	1	2	1	1	1	1

C: catalanoparlants inicials; B: bilingües inicials; E: castellanoparlants inicials.

6.1.10. Estudi estadístic exploratori de les heterorepeticions (V6)

La Taula 38 recull les medianes corresponents a la darrera variable pragmàtica, l'ús d'heterorepeticions (V6). Com s'hi pot observar, en general, els valors dels grups als 12 anys són inferiors que als 16 anys (amb l'excepció dels nois bilingües inicials). En aquest cas, són les noies als 12 anys les que mostren unes medianes més baixes i els nois monolingües inicials als 16 els que les presenten més elevades.

Taula 38. Mediane de realització de la variable V6.

12 anys						16 anys					
Masculí			Femení			Masculí			Femení		
C	B	E	C	B	E	C	B	E	C	B	E
4	5,5	3,5	3	3	2	8	4	7	4	5	5

C: catalanoparlants inicials; B: bilingües inicials; E: castellanoparlants inicials.

6.1.11. Estudi estadístic exploratori de la posició dels pronoms clítics en construccions perifràstiques (V14)

Tot seguit, ens centrem en les variables morfosintàctiques de l'estudi, començant per la posició dels pronoms clítics en construccions perifràstiques (V14). La Taula 39 en recull les medianes de realització. Pel disseny d'aquesta variable, que ja inclou els recomptes totals de les realitzacions proclítics (V14a), enclítics (V14b) i pleonàsmiques (V14c), no es va incorporar V14 a l'estudi univariant. En relació a V14a i V14b, s'observa que, independentment de la posició en què s'utilitzen els pronoms febles, les medianes més elevades corresponen a la perífrasi formada pel verb *poder* i formes d'infinitiu. Tot i això, en termes generals, la posició enclítica és força menys present al corpus que l'enclítica i la duplicació de pronoms clítics tant abans com després de la perífrasi

verbal només mostra una mediana superior a zero en el cas dels nois bilingües inicials als 12 anys, que alhora són els que tenen una mediana més elevada, d'una banda, de pronoms proclítics i, de l'altra, d'enclítics (juntament amb els bilingües inicials als 16 anys, tant nois com noies).

Taula 39. Mediane de realització de la variable V14.

Estudi multivariant		12 anys						16 anys					
		Masculí			Femení			Masculí			Femení		
Variable	Variant	C	B	E	C	B	E	C	B	E	C	B	E
V14a	PC + <i>tenir que / de</i> + INF	0	6	1	0	1	0	0	1	0	0	0	0
	PC + <i>haver de</i> + INF	3	1	1	2	2	3	3	2	2	2	4	3
	PC + <i>poder</i> + INF	9	19	5	7	5	8	9	11	7	8	11	9
	PC + <i>anar a</i> + INF	0	0	0	0	0	0	0	0	0	0	0	0
Total proclisi		14	26	8	14	9	12	12	13	10	10	17	13
V14b	<i>tenir que / de</i> + INF + PC	0	1	0	0	0	0	0	1	0	0	0	0
	<i>haver de</i> + INF + PC	0	0	0	0	0	0	0	0	0	0	0	0
	<i>poder</i> + INF + PC	0	2	1	0	1	0	1	2	1	1	2	1
Total enclisi		0	2	1	0	1	1	1	2	1	1	2	1
V14c	Total PC + PV + PC	0	1	0	0	0	0	0	0	0	0	0	0

C: catalanoparlants inicials; B: bilingües inicials; E: castellanoparlants inicials.

6.1.12. Estudi estadístic exploratori del nombre (V16) i el tipus de preguntes (V17)

Les dues variables que tractem a continuació, el nombre de preguntes no retòriques formulades pels informants (V16) i el tipus de preguntes (V17) són complementàries, ja que la primera correspondria a la variable germana de la segona que s'hauria inclòs a l'estudi univariant, en tant que comprèn el recompte total de preguntes formulades, independentment del tipus. Si ens fixem primer en les medianes de V16, observem que els grups que tenen un valor central més elevat de preguntes són el de nois bilingües inicials, tant als 12 com als 16 anys, mentre que els que presenten un valor més baix són els castellanoparlants inicials, de nou, tant al primer moment temporal en què van dur a terme la prova com al segon.

Taula 40. Medianes de realització de la variable V16.

12 anys						16 anys					
Masculí			Femení			Masculí			Femení		
C	B	E	C	B	E	C	B	E	C	B	E
2,5	3,5	1	2	2	2,5	3	3,5	1	2	2,5	1,5

C: catalanoparlants inicials; B: bilingües inicials; E: castellanoparlants inicials.

Les dades de V17 ofereixen una visió complementària d'aquests resultats, ja que mostren que els nois bilingües inicials als 12 anys formulen una mediana de dues preguntes obertes i dues de tancades i als 16 anys en formulen més del segon tipus que cap altre grup. Quant als castellanoparlants inicials, la Taula 41 mostra que als 12 anys fan la mateixa mediana de preguntes obertes que tancades (una) i que als 16 anys només és de preguntes tancades que en fan una mediana superior a zero.

Taula 41. Medianes de realització de la variable V17.

Estudi multivariant		12 anys						16 anys					
		Masculí			Femení			Masculí			Femení		
Variable	Variant	C	B	E	C	B	E	C	B	E	C	B	E
V17	Oberta	0	2	1	0	0	0	1	1	0	0	0	0
	Tanca-da	2	2	1	2	2	1	2	3	1	2	2	2

C: catalanoparlants inicials; B: bilingües inicials; E: castellanoparlants inicials.

6.1.13. Estudi estadístic exploratori de les construccions condicionals (V18)

La següent variable morfosintàctica és el nombre de construccions condicionals (V18), com recull la Taula 42. L'estudi multivariant mostra que la partícula més habitual per introduir aquest tipus d'oracions és la conjunció *si*, els valors centrals de la qual no baixen de nou en cap grup d'informants. Només els nois bilingües inicials als 16 anys presenten una mediana superior a zero per una de les altres variants de V18, *en (tot) cas*. L'estudi univariant confirma que és aquest grup de parlants el que fa un ús global més elevat de construccions condicionals, en contrast amb els castellanoparlants inicials que són, independentment del sexe i l'edat, els que presenten medianes més baixes.

Taula 42. Mediane de realització de la variable V18.

Estudi multivariant		12 anys						16 anys					
		Masculí			Femení			Masculí			Femení		
Variable	Variant	C	B	E	C	B	E	C	B	E	C	B	E
V18	<i>si</i>	16	21	10	12	15	11	18	21	9	15	15	11
	<i>mentre (s)</i>	0	0	0	0	0	0	0	0	0	0	0	0
	<i>en (tot) cas</i>	0	0	0	0	0	0	0	3	0	0	0	0
	<i>a no ser aue</i>	0	0	0	0	0	0	0	0	0	0	0	0
Estudi univariant													
Total V18		16	21	10	12	15	11	19,5	24	9	15	15	11

C: catalanoparlants inicials; B: bilingües inicials; E: castellanoparlants inicials.

6.1.14. Estudi estadístic exploratori de les perífrasis d'obligació (V22)

Quant a les perífrasis d'obligació (V22), els valors de la Taula 43 mostren que la variant més utilitzada és *haver de*, seguida de la no normativa *tenir que*, a força distància, si ens guiem per la mediana de realització, de *caldre* i *tenir de*, que només mostren medianes superiors a zero en els nois bilingües inicials (tant als 12 com als 16 anys). A més, també destaquen les medianes d'*haver de* i *tenir que* dels nois bilingües inicials als 12 anys. En el primer cas, a la baixa respecte la resta de grups, acompanyats dels nois castellanoparlants inicials a la mateixa edat, i en el segon cas, a l'alça de forma notable, ja que mostren una mediana de *tenir que* que quadruplica les dels grups que els segueixen (nois castellanoparlants inicials i noies bilingües inicials als 12 anys i nois bilingües inicials als 16 anys). Les dades de l'estudi univariant subratllen igualment la diferència dels nois bilingües inicials als 12 anys per la seva mediana d'ús de V22 considerablement més elevada que la resta de grups i la dels nois castellanoparlants inicials als 12 anys justament per la raó contrària.

Taula 43. Mediane de realització de la variable V22.

Estudi multivariant		12 anys						16 anys					
		Masculí			Femení			Masculí			Femení		
Variable	Variant	C	B	E	C	B	E	C	B	E	C	B	E
V22	<i>tenir que</i>	0	12	3	0	3	1	0	3	1	0	0	0
	<i>haver de</i>	7	1	2	6	2	7	6	3	3	5	10	5
	<i>tenir de</i>	0	3	0	0	0	0	0	1	0	0	0	0
	<i>caldre</i>	0	3	0	0	0	0	0	2	0	0	0	0
Estudi univariant													
Total V22		9	18	5	10	8,5	12	7	7,5	7	10	13,5	7,5

C: catalanoparlants inicials; B: bilingües inicials; E: castellanoparlants inicials.

6.1.15. Estudi estadístic exploratori dels determinants demostratius (V23)

La següent variable morfosintàctica és la dels determinants demostratius (V23). Com es recull a la Taula 44, una vegada més s'observen dues variants la distribució de les quals al llarg del corpus és més equilibrada que la resta. En aquest cas, es tracta de les formes adjectives i pronominals de proximitat, és a dir, *aquest* (i la resta de formes flexionades) seguit o no d'un sintagma nominal. Concretament, en les dues variants són de nou els nois bilingües inicials els que destaquen per mostrar unes medianes més elevades tant als 12 com als 16 anys. En canvi, els valors centrals més baixos d'aquestes dues variants es troben en grups diferents. Mentre que en el cas del pronom *aquest* són els informants (nois i noies) castellanoparlants inicials als 16 anys els que mostren un valor més baix, en el d'*aquest* + N les medianes més baixes són les de les noies bilingües i castellanoparlants inicials als 12 anys. Els resultats de l'estudi univariant mostren que, en conjunt, són els nois bilingües inicials als 16 anys els que tenen una mediana més elevada i les noies bilingües inicials als 12 anys les que tenen una mediana més baixa de V23.

Taula 44. Medianes de realització de la variable V23.

Estudi multivariant		12 anys						16 anys					
		Masculí			Femení			Masculí			Femení		
Variable	Variant	C	B	E	C	B	E	C	B	E	C	B	E
V23	<i>aquest</i> + N	1	2	1	1	0	0	2	3	1	1	2	1
	<i>aquell</i> + N	0	0	0	0	0	0	0	0	0	0	0	0
	<i>el</i> + N + <i>aquest</i>	0	1	0	0	0	0	1	1	0	0	0	0
	<i>el</i> + N + <i>aquell</i>	0	0	0	0	0	0	0	0	0	0	0	0
	<i>aquest</i>	2	5	2	2	3	2	2	6	1	2	2	1
	<i>aquell</i>	0	0	0	0	0	0	0	0	0	0	0	0
	(<i>un</i>) + N + <i>d'aquests</i>	0	0	1	0	0	0	1	0	1	0	2	0
	<i>un</i> + N + <i>d'aquells</i>	0	0	0	0	0	0	0	1	0	0	0	0
Estudi univariant													
Total V23		7	8	4,5	4	3	4	7	9,5	4,5	4	8	4

C: catalanoparlants inicials; B: bilingües inicials; E: castellanoparlants inicials.

6.1.16. Estudi estadístic exploratori dels pronoms clítics de datiu plural (V25)

La darrera variable morfosintàctica és la dels pronoms clítics de datiu plural (V25), les medianes de la qual es mostren a la Taula 45. La variant no sempre normativa *els hi* és la que presenta més grups de parlants (5/12) amb una mediana de realització diferent a zero, seguida de *li* (3/12). Les variants *els* i *lis* només mostren una variant superior a zero en el cas dels nois bilingües inicials als 12 anys, cosa que es reflecteix clarament en el recompte de l'estudi univariant, en què aquest grup d'informants és el que té una mediana més elevada, seguit de les noies bilingües inicials als 16 anys.

Taula 45. Medianes de realització de la variable V25.

Estudi multivariant		12 anys						16 anys					
		Masculí			Femení			Masculí			Femení		
Variable	Variant	C	B	E	C	B	E	C	B	E	C	B	E
V25	<i>els hi</i>	1	2	0	0	1	0	1	0	0	1	0	0
	<i>els</i>	0	1	0	0	0	0	0	0	0	0	0	0
	<i>li</i>	0	1	0	0	0	0	0	1	0	0	1	0
	<i>lis</i>	0	1	0	0	0	0	0	0	0	0	0	0
Estudi univariant													
Total V25		1	3,5	1	0	1	1	1	,5	0	1	1,5	1

C: catalanoparlants inicials; B: bilingües inicials; E: castellanoparlants inicials.

6.1.17. Estudi estadístic exploratori dels fenòmens de contacte de llengües (V38)

En darrer lloc, ens ocupem a continuació de les tres variables lèxiques de l'estudi: V38, V41 i V45. La primera, com s'ha explicat i es pot observar a la Taula 46, recull les unitats lèxiques degudes al contacte de llengües (V38), dividides en les categories d'importacions, híbrids, calcs i altres. Amb una gran distància, els valors centrals més elevats corresponen a la primera d'aquestes variants, les medianes més baixes de la qual les trobem als grup de noies castellanoparlants inicials als 12 anys i de nois catalanoparlants inicials als 12 i als 16 anys. Per contra, els valors més alts se situen en les noies castellanoparlants inicials als 16 anys i als nois bilingües inicials als 12 anys. La segona variant amb més grups de parlants amb medianes diferents a zero (3/12) és la de calcs, que presenta una mediana d'una realització en els grups de nois i noies bilingües i de nois castellanoparlants inicials als 12 anys. Finalment, els

nois bilingües inicials als 12 anys també mostren una mediana d'una realització d'unitats lèxiques híbrides. Tot i ser el grup amb més diversitat de variants a la Taula 46, però, els nois bilingües inicials als 12 anys no mostren la mediana més elevada als resultats de l'estudi univariant, sinó que, coincidint amb les dades de l'estudi multivariant, aquesta posició l'ocupen les noies castellanoparlants inicials als 16 anys.

Taula 46. Mediane de realització de la variable V38.

Estudi multivariant		12 anys						16 anys					
		Masculí			Femení			Masculí			Femení		
Variable	Variant	C	B	E	C	B	E	C	B	E	C	B	E
V38	Importacions	29	43	31	41	28	36	29	38	33	42	36	48
	Híbrids	0	1	0	0	0	0	0	0	0	0	0	0
	Cals	0	1	1	0	1	0	0	0	0	0	0	0
	Altres	0	0	0	0	0	0	0	0	0	0	0	0
Estudi univariant													
Total V38		29,5	43,5	35,5	42	28	37,5	30,5	38	33	43	35,5	48

C: catalanoparlants inicials; B: bilingües inicials; E: castellanoparlants inicials.

6.1.18. Estudi estadístic exploratori dels colors (V41)

La segona variable lèxica comprèn les referències que fan els informants als colors (V41). Com s'observa a la Taula 47, els valors centrals d'aquesta variable són, en general, de zero realitzacions, amb l'excepció de la variant sense el substantiu *color* al grup de nois bilingües inicials als 12 anys.

Taula 47. Mediane de realització de la variable V41.

Estudi multivariant		12 anys						16 anys					
		Masculí			Femení			Masculí			Femení		
Variable	Variant	C	B	E	C	B	E	C	B	E	C	B	E
V41	<i>color</i> [blau]	0	0	0	0	0	0	0	0	0	0	0	0
	[blau]	0	1	0	0	0	0	0	0	0	0	0	0
Estudi univariant													
Total V41		0	,5	0	0	0	0	0	0	0	0	0	0

C: catalanoparlants inicials; B: bilingües inicials; E: castellanoparlants inicials.

6.1.19. Estudi estadístic exploratori del nombre d'adverbis acabats en *-ment* (V45)

L'última variable de l'estudi és l'ús d'adverbis acabats en *-ment* (V45). Com mostra la Taula 48, els grups que presenten medianes més elevades corresponen als 16 anys: d'una banda, les noies bilingües inicials i, de l'altra, els nois catalanoparlants i bilingües inicials. En canvi, els grups amb medianes més baixes corresponen tots a l'edat de 12 anys (els tres grups de noies i els nois bilingües inicials).

Taula 48. Medianes de realització de la variable V45.

12 anys						16 anys					
Masculí			Femení			Masculí			Femení		
C	B	E	C	B	E	C	B	E	C	B	E
1	0	2	0	0	0	3,5	3,5	1,5	2	5,5	2

C: catalanoparlants inicials; B: bilingües inicials; E: castellanoparlants inicials.

En resum, aquesta primera aproximació als resultats sembla indicar que certs grups d'informants tenen un comportament respecte a les variables lingüístiques que analitzem en aquest estudi que els diferencia de la resta de grups que conformen el corpus d'anàlisi. Per exemple, hem observat que els nois bilingües inicials a l'edat de 12 anys presenten medianes més elevades de diverses variables i variants que la resta de grups. Més endavant, ens ocuparem dels resultats de les anàlisis de significació estadística per poder valorar el pes que cal atribuir a les relacions entre les variables lingüístiques i els grups d'edat, sexe i llengua inicial. Abans, cal que explorem, amb més detall que el que ofereixen les observacions sobre les medianes de realització dels diferents grups de parlants, precisament quines relacions s'estableixen entre les variables de l'estudi.

6.2. Anàlisi de Components Principals: Exploració de les relacions globals entre les variables

Anteriorment (al subapartat 5.6.1), s'ha explicat que l'Anàlisi de Components Principals (ACP) és una prova estadística multivariant que permet explicar gran part de la variància de les dades mitjançant només dues dimensions (també anomenades factors), que crea a partir de les variables d'un estudi. D'aquesta manera, l'ACP aporta informació sobre quines correlacions existeixen entre les variables.

A continuació, doncs, tractem els productes principals que genera aquesta prova estadística per a cadascuna de les variables lingüístiques: el resum del model estadístic, la matriu de correlacions de les variables transformades, la matriu de saturacions als components rotats, els gràfics de transformació de les variables independents i el gràfic de components rotats. Comencem pels resultats de l'estudi univariant, que engloba les variables V3b, V3c, V3f, V3h, V3i, V3k, V3l, V5, V6, V16, V22, V23, V25, V41 i V45.

6.2.1. Anàlisi de Components Principals de l'estudi univariant

Com mostra la Taula 49, el model estadístic generat per a l'estudi univariant conserva, amb només dues dimensions, la capacitat d'explicar més d'un 30,5 % de la variància de les dades corresponents a les 15 variables dependents que acabem d'enumerar. Concretament, la primera dimensió explica gairebé el 18 % de la variància de les dades i la segona, el 12,6 %.

Taula 49. Resum del model generat per l'Anàlisi de Components Principals per a les variables incloses a l'estudi univariant.

Resum model multivariant rotat ^a			
Dimensió	Alfa de Cronbach	Variància explicada	
		Total (valors propis)	% Variància
1	,794	3,757	17,891
2	,716	2,648	12,611
Total	,886 ^b	6,405	30,501

^a Mètode de rotació: Varimax amb Normalització Kaiser.

^b Alfa de Cronbach total basat en el total de valors propis.

El següent producte de l'ACP que ens interessa en relació a les variables de l'estudi univariant és la matriu de correlacions de les variables transformades. Ara bé, per la gran quantitat de variables que conté, dirigim els lectors interessats a consultar-la a l'**Annex 5**. Les tres correlacions més destacables són les que s'estableixen entre la variable V38 total de fenòmens de contacte de llengües i V3c total de *bueno* i *bé* (que obté un valor de 0.694), entre la variable V22 total de perífrasis d'obligació i la variable V3j total de *pero*, *però*, *p(e)rò* i *p(er)ò* (0.509) i entre la variable V18 total de construccions condicionals i la variable V3j (0.5).

Pel que fa a la matriu de saturacions als components rotats corresponent a l'estudi univariant, hi veiem que les 20 variables es distribueixen de forma igualitària entre els dos components. Les deu que tenen un valor de saturació més elevat a la primera dimensió apareixen a la Taula 50 sombreades en gris, mentre que les altres deu tenen un pes més gran a la segona dimensió. Les primeres constitueixen un conjunt de dades que permet explicar millor el comportament dels informants que el format per la resta de variables.

Taula 50. Matriu de saturacions als components rotats per a les variables incloses a l'estudi univariant.

Variables transformades	Dimensió	
	1	2
Llengua inicial	,028	-,436
Edat	,001	,510
Sexe	,340	-,364
V3b Total <i>a ver</i> / <i>a veure</i>	,256	,152
V3c Total <i>bueno</i> / <i>bé</i>	,604	,125
V3f Total <i>después</i> / <i>després</i>	,348	,142
V3h Total <i>hombre</i> / <i>home</i>	-,124	,659
V3i Total <i>o sea</i> / <i>o sigui</i>	,374	,512
V3j Total <i>pero</i> / <i>però</i>	,610	,487
V3k Total <i>pues</i> / <i>doncs</i>	,584	-,058
V3l Total <i>vale</i> / <i>d'acord</i>	,652	,059
V5 Total salutacions	-,028	-,131
V6 Heterorepetició	,067	,453
V16 Total preguntes	,166	,491
V18 Total condicionals	,446	,511
V22 Total perífrasis obligació	,558	,151

Resultats

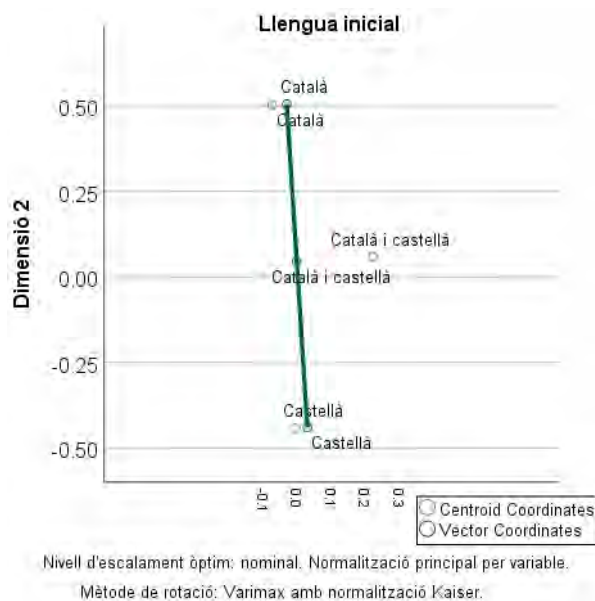
V23 Total determinants demostratius	,516	,272
V25 Total clítics datius	,541	-,038
V38 Total fenòmens contacte	,765	,054
V41 Total colors	,205	-,029
V45 Total adverbis <i>-ment</i>	,254	,491

Normalització principal per variable.

^a Mètode de rotació: Varimax amb Normalització Kaiser. La rotació no ha convergit en 3 iteracions. (Convergència = ,000).

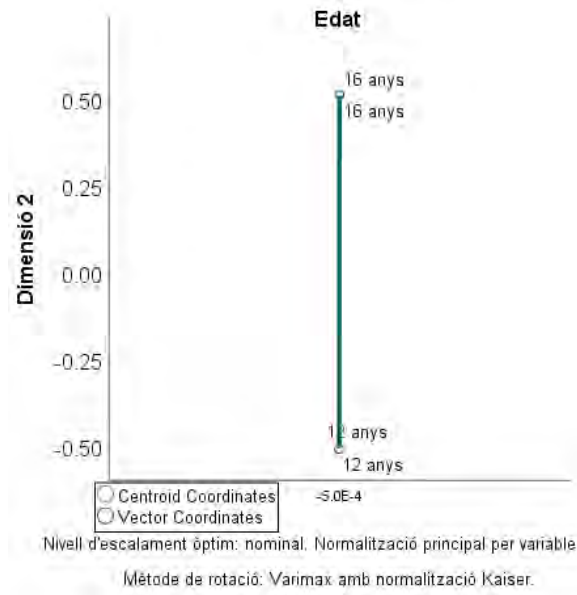
Tot seguit, els gràfics de transformació de les variables edat, sexe i llengua inicial mostren la reordenació dels grups d'informants als dos components principals. Com hem vist, a l'estudi univariant, les tres variables independents tenen un valor de saturació més elevat a la dimensió 2 (si bé la diferència entre els valors als dos components en el cas del sexe és força minça). Per tant, si ens fixem en l'eix d'ordenades de la Figura 34, corresponent al segon component, hi veiem que el grup de catalanoparlants inicials es projecta al valor 0,5, el de bilingües inicials se situa prop de zero i el de castellanoparlants inicials, vora -0,5.

Figura 34. Gràfic de transformació de la variable independent llengua inicial a l'Anàlisi de Components Principals de les variables univariants.



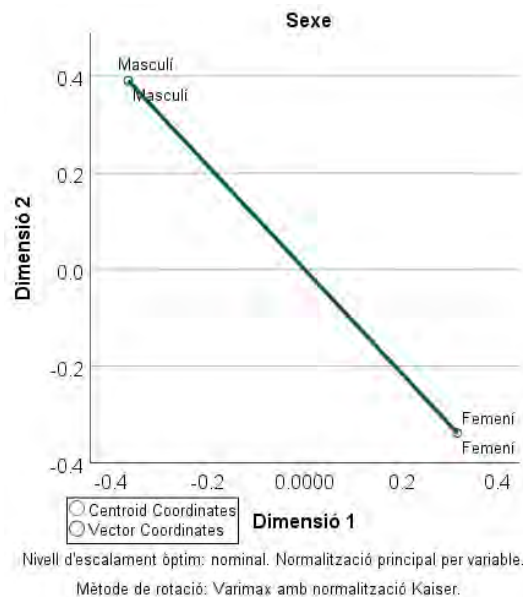
Pel que fa als grups d'edat, la Figura 35 mostra que el valor 0,5 també es correspon amb l'edat de 16 anys, mentre que el de -0,5 ho fa amb l'edat de 12 anys.

Figura 35. Gràfic de transformació de la variable independent edat a l'Anàlisi de Components Principals de les variables univariants.



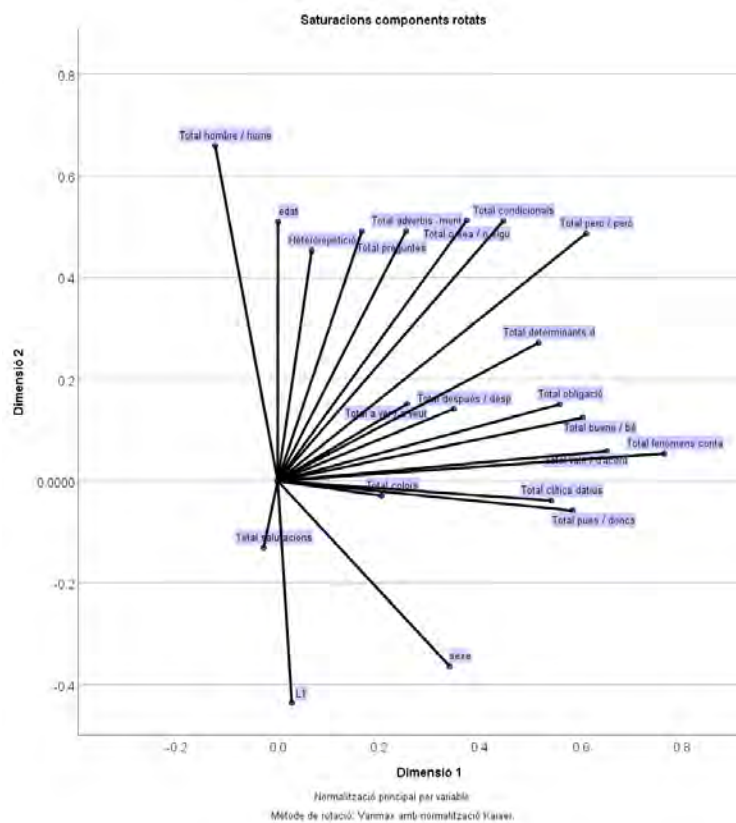
Quant al sexe dels informants, observem a la Figura 36 que els nois se situen prop de 0,4 i les noies més aviat prop de -0,4. A més, en aquest cas els dos grups també se situen força distanciats a l'eix d'abscisses (dimensió 1), la qual cosa representa la proximitat en els valors de saturació per als dos components que hem observat a la matriu anterior.

Figura 36. Gràfic de transformació de la variable independent sexe a l'Anàlisi de Components Principals de les variables univariants.



A partir de la informació que hem tractat fins ara, doncs, podem interpretar a continuació el gràfic de saturacions als components rotats, en el qual, per exemple, l'àrea propera al valor de 0,5 de l'eix d'ordenades correspon, segons els gràfics anteriors, als grups de catalanoparlants inicials i de 16 anys, de la mateixa manera que la zona pròxima al valor de 0,4 es relaciona amb els informants nois. Com mostra la Figura 37, entre aquests dos valors s'hi situen les variables dependents total d'heterorepeticions (V6), total de preguntes no retòriques (V16), total d'adverbis acabats en *-ment* (V45), total de *o sea / o sigui* (V3i), total de construccions condicionals (V18) i total de *pero / però* (V3j). A més, la distància entre les línies que representen cada variable indica el grau de correlació que mantenen. Així doncs, les correlacions més destacades es troben entre el total de referències a colors i el del doblet *pues / doncs*, entre aquesta darrera i el total de pronoms clítics de datiu plural, entre el total de *vale / d'acord* i el de fenòmens de contacte de llengües i entre el total d'*a ver / a veure* i el de determinants demostratius.

Figura 37. Gràfic de saturacions als components rotats per a les variables de l'estudi univariànt.



6.2.2. Anàlisi de Components Principals del doblat *a ver* / *a veure* (V3b)

A continuació, ens ocupem dels resultats de l'estudi multivariant, en el qual s'ha dut a terme una Anàlisi de Components Principals per cada variable. Com sempre, comencem per les variables pragmàtiques i, concretament, la V3b, que comprèn les partícules discursives *a ver*, *a veure*, *a ve(u)re* i *aviam*. El resum del model generat per a aquesta variable, com es desprèn de la Taula 51, és capaç d'explicar el 41,5 % de les dades. El primer component dona compte del 21,1 % i el segon, d'aproximadament el 20,4 % de la variància que presenten.

Taula 51. Resum del model generat per l'Anàlisi de Components Principals per a V3b.

Resum model multivariant rotat ^a			
Dimensió	Alfa de Cronbach	Variància explicada	
		Total (valors propis)	% Variància
1	,384	1,480	21,150
2	,355	1,426	20,371
Total	,765 ^b	2,906	41,521

^a Mètode de rotació: Varimax amb Normalització Kaiser.

^b Alfa de Cronbach total basat en el total de valors propis.

Pel que fa a la matriu de correlació de les variables (Taula 52), hi observem que cap de les parelles de variables mostra valors especialment pròxims a 1 (tots els valors són inferiors a 0,3), de manera que les correlacions entre les variables independents i les variants de V3b són més aviat febles.

Taula 52. Matriu de correlacions entre variables per a V3b.

	L1	Edat	Sexe	<i>a ver</i>	<i>a veure</i>	<i>a ve(u)re</i>	<i>aviam</i>
L1	1,000	,000	,250	,152	-,086	,067	,007
Edat	,000	1,000	,000	-,095	,074	,198	,060
Sexe	,250	,000	1,000	,169	,082	,220	-,076
<i>a ver</i>	,152	-,095	,169	1,000	-,086	,164	-,062
<i>a veure</i>	-,086	,074	,082	-,086	1,000	,276	-,090
<i>a ve(u)re</i>	,067	,198	,220	,164	,276	1,000	-,094
<i>aviam</i>	,007	,060	-,076	-,062	-,090	-,094	1,000
Dimensió	1	2	3	4	5	6	7
Valors propis	1,584	1,323	1,079	,878	,855	,695	,587

Tot seguit, observem a la matriu de saturacions als components rotats que una de les variants de V3b (*a ver*), juntament amb la llengua inicial i el sexe dels informants tenen més pes a la primera que a la segona dimensió. En canvi, tant l'edat com la resta de variants contribueixen més al segon component i, per tant, expliquen pitjor les dades que el conjunt de llengua inicial, sexe i *a ver*.

Taula 53. Matriu de saturacions als components rotats per a V3b.

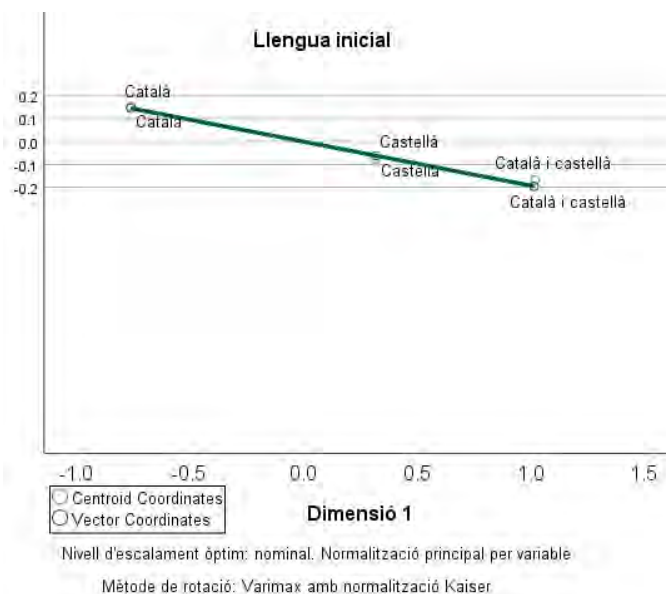
a	Dimensió	
	1	2
Llengua inicial	,650	-,125
Edat	-,167	,504
Sexe	,667	,239
<i>a ver</i>	,654	-,095
<i>a veure</i>	-,113	,725
<i>a ve(u)re</i>	,330	,720
<i>aviam</i>	-,188	-,215

Normalització principal per variable.

^a Mètode de rotació: Varimax amb Normalització Kaiser. La rotació no ha convergit en 3 iteracions. (Convergència = ,000).

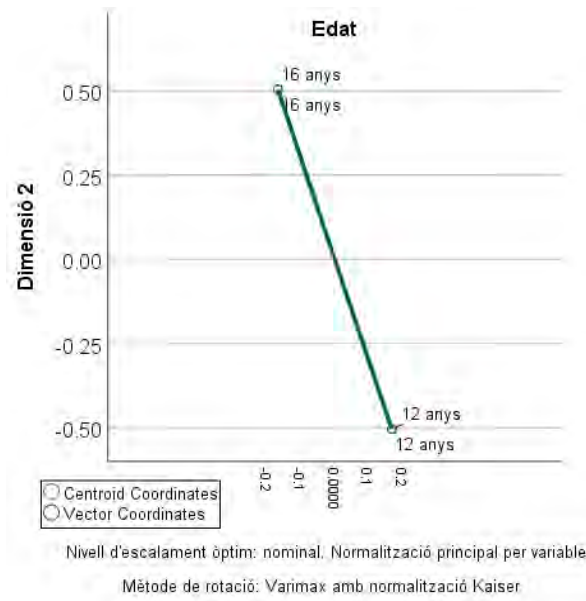
El primer gràfic de transformació per a la variable V3b mostra la distribució dels grups de llengua inicial al llarg de l'eix d'abscisses (dimensió 1). Els catalanoparlants inicials es projecten entre els valors -1 i -0,5; els castellanoparlants inicials, prop de 0,5, i els bilingües inicials apareixen vora 1.

Figura 38. Gràfic de transformació de la variable independent llengua inicial a l'Anàlisi de Components Principals de V3b.



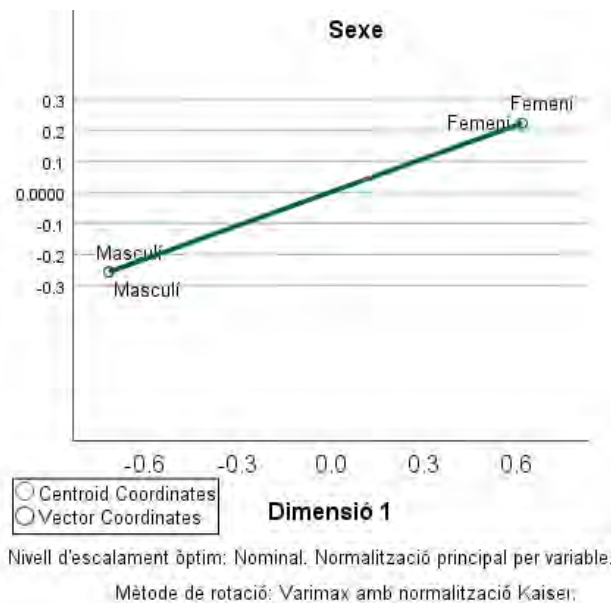
El segon gràfic de transformació indica que, per a aquesta variable, l'edat de 16 anys es troba al voltant del valor 0,5 de l'eix d'ordenades (dimensió 2) i l'edat de 12 anys es troba a -0,5.

Figura 39. Gràfic de transformació de la variable independent edat a l'Anàlisi de Components Principals de V3b.



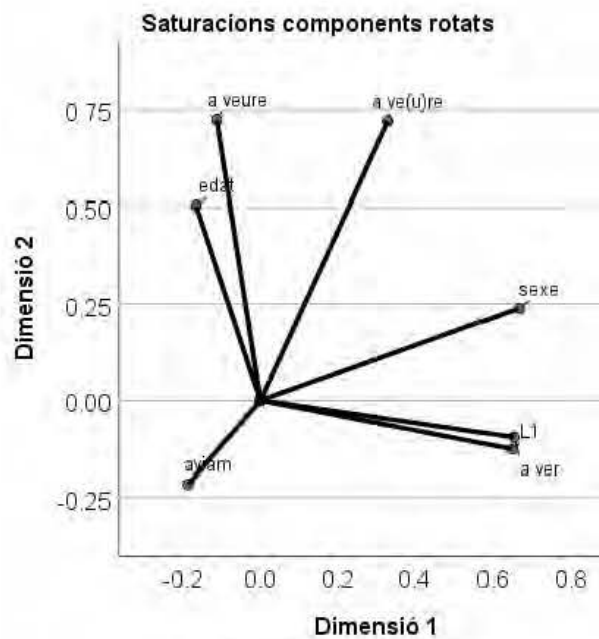
El tercer gràfic de transformació mostra que, mentre que els nois es projecten vora -0,6 a l'eix d'abscisses, les noies ho fan al 0,6.

Figura 40. Gràfic de transformació de la variable independent sexe a l'Anàlisi de Components Principals de V3b.



Així doncs, amb la informació anterior, podem observar al gràfic de saturacions als components rotats (Figura 41) que les variants de V3b i les variables independents no se situen especialment a prop de cap dels grups de llengua inicial (que corresponen als valors -1, 0,5 i 1 de la dimensió 1). En canvi, la variant *a ver* apareix fortament relacionada amb la llengua inicial i vora 0,6 a la dimensió 1 que, recordem, correspondria al sexe femení. Quant a l'edat, es projecta al valor 0,5 de l'eix d'ordenades, que correspon als 16 anys i els angles entre les línies mostren que, d'entre la resta de factors del gràfic, manté més correlació amb la variant *a veure*. Les altres dues variants de V3b, però, no apareixen especialment a prop de cap dels grups de parlants ni estretament correlacionades amb cap variable independent.

Figura 41. Gràfic de saturacions als components rotats per a la variable V3b.



Normalització principal per variable.

Mètode de rotació: Varimax amb normalització Kaiser.

6.2.3. Anàlisi de Components Principals del doblet *bueno* / *bé* (V3c)

Ens fixem tot seguit en el resum del model generat per a la variable V3c, que considera les partícules discursives *bueno* i *bé*, mostra que assoleix la capacitat d'explicar pràcticament el 50 % de la variància de les dades. En aquest cas, la primera dimensió n'explica el 25,7 % i la segona, més del 24,2 %.

Taula 54. Resum del model generat per l'Anàlisi de Components Principals per a V3c.

Resum model multivariant rotat ^a			
Dimensió	Alfa de Cronbach	Variància explicada	
		Total (valors propis)	% Variància
1	,280	1,288	25,759
2	,220	1,212	24,238
Total	,750 ^b	2,500	49,997

^a Mètode de rotació: Varimax amb Normalització Kaiser.

^b Alfa de Cronbach total basat en el total de valors propis.

Quant a la matriu de correlació per a V3c, com en el cas anterior, no s'hi observen valors propers a 1, per la qual cosa sembla que les correlacions entre les variables no són especialment fortes.

Taula 55. Matriu de correlacions entre variables per a V3c.

	L1	Edat	Sexe	<i>bueno</i>	<i>bé</i>
Llengua inicial	1,000	,000	,258	-,084	,087
Edat	,000	1,000	,000	,184	-,067
Sexe	,258	,000	1,000	,050	,049
<i>bueno</i>	-,084	,184	,050	1,000	-,029
<i>bé</i>	,087	-,067	,049	-,029	1,000
Dimensió	1	2	3	4	5
Valors propis	1,305	1,195	,959	,849	,693

La matriu de saturacions als components rotats (Taula 56) mostra que la llengua inicial, el sexe i la variant *bé* contribueixen més al primer component, mentre que l'edat i *bueno* tenen més pes al segon. Així doncs, la combinació de *bé*, llengua inicial i sexe explica millor les dades de V3c que l'edat i *bueno*.

Taula 56. Matriu de saturacions als components rotats per a V3c.

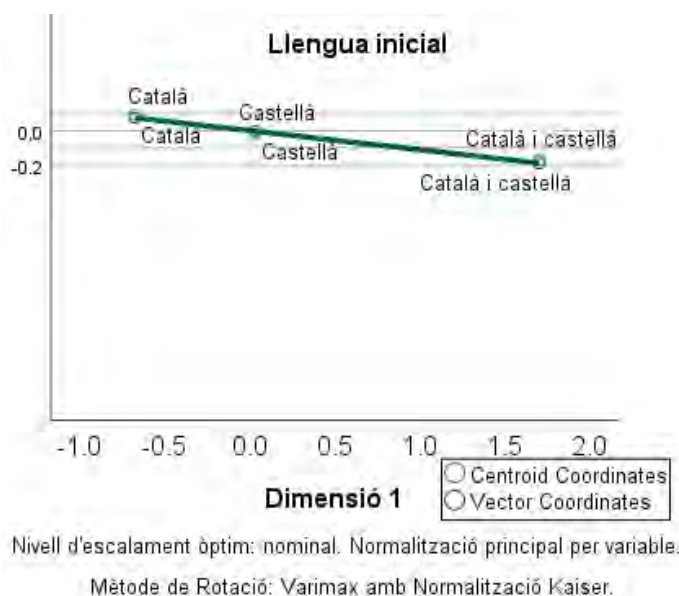
a	Dimensió	
	1	2
Llengua inicial	,775	-,087
Edat	,019	,739
Sexe	,766	,147
<i>bueno</i>	-,014	,746
<i>bé</i>	,317	-,281

Normalització principal per variable.

^a Mètode de rotació: Varimax amb normalització Kaiser. La rotació no ha convergit en 3 iteracions. (Convergència = ,000).

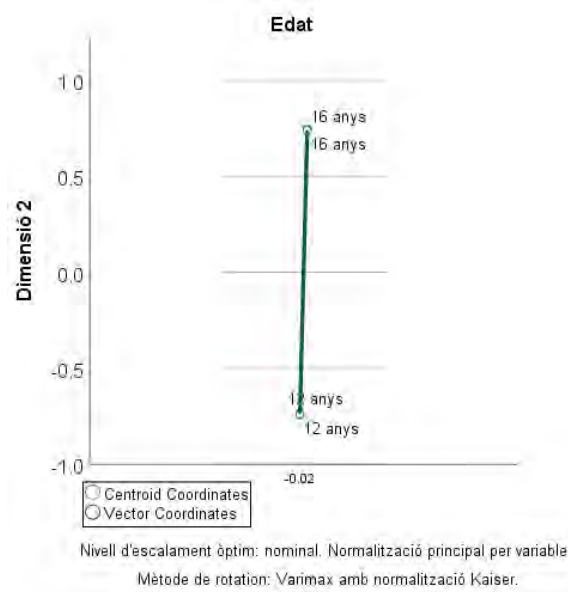
A continuació observem els tres gràfics de transformació per a V3c. El de la llengua inicial (Figura 42), representada a l'eix d'abscisses, mostra que els catalanoparlants inicials se situen prop de -0,5, els castellanoparlants inicials es troben prop de 0 i els bilingües inicials més aviat vora 2.

Figura 42. Gràfic de transformació de la variable independent llengua inicial a l'Anàlisi de Components Principals de V3c.



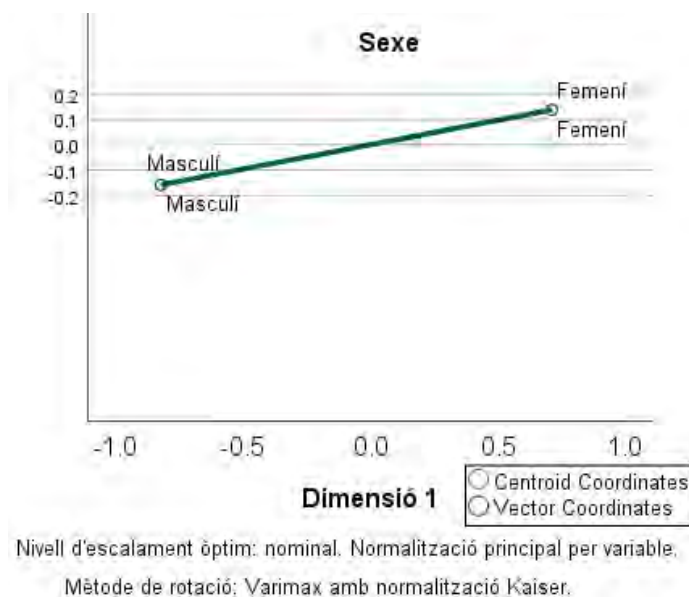
El gràfic de transformació de la variable edat per a V3c (Figura 43) projecta les dades corresponents a l'edat de 12 anys entre els valors -1 i -0,5 i les corresponents a l'edat de 16 anys entre 1 i 0,5 de l'eix d'ordenades.

Figura 43. Gràfic de transformació de la variable independent edat a l'Anàlisi de Components Principals de V3c.



El tercer gràfic de transformació per a V3c, el de la variable sexe (Figura 44), mostra que, a la dimensió 1, els nois se situen entre -1 i -0,5, mentre que les noies es troben entre els valors de 0,5 i 1.

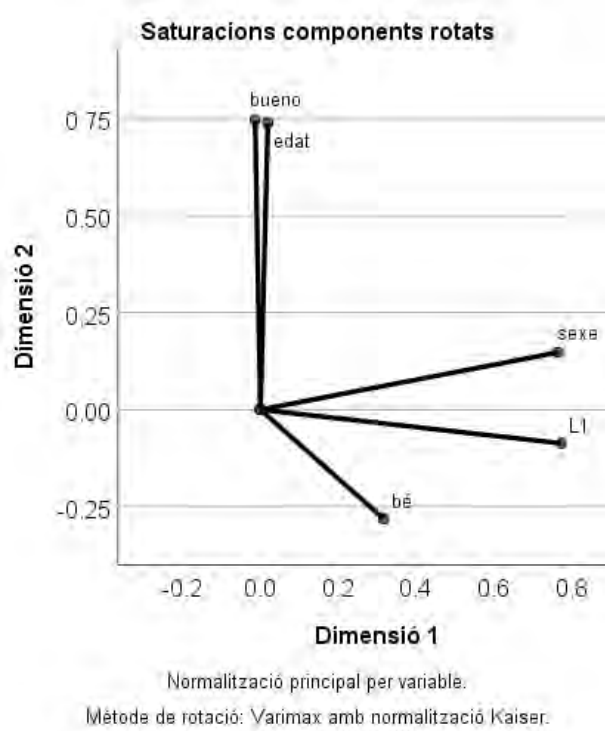
Figura 44. Gràfic de transformació de la variable independent sexe a l'Anàlisi de Components Principals de V3c.



A partir de la informació anterior, doncs, podem interpretar el gràfic de saturacions als components rotats de V3c. Hi observem que les línies més pròximes entre elles són les de la variant *bueno* i l'edat dels informants, la qual cosa indica que aquests dos factors podrien estar correlacionats. A més, es

troben, d'una banda, al bell mig de l'àrea que el gràfic de transformació de la variable edat atribueix als 16 anys (entre 1 i 0,5 de la dimensió 2) i, de l'altra, prop de 0 de l'eix horitzontal, àrea que correspon als castellanoparlants inicials, segons el gràfic de transformació de la llengua inicial. La variant *bé*, en canvi, no sembla mantenir cap relació destacable amb la resta de factors.

Figura 45. Gràfic de saturacions als components rotats per a la variable V3c.



6.2.4. Anàlisi de Components Principals del doblet *después* / *després* (V3f)

Quant a la següent parella de partícules discursives (*después* i *després*, comptabilitzades com a V3f), el model generat per l'ACP arriba a explicar prop del 50 % de la variància de les dades introduïdes. Concretament, la capacitat explicativa del primer component sobrepassa el 29,2 % i la del segon component supera lleugerament el 20,1 %.

Taula 57. Resum del model generat per l'Anàlisi de Components Principals per a V3f.

Resum model multivariant rotat ^a			
Dimensió	Alfa de Cronbach	Variància explicada	
		Total (valors propis)	% Variància
1	,395	1,462	29,245
2	,008	1,006	20,118
Total	,744 ^b	2,468	49,362

^a Mètode de rotació: Varimax amb Normalització Kaiser.

^b Alfa de Cronbach total basat en el total de valors propis.

En el cas d'aquesta variable, la matriu de correlacions presenta, de nou, valors considerablement per sota de la unitat. Això indica, com s'ha observat per a les dues variables anteriors, que les correlacions entre les variables no són especialment destacables.

Taula 58. Matriu de correlacions entre variables per a V3f.

	L1	Edat	Sexe	<i>después</i>	<i>després</i>
Llengua inicial	1,000	,000	,258	,159	-,108
Edat	,000	1,000	,000	-,013	-,016
Sexe	,258	,000	1,000	,176	-,090
<i>después</i>	,159	-,013	,176	1,000	-,103
<i>després</i>	-,108	-,016	-,090	-,103	1,000
Dimensió	1	2	3	4	5
Valors propis	1,464	1,005	,936	,856	,740

La matriu de saturacions als components rotats per a la variable V3f (Taula 59) indica que totes les variables i variants tenen força més pes al primer component, excepte l'edat dels informants, que contribueix més al segon i, per tant, no es mostra tan rellevant per explicar les dades.

Taula 59. Matriu de saturacions als components rotats per a V3f.

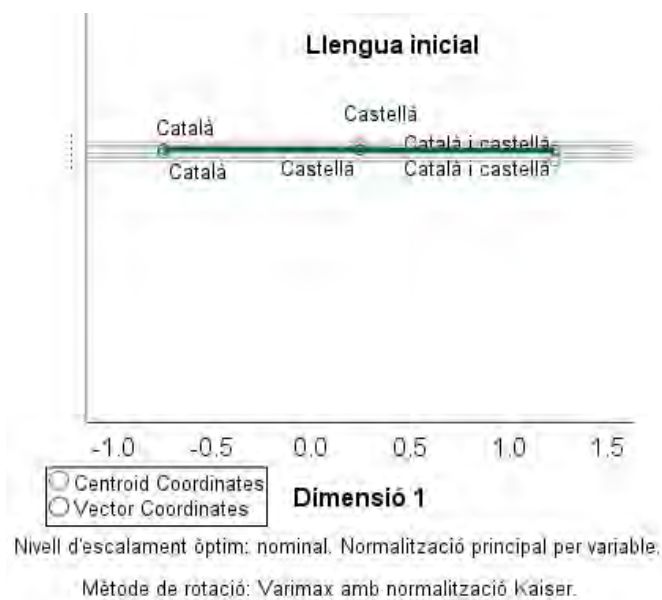
a	Dimensió	
	1	2
Llengua inicial	,684	-,002
Edat	-,059	,963
Sexe	,688	-,020
<i>después</i>	,592	-,025
<i>després</i>	-,408	-,279

Normalització principal per variable.

^a Mètode de rotació: Varimax amb normalització Kaiser. La rotació no ha convergit en 3 iteracions. (Convergència = ,000).

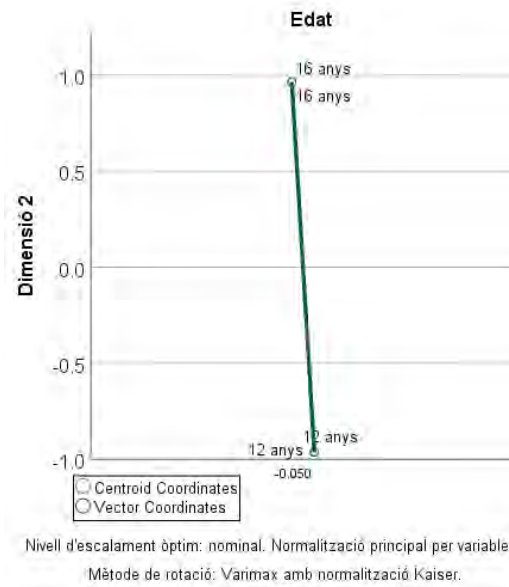
Pel que fa als gràfics de transformació de V3f, el de la llengua inicial (Figura 46) projecta el grup de catalanoparlants inicials entre -1 i -0,5 de la dimensió 1; el de castellanoparlants inicials, prop de 0, i el de bilingües inicials, entre 1 i 1,5.

Figura 46. Gràfic de transformació de la variable independent llengua inicial a l'Anàlisi de Components Principals de V3f.



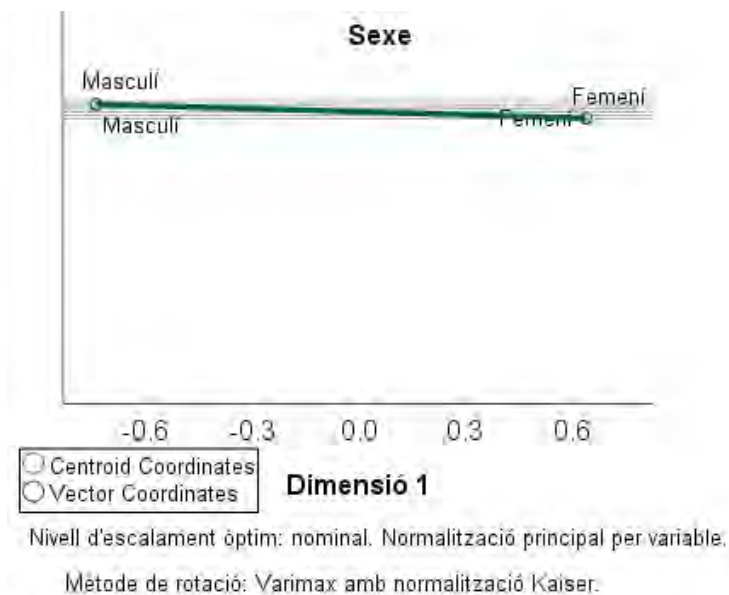
El gràfic de transformació de la variable edat per a V3f situa les dades corresponents a l'edat de 12 anys vora -1 i les de 16 anys vora 1 de l'eix d'ordenades (dimensió 2).

Figura 47. Gràfic de transformació de la variable independent edat a l'Anàlisi de Components Principals de V3f.



El gràfic de transformació de la variable sexe per a V3f, al seu torn, mostra que aquests grups se situen al llarg de l'eix corresponent a la dimensió 1. Més concretament, els nois es projecten prop de -0,6 i les noies, prop de 0,6.

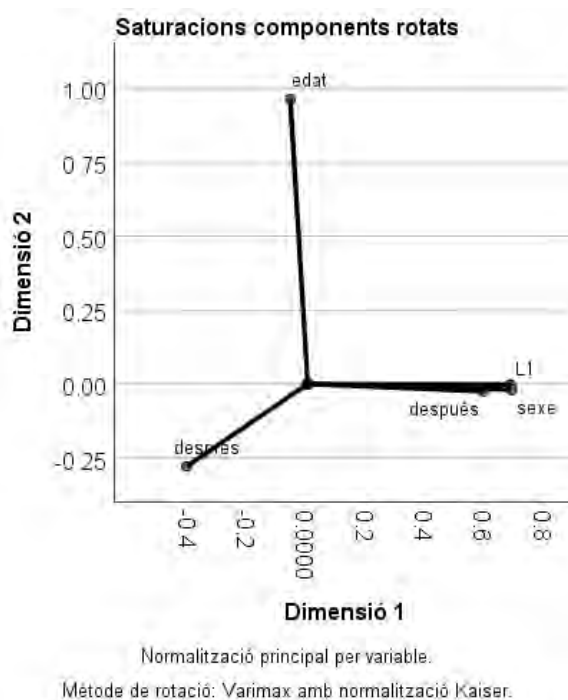
Figura 48. Gràfic de transformació de la variable independent sexe a l'Anàlisi de Components Principals de V3f.



Finalment, el gràfic de saturacions als components rotats per a V3f mostra una forta correlació entre la llengua inicial, el sexe i la variant *después*. Tots tres factors apareixen prop de 0 a l'eix d'ordenades i entre 0,6 i 0,8 al d'abscisses. Segons el gràfic anterior, aquesta situació correspon al grup de noies. En canvi,

ni l'edat ni la variant *després* semblen mantenir cap correlació particularment destacable entre elles ni amb els altres factors.

Figura 49. Gràfic de saturacions als components rotats per a la variable V3f.



6.2.5. Anàlisi de Components Principals del doblet *hombre / home* (V3h)

La quarta variable pragmàtica és el doblet de partícules discursives *hombre* i *home* (V3h). En aquesta ocasió, el model estadístic generat per l'ACP presenta una capacitat explicativa de més del 51,2 % de la variància de les dades (la tercera més elevada de l'estudi), que es reparteix entre les dues dimensions amb un 26,4 % i un 24,7 %, respectivament.

Taula 60. Resum del model generat per l'Anàlisi de Components Principals per a V3h.

Resum model multivariant rotat ^a			
Dimensió	Alfa de Cronbach	Variància explicada	
		Total (valors propis)	% Variància
1	,317	1,322	26,437
2	,256	1,238	24,768
Total	,762 ^b	2,560	51,205

^aMètode de rotació: Varimax amb Normalització Kaiser.

^bAlfa de Cronbach total basat en el total de valors propis.

Pel que fa a la matriu de correlacions entre les variables (Taula 61), en el cas de V3h, com en els anteriors, només hi apareixen valors baixos (inferiors a 0,3). Per tant, no s'hi identifiquen correlacions fortes entre aquestes variables.

Taula 61. Matriu de correlacions entre variables per a V3h.

	L1	Edat	Sexe	<i>hombre</i>	<i>home</i>
Llengua inicial	1,000	,000	,257	,000	-,201
Edat	,000	1,000	,000	,041	,203
Sexe	,257	,000	1,000	-,053	-,111
<i>hombre</i>	,000	,041	-,053	1,000	,116
<i>home</i>	-,201	,203	-,111	,116	1,000
Dimensió	1	2	3	4	5
Valors propis	1,440	1,121	,972	,789	,679

La Taula 62 mostra la matriu de saturacions als components rotats per a la variable V3h. S'hi pot observar que, mentre que la llengua inicial i el sexe tenen valors més elevats per a la primera dimensió, l'edat i les dues variants d'aquesta variable tenen més pes al segon factor, és a dir, el conjunt format per la llengua inicial i sexe expliquen millor les dades que el format per l'edat i les dues variants de V3h.

Taula 62. Matriu de saturacions als components rotats per a V3h.

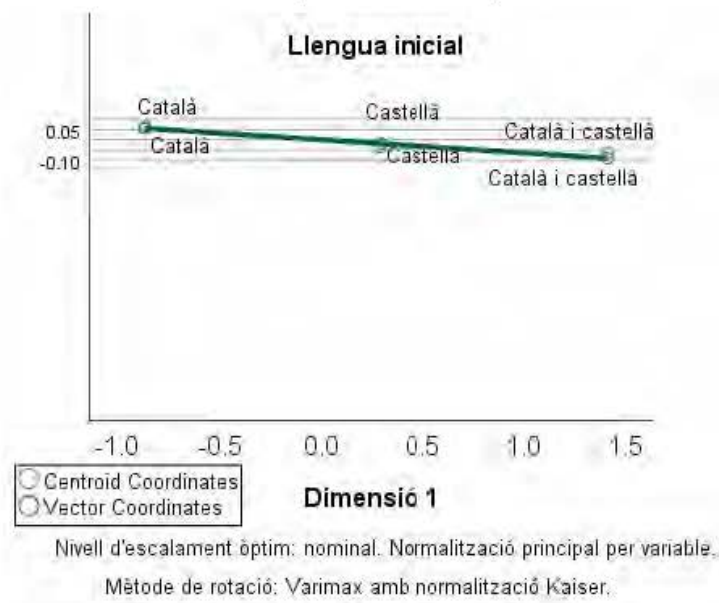
a	Dimensió	
	1	2
Llengua inicial	,788	-,051
Edat	,143	,748
Sexe	,752	,013
<i>hombre</i>	-,003	,460
<i>home</i>	-,337	,681

Normalització principal per variable.

^a Mètode de rotació: Varimax amb normalització Kaiser. La rotació no ha convergit en 3 iteracions. (Convergència = ,000).

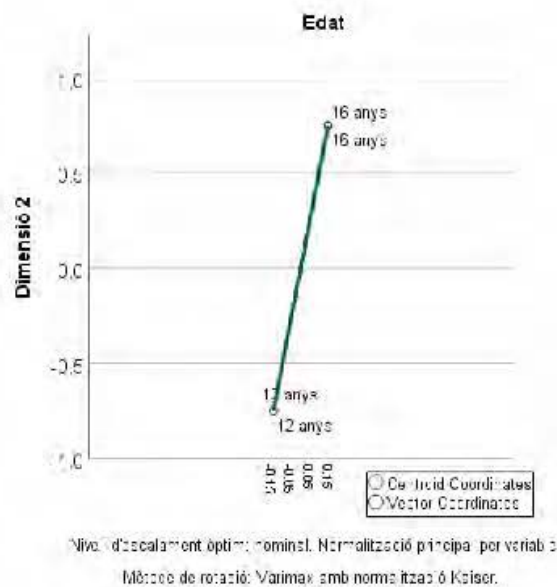
El primer gràfic de transformació per a V3h indica que la dimensió rellevant per a la llengua inicial en aquesta ocasió, com hem vist, és la primera i que els grups de catalanoparlants inicials es troben entre els valors de -1 i -0,5, a força distància dels de castellanoparlants (que se situen vora 0,5) i els de bilingües inicials (1,5).

Figura 50. Gràfic de transformació de la variable independent llengua inicial a l'Anàlisi de Components Principals de V3h.



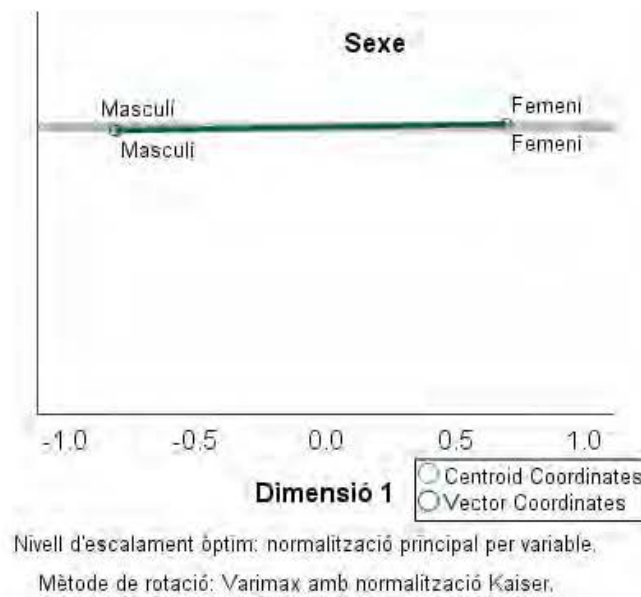
El gràfic de transformació de la variable edat per a V3h projecta els dos grups d'edat sobre la dimensió 2: l'edat de 16 anys se situa entre 1 i 0,5 i la de 12 anys entre -0,5 i -1.

Figura 51. Gràfic de transformació de la variable independent edat a l'Anàlisi de Components Principals de V3h.



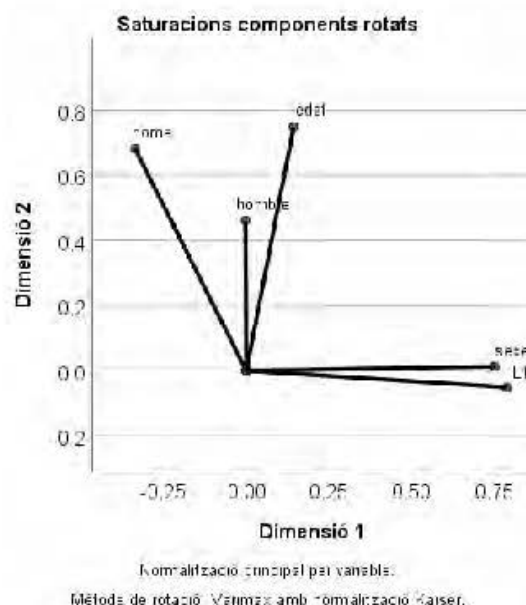
La Figura 52 mostra que la variable sexe es projecta sobre l'eix d'abscisses (dimensió 1) de manera que els nois se situen entre els valors de -1 i -0,5, mentre que les noies es troben entre els valors positius de 0,5 i 1.

Figura 52. Gràfic de transformació de la variable independent sexe a l'Anàlisi de Components Principals de V3h.



Per a la interpretació del gràfic de saturacions als components rotats per a V3h, la informació dels gràfics de transformació anteriors ens indica, en primer lloc, que la variant *home* i la variable edat es projecten a l'àrea assignada a l'edat de 16 anys (entre 0,5 i 1 de la dimensió 2) i, en segon, que l'àrea en què apareixen les variables sexe i llengua inicial es correspon als grups d'informants noies. L'observació dels angles que separen les línies de cada factor ens mostra, a més, que les correlacions més destacables operen entre la variant *home* i l'edat i entre les variables sexe i llengua inicial.

Figura 53. Gràfic de saturacions als components rotats per a la variable V3h.



6.2.6. Anàlisi de Components Principals del doblat *o sea / o sigui* (V3i)

La Taula 63 mostra, tot seguit, que el model generat per l'ACP per a la variable V3i, que considera el doblat *o sea / o sigui*, supera el 52,4 % de variància explicada (la segona més elevada de tot l'estudi), ja que tots dos components arriben a explicar més del 26 % de la variabilitat de les dades.

Taula 63. Resum del model generat per l'Anàlisi de Components Principals per a V3i.

Resum model multivariant rotat ^a			
Dimensió	Alfa de Cronbach	Variància explicada	
		Total (valors propis)	% Variància
1	,301	1,316	26,321
2	,295	1,307	26,143
Total	,773 ^b	2,623	52,464

^a Mètode de rotació: Varimax amb Normalització Kaiser.

^b Alfa de Cronbach total basat en el total de valors propis.

La matriu de correlacions entre variables corresponent a la variable V3i es mostra a la Taula 64. Una vegada més, cap dels valors d'aquesta matriu s'apropa a 1, de manera que no sembla que les correlacions entre les variables siguin especialment fortes.

Taula 64. Matriu de correlacions entre variables per a V3i.

	L1	Edat	Sexe	<i>o sea</i>	<i>o sigui</i>
Llengua inicial	1,000	,000	,076	,245	-,213
Edat	,000	1,000	,000	,117	,254
Sexe	,076	,000	1,000	,051	,062
<i>o sea</i>	,245	,117	,051	1,000	-,016
<i>o sigui</i>	-,213	,254	,062	-,016	1,000
Dimensió	1	2	3	4	5
Valors propis	1,356	1,267	,992	,742	,643

Pel que fa a la matriu de saturacions als components rotats per a V3i, hi observem que l'edat i la variant *o sigui* tenen més pes al primer component, a diferència de les altres dues variables independents i la variant *o sea*, que

presenten valors més elevats al segon component. El primer bloc de factors es mostra, així, més rellevant per diferenciar entre els informants.

Taula 65. Matriu de saturacions als components rotats per a V3i.

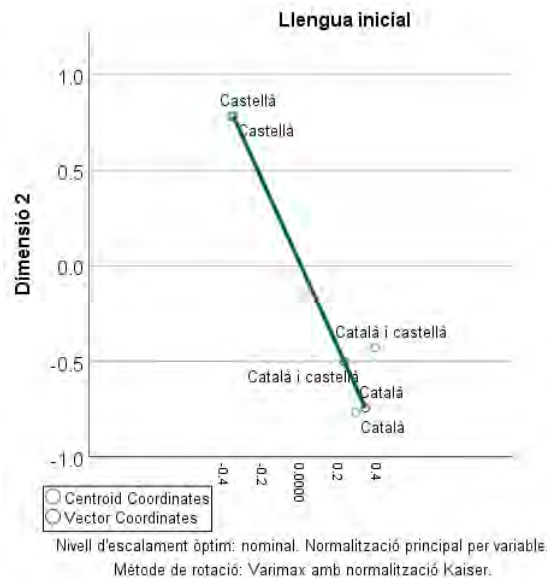
a	Dimensió	
	1	2
Llengua inicial	-,333	,734
Edat	,724	,276
Sexe	,136	,304
<i>o sea</i>	,097	,761
<i>o sigui</i>	,808	-,142

Normalització principal per variable.

^a Mètode de rotació: Varimax amb normalització Kaiser. La rotació no ha convergit en 3 iteracions. (Convergència = ,000).

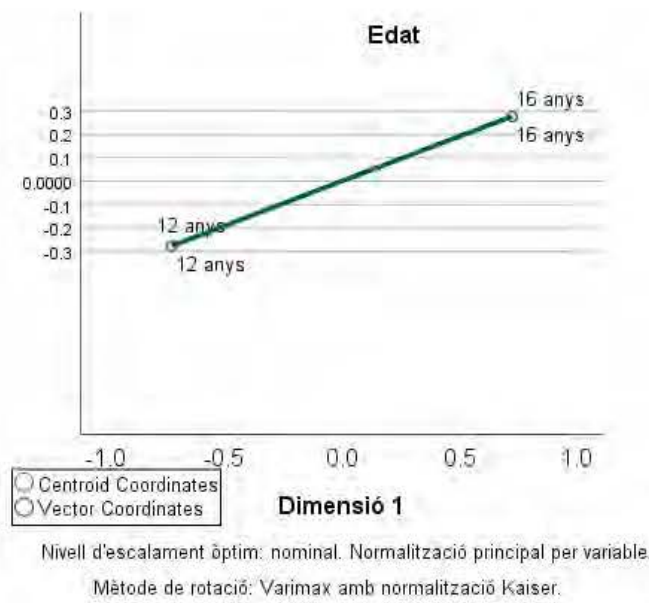
Com s'observa al gràfic de transformació de la llengua inicial per a V3i (Figura 54), la distància entre el grup de castellanoparlants inicials (situats entre 0,5 i 1 a la dimensió 2) i el de bilingües inicials (que es troben al voltant de -0,5) és força més gran que entre aquest últim i el de catalanoparlants inicials (que apareixen entre -0,5 i -1).

Figura 54. Gràfic de transformació de la variable independent llengua inicial a l'Anàlisi de Components Principals de V3i.



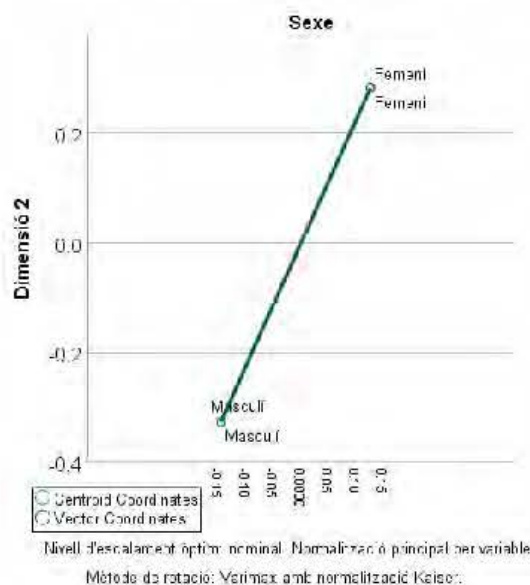
El gràfic de transformació de l'edat per a V3i (Figura 55), situa, d'una banda, els 12 anys entre -1 i -0,5 i, de l'altra, els 16 anys entre 0,5 i 1 a la dimensió 1 (eix d'abscisses).

Figura 55. Gràfic de transformació de la variable independent edat a l'Anàlisi de Components Principals de V3i.



El darrer gràfic de transformació, el de la variable sexe, mostra que, en el cas de V3i, les noies es projecten entre els valors 0,2 i 0,4 i els nois entre -0,2 i -0,4 de l'eix d'ordenades.

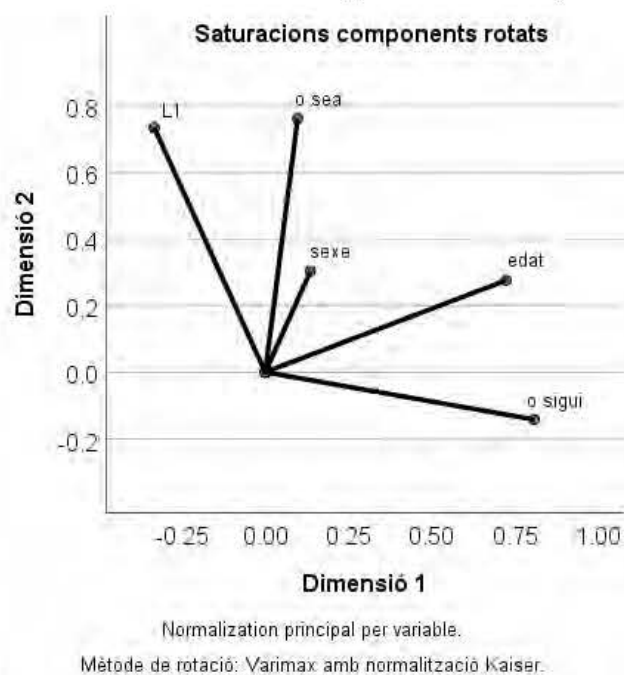
Figura 56. Gràfic de transformació de la variable independent sexe a l'Anàlisi de Components Principals de V3i.



Per tant, amb la informació anterior observem ara el gràfic de saturacions als components rotats per a V3i. Segons les dades relatives a la dimensió 1, hi veiem que la variable edat i la variant o sigui es troben a l'àrea dels 16 anys (entre 0,5 i 1). Si considerem, a més, el que ens indiquen els gràfics de transformació sobre

la dimensió 2, descobrim que la variant *o sea* i la variable llengua inicial se situen a l'àrea corresponent als castellanoparlants inicials (entre 0,5 i 1 de l'eix d'ordenades), que les variables sexe i edat es troben a la zona de les noies (entre 0,2 i 0,4) del mateix eix) i que la variant *o sigui* s'aproxima, tot i que no arriba a situar-s'hi, a l'àrea dels nois (entre -0,2 i -0,4). Pel que fa a l'amplitud dels angles entre les línies que representen els factors, els dos més propers entre ells són la variant *o sea* i la variable sexe.

Figura 57. Gràfic de saturacions als components rotats per a la variable V3i.



6.2.7. Anàlisi de Components Principals del doblet *pero* / *però* (V3j)

A continuació, tractem l'ACP corresponent a la variable V3j, que engloba les partícules discursives adversatives *pero* i *però* (aquesta darrera en diferents realitzacions formals). El resum del model estadístic (Taula 66) indica que la primera dimensió explica més del 22,3 % de la variància de les dades, mentre que la segona dimensió n'explica més del 19,6 %. En conjunt, per tant, el model assoleix pràcticament el 42,0 % de variància explicada.

Taula 66. Resum del model generat per l'Anàlisi de Components Principals per a V3j.

Resum model multivariant rotat ^a			
Dimensió	Alfa de Cronbach	Variància explicada	
		Total (valors propis)	% Variància
1	,421	1,564	22,339
2	,318	1,375	19,647
Total	,770 ^b	2,939	41,986

^a Mètode de rotació: Varimax amb Normalització Kaiser.

^b Alfa de Cronbach total basat en el total de valors propis.

La Taula 67 mostra la matriu de correlacions per a V3j, en la qual la correlació més destacable s'estableix, en sentit negatiu, entre la llengua inicial i la variant *p(er)ò* (-,392).

Taula 67. Matriu de correlacions entre variables per a V3j.

	L1	Edat	Sexe	<i>pero</i>	<i>però</i>	<i>p(e)rò</i>	<i>p(er)ò</i>
Llengua inicial	1,000	,000	,176	,039	-,082	-,096	-,392
Edat	,000	1,000	,000	-,055	-,087	,209	,082
Sexe	,176	,000	1,000	,013	,036	,176	-,209
<i>pero</i>	,039	-,055	,013	1,000	-,021	-,165	-,042
<i>però</i>	-,082	-,087	,036	-,021	1,000	-,148	,107
<i>p(e)rò</i>	-,096	,209	,176	-,165	-,148	1,000	-,038
<i>p(er)ò</i>	-,392	,082	-,209	-,042	,107	-,038	1,000
Dimensió	1	2	3	4	5	6	7
Valors propis	1,564	1,375	1,051	,936	,873	,657	,544

Quant a la matriu de saturacions als components rotats per a aquesta variable, la Taula 68 indica que la llengua inicial i el sexe dels informants, a més de la variant *p(er)*ò són els factors que contribueixen més a la primera dimensió i, per tant, diferencien millor que la resta de dades els informants del corpus. En canvi, l'edat i les altres tres variants de V3j tenen més pes a la segona.

Taula 68. Matriu de saturacions als components rotats per a V3j.

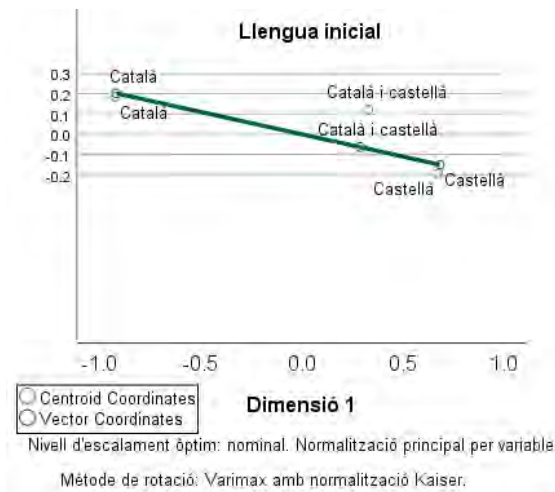
a	Dimensió	
	1	2
Llengua inicial	,747	-,165
Edat	-,057	,604
Sexe	,545	,196
<i>pero</i>	,111	-,435
<i>però</i>	-,241	-,350
<i>p(e)rò</i>	,099	,792
<i>p(er)ò</i>	-,791	,074

Normalització principal per variable.

^a Mètode de rotació: Varimax amb normalització Kaiser. La rotació no ha convergit en 3 iteracions. (Convergència = ,000).

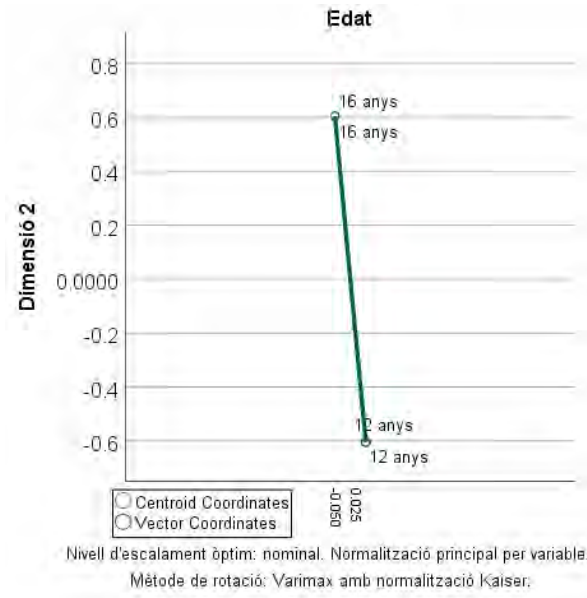
En consonància amb aquestes dades, els gràfics de transformació per a V3j mostren que la llengua inicial es projecte al llarg de la dimensió 1 (eix d'abscisses), l'edat a la dimensió 2 (eix d'ordenades) i el sexe a la dimensió 1. El primer d'aquests gràfics (Figura 58) indica que els catalanoparlants inicials se situen prop de -1; els bilingües inicials, entre 0 i 0,5, i els castellanoparlants inicials, entre 0,5 i 1 a l'eix d'abscisses.

Figura 58. Gràfic de transformació de la variable independent llengua inicial a l'Anàlisi de Components Principals de V3j.



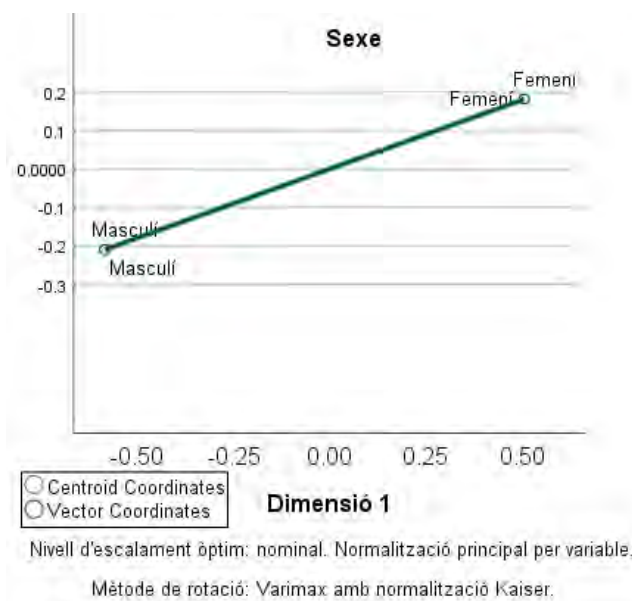
El segon gràfic de transformació per a V3j mostra que les dades corresponents a l'edat de 12 anys es representen vora -0,6 i la de 16 anys, prop de 0,6 a l'eix d'ordenades.

Figura 59. Gràfic de transformació de la variable independent edat a l'Anàlisi de Components Principals de V3j.



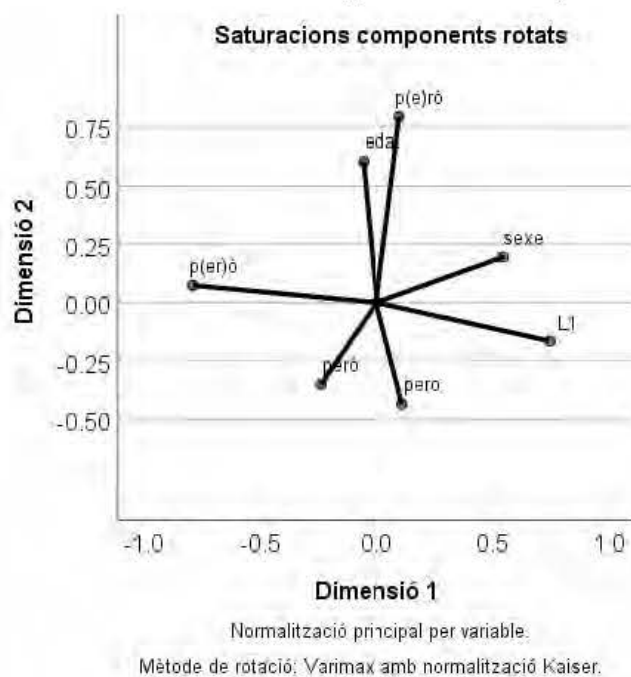
El tercer gràfic de transformació per a V3j és el del sexe dels informants i, com s'observa a la Figura 60, indica que els nois es representen vora -0,5 i les noies prop de 0,5 a l'eix d'abscisses.

Figura 60. Gràfic de transformació de la variable independent sexe a l'Anàlisi de Components Principals de V3j.



Pel que fa al gràfic de saturacions als components rotats per a V3j, segons la informació anterior sobre l'eix corresponent a la dimensió 1, observem a la Figura 61, que la variant *p(er)ò* s'aproxima a l'àrea dels catalanoparlants inicials (-1), que les variants *p(e)rò* i *pero* es troben a la zona de bilingües inicials (entre 0 i 0,5), que les variables llengua inicial i sexe se situen a l'àrea dels castellanoparlants inicials (entre 0,5 i 1), que la variant *però* s'aproxima a la zona dels informants nois (-0,5) i que la variable sexe s'acosta a la de les noies (0,5). La dimensió 2 ens informa que l'edat es representa a l'àrea assignada als 16 anys al gràfic de transformació i que la variant *p(e)rò* s'hi aproxima. Finalment, l'angle més estret d'aquest gràfic de saturacions és precisament el que separa l'edat i *p(e)rò*, de forma que podrien presentar certa correlació.

Figura 61. Gràfic de saturacions als components rotats per a la variable V3j.



6.2.8. Anàlisi de Components Principals del doblet *pues* / *doncs* (V3k)

El penúltim doblet de partícules discursives és el de *pues* (en dues realitzacions diferents) i *doncs*, és a dir, V3k. Aquesta vegada, el model estadístic de l'ACP assoleix en conjunt més del 47,2 % de variància explicada, ja que el primer component arriba, aproximadament, al 24,8 % i el segon, al 22,6 %.

Taula 69. Resum del model generat per l'Anàlisi de Components Principals per a V3k.

Resum model multivariant rotat ^a			
Dimensió	Alfa de Cronbach	Variància explicada	
		Total (valors propis)	% Variància
1	,396	1,486	24,763
2	,315	1,348	22,460
Total	,776 ^b	2,833	47,223

^a Mètode de rotació: Varimax amb Normalització Kaiser.

^b Alfa de Cronbach total basat en el total de valors propis.

Pel que fa a la matriu de correlació per a aquesta variable, tot i que cap dels valors s'acosta a 1, la correlació més forta que s'observa a la Taula 70 és una de negativa entre les variants *pues* i *doncs* (de -0,334 punts).

Taula 70. Matriu de correlacions entre variables per a V3k.

	L1	Edat	Sexe	<i>pues</i>	<i>pos</i>	<i>doncs</i>
Llengua inicial	1,000	,000	,261	,115	,058	,034
Edat	,000	1,000	,000	-,180	,149	-,005
Sexe	,261	,000	1,000	,086	,153	,101
<i>pues</i>	,115	-,180	,086	1,000	,209	-,334
<i>pos</i>	,058	,149	,153	,209	1,000	-,188
<i>doncs</i>	,034	-,005	,101	-,334	-,188	1,000
Dimensió	1	2	3	4	5	6
Valors propis	1,543	1,290	1,144	,799	,652	,572

La Taula 71 correspon a la matriu de saturacions als components rotats per a la variable V3k i mostra que l'edat i les tres partícules discursives tenen valors de saturació més elevats per a la primera dimensió, mentre que la llengua inicial i el sexe dels informants contribueixen més a la segona. Així doncs, aquestes dues no

es mostren tan rellevants per explicar el comportament lingüístic dels informants en relació a V3k.

Taula 71. Matriu de saturacions als components rotats per a V3k.

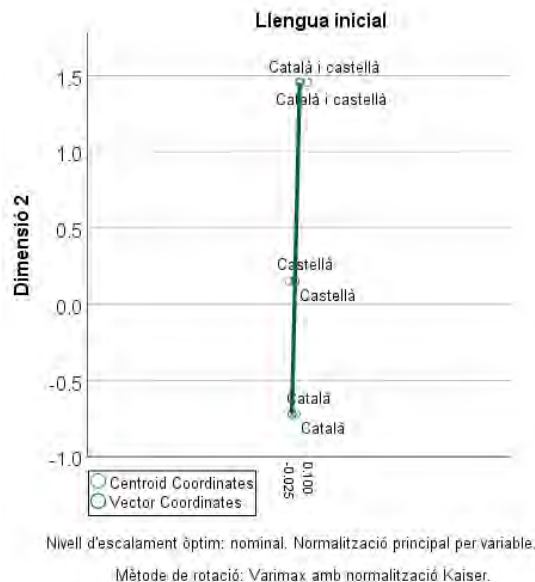
a	Dimensió	
	1	2
Llengua inicial	,017	,722
Edat	-,142	,099
Sexe	-,029	,798
<i>pues</i>	,773	,155
<i>pos</i>	,511	,337
<i>doncs</i>	-,778	,208

Normalització principal per variable.

^a Mètode de rotació: Varimax amb normalització Kaiser. La rotació no ha convergit en 3 iteracions. (Convergència = ,000).

Per això, el gràfic de transformació de la llengua inicial per a V3k mostra la distribució dels grups d'informants al llarg de l'eix d'ordenades: els bilingües inicials es representen prop del valor 1,5; els castellanoparlants inicials, entre 0,5 i 0, i els catalanoparlants inicials, entre -0,5 i -1.

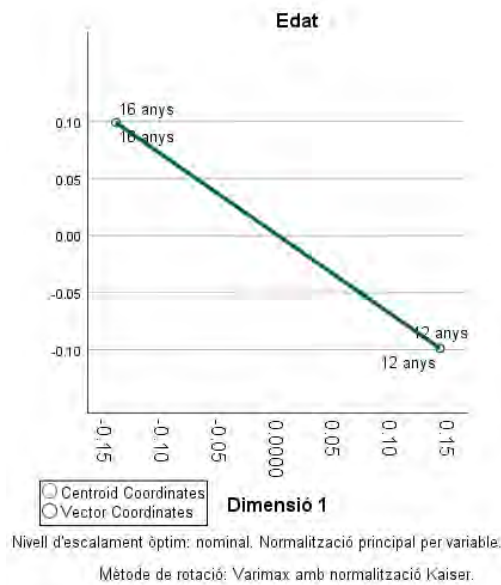
Figura 62. Gràfic de transformació de la variable independent llengua inicial a l'Anàlisi de Components Principals de V3k.



El gràfic de transformació de la variable edat per a V3k (Figura 63), al seu torn, mostra que l'edat de 16 anys es troba prop de -0,15 i la de 12 anys prop de 0,15 a la dimensió 1.

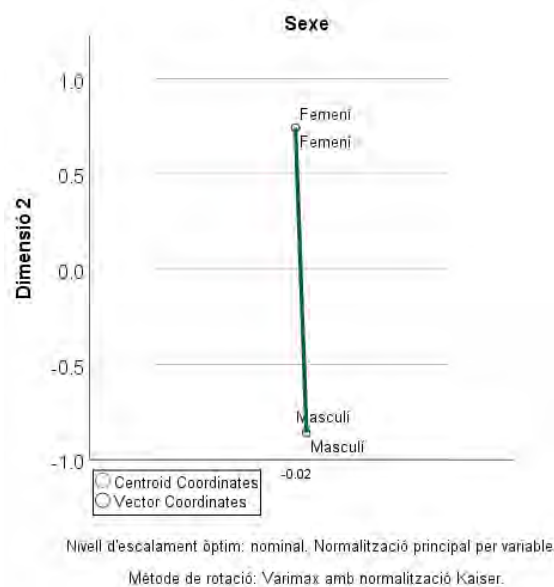
Resultats

Figura 63. Gràfic de transformació de la variable independent edat a l'Anàlisi de Components Principals de V3k.



El tercer gràfic de transformació, el de la variable sexe, indica que les noies se situen entre els valors d'1 i de 0,5 i els nois entre -0,5 i -1 (tot i que més aviat més a prop d'aquest últim) a l'eix d'ordenades (dimensió 2).

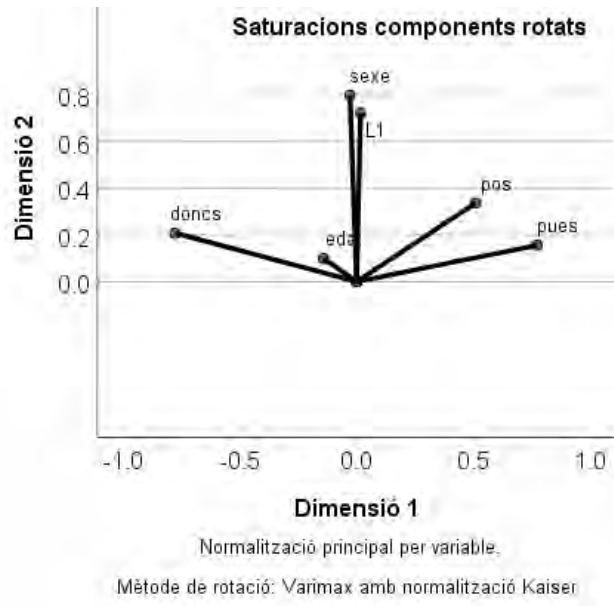
Figura 64. Gràfic de transformació de la variable independent sexe a l'Anàlisi de Components Principals de V3k.



Al gràfic de saturacions als components rotats per a V3k hi veiem que les dues correlacions més destacades es troben, primer, entre les variables sexe i llengua inicial i, segon, entre la variant *doncs* i la variable edat. Quant a la informació que aporten els gràfics anteriors, podem veure, d'una banda, que les variables

sexe i llengua inicial es troben a l'àrea de les noies (entre 1 i 0,5 a la dimensió 2) i que les tres variants i l'edat ocupen la zona dels castellanoparlants inicials (entre 0 i 0,5 encara a l'eix d'ordenades). De l'altra, si observem l'eix d'abscisses, hi veiem que l'edat se situa prop de l'àrea dels 16 anys i la llengua inicial, prop de la dels 12 anys (-0,15 i 0,15, respectivament).

Figura 65. Gràfic de saturacions als components rotats per a la variable V3k.



6.2.9. Anàlisi de Components Principals del doblet *vale / d'acord* (V3l)

La darrera variable que comptabilitza partícules discursives és V3l, que s'ocupa de *vale*, *val* i *d'acord*. Les dimensions del model generat per l'ACP per aquesta variable arriben a explicar, respectivament, el 22,7 % i el 21,9 % de la variància de les dades. Així doncs, en conjunt, el model generat per V3l aconsegueix explicar-ne aproximadament el 44,7 %.

Taula 72. Resum del model generat per l'Anàlisi de Components Principals per a V3l.

Resum model multivariant rotat ^a			
Dimensió	Alfa de Cronbach	Variància explicada	
		Total (valors propis)	% Variància
1	,326	1,365	22,754
2	,293	1,316	21,926
Total	,752 ^b	2,681	44,680

^a Mètode de rotació: Varimax amb Normalització Kaiser.

^b Alfa de Cronbach total basat en el total de valors propis.

La matriu de correlacions entre les variables per a V3l (Taula 73) no mostra cap valor realment proper a 1. El més elevat és el que indica que existeix una correlació negativa de -0,284 punts entre la variant *val* i l'edat dels informants.

Taula 73. Matriu de correlacions entre variables per a V3l.

	L1	Edat	Sexe	<i>vale</i>	<i>val</i>	<i>d'acord</i>
Llengua inicial	1,000	,000	,258	,009	,045	,007
Edat	,000	1,000	,000	,069	-,284	,156
Sexe	,258	,000	1,000	,111	,079	-,182
<i>vale</i>	,009	,069	,111	1,000	,046	-,052
<i>val</i>	,045	-,284	,079	,046	1,000	-,056
<i>d'acord</i>	,007	,156	-,182	-,052	-,056	1,000
Dimensió	1	2	3	4	5	6
Valors propis	1,442	1,240	1,023	,956	,704	,635

La Taula 74 correspon a la matriu de saturacions als components rotats per a V3l. Hi observem que, mentre que la llengua inicial, el sexe i la variant *vale* contribueixen més al segon factor que al primer, el cas de l'edat i les variants *val* i *d'acord* és el contrari. És a dir, l'edat i dues de les variants resulten més rellevants per explicar les dades de V3l que la resta de factors.

Taula 74. Matriu de saturacions als components rotats per a V3l.

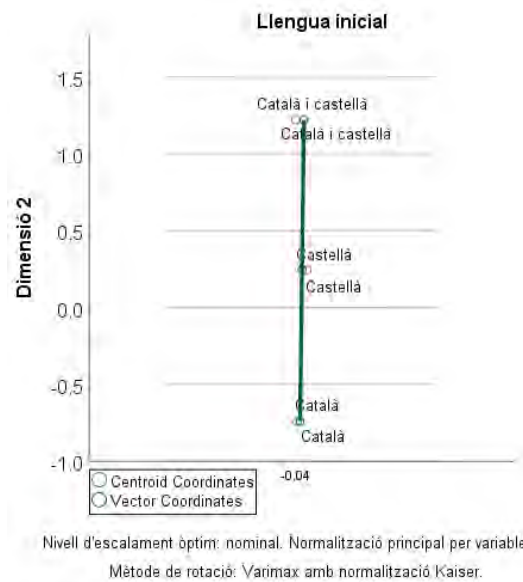
^a	Dimensió	
	1	2
Llengua inicial	,009	,677
Edat	,795	,172
Sexe	-,149	,792
<i>vale</i>	,029	,357
<i>val</i>	-,712	,018
<i>d'acord</i>	,450	-,268

Normalització principal per variable.

^a Mètode de rotació: Varimax amb normalització Kaiser. La rotació no ha convergint en 3 iteracions. (Convergència = ,000).

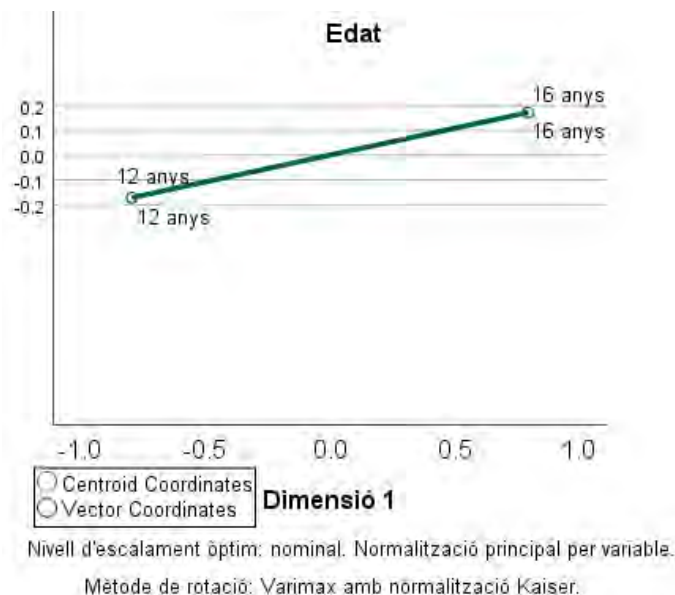
Tot seguit, el gràfic de transformació de la llengua inicial per a la variable V31 (Figura 66) mostra que els informants bilingües inicials es representen entre els valors 1 i 1,5 de la dimensió 2, els castellanoparlants inicials apareixen entre 0 i 0,5, i els catalanoparlants inicials, entre -0,5 i -1.

Figura 66. Gràfic de transformació de la variable independent llengua inicial a l'Anàlisi de Components Principals de V31.



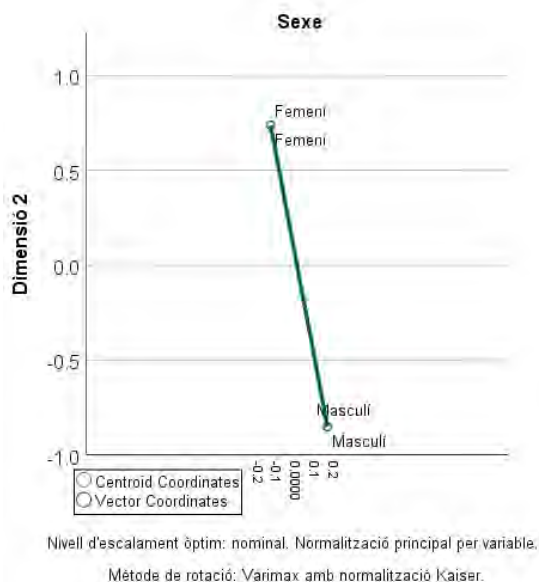
El gràfic de transformació de l'edat per a V31 (Figura 67) indica que, a la dimensió 1, l'edat de 12 anys es mostra entre els valors de -1 i -0,5 i la de 16 anys, entre 0,5 i 1.

Figura 67. Gràfic de transformació de la variable independent edat a l'Anàlisi de Components Principals de V31.



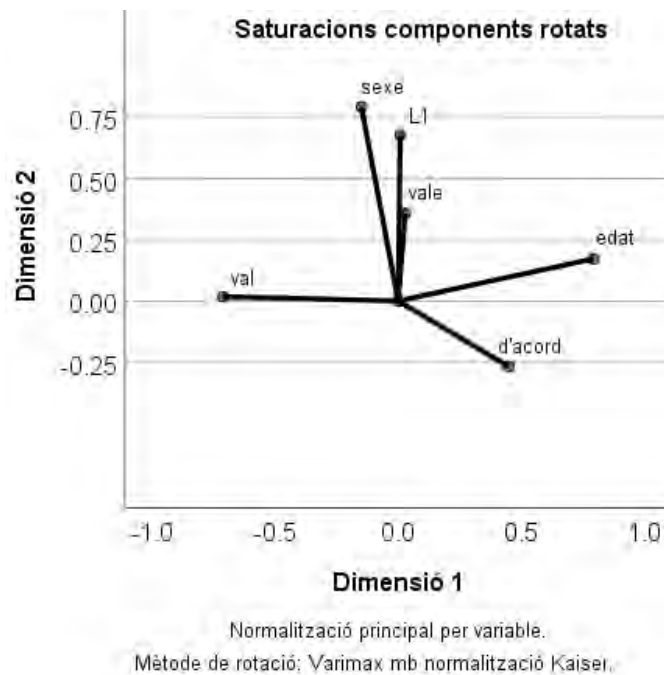
El darrer gràfic de transformació per a V3l, corresponent a la variable sexe (Figura 68), projecta a la dimensió 2 les noies entre els valors positius de 0,5 i 1 i els nois entre els negatius de -0,5 i -1.

Figura 68. Gràfic de transformació de la variable independent sexe a l'Anàlisi de Components Principals de V3l.



Com mostra el gràfic de saturacions als components rotats per a V3l a la Figura 69, les dues correlacions més destacades són les que s'estableixen entre la variant *vale* i la llengua inicial i entre la llengua inicial i el sexe. Tenint en compte la informació recollida als gràfics anteriors per a la dimensió 2, podem observar que les variables sexe i llengua inicial es troben a l'àrea corresponent a les informants noies (entre 0,5 i 1) i que la variable edat i les variants *val* i *vale* se situen a la zona dels catalanoparlants inicials (entre 0 i 0,5 a l'eix d'ordenades). A més, la informació anterior sobre la distribució de factors a la dimensió 1 indica que la variant *val* es troba a la zona que es relaciona amb l'edat dels 12 anys, mentre que la variable edat i la variant *d'acord* es mostren a l'àrea dels 16 anys.

Figura 69. Gràfic de saturacions als components rotats per a la variable V31.



6.2.10. Anàlisi de Components Principals de les fórmules d'obertura de la interacció (V5)

A continuació, ens ocupem dels resultats de l'ACP per a la variable V5. En primer lloc, podem observar a la Taula 75 que el model arriba a explicar més del 45,4 % de la variància del corpus. La primera dimensió comprèn un 24,1 % i la segona més d'un 21,3 %.

Taula 75. Resum del model generat per l'Anàlisi de Components Principals per a V5.

Resum model multivariant rotat ^a			
Dimensió	Alfa de Cronbach	Variància explicada	
		Total (valors propis)	% Variància
1	,481	1,687	24,104
2	,394	1,495	21,355
Total	,800 ^b	3,182	45,458

^aMètode de rotació: Varimax amb Normalització Kaiser.

^bAlfa de Cronbach total basat en el total de valors propis.

En segon lloc, la matriu de correlacions entre les variables independents i les variants de V5 ens informa que les més potents són correlacions negatives entre diverses variants: *hola* i *bon dia* (amb un valor de -0,584), *bon dia* i *bona tarda* (-0,405) i entre *bon dia* i *hola bon dia* (-0,337).

Taula 76. Matriu de correlacions entre variables per a V5.

	L1	Edat	Sexe	<i>hola</i>	<i>bon dia</i>	<i>bona tarda</i>	<i>hola bon dia</i>
Llengua inicial	1,000	,000	,239	-,088	-,053	,297	-,077
Edat	,000	1,000	,000	-,128	,090	,025	,112
Sexe	,239	,000	1,000	,056	-,042	,001	,092
<i>hola</i>	-,088	-,128	,056	1,000	-,584	-,097	,003
<i>bon dia</i>	-,053	,090	-,042	-,584	1,000	-,405	-,337
<i>bona tarda</i>	,297	,025	,001	-,097	-,405	1,000	-,090
<i>hola bon dia</i>	-,077	,112	,092	,003	-,337	-,090	1,000
Dimensió	1	2	3	4	5	6	7
Valors propis	1,769	1,413	1,164	1,067	,825	,611	,150

En tercer lloc, la matriu de saturacions als components rotats per a V5 (Taula 77) mostra que l'edat i tres de les variants (*hola*, *bon dia* i *hola bon dia*) tenen un valor més elevat a la primera dimensió, al contrari que la llengua inicial, el sexe i la variant *bona tarda*, que no es mostren tan rellevants per explicar les dades de V5.

Taula 77. Matriu de saturacions als components rotats per a V5.

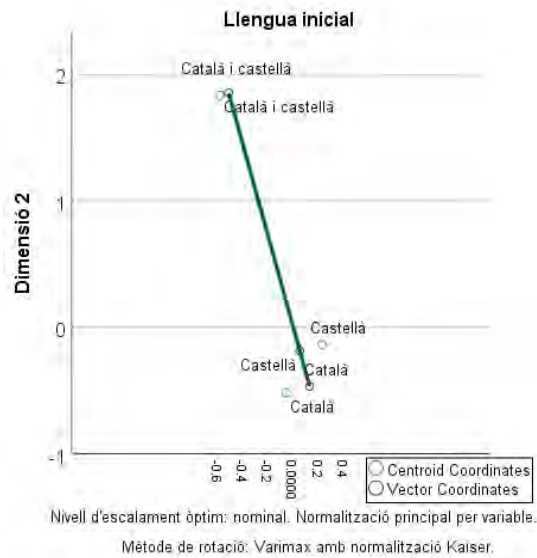
a	Dimensió	
	1	2
Llengua inicial	-,215	,774
Edat	-,202	,036
Sexe	,016	,389
<i>hola</i>	,810	-,057
<i>bon dia</i>	-,870	-,376
<i>bona tarda</i>	,075	,772
<i>hola bon dia</i>	,425	-,050

Normalització principal per variable.

^a Mètode de rotació: Varimax amb normalització Kaiser. La rotació no ha convergint en 3 iteracions. (Convergència = ,000).

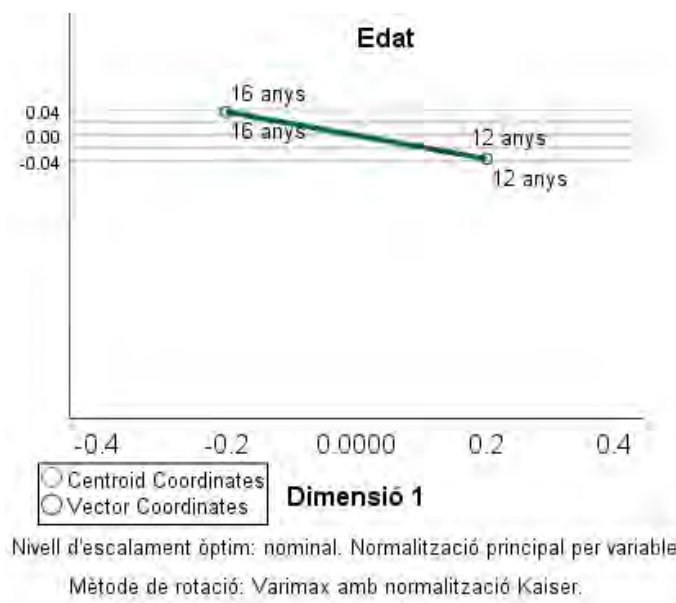
En quart lloc, observem els gràfics de transformació per aquesta variable. El primer indica que els bilingües inicials es projecten a l'eix d'ordenades (dimensió 2) vora el valor de 2, força distanciats dels castellanoparlants inicials (prop de 0) i dels catalanoparlants inicials (aproximadament a -0,5).

Figura 70. Gràfic de transformació de la variable independent llengua inicial a l'Anàlisi de Components Principals de V5.



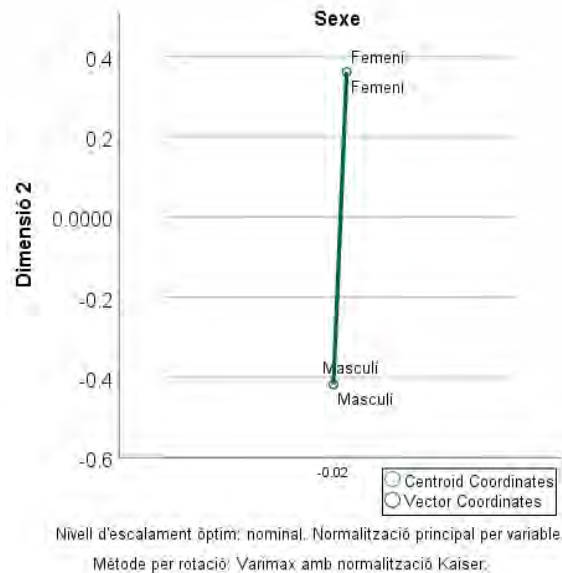
El gràfic de transformació següent mostra els grups d'edat a l'eix d'abscisses (dimensió 1): els 16 anys prop de -0,2 i els 12 anys prop de 0,2.

Figura 71. Gràfic de transformació de la variable independent edat a l'Anàlisi de Components Principals de V5.



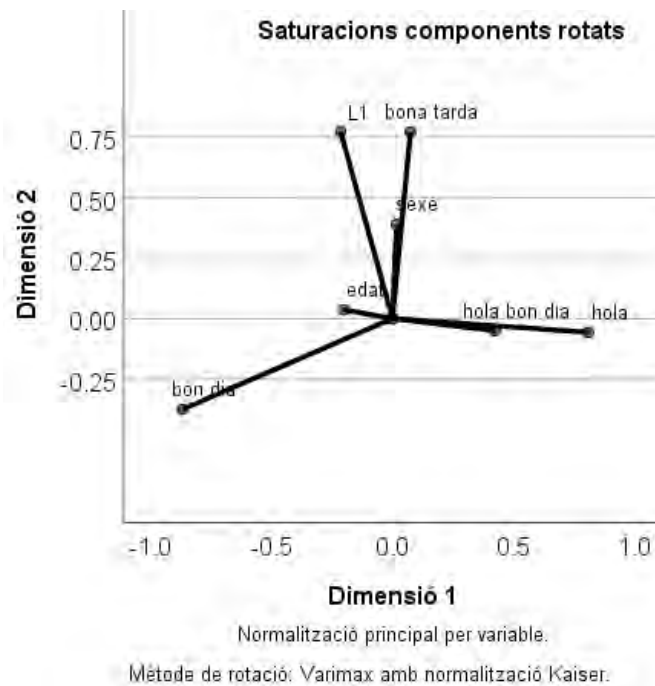
El tercer i últim gràfic d'aquest tipus mostra que, en el cas de la variable V5, les noies es representen cap al valor 0,4 i els nois cap al -0,4 a la dimensió 2.

Figura 72. Gràfic de transformació de la variable independent sexe a l'Anàlisi de Components Principals de V5.



Així doncs, mitjançant la informació que acabem de resumir sobre les dues dimensions, veiem a la Figura 73 que la variable edat es troba, pel que fa a l'eix d'abscisses i juntament amb la llengua inicial, a l'àrea dels 16 anys (-0,2). Pel que fa a l'eix d'ordenades, la variable edat es situa a la zona dels castellanoparlants inicials (vora 2). En aquest mateix eix, hi veiem que la variable sexe apareix aproximadament a l'àrea de les informants noies (0,4). Quant a les variants, a l'àrea dels castellanoparlants inicials s'hi troben *hola bon dia* i *hola*, i *bona tarda* se situa a la zona dels 12 anys (0,2). L'estretor dels angles entre els factors és especialment acusada entre *bona tarda* i la variable sexe i entre *hola bon dia* i *hola*, si bé les línies de *bon dia* i de la variable edat, d'una banda, i les de *bona tarda* i la llengua inicial, de l'altra, tampoc són molt distants entre elles.

Figura 73. Gràfic de saturacions als components rotats per a la variable V5.



6.2.11. Anàlisi de Components Principals de la posició dels pronoms clítics en construccions perifràstiques (V14)

La primera variable morfosintàctica és V14, que examina la posició dels pronoms clítics en construccions perifràstiques, com s'ha explicat anteriorment (als subapartats 5.3.2.3.2 i 5.5.4). Per a l'Anàlisi de Components Principals, es va dividir en les tres variables complementàries que es tracten a continuació.

6.2.11.1. Anàlisi de Components Principals de la posició proclítica (V14a)

La primera d'aquestes variables (V14a) comptabilitza les realitzacions proclítics dels pronoms febles en les construccions perifràstiques d'infinitiu amb *tenir que*, *haver de*, *poder* i *anar a*. El model estadístic generat per l'ACP en aquest cas assoleix una variància explicada del 40,4 %, que es reparteixen la primera dimensió amb un 21,8 % i la segona amb un 18,5 %.

Taula 78. Resum del model generat per l'Anàlisi de Components Principals per a V14a.

Resum model multivariant rotat ^a			
Dimensió	Alfa de Cronbach	Variància explicada	
		Total (valors propis)	% Variància
1	,404	1,530	21,851
2	,269	1,298	18,549
Total	,754 ^b	2,828	40,401

^a Mètode de rotació: Varimax amb Normalització Kaiser.

^b Alfa de Cronbach total basat en el total de valors propis.

La matriu de correlacions no mostra cap correlació propera a 1 entre les variables independents i les variants de V14a. Ara bé, la més destacada és una correlació negativa de -0,33 punts entre les variants amb *haver de* i *tenir que*, que és fàcilment interpretable si tenim en compte que són les dues úniques variants que comparteixen significat.

Taula 79. Matriu de correlacions entre variables per a V14a.

	L1	Edat	Sexe	PC + <i>tenir que</i> / <i>de</i> + INF	PC + <i>haver de</i> + INF	PC + <i>poder</i> + INF	PC+ <i>anar</i> <i>a</i> + INF
Llengua inicial	1,000	,000	,137	-,069	-,062	-,096	-,145
Edat	,000	1,000	,000	-,187	,057	,130	-,187
Sexe	,137	,000	1,000	,036	,154	-,032	-,064
PC + <i>tenir que</i> / <i>de</i> + INF	-,069	-,187	,036	1,000	-,330	,001	-,034
PC + <i>haver de</i> + INF	-,062	,057	,154	-,330	1,000	,282	-,076
PC + <i>poder</i> + INF	-,096	,130	-,032	,001	,282	1,000	,066
PC+ <i>anar a</i> + INF	-,145	-,187	-,064	-,034	-,076	,066	1,000
Dimensió	1	2	3	4	5	6	7
Valors propis	1,535	1,293	1,078	1,015	,835	,799	,445

La Taula 80 mostra la matriu de saturacions als components rotats per a aquesta variable, on es pot veure que l'edat i totes les variants excepte *PC+ anar a + INF* tenen més pes al primer component que al segon i, doncs, resulten rellevants per explicar les dades. En canvi, la llengua inicial, el sexe i aquesta variant contribueixen més al segon component.

Taula 80. Matriu de saturacions als components rotats per a V14a.

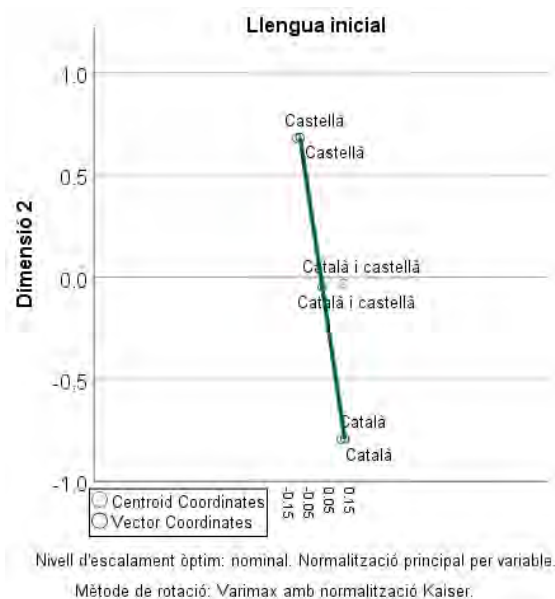
a	Dimensió	
	1	2
Llengua inicial	-,101	,680
Edat	,458	,247
Sexe	,104	,482
<i>PC + tenir que / de + INF</i>	-,614	-,106
<i>PC + haver de + INF</i>	,781	,005
<i>PC + poder + INF</i>	,549	-,357
<i>PC+ anar a + INF</i>	-,107	-,635

Normalització principal per variable.

^a Mètode de rotació: Varimax amb normalització Kaiser. La rotació no ha convergit en 3 iteracions. (Convergència = ,000).

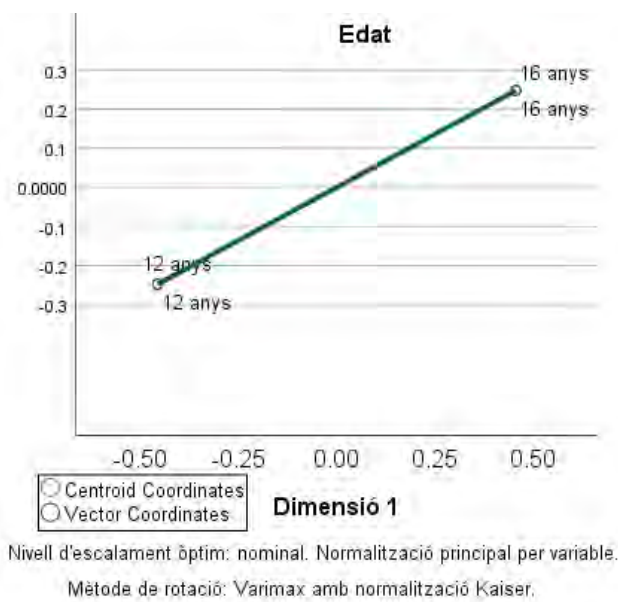
Així, el gràfic de transformació de la variable llengua inicial projecta els tres grups que la componen sobre la dimensió 2. Els castellanoparlants inicials es representen entre 1 i 0,5, els bilingües inicials a l'equador de l'eix i els catalanoparlants inicials entre -0,5 i -1.

Figura 74. Gràfic de transformació de la variable independent llengua inicial a l'Anàlisi de Components Principals de V14a.



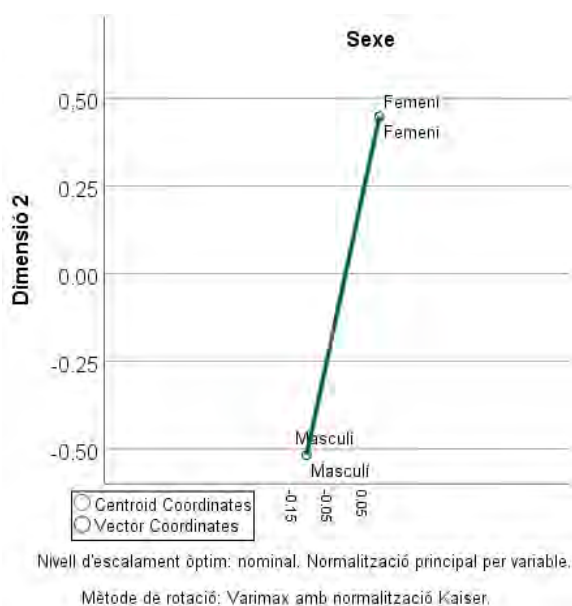
Per la seva banda, el següent gràfic de transformació mostra que l'edat de 12 anys es projecta prop de -0,5 i la de 16 anys, vora 0,5 a l'eix que correspon a la dimensió 1.

Figura 75. Gràfic de transformació de la variable independent edat a l'Anàlisi de Components Principals de V14a.



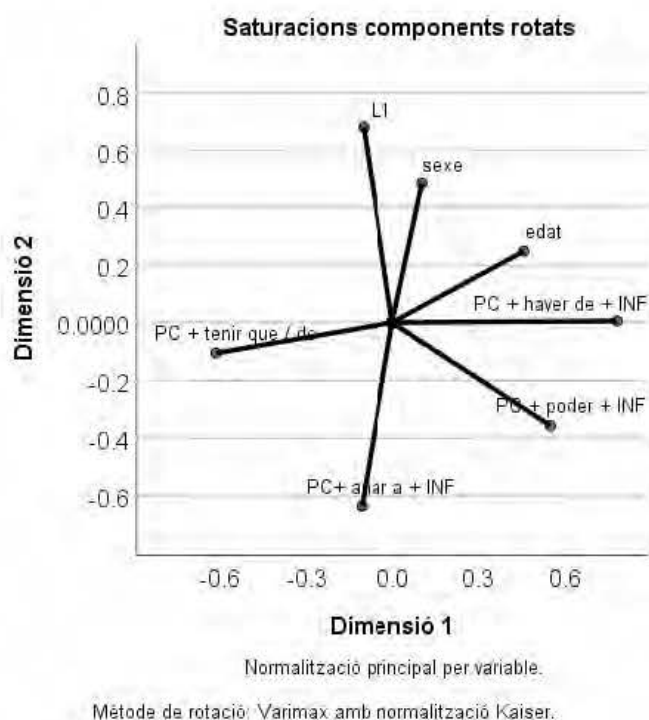
El darrer gràfic de transformació per a V14a indica que, a la dimensió 2, les noies se situen prop de 0,5 i els nois prop de -0,5.

Figura 76. Gràfic de transformació de la variable independent sexe a l'Anàlisi de Components Principals de V14a.



Si traslladem la informació anterior al gràfic següent, que mostra les saturacions als components rotats per a V14a, observem que, segons el gràfic de transformació de la variable edat, la variant *PC + tenir que / de + INF* es troba propera a la zona de 12 anys (-0,5) i la variable edat, així com també la variant *PC + poder + INF*, es troben a prop de l'àrea dels 16 anys (0,5). Quant a la dimensió 2, hi veiem que les variables llengua inicial i sexe es troben a l'àrea dels catalanoparlants inicials (entre 1 i 0,5) i que la darrera també se situa a la zona corresponent a les informants noies (0,5). Les variants amb les perífrasis d'obligació *haver de* i *tenir que / de* apareixen pròximes a l'àrea dels parlants bilingües inicials (vora 0) i la variant amb *anar a* se situa més aviat a prop de la zona dels nois (-0,5) i dels castellanoparlants inicials (entre -0,5 i -1). Cap dels angles entre els factors d'aquest gràfic sembla prou estret com per indicar una correlació destacable.

Figura 77. Gràfic de saturacions als components rotats per a la variable V14a.



6.2.11.2. Anàlisi de Components Principals de la posició enclítica (V14b)

La segona variable d'aquest grup és la que engloba les construccions perifràstiques que presenten els pronoms febles en posició enclítica (V14b). El recompte de pronoms enclítics amb la construcció amb *anar a* seguit d'infinitiu no va superar el mínim de realitzacions establert, per la qual cosa no es té en compte aquesta variant per a l'Anàlisi de Components Principals. En qualsevol cas, el primer component del model estadístic per a V14b explica un 23,8 % de la variància de les dades i el segon, més del 22,4 %, de manera que la variància explicada total del model arriba al 46,2 %.

Taula 81. Resum del model generat per l'Anàlisi de Components Principals per a V14b.

Resum model multivariant rotat ^a			
Dimensió	Alfa de Cronbach	Variància explicada	
		Total (valors propis)	% Variància
1	,362	1,431	23,857
2	,308	1,345	22,419
Total	,768 ^b	2,777	46,277

^a Mètode de rotació: Varimax amb Normalització Kaiser.

^b Alfa de Cronbach total basat en el total de valors propis.

La Taula 82 mostra que les variants de V14b no mantenen una correlació forta entre elles ni amb les variables independents, ja que el valor més elevat de la matriu de correlacions és la que opera entre el sexe i la llengua inicial dels informants, amb un 0,258, un valor que se situa lluny de 1.

Taula 82. Matriu de correlacions entre variables per a V14b.

	L1	Edat	Sexe	<i>tenir que</i> / <i>de</i> + INF + PC	<i>haver de</i> + INF + PC	<i>poder</i> + INF + PC
Llengua inicial	1,000	,000	,258	,193	,105	,095
Edat	,000	1,000	,000	-,119	,134	,197
Sexe	,258	,000	1,000	,161	,005	-,103
<i>tenir que</i> / <i>de</i> + INF + PC	,193	-,119	,161	1,000	,042	,006
<i>haver de</i> + INF + PC	,105	,134	,005	,042	1,000	,138
<i>poder</i> + INF + PC	,095	,197	-,103	,006	,138	1,000
Dimensió	1	2	3	4	5	6
Valors propis	1,433	1,343	,940	,865	,766	,652

Pel que fa a la matriu de saturacions als components rotats, la Taula 83 mostra que, en el cas de V14b, la llengua inicial, el sexe i la variant amb *tenir que* o *de* contribueixen més al primer component que al segon, a diferència de l'edat i la resta de variants, que no es mostren tan rellevants.

Taula 83. Matriu de saturacions als components rotats per a V14b.

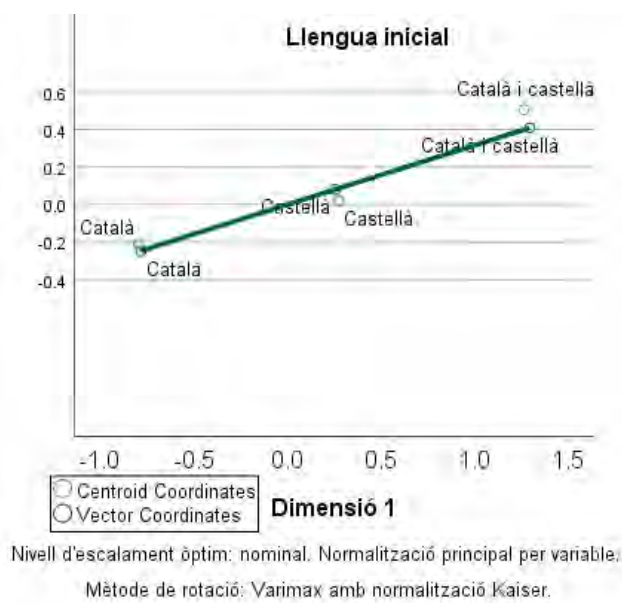
a	Dimensió	
	1	2
Llengua inicial	,716	,226
Edat	-,158	,661
Sexe	,678	-,098
<i>tenir que</i> / <i>de</i> + INF + PC	,636	-,084
<i>haver de</i> + INF + PC	,169	,588
<i>poder</i> + INF + PC	-,035	,703

Normalització principal per variable.

^a Mètode de rotació: Varimax amb normalització Kaiser. La rotació no ha convergit en 3 iteracions. (Convergència = ,000).

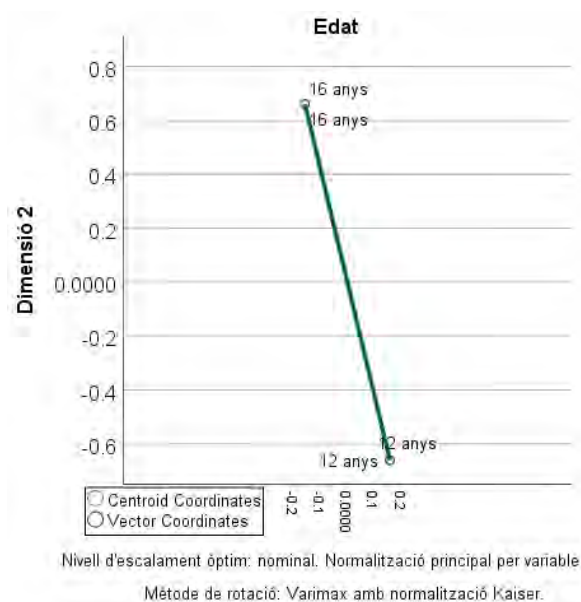
Així, el gràfic de transformació corresponent a la variable llengua inicial, mostra els grups d'informants a l'eix d'abscisses (dimensió 1). Aquesta vegada, el de catalanoparlants inicials apareix entre els valors de -1 i -0,5; el de castellanoparlants inicials, entre 0 i 0,5, i el de bilingües inicials, entre 1 i 1,5.

Figura 78. Gràfic de transformació de la variable independent llengua inicial a l'Anàlisi de Components Principals de V14b.



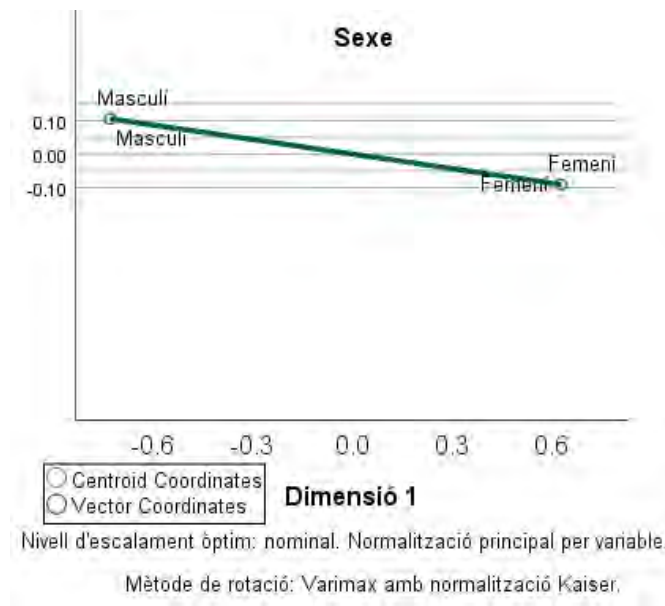
En canvi, al següent gràfic de transformació, la variable edat es distribueix al llarg de l'eix d'ordenades (dimensió 2): els 12 anys se situen prop de -0,6 i els 16, entre 0,6 i 0,8.

Figura 79. Gràfic de transformació de la variable independent edat a l'Anàlisi de Components Principals de V14b.



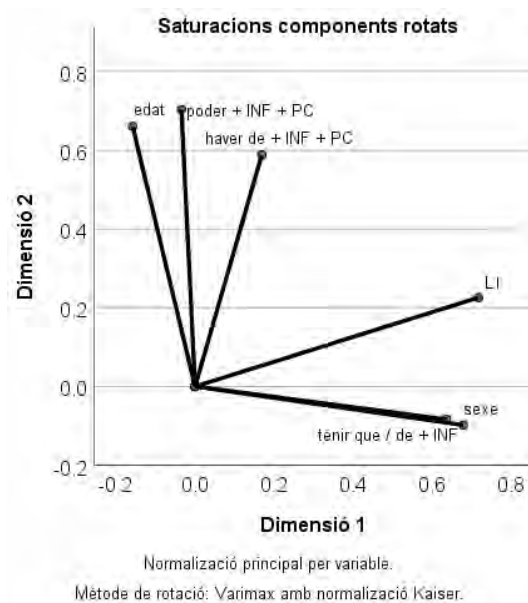
A l'últim gràfic de transformació per a V14b, s'hi mostren els grups de sexe a la dimensió 1. Els grups de nois es troben prop de -0,6 i els de noies, prop de 0,6.

Figura 80. Gràfic de transformació de la variable independent sexe a l'Anàlisi de Components Principals de V14b.



Tenint en compte aquesta informació, veiem al gràfic de saturacions als components rotats per a aquesta variable que la correlació més forta que s'hi representa es troba entre la variant *tenir que / de + INF + PC* i la variable *sexe*, que se situen a l'àrea de les noies a la dimensió 1 (vora 0,6). A més, també podria haver-hi certa correlació entre l'edat i la variant *poder + INF + PC*, que, juntament amb la variant *haver de + INF + PC*, es troben a l'àrea dels 16 anys a la dimensió 2. La informació anterior respecte els grups de llengua inicial no sembla rellevant per a la interpretació d'aquest gràfic.

Figura 81. Gràfic de saturacions als components rotats per a la variable V14b.



6.2.11.3. Anàlisi de Components Principals de la duplicació de pronoms clítics (V14c)

La tercera variable sobre la posició dels pronoms clítics en construccions perifràstiques contempla el total de realitzacions proclítics, enclítics i duplicades (en què es realitzen pronoms febles tant abans com després de les formes verbals). Com mostra la Taula 84, el model estadístic generat per l'ACP en aquest cas també supera el 40,0 % de variància explicada. Concretament, la primera dimensió arriba a explicar el 22,8 % de la variància de les dades i la segona, més del 21,1 %.

Taula 84. Resum del model generat per l'Anàlisi de Components Principals per a V14c.

Resum model multivariant rotat ^a			
Dimensió	Alfa de Cronbach	Variància explicada	
		Total (valors propis)	% Variància
1	,328	1,373	22,884
2	,258	1,270	21,166
Total	,746 ^b	2,643	44,050

^a Mètode de rotació: Varimax amb Normalització Kaiser.

^b Alfa de Cronbach total basat en el total de valors propis.

La matriu de correlacions per a V14 (Taula 85) no en mostra cap de destacada, ja que el valor més elevat és de 0,261 punts (entre les variants total de realitzacions enclítics i total de realitzacions duplicades).

Taula 85. Matriu de correlacions entre variables per a V14.

	L1	Edat	Sexe	Total proclisi	Total enclisi	Total doble
Llengua inicial	1,000	,000	,260	,035	,154	,024
Edat	,000	1,000	,000	,047	,098	,095
Sexe	,260	,000	1,000	,025	-,033	-,021
Total proclisi	,035	,047	,025	1,000	,082	,062
Total enclisi	,154	,098	-,033	,082	1,000	,261
Total doble	,024	,095	-,021	,062	,261	1,000
Dimensió	1	2	3	4	5	6
Valors propis	1,406	1,237	,969	,945	,793	,650

La majoria de factors de V14, com mostra la Taula 86, tenen un valor de saturació més elevat per al primer component que per al segon. Només la llengua inicial i el sexe dels informants tenen un pes més important al segon component, és a dir, resulten menys rellevants per diferenciar els informants.

Taula 86. Matriu de saturacions als components rotats per a V14c.

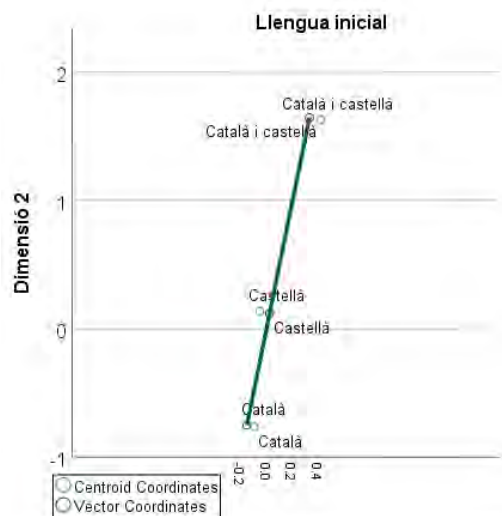
a	Dimensió	
	1	2
Llengua inicial	,160	,787
Edat	,428	-,074
Sexe	-,112	,787
Total proclisi	,327	,093
Total enclisi	,737	,107
Total doble	,708	-,069

Normalització principal per variable.

^a Mètode de rotació: Varimax amb normalització Kaiser. La rotació no ha convergit en 3 iteracions. (Convergència = ,000).

En consonància amb els valors de saturació de la variable llengua inicial, doncs, el gràfic de transformació d'aquesta variable per a V14 mostra els grups d'informants més distanciat a la dimensió 2 (eix d'ordenades). El de bilingües inicials se situa entre els valors d'1 i 2; el de castellanoparlants inicials, prop de 0, i el de catalanoparlants inicials, vora -1.

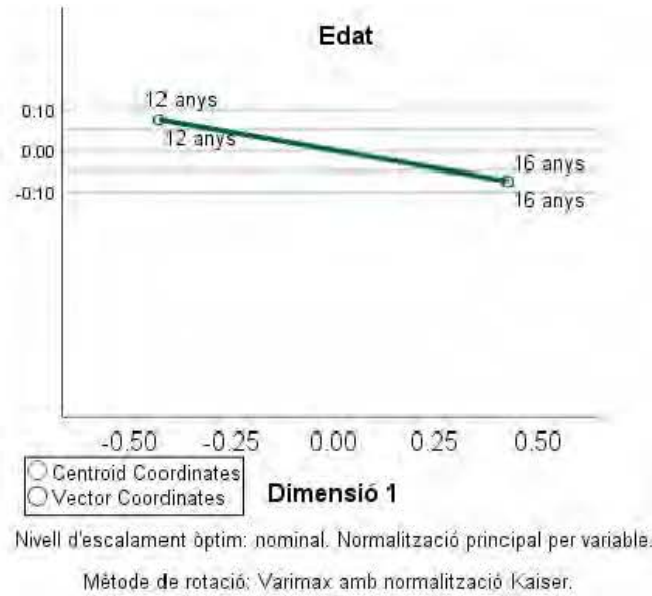
Figura 82. Gràfic de transformació de la variable independent llengua inicial a l'Anàlisi de Components Principals de V14.



Nivell d'escalament òptim: nominal. Normalització principal per variable.
Mètode de rotació: Varimax amb normalització Kaiser.

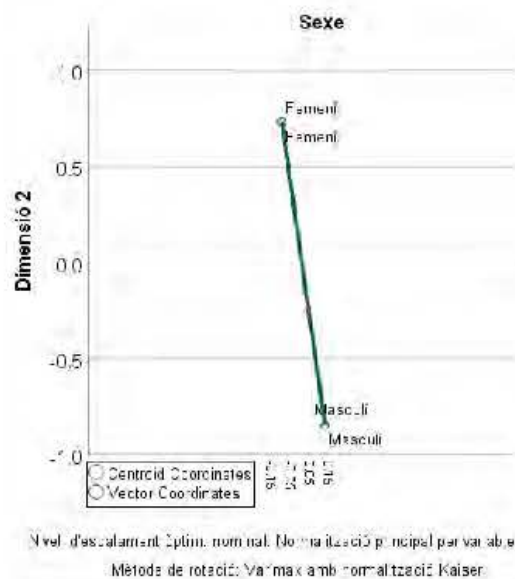
La variable edat, en canvi, es distribueix al llarg de la dimensió 1 (eix d'abscisses), ja que hi té més pes, com acabem de veure. El grup de 12 anys es projecta entre -0,5 i -0,25 i el de 16 anys, vora 0,5.

Figura 83. Gràfic de transformació de la variable independent edat a l'Anàlisi de Components Principals de V14.



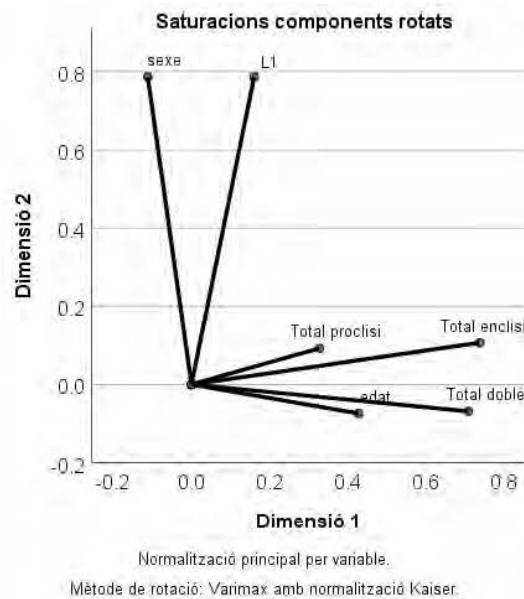
Finalment, els grups de la variable sexe es projecten a la dimensió 2: les noies, entre 1 i 0,5 i els nois, vora -1.

Figura 84. Gràfic de transformació de la variable independent sexe a l'Anàlisi de Components Principals de V14.



El gràfic de saturacions als components rotats per V14 (Figura 85) mostra, a la dimensió 2, en primer lloc, les variables sexe i llengua inicial a la zona assignada al gràfic anterior a les noies (entre 0,5 i 1) i, en segon lloc, la variable edat i la variant total de realitzacions duplicades prop de la zona assignada anteriorment als castellanoparlants inicials (0). A més, els angles entre les línies de les variables i variants mostren que les correlacions més destacables es troben entre les variants total de realitzacions proclítics i total de realitzacions enclítics i entre la variable edat i la variant total de realitzacions duplicades.

Figura 85. Gràfic de saturacions als components rotats per a la variable V14.



6.2.12. Anàlisi de Components Principals del tipus de preguntes (V17)

La següent variable morfosintàctica inclosa a l'estudi multivariant és el tipus de preguntes formulades pels informants (V17), que està formada per dues variants: preguntes obertes i preguntes tancades. Com recull la Taula 87, cadascuna de les dimensions del model estadístic generat per l'ACP en relació a aquesta variable aconsegueix explicar més del 25 % de la variància de les dades, de manera que, en conjunt, el model supera el 50,8 % de variància explicada i assoleix la quarta posició en relació a aquest paràmetre de tot l'estudi.

Taula 87. Resum del model generat per l'Anàlisi de Components Principals per a V17.

Resum model multivariant rotat ^a			
Dimensió	Alfa de Cronbach	Variància explicada	
		Total (valors propis)	% Variància
1	,271	1,275	25,499
2	,266	1,269	25,377
Total	,759 ^b	2,544	50,876

^a Mètode de rotació: Varimax amb Normalització Kaiser.

^b Alfa de Cronbach total basat en el total de valors propis.

La matriu de correlacions per a V17 (Taula 88) segueix la tendència de les que s'han observat fins ara en el sentit que no mostra cap valor de correlació particularment proper a 1. Això no obstant, crida l'atenció que el valor més elevat de la matriu (0,245) es trobi entre les dues variants d'aquesta variable, la qual cosa apunta que els informants que formulen preguntes d'un tipus també tendeixen, en certa mesura, a formular-ne de l'altre.

Taula 88. Matriu de correlacions entre variables per a V17.

	L1	Edat	Sexe	Oberta	Tancada
Llengua inicial	1,000	,000	,239	,022	,052
Edat	,000	1,000	,000	-,118	,120
Sexe	,239	,000	1,000	-,180	,024
Oberta	,022	-,118	-,180	1,000	,245
Tancada	,052	,120	,024	,245	1,000
Dimensió	1	2	3	4	5
Valors propis	1,315	1,229	1,075	,777	,604

Pel que fa a la matriu de saturacions als components rotats, en aquesta ocasió veiem a la Taula 89 que els dos únics factors que tenen més pes a la primera dimensió són precisament les dues variants de V17, mentre que les variables independents mostren valors més destacats a la segona dimensió. Així doncs, conjuntament, les variants de V17 permeten diferenciar millor els informants que la llengua inicial, l'edat i el sexe.

Taula 89. Matriu de saturacions als components rotats per a V17.

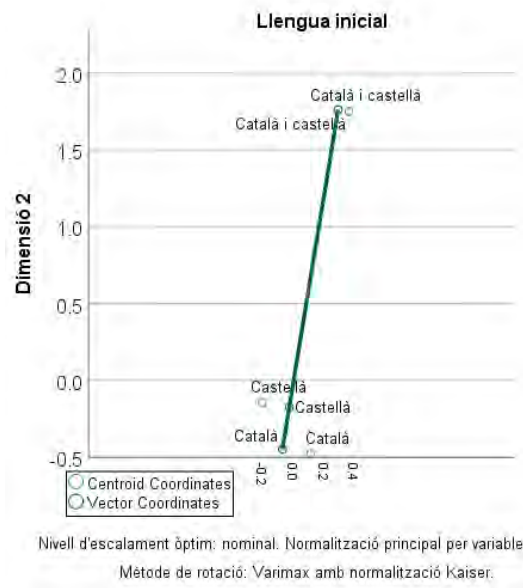
a	Dimensió	
	1	2
Llengua inicial	,121	,736
Edat	,028	,198
Sexe	-,227	,769
Oberta	,793	-,190
Tancada	,761	,247

Normalització principal per variable.

^a Mètode de rotació: Varimax amb normalització Kaiser. La rotació no ha convergit en 3 iteracions. (Convergència = ,000).

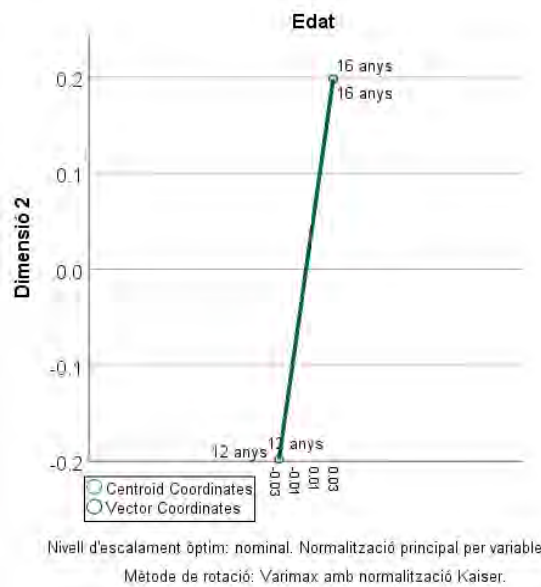
Per això, els tres gràfics de transformació de les variables que es presenten a continuació projecten els diversos grups de parlants sobre l'eix d'ordenades (component 2). El primer mostra que els bilingües inicials (situats entre 1,5 i 2) es distancien considerablement dels monolingües inicials. Els castellanoparlants inicials apareixen entre 0 i -0,5 (tot i que més aviat a prop de 0) i els catalanoparlants inicials, vora -0,5.

Figura 86. Gràfic de transformació de la variable independent llengua inicial a l'Anàlisi de Components Principals de V17.



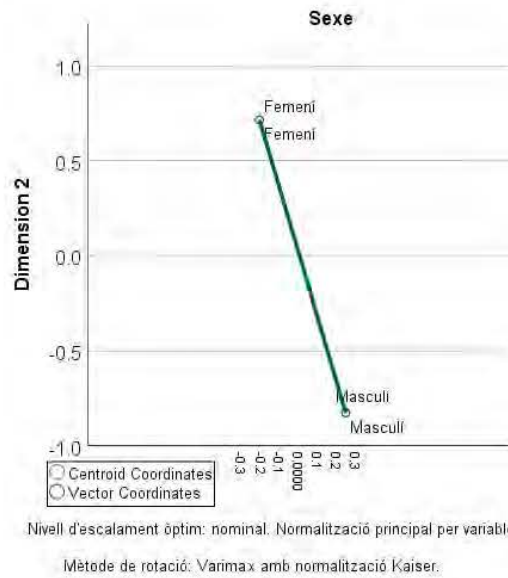
El segon gràfic de transformació situa l'edat de 16 anys al valor de 2 i l'edat de 12 anys al de -0,2, tal com s'observa a la Figura 87.

Figura 87. Gràfic de transformació de la variable independent edat a l'Anàlisi de Components Principals de V17.



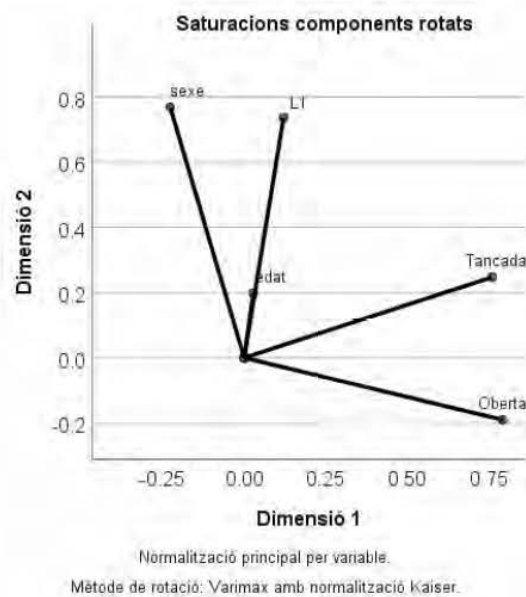
El tercer i últim gràfic de transformació per a la variable V17 indica que les noies se situen entre 0,5 i 1, mentre que els nois es troben entre aquests mateixos valors, però de la meitat negativa del gràfic.

Figura 88. Gràfic de transformació de la variable independent sexe a l'Anàlisi de Components Principals de V17.



Finalment, tenint en compte la informació que s'ha tractat fins ara, és possible observar al gràfic de saturacions als components rotats per a aquesta variable que les variables sexe i llengua inicial es troben a l'àrea assignada a les noies (entre 0,5 i 1); que la variant preguntes tancades i la variable edat se situen a la zona dels castellanoparlants (entre 0 i 0,5) i la variant preguntes obertes, a la de 12 anys (al 0,2). A més, els angles entre les línies de cada factor indiquen que la correlació més destacada s'estableix entre les variables llengua inicial i edat.

Figura 89. Gràfic de saturacions als components rotats per a la variable V17.



6.2.13. Anàlisi de Components Principals de les construccions condicionals (V18)

Quant a les construccions condicionals (V18), el model generat per l'ACP aconsegueix explicar més del 40,6 % de la variància de les dades, ja que el primer component supera el 21,5 % i el segon arriba al 18,9 %, com s'observa a la Taula 90.

Taula 90. Resum del model generat per l'Anàlisi de Components Principals per a V18.

Resum model multivariant rotat ^a			
Dimensió	Alfa de Cronbach	Variància explicada	
		Total (valors propis)	% Variància
1	,396	1,514	21,632
2	,289	1,329	18,980
Total	,756 ^b	2,843	40,612

^a Mètode de rotació: Varimax amb Normalització Kaiser.

^b Alfa de Cronbach total basat en el total de valors propis.

Tot seguit, la Taula 91 mostra la matriu de correlacions entre les variables obtinguda per a V18. Els valors de correlació més elevats, de nou, se situen al voltant de (-)0,2, de manera que disten força de ser absoluts. Tot i això, les correlacions més destacades operen entre *si* i la llengua inicial (0,271), d'una banda, i *si* i *a no ser que* (0,232), de l'altra.

Taula 91. Matriu de correlacions entre variables per a V18.

	L1	Edat	Sexe	<i>si</i>	<i>mentre(s)</i>	<i>a no ser que</i>	<i>en (tot) cas</i>
Llengua inicial	1,000	,000	,050	,271	-,210	,134	,044
Edat	,000	1,000	,000	,093	,051	-,016	-,029
Sexe	,050	,000	1,000	-,138	-,209	-,099	,014
<i>si</i>	,271	,093	-,138	1,000	,085	,232	,167
<i>mentre(s)</i>	-,210	,051	-,209	,085	1,000	-,032	,009
<i>a no ser que</i>	,134	-,016	-,099	,232	-,032	1,000	,081
<i>en (tot) cas</i>	,044	-,029	,014	,167	,009	,081	1,000
Dimensió	1	2	3	4	5	6	7
Valors propis	1,518	1,325	1,028	,975	,826	,742	,586

Pel que fa a la composició dels components principals per a V18, la matriu de saturacions als components rotats (Taula 92) mostra que la llengua inicial i les variants *si*, *a no ser que* i *en (tot) cas* contribueixen més a la primera dimensió que no pas a la segona. En canvi, la variant *mentre(s)* i les variables edat i sexe tenen més pes al segon component (no es mostren tan rellevants per diferenciar els informants).

Taula 92. Matriu de saturacions als components rotats per a V18.

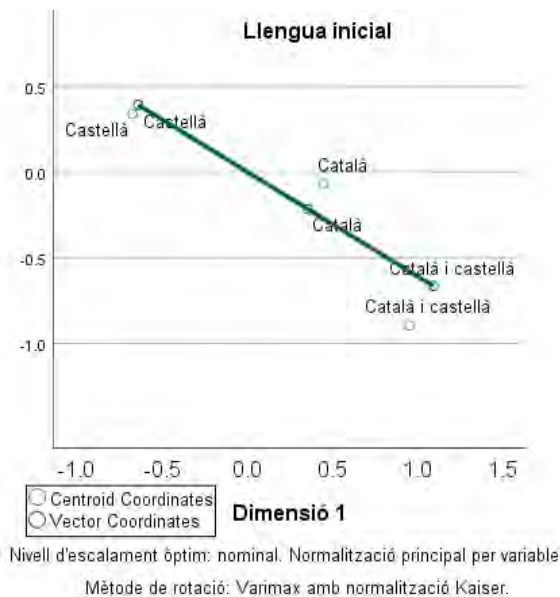
a	Dimensió	
	1	2
Llengua inicial	,642	-,392
Edat	,072	,176
Sexe	-,150	-,680
<i>si</i>	,747	,254
<i>mentre(s)</i>	-,140	,777
<i>a no ser que</i>	,595	,106
<i>en (tot) cas</i>	,378	,049

Normalització principal per variable.

^a Mètode de rotació: Varimax amb normalització Kaiser. La rotació no ha convergit en 3 iteracions. (Convergència = ,000).

Així doncs, el gràfic de transformació de la llengua inicial mostra, a l'eix corresponent a la dimensió 1, que els castellanoparlants inicials se situen prop de -0,5, els catalanoparlants inicials, vora 0,5 i els bilingües inicials, al voltant d'1.

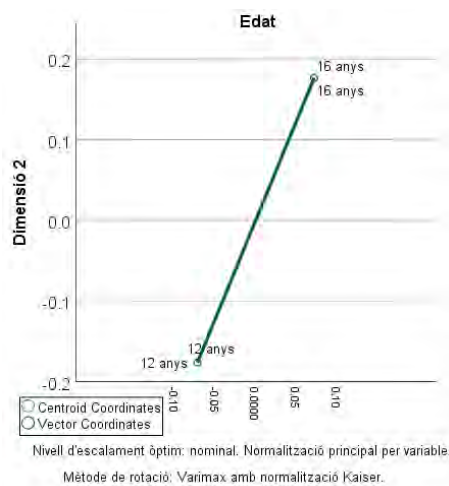
Figura 90. Gràfic de transformació de la variable independent llengua inicial a l'Anàlisi de Components Principals de V18.



Resultats

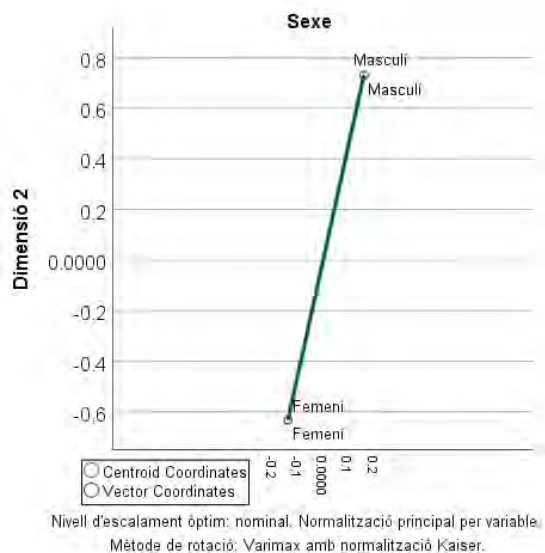
El gràfic de transformació de la variable edat, en canvi, pren com a referència l'eix corresponent a la dimensió 2. Mostra que els 12 anys es representen vora -0,2 i els 16 anys, prop de 0,2.

Figura 91. Gràfic de transformació de la variable independent edat a l'Anàlisi de Components Principals de V18.



El tercer gràfic de transformació, que mostra el sexe dels informants, situa les noies aproximadament a -0,6 i els nois prop de 0,8.

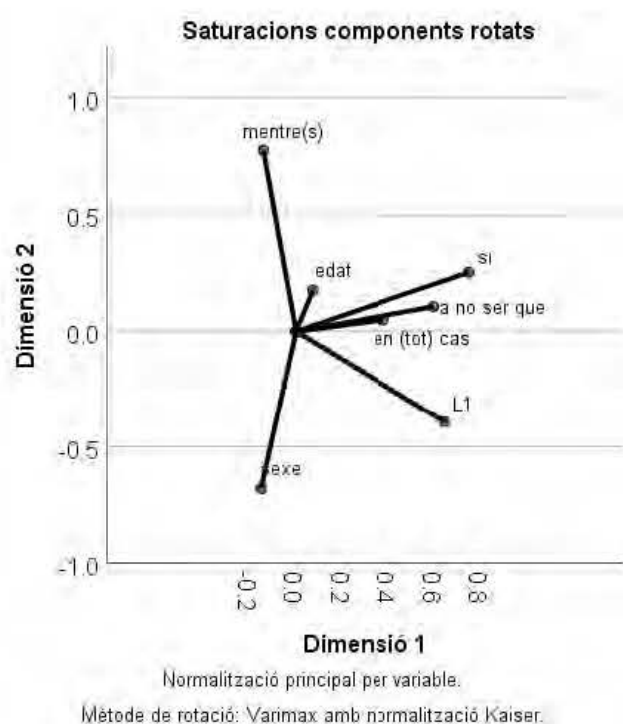
Figura 92. Gràfic de transformació de la variable independent sexe a l'Anàlisi de Components Principals de V18.



Tenint en compte aquestes dades, el gràfic de saturacions als components rotats mostra el següent. Respecte la dimensió 2, primer, que la variant *mentre(s)* se situa a la zona dels informants nois (0,8); segon, que la variable edat i les variants *a no ser que* i *en (tot) cas* s'aproximen a l'àrea dels 16 anys (0,2); tercer,

que la llengua inicial s'apropa a la zona de 12 anys (-0,2), i, quart, que la variable sexe es troba a la de les noies (-0,6). Pel que fa a la dimensió 1, hi podem observar que les variants *si* i *a no ser que*, així com la variable llengua inicial, apareixen entre els grups de catalanoparlants (0,5) i bilingües (1) inicials. Quant a la proximitat de les línies que representen cada factor, la Figura 93 mostra una forta correlació entre les variants *a no ser que* i *en (tot) cas*. A més aquestes variants (en especial, la primera) es mostren força properes a *si*. També és possible que *mentre(s)* i la variable edat mostrin certa correlació.

Figura 93. Gràfic de saturacions als components rotats per a la variable V18.



6.2.14. Anàlisi de Components Principals de les perífrasis d'obligació (V22)

La variable següent considera quatre perífrasis d'obligació presents al corpus (V22). El model bidimensional generat per l'ACP per a aquesta variable explica més del 40,4 % de la variància de les dades, com mostra la Taula 93. Específicament, el primer component assoleix la capacitat d'explicar més del 21,5 % de la variància i el segon, més del 18,8 %.

Taula 93. Resum del model generat per l'Anàlisi de Components Principals per a V22.

Resum model multivariant rotat ^a			
Dimensió	Alfa de Cronbach	Variància explicada	
		Total (valors propis)	% Variància
1	,401	1,510	21,574
2	,295	1,323	18,898
Total	,755 ^b	2,833	40,473

^a Mètode de rotació: Varimax amb Normalització Kaiser.

^b Alfa de Cronbach total basat en el total de valors propis.

Pel que fa a les correlacions entre les variables, resumides a la Taula 94, s'hi observa que la relació més forta és la que uneix les variants *haver de* i *tenir que* (-0,354).

Taula 94. Matriu de correlacions entre variables per a V22.

	L1	Edat	Sexe	<i>tenir que</i>	<i>haver de</i>	<i>tenir de</i>	<i>caldre</i>
Llengua inicial	1,000	,000	,259	,162	-,101	-,023	,061
Edat	,000	1,000	,000	-,225	,158	-,012	,037
Sexe	,259	,000	1,000	,123	,056	,043	-,070
<i>tenir que</i>	,162	-,225	,123	1,000	-,354	-,086	,014
<i>haver de</i>	-,101	,158	,056	-,354	1,000	-,172	,075
<i>tenir de</i>	-,023	-,012	,043	-,086	-,172	1,000	-,101
<i>caldre</i>	,061	,037	-,070	,014	,075	-,101	1,000
Dimensió	1	2	3	4	5	6	7
Valors propis	1,590	1,243	1,147	,954	,844	,706	,516

Per la seva banda, la matriu de saturacions als components rotats per a aquesta variable indica que tan sols la llengua inicial i el sexe tenen més pes a la segona dimensió. Totes les variants de V22, juntament amb la variable edat,

contribueixen més a la primera dimensió, és a dir, són rellevants per explicar les dades de V22.

Taula 95. Matriu de saturacions als components rotats per a V22.

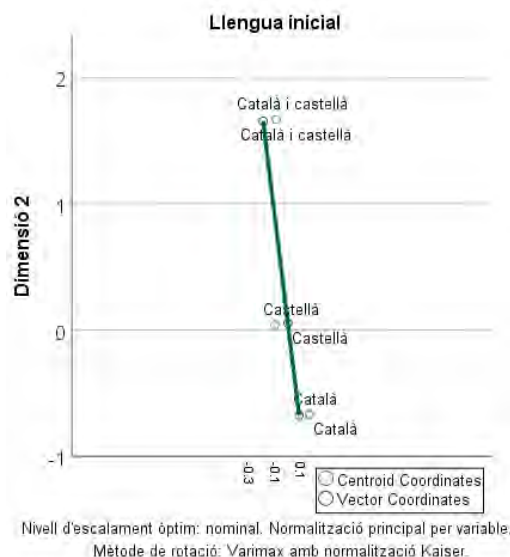
a	Dimensió	
	1	2
Llengua inicial	-,093	,759
Edat	,564	,031
Sexe	,023	,693
<i>tenir que</i>	-,652	,387
<i>haver de</i>	,786	,021
<i>tenir de</i>	-,286	-,270
<i>caldre</i>	,239	,207

Normalització principal per variable.

^a Mètode de rotació: Varimax amb normalització Kaiser. La rotació no ha convergit en 3 iteracions. (Convergència = ,000).

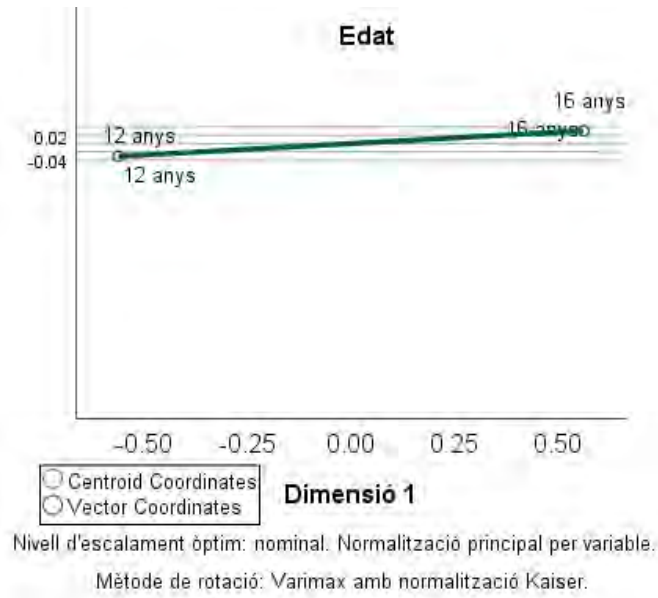
Per tant, el gràfic de transformació de la llengua inicial ubica els tres grups en què es divideix el corpus respecte a aquesta variable al llarg de l'eix d'ordenades (que representa el segon component). Ho fa de la manera següent: els catalanoparlants inicials se situen prop de -1; els castellanoparlants inicials, vora 0, i els bilingües inicials, entre 1 i 2, tot i que més aviat prop del darrer.

Figura 94. Gràfic de transformació de la variable independent llengua inicial a l'Anàlisi de Components Principals de V22.



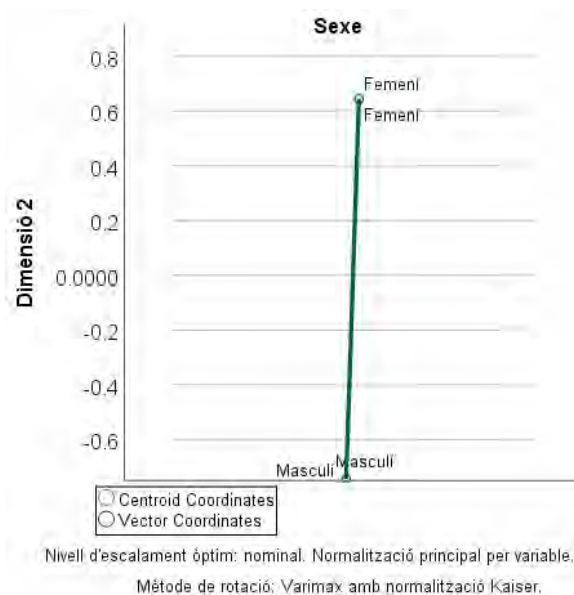
En canvi, el gràfic de transformació de l'edat, utilitza principalment l'eix d'abscisses (visualització del primer component) i mostra que l'edat de 12 anys es projecta cap a -0,5 i la de 16 anys, cap a 0,5.

Figura 95. Gràfic de transformació de la variable independent edat a l'Anàlisi de Components Principals de V22.



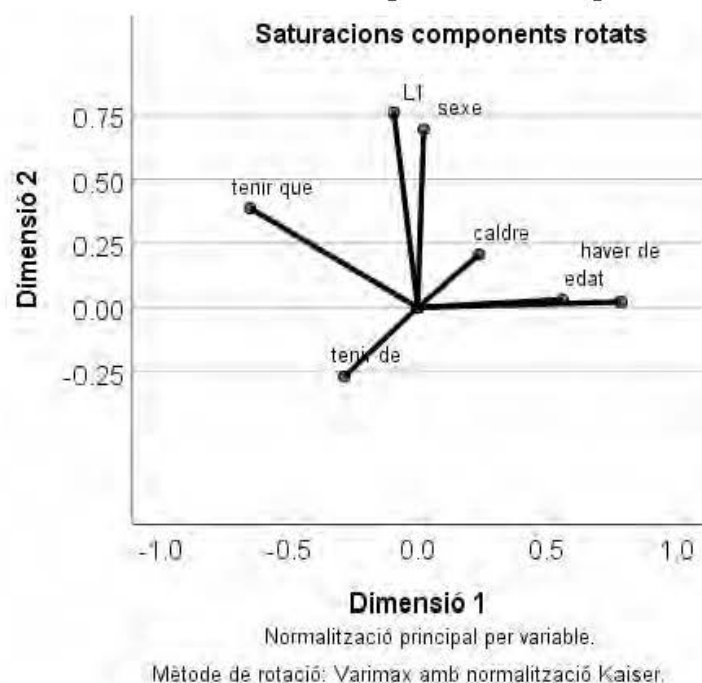
A continuació, la Figura 96 mostra el darrer gràfic de transformació, que es correspon amb la variable sexe. A la dimensió 2, s'hi pot observar que els nois es projecten vora -0,8 (o, en tot cas, més enllà de -0,6), mentre que les noies se situen just per sobre de 0,6.

Figura 96. Gràfic de transformació de la variable independent sexe a l'Anàlisi de Components Principals de V22.



Amb les informacions anteriors, podem veure al gràfic de saturacions als components rotats per a V22 que l'àrea de l'eix d'ordenades en què es projecten la llengua inicial i el sexe s'aproxima a la de les noies (aproximadament 0,7) i que l'àrea en què apareixen l'edat i *haver de* (0) és la que s'ha assignat anteriorment als castellanoparlants inicials. A l'eix d'abscisses, hi veiem que *tenir de* apareix proper a l'àrea de 12 anys (-0,5) i l'edat, propera als 16 anys (0,5). Pel que fa a l'estretor de les correlacions entre les variables, la més destacada és la correlació entre l'edat i *haver de*, seguida de la que uneix la llengua inicial i el sexe dels informants.

Figura 97. Gràfic de saturacions als components rotats per a la variable V22.



6.2.15. Anàlisi de Components Principals dels determinants demostratius (V23)

La penúltima variable morfosintàctica de l'estudi multivariant és la que contempla diverses formes dels determinants demostratius (V23). En aquest cas, el model assoleix una variància explicada de prop del 30 %. El primer component arriba a explicar més del 16,7 % i el segon, més del 13,1 % de la variància de les dades.

Taula 96. Resum del model generat per l'Anàlisi de Components Principals per a V23.

Resum model multivariant rotat ^a			
Dimensió	Alfa de Cronbach	Variància explicada	
		Total (valors propis)	% Variància
1	,504	1,840	16,729
2	,343	1,448	13,167
Total	,766 ^b	3,289	29,896

^a Mètode de rotació: Varimax amb Normalització Kaiser.

^b Alfa de Cronbach total basat en el total de valors propis.

Quant a la matriu de correlacions de la variable V23, les dues més destacades es troben entre variants que presenten el demostratiu de distància: *aquell* i *el + N + aquell* (0,371) i *un + N + d'aquells* i *el + N + aquell* (0,348).

Taula 97. Matriu de correlacions entre variables per a V23.

	L1	Edat	Sexe	V23a	V23b	V23c	V23d	V23e	V23f	V23g	V23h
L1	1	0	0,2	-0,2	-0,2	0	-0,2	0,1	-0,1	0,2	-0,1
Edat	0	1	0	0,2	0,1	0	0	-0,1	0	0	0
Sexe	0,2	0	1	-0,1	0	-0,1	0	0	-0,1	0	-0,1
V23a	-0,2	0,2	-0,1	1	0,1	0,1	0	0,2	0	0	-0,1
V23b	-0,2	0,1	0	0,1	1	-0,1	0,2	0,1	0,2	-0,1	0,1
V23c	0	0	-0,1	0,1	-0,1	1	0,1	0,1	0	0,2	-0,1
V23d	-0,2	0	0	0	0,2	0,1	1	0,1	0,4	0,1	0,4
V23e	0,1	-0,1	0	0,2	0,1	0,1	0,1	1	0,1	0,1	-0,1
V23f	-0,1	0	0	0	0,2	0	0,4	0,1	1	0	0,2
V23g	0,2	0	0	0	-0,1	0,2	0,1	0,1	0	1	0

V23h	-0,1	0	-0,1	-0,1	0,1	-0,1	0,4	-0,1	0,2	0	1
D	1	2	3	4	5	6	7	8	9	10	11
V. P.	1,9	1,4	1,4	1,2	1	0,9	0,8	0,7	0,7	0,6	0,5
V23a: <i>aquest</i> + N; V23b: <i>aquell</i> + N; V23c: <i>el</i> + N + <i>aquest</i> ; V23d: <i>el</i> + N + <i>aquell</i> ; V23e: <i>aquest</i> ; V23f: <i>aquell</i> ; V23g: (<i>un</i> +) N + <i>d'aquests</i> ; V23h: (<i>un</i> +) N + <i>d'aquells</i> D: Dimensió; V. P.: Valors propis Valors arrodonits per qüestions d'espai.											

La matriu de saturacions als components rotats mostra que quatre de les variants de V23, les que presenten el demostratiu de distància (*aquell*), i la variable llengua inicial contribueixen més al primer component i, per tant, són més rellevants que la resta per diferenciar els informants. En contrapartida, les variants que inclouen el demostratiu de proximitat (*aquest*) i les variables edat i sexe tenen més pes al segon component del model.

Taula 98. Matriu de saturacions als components rotats per a V23.

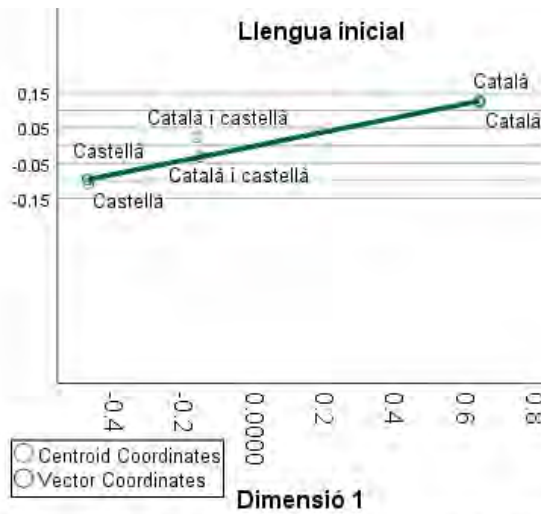
a	Dimensió	
	1	2
Llengua inicial	-,519	-,104
Edat	,070	,202
Sexe	-,172	-,362
<i>aquest</i> + N	,034	,527
<i>aquell</i> + N	,469	-,072
<i>el</i> + N + <i>aquest</i>	-,094	,683
<i>el</i> + N + <i>aquell</i>	,708	,243
<i>aquest</i>	,025	,438
<i>aquell</i>	,639	,046
(<i>un</i>) + N + <i>d'aquests</i>	-,247	,507
<i>un</i> + N + <i>d'aquells</i>	,578	-,080

Normalització principal per variable.

^a Mètode de rotació: Varimax amb normalització Kaiser. La rotació no ha convergint en 3 iteracions. (Convergència = ,000).

A la dimensió 1, per tant, del primer gràfic de transformació, s'hi pot observar la distribució dels grups d'informants segons la seva llengua inicial. Com mostra la Figura 98, els castellanoparlants inicials es projecten aproximadament a -0,5; els bilingües inicials vora -0,1, i els catalanoparlants, cap a 0,7, a una distància considerable dels dos anteriors.

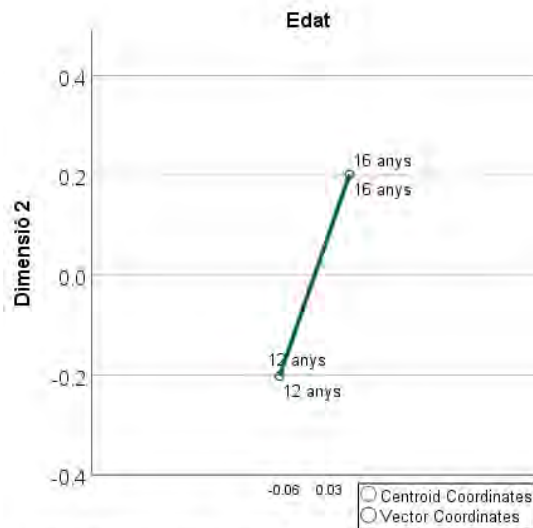
Figura 98. Gràfic de transformació de la variable independent llengua inicial a l'Anàlisi de Components Principals de V23.



Nivell d'escalament òptim: nominal. Normalització principal per variable.
Mètode de rotació: Varimax amb normalització Kaiser.

Al segon gràfic de transformació s'hi observa, a la dimensió 2, la distribució de les dues edats per a les quals tenim dades dels informants. Els 12 anys es mostren al valor de -0,2 i els 16 anys, al de 0,2.

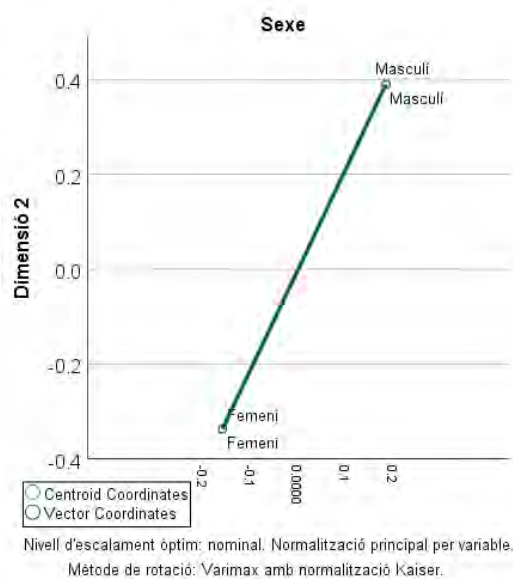
Figura 99. Gràfic de transformació de la variable independent edat a l'Anàlisi de Components Principals de V23.



Nivell d'escalament òptim: nominal. Normalització principal per variable.
Mètode de rotació: Varimax amb normalització Kaiser.

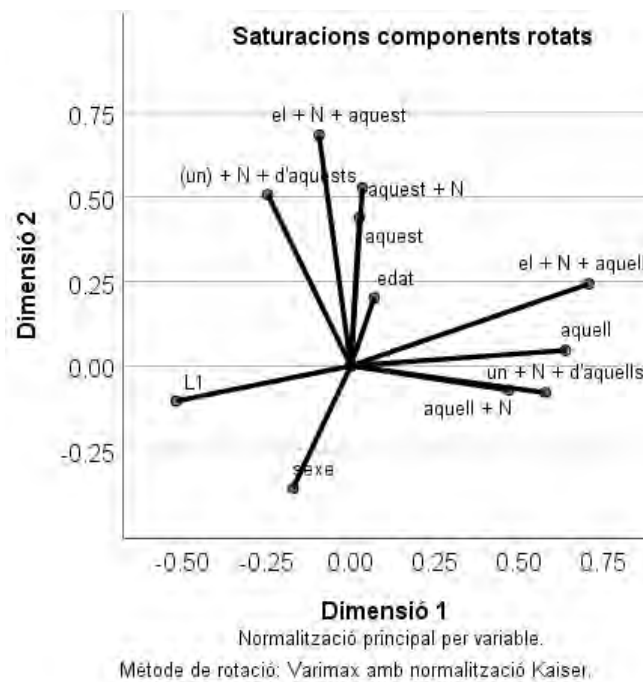
La variable sexe també es projecta al llarg de la dimensió 2, com mostra el tercer gràfic de transformació per a la variable V23. Els nois s'hi projecten prop de 0,4, mentre que les noies hi apareixen vora el valor de -0,4.

Figura 100. Gràfic de transformació de la variable independent sexe a l'Anàlisi de Components Principals de V23.



Si recollim la informació anterior sobre les relacions entre les variables independents i V23 al gràfic de saturacions als components rotats, hi podem veure, fixant-nos en primer lloc en la dimensió 1, que a l'àrea assignada als castellanoparlants inicials (-0,5) s'hi projecta la variable llengua inicial, a l'assignada als bilingües inicials (-0,1) hi trobem la variant *el + N + aquell* i, a la dels catalanoparlants inicials (0,7), les variants *aquell* i *el + N + aquell*. Pel que fa a la dimensió 2, aquesta darrera variant s'aproxima a l'àrea dels 16 anys (0,2), on també s'hi situa la representació de la variable edat, mentre que a la zona dels parlants nois (0,4) hi trobem la variant *aquest*. Finalment, els angles entre els factors que apareixen al gràfic ens indiquen que les correlacions més fortes operen, d'una banda, entre les variants *aquest + N* i *aquest* i, de l'altra, entre *un + N + d'aquells* i *aquell + N*. També s'observen certes correlacions entre les variants *el + N + aquest* i *aquest + N*, la primera i *aquest*, *aquest* i la variable edat, *aquell* i *un + N + d'aquells* i *aquell* i *aquell + N*. En menor mesura, també pot ser que existeixin certes correlacions entre (*un*) + *N + d'aquests* i *el + N + aquest* i entre *el + N + aquell* i *aquell*. És possible que la gran quantitat de correlacions aparents tingui relació amb la proximitat semàntica i funcional (com a elements referencials) entre aquestes variants.

Figura 101. Gràfic de saturacions als components rotats per a la variable V23.



6.2.16. Anàlisi de Components Principals dels pronoms clítics de datiu plural (V25)

La darrera variable morfosintàctica de l'estudi multivariant és la que recull les realitzacions dels pronoms clítics que funcionen de datiu plural (V25). La Taula 99 mostra el percentatge de variància de les dades que arriba a explicar el model estadístic obtingut mitjançant l'ACP, el 45 %. El primer component aporta la capacitat d'explicació de més del 25,3 % de la variància i el segon, de més del 19,6 %.

Taula 99. Resum del model generat per l'Anàlisi de Components Principals per a V25.

Resum model multivariant rotat ^a			
Dimensió	Alfa de Cronbach	Variància explicada	
		Total (valors propis)	% Variància
1	,516	1,777	25,389
2	,329	1,374	19,630
Total	,796 ^b	3,151	45,019

^a Mètode de rotació: Varimax amb Normalització Kaiser.

^b Alfa de Cronbach total basat en el total de valors propis.

Pel que fa al pes de les correlacions entre les variables en el cas de V25, com mostra la Taula 100, els valors més elevats es troben al voltant de 0,3: en primer lloc, se situa la correlació entre els pronoms *els* i *lis* (amb precisament aquest valor) i, en segon, la que existeix entre *li* i *els* (amb un 0,294).

Taula 100. Matriu de correlacions entre variables per a V25.

	L1	Edat	Sexe	<i>els hi + V / V + -els-hi / V + 'ls-hi / V + -los-hi</i>	<i>els + V / V + -los / V + 'ls</i>	<i>li + V / V + -li</i>	<i>lis + V / V + -lis</i>
Llengua inicial	1,000	,000	,207	-,154	,251	,144	,243
Edat	,000	1,000	,000	,058	,010	,134	-,114
Sexe	,207	,000	1,000	-,016	,131	,120	,130
<i>els hi + V / V + -els-hi / V + 'ls-hi / V + -los-hi</i>	-,154	,058	-,016	1,000	,148	,245	-,067
<i>els + V / V + -los / V + 'ls</i>	,251	,010	,131	,148	1,000	,294	,300
<i>li + V / V + -li</i>	,144	,134	,120	,245	,294	1,000	,199
<i>lis + V / V + -lis</i>	,243	-,114	,130	-,067	,300	,199	1,000
Dimensió	1	2	3	4	5	6	7
Valors propis	1,834	1,318	1,019	,885	,707	,659	,578

Tot seguit, la matriu de saturacions als components rotats indica que dues de les variants de V25 (*els* i *lis*), així com la llengua inicial i el sexe dels informants contribueixen principalment a la primera dimensió, a diferència de les altres dues variants (*els hi* i *li*) i l'edat, que resulten menys rellevants.

Taula 101. Matriu de saturacions als components rotats per a V25.

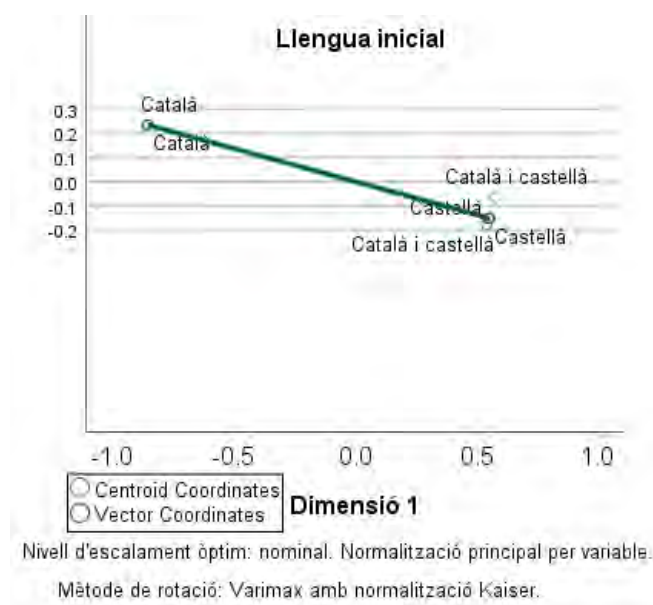
a	Dimensió	
	1	2
Llengua inicial	,687	-,188
Edat	-,118	,433
Sexe	,481	-,037
<i>els hi + V / V + -els-hi / V + 'ls-hi / V + -los-hi</i>	-,122	,774
<i>els + V / V + -los / V + 'ls</i>	,628	,362
<i>li + V / V + -li</i>	,423	,644
<i>lis + V / V + -lis</i>	,687	-,067

Normalització principal per variable.

^a Mètode de rotació: Varimax amb normalització Kaiser. La rotació no ha convergit en 3 iteracions. (Convergència = ,000).

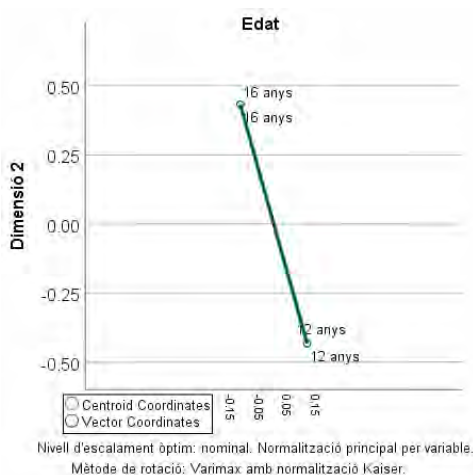
Com s'observa a la Figura 102, el gràfic de transformació de la variable llengua inicial mostra que els catalanoparlants es projecten vora -1 a la primera dimensió i els castellanoparlants i els bilingües inicials es concentren al voltant de 0,5.

Figura 102. Gràfic de transformació de la variable independent llengua inicial a l'Anàlisi de Components Principals de V25.



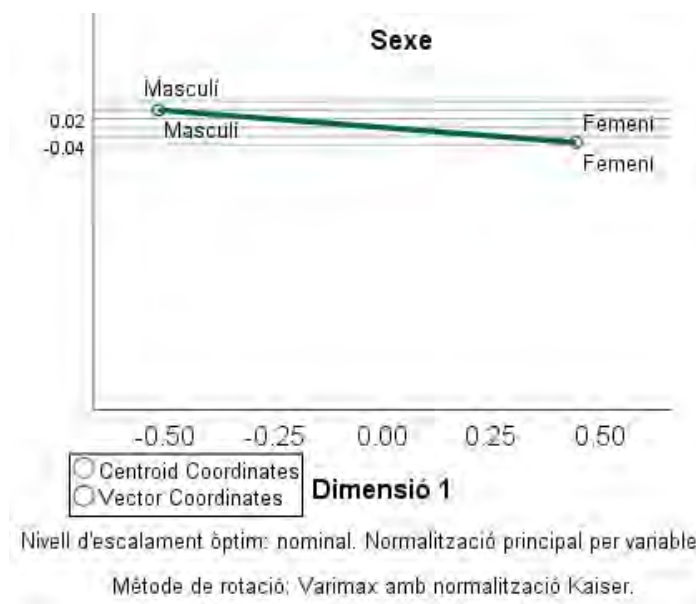
Quant a les dues edats en què es van recollir dades dels informants, el gràfic de transformació de la variable edat mostra que la representació dels 12 anys es troba prop de -0,5 a la segona dimensió, mentre que la dels 16 anys s'ubica prop de 0,5.

Figura 103. Gràfic de transformació de la variable independent edat a l'Anàlisi de Components Principals de V25.



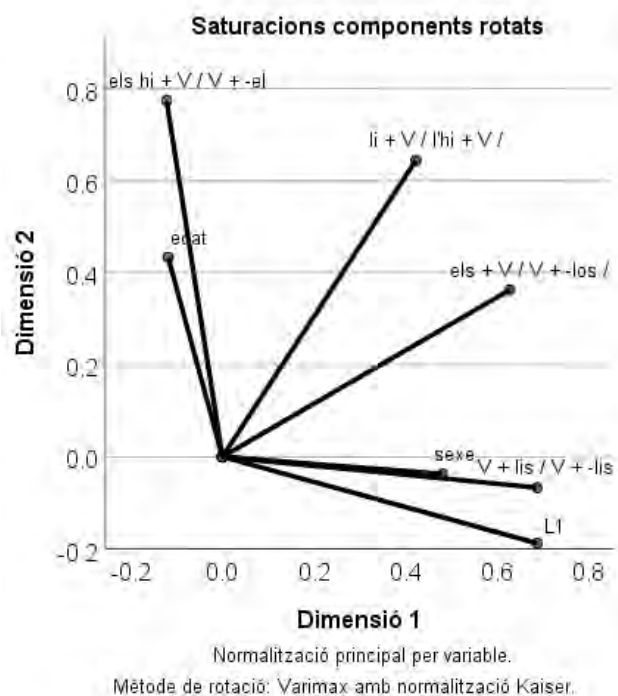
I pel que fa a la variable sexe, el gràfic de transformació d'aquesta variable per a V25 (Figura 104) mostra, de nou a la primera dimensió, que els nois es projecten a -0,5 i les noies a 0,5.

Figura 104. Gràfic de transformació de la variable independent sexe a l'Anàlisi de Components Principals de V25.



Tenint en compte les observacions sobre V25 realitzades fins ara, si observem el gràfic de saturacions als components rotats generat per l'ACP, hi podrem veure, en primer lloc, que la variable sexe se situa aproximadament sobre 0,5 de l'eix d'abscisses (component 1), valor associat anteriorment als castellanoparlants inicials, als bilingües inicials i a les noies i que, a l'eix d'ordenades (component 2) el factor que més s'aproxima a alguna de les zones descrites als gràfics de transformació és l'edat, que apareix prop de la ubicació dels 16 anys (0,5). En segon lloc, també s'observa en aquest últim gràfic relatiu a V25 que les correlacions més destacades, representades per poca distància entre les línies dels factors, es donen entre la variable sexe i la variant amb el pronom *lis*, entre la variable edat i la variant amb *els hi*, i entre *lis* i la variable llengua inicial.

Figura 105. Gràfic de saturacions als components rotats per a la variable V25.



6.2.17. Anàlisi de Components Principals dels fenòmens de contacte de llengües (V38)

A continuació abordem els resultats de l'ACP per a les dues variables lèxiques de l'estudi multivariant, començant pels fenòmens de contacte de llengües (V38). El primer component del model generat per a aquesta variable explica pràcticament el 19,0 % de la variància de les dades, a poca distància del segon component, que n'explica més del 18,2 %. D'aquesta manera, el model explica, globalment, més del 37,2 % de la variància.

Taula 102. Resum del model generat per l'Anàlisi de Components Principals per a V38.

Resum model multivariant rotat ^a			
Dimensió	Alfa de Cronbach	Variància explicada	
		Total (valors propis)	% Variància
1	,290	1,330	18,999
2	,253	1,277	18,245
Total	,719 ^b	2,607	37,244

^a Mètode de rotació: Varimax amb Normalització Kaiser.

^b Alfa de Cronbach total basat en el total de valors propis.

Pel que fa al pes de les correlacions entre les variants i les variables, com es pot observar a la Taula 103, cap dels valors s'acosta a l'1. La correlació més forta, com en ocasions anteriors, s'atribueix al sexe i la llengua inicial (0,26).

Taula 103. Matriu de correlacions entre variables per a V38.

	L1	Edat	Sexe	Impor- tacions	Híbrids	Calcs	Altres
Llengua inicial	1,000	,000	,260	-,072	,078	,016	-,002
Edat	,000	1,000	,000	,045	-,101	,027	,000
Sexe	,260	,000	1,000	,030	-,018	-,053	-,094
Importacions	-,072	,045	,030	1,000	,120	-,027	,053
Híbrids	,078	-,101	-,018	,120	1,000	-,059	,262
Calcs	,016	,027	-,053	-,027	-,059	1,000	,052
Altres	-,002	,000	-,094	,053	,262	,052	1,000
Dimension	1	2	3	4	5	6	7
Valors propis	1,335	1,272	1,058	1,047	,911	,718	,659

La matriu de saturacions als components rotats, d'altra banda, mostra que només la variant calcs acompanya les variables llengua inicial i sexe amb una contribució més important al segon component que al primer. L'edat i les altres tres variants de V38, en canvi, tenen més pes al primer component, com indica la Taula 104, de manera que són les més rellevants per diferenciar els informants.

Taula 104. Matriu de saturacions als components rotats per a V38.

a	Dimensió	
	1	2
Llengua inicial	,030	,778
Edat	-,204	-,097
Sexe	-,145	,771
Importacions	,357	-,048
Híbrids	,792	,168
Calcs	-,058	-,163
Altres	,713	-,103

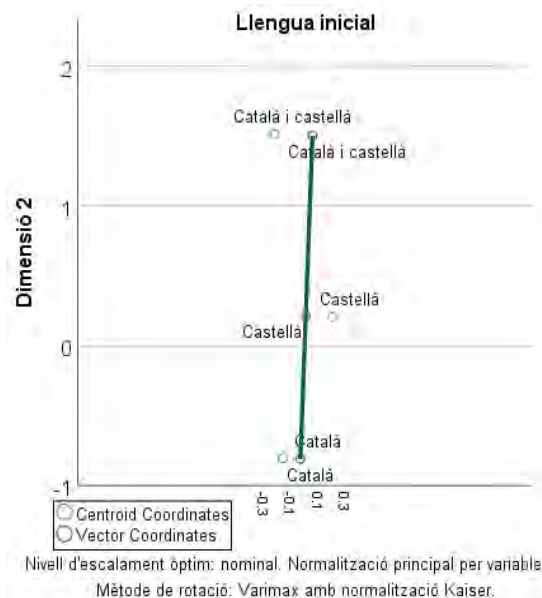
Normalització principal per variable.

^a Mètode de rotació: Varimax amb normalització Kaiser. La rotació no ha convergit en 3 iteracions. (Convergència = ,000).

Així doncs, dos dels gràfics de transformació de les variables per a V38 tenen com a dimensió més rellevant la segona: el de la llengua inicial i el del sexe. Al primer dels tres gràfics següents, hi observem que els bilingües inicials apareixen

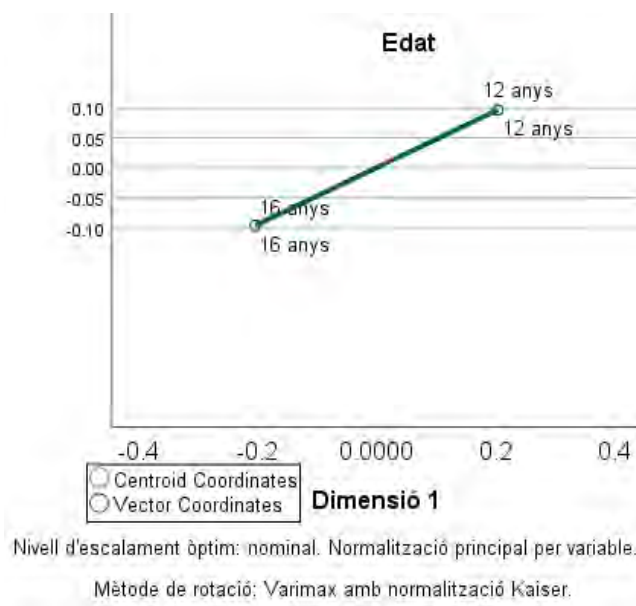
entre els valors d'1 i 2; els castellanoparlants, prop de 0, i els catalanoparlants, vora -1 a l'eix d'ordenades.

Figura 106. Gràfic de transformació de la variable independent llengua inicial a l'Anàlisi de Components Principals de V38.



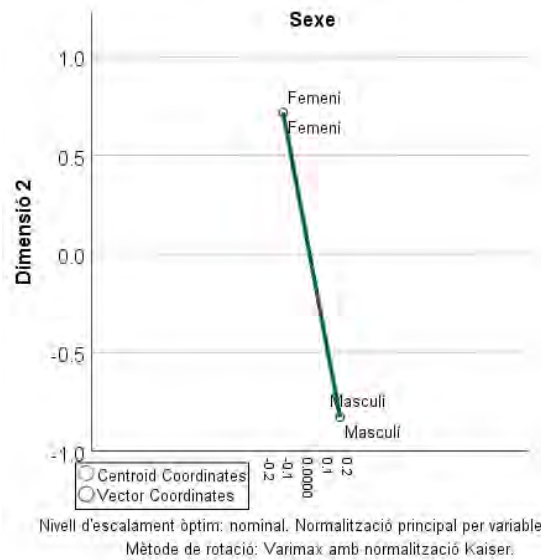
Al segon gràfic de transformació, en canvi, ens fixem en la segona dimensió (eix d'abscisses) i hi observem que l'edat de 12 anys s'hi situa prop de -0,2 i la de 16 anys, vora 0,2.

Figura 107. Gràfic de transformació de la variable independent edat a l'Anàlisi de Components Principals de V38.



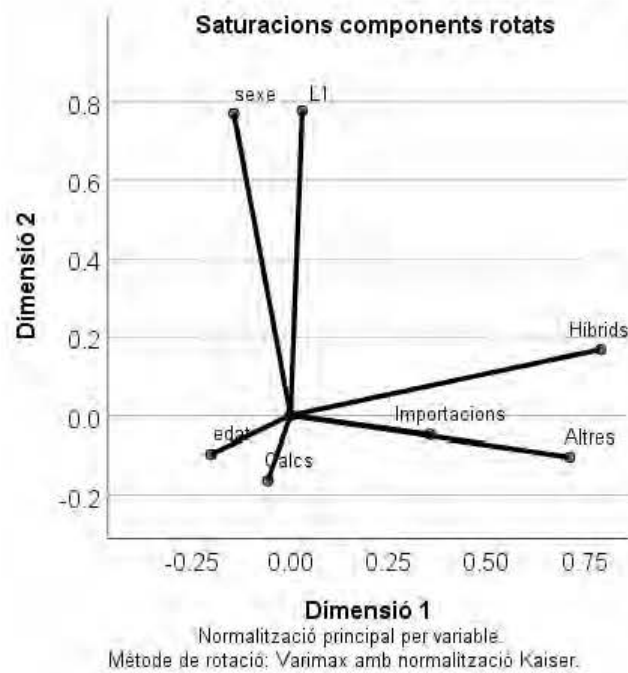
El darrer gràfic de transformació, que correspon a la variable sexe en relació a V38, ubica les noies entre els valors de 0,5 i 1 de la dimensió 2 (eix d'ordenades) i els nois entre -0,5 i -1 (però més propers a aquest últim).

Figura 108. Gràfic de transformació de la variable independent sexe a l'Anàlisi de Components Principals de V38.



Finalment, el gràfic de saturacions als components rotats per a V38 situa, a l'eix d'abscisses, l'edat a prop de la zona atribuïda anteriorment als 12 anys (-0,2). A l'eix d'ordenades, en primer lloc, la variant Importacions apareix aproximadament a l'àrea dels castellanoparlants inicials (0) i, en segon, la variant Altres i l'edat se situen vora la zona de catalanoparlants inicials (1), de la qual també es projecten a poca distància les variants Importacions i Calcs. Pel que fa a les correlacions entre els factors que es representen al gràfic, els angles indiquen que les variants Importacions i Altres es troben fortament relacionades i que també hi podria haver correlacions entre l'edat i la variant Calcs i entre el sexe i la llengua inicial, com ja havíem vist en gràfics de saturacions corresponents a altres variables.

Figura 109. Gràfic de saturacions als components rotats per a la variable V38.



6.2.18. Anàlisi de Components Principals dels colors (V41)

Per últim, resumim a continuació els resultats de l'ACP per a la darrera variable de l'estudi multivariant, les referències als colors (V41). El model generat per aquesta prova presenta el percentatge de variància explicada més elevat de tot el projecte, el 56,3 %, repartit entre les dues dimensions com mostra la Taula 105: el primer component assoleix el 30,8 % i el segon, el 25,5 %.

Taula 105. Resum del model generat per l'Anàlisi de Components Principals per a V41.

Resum model multivariant rotat ^a			
Dimensió	Alfa de Cronbach	Variància explicada	
		Total (valors propis)	% Variància
1	,442	1,540	30,791
2	,278	1,276	25,527
Total	,806 ^b	2,816	56,318

^a Mètode de rotació: Varimax amb Normalització Kaiser.

^b Alfa de Cronbach total basat en el total de valors propis.

Pel que fa a la matriu de correlacions (Taula 106), hi destaca la correlació entre les dues variants de V41 amb un 0,517, molt per sobre de la resta.

Taula 106. Matriu de correlacions entre variables per a V41.

	L1	Edat	Sexe	color [blau]	[blau]
Llengua inicial	1,000	,000	,259	-,026	-,210
Edat	,000	1,000	,000	-,152	,051
Sexe	,259	,000	1,000	-,008	,045
color [blau]	-,026	-,152	-,008	1,000	,517
[blau]	-,210	,051	,045	,517	1,000
Dimension	1	2	3	4	5
Valors propis	1,579	1,237	1,031	,751	,401

La matriu següent mostra els valors de saturació als components rotats de cadascuna de les variables independents i de les dues variants. S'hi observa que tant l'edat com les dues formes de realització de V41 contribueixen més a la primera dimensió i que, en canvi, la llengua inicial i el sexe tenen valors de saturació més elevats a la segona. Tot i això, en el cas de la variable edat, la diferència entre els dos valors és força minça, fet que es reflectirà al gràfic de transformació d'aquesta variable. De totes maneres, els factors sombrejats a la Taula 107 resulten més rellevants que els altres per diferenciar els informants.

Taula 107. Matriu de saturacions als components rotats per a V41.

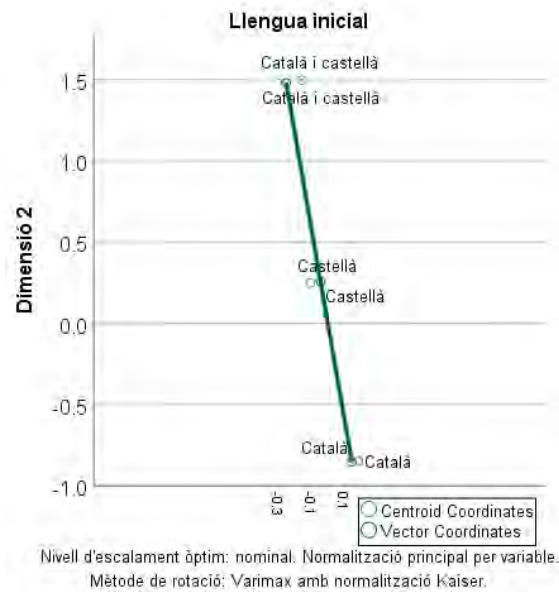
a	Dimensió	
	1	2
Llengua inicial	-,138	,797
Edat	-,188	-,111
Sexe	,148	,777
color [blau]	,866	,003
[blau]	,844	-,163

Normalització principal per variable.

^a Mètode de rotació: Varimax amb normalització Kaiser. La rotació no ha convergit en 3 iteracions. (Convergència = ,000).

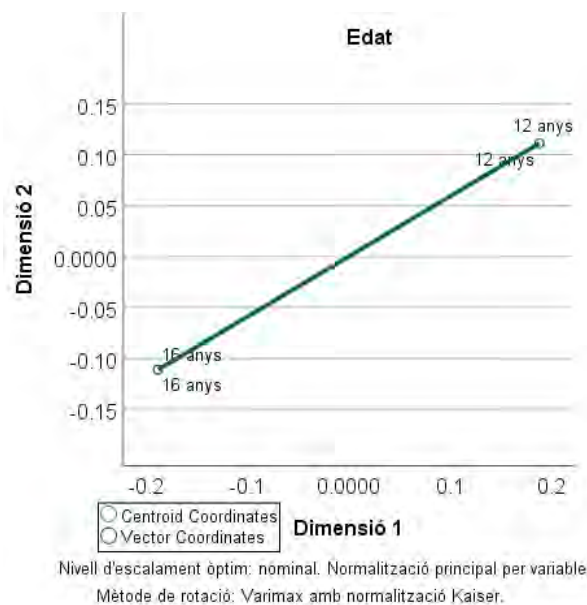
Així, el gràfic de transformació de la variable llengua inicial mostra com es distribueixen els tres grups que la componen a la segona dimensió. Els catalanoparlants inicials es mostren propers a -1, els castellanoparlants inicials apareixen entre 0 i 0,5 i els bilingües inicials se situen vora 1,5.

Figura 110. Gràfic de transformació de la variable independent llengua inicial a l'Anàlisi de Components Principals de V41.



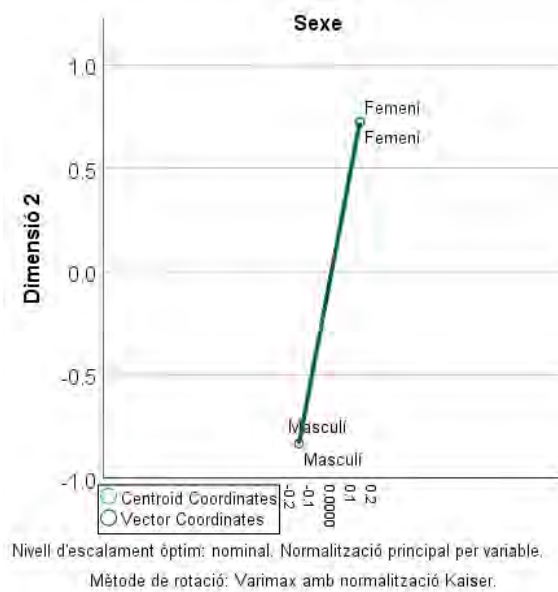
D'altra banda, com avançàvem, al gràfic corresponent a la variable edat els valors de les dues dimensions es mostren prou espaiats com per localitzar-hi els dos grups que conformen aquesta variable. Això és deu a la poca diferència en els valors de saturació observats anteriorment. Malgrat tot, utilitzem com a referència els del component en el qual l'edat té més pes, el primer. Per tant, observem que la Figura 111 mostra els 16 anys al valor de -0,2 de l'eix d'abscisses i els 12 anys al de 0,2.

Figura 111. Gràfic de transformació de la variable independent edat a l'Anàlisi de Components Principals de V41.



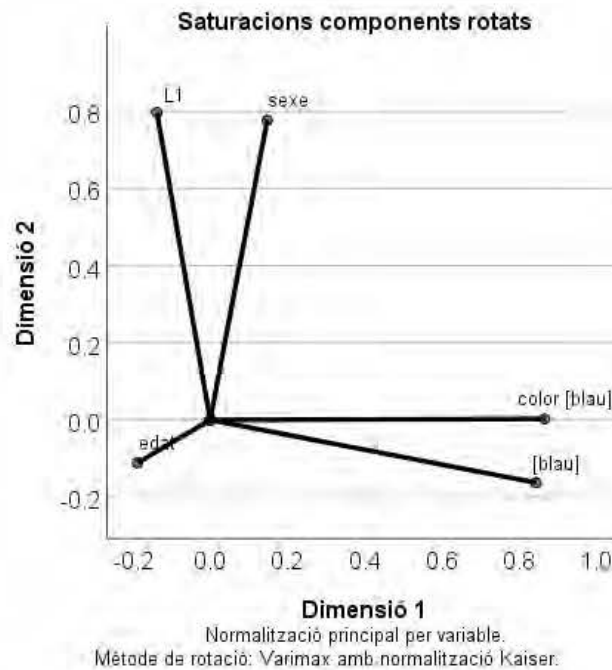
Quant al gràfic de transformació de la variable sexe per a V41, la Figura 112 mostra que les noies s'ubiquen entre els valors de 0,5 i 1 a l'eix d'ordenades, mentre que els nois hi apareixen entre -0,5 i -0,1 (si bé tendeixen més cap a aquest últim).

Figura 112. Gràfic de transformació de la variable independent sexe a l'Anàlisi de Components Principals de V41.



Finalment, el gràfic de saturacions als components rotats per a V41 condensa la informació anterior sobre les relacions que mantenen els diversos factors. Respecte la dimensió 1, hi podem observar que l'edat es projecta a l'àrea anteriorment assignada als 16 anys (-0,2) i el sexe, a l'atribuïda als 12 (0,2). Pel que fa a la dimensió 2, la llengua inicial i el sexe se situen entre 0,5 i 1, és a dir, el rang en què en un dels gràfics anteriors es trobava el grup d'informants noies. La separació de les línies que representen els cinc factors indica que formen tres grups diferents: les dues variants de V41 mantenen certa correlació entre elles, així com la llengua inicial i el sexe dels informants, i totes dues parelles es troben força distanciades de la variable edat.

Figura 113. Gràfic de saturacions als components rotats per a la variable V41.



Recapitulant, malgrat que els resultats obtinguts mitjançant les matrius de correlacions entre les variables són limitats en tant que els valors, amb poques excepcions, disten considerablement de mostrar correlacions fortes, les anàlisis de components principals realitzades per a les variables de l'estudi mostren, aparentment, diverses relacions entre les variants lingüístiques i les variables independents. A més, cal tenir en compte que una de les correlacions aparents que s'ha repetit a diverses ACP, la que existiria entre el sexe i la llengua inicial, possiblement té a veure amb la composició del corpus, en què hi ha cinc noies menys que nois que tenen el català com a llengua inicial, quatre noies més que nois que tenen el castellà com a llengua inicial i sis noies més que nois amb totes dues llengües com a inicials (vegeu apartat 5.2). En qualsevol cas, per tal de comprovar quines de les tendències de correlació observades mitjançant les ACP entre les variables dependents i les independents es confirmen com a estadísticament significatives, es van dur a terme les anàlisis detallades al subapartat 5.6.2, com s'explica a continuació.

6.3. Anàlisi de significació estadística de les dades

Com s'explica al capítol sobre els objectius i les hipòtesis de partida d'aquest estudi (capítol 3), l'etapa central de l'anàlisi estadística de les dades consisteix a posar a prova la significació de les relacions detectades entre les variables. Per fer-ho, com s'ha explicat a l'apartat 5.6, s'han aplicat tres proves de significació estadística a les dades obtingudes a les etapes anteriors, una per cada variable independent. El nivell de significació estadística (valor p) utilitzat és de 0,05 (o 5 %), l'habitual en estudis sociolingüístics. En aquest estudi, la hipòtesi nul·la que s'intenta rebutjar és que les variables independents (edat, sexe i llengua inicial) no estan correlacionades amb les dependents. Els resultats majors de 0,05 indiquen que cal mantenir aquesta hipòtesi nul·la (H_0), ja que el risc de rebutjar-la erròniament és més del 95 %. Per tant, casos així, la relació entre dues variables no és estadísticament significativa. En canvi, quan la prova estadística retorna un resultat igual o menor a 0,05, podem rebutjar la H_0 i verificar la H_1 o hipòtesi alternativa, és a dir, que sí que existeix una relació entre les variables, amb un risc acceptable (equivalent o inferior a 5 %) que la relació identificada sigui casual.

Tot seguit, doncs, tractem els resultats de les proves estadístiques per a les variables pragmàtiques, morfosintàctiques i lèxiques. Per facilitar la lectura d'aquest apartat, evitem mencionar repetidament les proves estadístiques utilitzades per calcular els valors p per a l'edat (estudi apariat de les dades per mitjà de la prova dels rangs amb signe de Wilcoxon), el sexe (prova U de Mann-Whitney) i la llengua inicial (prova de Kruskal-Wallis) en cadascun dels subapartats següents. No obstant això, val la pena recordar el funcionament d'aquestes proves explicat anteriorment (a l'apartat 5.6.2).

6.3.1. Significació estadística de les variables pragmàtiques

En primer lloc, pel que fa als resultats corresponents a les variables pragmàtiques incloses a l'estudi univariant, observem a la Taula 108 que la variable independent correlacionada de forma estadísticament significativa amb un nombre més elevat de variables pragmàtiques (6/10) és l'edat dels informants, seguida del sexe (2/10) i de la llengua inicial (1/10). Concretament, l'edat mostra un valor p igual o superior a 0,05 per als dobles de partícules discursives V3b, V3c, V3h, V3i i V3j i també per a les heterorepeticions (V6), mentre que el sexe ho fa en el cas de les dades de V3b corresponents a l'edat de 12 anys i de les heterorepeticions tant als 12 com als 16 anys. La llengua inicial només es correlaciona de forma estadísticament significativa amb les dades del doblet de partícules discursives *hombre / home* (V3h) de l'edat de 16 anys.

Taula 108. Resultats de les proves de significació estadística per a les variables pragmàtiques de l'estudi univariant.

Variable	Variant	Edat	Sexe		Llengua inicial	
			12 anys	16 anys	12 anys	16 anys
V3b	Total <i>a ver / a veure</i>	0,005*	0,046*	0,350	0,960	0,264
V3c	Total <i>bueno / bé</i>	0,013*	0,320	0,161	0,240	0,965
V3f	Total <i>después / després</i>	0,379	0,090	0,908	0,903	0,507
V3h	Total <i>hombre / home</i>	0,016*	0,218	0,497	0,708	0,042*
V3i	Total <i>o sea / o sigui</i>	0,000*	0,374	0,159	0,096	0,399
V3j	Total <i>pero / però</i>	0,000*	0,247	0,275	0,173	0,442
V3k	Total <i>pues / doncs</i>	0,288	0,160	0,177	0,679	0,803
V3l	Total <i>vale / d'acord</i>	0,938	0,337	0,342	0,920	0,892
V5	Total salutacions	0,393	0,598	0,475	0,829	0,832
V6	Heterorepetició	0,000*	0,049*	0,002*	0,308	0,788

*Valor p igual o inferior al nivell de significació (0,05).

L'estudi multivariant de V3 i V5 mostra amb més detall les correlacions entre les variants i les variables independents. Hi podem veure que només algunes de les variants de cada variable dependent mostren valors p iguals o inferiors a 0,05. De les 26 variants que conformen les variables dels dobles de partícules discursives (V3) i les fórmules d'obertura de la interacció (V5), un total de 10 (el 38,46 %) presenten correlacions estadísticament significatives amb l'edat dels informants: *a ve(u)re* (V3b), *bueno* (V3c), *home* (V3h), *o sea* i *o sigui* (V3i),

p(e)rò i *p(er)ò* (V3j), *pos* (V3k), *val* i *d'acord* (V3l). Fins a 6 variants (el 23,08 % del total) es correlacionen de manera estadísticament significativa amb la llengua inicial dels parlants: *home* (als 16 anys), *o sea* (als 16 anys), *o sigui* (a les dues edats), *p(er)ò* (a les dues edats), *bona tarda* (als 12 anys) i *hola bon dia* (als 16 anys). A més, 4 variants (el 15,39 % del total) mostren valors p iguals o inferiors a 0,05 en relació amb el sexe: de V3b, *a ver* (als 16 anys) i *a ve(u)re* (als 12 anys); de V3f, *després* (als 12 anys), i de V3j, *p(er)ò* (també als 12 anys).

Taula 109. Resultats de les proves de significació estadística per a les variables pragmàtiques de l'estudi multivariant.

Variable	Variant	Edat	Sexe		Llengua inicial	
			12 anys	16 anys	12 anys	16 anys
V3b	V3b1: <i>a ver</i>	0,429	0,312	0,057*	0,316	0,260
	V3b2: <i>a veure</i>	0,524	0,618	0,947	0,188	0,496
	V3b3: <i>a ve(u)re</i>	0,004*	0,020*	0,446	0,372	0,727
	V3b4: <i>aviam</i>	0,438	0,901	0,825	0,259	0,412
V3c	V3c1: <i>bueno</i>	0,018*	0,265	0,148	0,224	0,968
	V3c2: <i>bé</i>	0,583	0,121	0,114	0,352	0,936
V3f	V3f1: <i>después</i>	1	0,57	0,185	0,124	0,287
	V3f2: <i>després</i>	0,417	0,029*	0,714	0,824	0,301
V3h	V3h1: <i>hombre</i>	0,754	0,901	0,901	0,309	0,837
	V3h2: <i>home</i>	0,027*	0,170	0,538	0,596	0,033*
V3i	V3i1: <i>o sea</i>	0,058*	0,143	0,515	0,173	0,027*
	V3i2: <i>o sigui</i>	0,000*	0,636	0,150	0,013*	0,054*
V3j	V3j1: <i>pero</i>	0,596	0,927	0,105	0,099	0,600
	V3j2: <i>però</i>	0,424	0,574	0,634	0,69	0,682
	V3j3: <i>p(e)rò</i>	0,000*	0,068	0,152	0,732	0,715
	V3j4: <i>p(er)ò</i>	0,056*	0,059*	0,299	0,004*	0,009*
V3k	V3k1: <i>pues</i>	0,074	0,629	0,687	0,670	0,826
	V3k2: <i>pos</i>	0,022*	0,904	0,071	0,253	0,358
	V3k3: <i>doncs</i>	0,461	0,975	0,489	0,928	0,506
V3l	V3l1: <i>vale</i>	0,134	0,393	0,146	0,725	0,695
	V3l2: <i>val</i>	0,000*	0,574	0,535	0,971	0,185
	V3l3: <i>d'acord</i>	0,007*	0,324	0,60	0,404	0,990
V5	V5a: <i>hola</i>	0,168	0,535	0,073	0,136	0,771
	V5b: <i>bon dia</i>	0,326	0,840	0,321	0,769	0,059
	V5c: <i>bona tarda</i>	1	0,506	0,550	0,035*	0,064
	V5d: <i>hola bon dia</i>	0,222	0,646	0,323	0,688	0,012*

*Valor p igual o inferior al nivell de significació (0,05).

Per tant, segons l'estudi univariant, V3f, V3k, V3l i V5 no es correlacionen de forma estadísticament significativa amb cap de les tres variables independents. Ara bé, segons l'estudi multivariant, totes les variables pragmàtiques inclouen com a mínim una variant que sí que es correlaciona de manera estadísticament significativa amb una o més variables independents.

6.3.2. Significació estadística de les variables morfosintàctiques

En segon lloc, com reflecteix la Taula 110, dues de les cinc variables morfosintàctiques de l'estudi univariant (el 40 %) presenten correlacions estadísticament significatives amb alguna de les variables independents. Aquestes dues variables són el total de construccions condicionals (V18), que es correlaciona amb l'edat i amb la llengua inicial (als 12 anys), i el total de perífrasis verbals d'obligació (V22), que ho fa amb el sexe (a l'edat de 12 anys).

Taula 110. Resultats de les proves de significació estadística per a les variables morfosintàctiques de l'estudi univariant.

Variable	Variant	Edat	Sexe		Llengua inicial	
			12 anys	16 anys	12 anys	16 anys
V16	Total preguntes	0,282	0,704	0,554	0,397	0,120
V18	Total condicionals	0,050*	0,237	0,952	0,051*	0,062
V22	Total perífrasis d'obligació	0,983	0,022*	0,306	0,549	0,272
V23	Total determinants demostratius	0,305	0,267	0,518	0,991	0,120
V25	Total clítics datiu plural	0,849	0,681	0,179	0,406	0,973

*Valor *p* igual o inferior al nivell de significació (0,05).

L'estudi multivariant de les mateixes variables mostra un panorama més detallat de les correlacions estadísticament significatives entre les variants que les formen i les tres variables independents. Com es pot observar a la Taula 111, 8 de les 32 variants morfosintàctiques (el 25 %) presenten una correlació estadísticament significativa amb la llengua inicial: de V14a, la variant PC + *anar a* + INF (als 12 anys); de V18, *si* (a les dues edats); de V22, *caldre* (als 12 anys); de V23, *el* + N + *aquell* (als 12 anys), i les quatre variants de V25 als 12 anys (i *lis* + V / V + *-lis* també als 16 anys). Pel que fa a la variable edat, 7 de les 32 variants morfosintàctiques (el 21,88 %) s'hi mostren correlacionades amb un

valor *p* igual o inferior al nivell de significació: PC + *tenir que / de* + INF i PC + *anar a* + INF (V14a), *poder* + INF + PC (V14b), *si* (V18), *tenir que* i *haver de* (V22), i *aquest* + N (V23). Finalment, dues variants morfosintàctiques es correlacionen de manera estadísticament significativa amb el sexe dels informants: les preguntes obertes (V17) i les construccions condicionals amb la conjunció *mentre(s)* (V18).

Taula 111. Resultats de les proves de significació estadística per a les variables morfosintàctiques de l'estudi multivariant.

Variable	Variant	Edat	Sexe		Llengua inicial	
			12 anys	16 anys	12 anys	16 anys
V14a	V14a1: PC + <i>tenir que / de</i> + INF	0,001*	0,597	0,754	0,739	0,961
	V14a2: PC + <i>haver de</i> + INF	0,080	0,467	0,355	0,397	0,311
	V14a3: PC + <i>poder</i> + INF	0,129	0,795	0,961	0,628	0,438
	V14a4: PC + <i>anar a</i> + INF	0,057*	0,175	0,352	0,036*	0,561
V14b	V14b1: <i>tenir que / de</i> + INF + PC	0,109	0,121	0,877	0,192	0,180
	V14b2: <i>haver de</i> + INF + PC	0,229	0,474	0,595	0,560	0,086
	V14b3: <i>poder</i> + INF + PC	0,017*	0,364	0,834	0,459	0,229
V14	Total proclisi	0,502	0,727	0,230	0,284	0,129
	Total enclisi	0,071	0,708	0,789	0,158	0,462
	Total doble	0,392	0,882	0,881	0,762	0,183
V17	V17a: Oberta	0,189	0,536	0,015*	0,777	0,305
	V17b: Tancada	0,069	0,892	0,932	0,098	0,105
V18	V18b: <i>si</i>	0,048*	0,235	0,952	0,034*	0,047*
	V18c: <i>mentre(s)</i>	0,429	0,058*	0,151	0,448	0,083
	V18f: <i>a no ser que</i>	0,821	0,157	0,485	0,593	0,246
	V18g: <i>en (tot) cas</i>	1	0,381	0,901	0,689	0,268
V22	V22a: <i>tenir que</i>	0,002*	0,133	0,520	0,115	0,651
	V22b: <i>haver de</i>	0,015*	0,961	0,603	0,073	0,389
	V22c: <i>tenir de</i>	0,375	0,827	0,589	0,968	0,970
	V22e: <i>caldre</i>	0,672	0,790	0,485	0,027*	0,519
V23	V23a: <i>aquest</i> + N	0,004*	0,139	0,343	0,181	0,308
	V23b: <i>aquell</i> + N	0,602	0,796	0,128	0,269	0,537
	V23c: <i>el</i> + N + <i>aquest</i>	0,545	0,352	0,248	0,620	0,346

Resultats

	V23d: <i>el + N + aquell</i>	0,826	0,825	0,862	0,007*	0,374
	V23e: <i>aquest</i>	0,402	0,932	0,714	0,517	0,241
	V23f: <i>aquell</i>	1	0,120	0,506	0,089	0,391
	V23g: <i>(un) + N + d'aquests</i>	0,752	0,588	0,835	0,397	0,131
	V23h: <i>un + N + d'aquells</i>	0,792	0,858	0,289	0,685	0,407
V25	V25a: <i>els hi + V / V + -els-hi / V + 'ls-hi / V + -los-hi</i>	0,394	0,109	0,318	0,054*	0,230
	V25b: <i>els + V / V + -los / V + 'ls</i>	0,984	0,609	0,411	0,021*	0,156
	V25c: <i>li + V / V + -li</i>	0,215	0,842	0,408	0,023*	0,615
	V25d: <i>lis + V / V + -lis</i>	0,142	0,530	0,223	0,057*	0,046*
*Valor <i>p</i> igual o inferior al nivell de significació (0,05).						

En resum, mentre que els resultats de l'estudi univariant mostren que el recompte total de V17, V23 i V25 no es correlacionen de forma estadísticament significativa amb cap variable independent, els de l'estudi multivariant demostren que l'única variable que no presenta cap variant que s'hi correlacioni és V14c, que, cal recordar, està formada pels totals de pronoms clítics segons la seva posició a les construccions perifràstiques

6.3.3. Significació estadística de les variables lèxiques

En darrer lloc, de les tres variables lèxiques incloses a l'estudi univariant, dues (el total de referències a colors i el total d'adverbis acabats en *-ment*) es mostren correlacionades de forma estadísticament significativa amb alguna de les variables independents. Concretament, com s'observa a la Taula 112, V41 mostra un valor *p* de 0,05 al resultat de la prova Kruskal Wallis (és a dir, en relació amb la variable independent llengua inicial) per a les dades recollides quan els informants tenien aproximadament 16 anys i V45 presenta un valor *p* inferior al nivell de significació a la prova dels rangs amb signe de Wilcoxon (en relació amb la variable independent edat).

Taula 112. Resultats de les proves de significació estadística per a les variables lèxiques de l'estudi univariant.

Variable	Variant	Edat	Sexe		Llengua inicial	
			12 anys	16 anys	12 anys	16 anys
V38	Total fenòmens contacte llengües	0,281	0,952	0,139	0,762	0,990
V41	Total colors	0,864	0,583	0,940	0,773	0,053*
V45	Total adverbis <i>-ment</i>	0,000*	0,069	0,747	0,426	0,066

*Valor *p* igual o inferior al nivell de significació (0,05).

Els resultats de l'estudi multivariant (Taula 113) complementen els anteriors i mostren que, tot i que el recompte total de la variable Fenòmens de contacte de llengües (V38) no presenta correlacions estadísticament significatives amb les variables independents, les dades corresponent als 12 anys de la variant Altres en realitat estan significativament correlacionades amb la llengua inicial dels informants. Aquesta variable independent també es correlaciona de manera estadísticament significativa amb les dades de la variable colors corresponents als 16 anys.

Taula 113. Resultats de les proves de significació estadística per a les variables lèxiques de l'estudi multivariant.

Variable	Variant	Edat	Sexe		Llengua inicial	
			12 anys	16 anys	12 anys	16 anys
V38	V38a: Importacions	0,435	0,890	0,136	0,800	0,990
	V38b: Híbrids	0,392	0,825	0,646	0,107	0,688
	V38c: Calcs	0,701	0,418	0,319	0,357	0,523
	V38d: Altres	1	0,512	0,825	0,046*	0,473
V41	V41a: <i>color [blau]</i>	0,062	0,851	0,951	0,666	0,815
	V41b: <i>[blau]</i>	0,488	0,615	0,785	0,449	0,032*

*Valor *p* igual o inferior al nivell de significació (0,05).

Així doncs, si bé de l'estudi univariant se'n desprèn que totes les variables lèxiques menys els fenòmens de contacte de llengües (V38) mostren correlacions estadísticament significatives amb alguna de les variables independents, l'estudi multivariant mostra que una de les variants que formen V38 també presenta un valor *p* significatiu.

Després de resumir l'estudi estadístic exploratori, revisar les relacions entre les variables i repassar els resultats de les proves d'anàlisi de significació, ens ocupem, finalment, de la caracterització del corpus d'anàlisi a partir de les variables dependents de l'estudi. Per fer-ho, ens servim d'una tècnica de mineria de dades, els arbres de classificació.

6.4. Arbres de classificació

Tot seguit explorem les relacions entre les variables des d'una perspectiva complementària a les adoptades fins ara per tal de caracteritzar els informants, dividits en grups d'edat, sexe i llengua inicial, a partir de les variables lingüístiques de l'estudi. Una de les eines estadístiques que permet dur a terme aquest objectiu són els arbres de classificació, un mètode de classificació no lineal els resultats del qual són fàcilment interpretables, com s'ha explicat anteriorment (5.6.1.2).

D'entre els diversos mètodes existents per a la creació d'arbres de classificació (vegeu Berlanga Silvente, Rubio Hurtado i Vilà Baños, 2013:68; Rokach & Maimon, 2015: 77-81), s'ha optat per dos dels mètodes més àmpliament utilitzats: CHAID (sigles de l'anglès *Chi-square Automatic Interaction Detection*, és a dir, detecció automàtica d'interaccions mitjançant khi quadrat) i CRT (*Classification and Regression Trees* o arbres de classificació i regressió). El primer escull la variable independent o predictora que mostra una interacció més potent amb la dependent i fa convergir les categories de cada variable predictora que no són significativament diferents de la variable dependent (IBM, 2021). El segon divideix les dades en segments segons el criteri que siguin el més homogeni possible en relació a la variable dependent (IBM, 2021). En cada cas, s'ha seleccionat el mètode que genera un model amb un percentatge de classificació correcta més elevat, CRT.

En primer lloc, ens ocupem dels arbres que caracteritzen les dues edats dels informants (12 i 16 anys) segons les variables lingüístiques de l'estudi univariant i, tot seguit, segons les de l'estudi multivariant. Com veurem, tots dos tenen una profunditat de 5 nivells per sota del node arrel. La Taula 114 mostra els percentatges de classificació correcta de cada categoria de la variable dependent

(edat) corresponents al primer d'aquests arbres. L'error estàndard d'aquest primer model és de 0,032.

Taula 114. Taula de classificació de la variable edat amb les dades de l'estudi univariant.

Observat	Predit		
	12 anys	16 anys	Percentatge correcte
12 anys	56	13	81,2 %
16 anys	10	59	85,5 %
Percentatge global	47,8 %	52,2 %	83,3 %

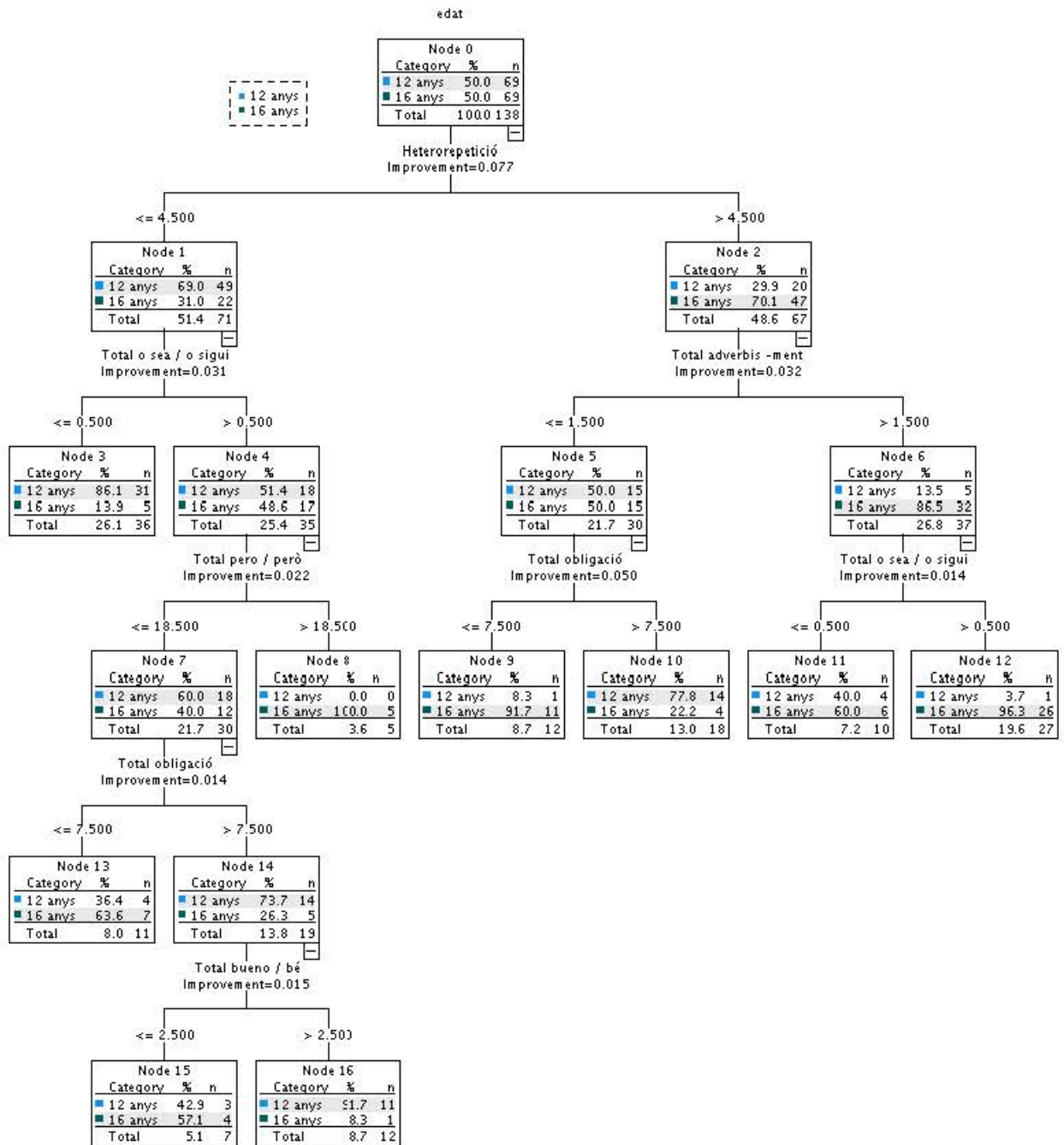
Mètode de creixement: CRT.

Variable dependent: edat.

Seguint les segmentacions del node 1 (costat esquerre), aquest arbre mostra que, de l'estudi univariant, la variable que discrimina millor les dues edats en què es va entrevistar els informants és el total d'heterorepeticions (V6): la majoria de transcripcions que en presenten 4,5 o menys corresponen a l'edat de 12 anys (node 1), mentre que la majoria de les que en presenten més d'aquest total corresponen als 16 anys (node 2). La primera segmentació del node 1 mostra que un total de 0,5 o menys partícules discursives *o sea / o sigui* (V3i) permet identificar la majoria de transcripcions de les proves realitzades quan els informants tenien 12 anys (node 3). D'entre les transcripcions que presenten 0,5 o més realitzacions de V3i, en canvi, aproximadament el mateix nombre corresponen a l'edat de 12 i de 16 anys (18 i 17, respectivament), per la qual cosa es genera una altra segmentació d'aquest node (el 4). Arribats a aquest punt, la variable més útil per discriminar les dades és V3j (total de *pero / però*): la majoria de les transcripcions que en contenen 18,5 realitzacions o menys pertanyen a informants del primer grup d'edat (node 7), mentre que totes les que en contenen més corresponen al segon (node 8). A continuació, el node 7 és segmentat per la variable V22 (total de perífrasis d'obligació), ja que la majoria de les transcripcions que presenten 7,5 perífrasis d'aquest tipus o menys pertanyen a la categoria de 16 anys (node 13) i la majoria de les que en contenen més de 7,5, pertanyen a la de 12 anys (node 14). Aquest darrer grup

de transcripcions encara es divideix una vegada més entre les que contenen 2,5 realitzacions de V3c (total de *bueno / bé*) o menys, la majoria de les quals són d'informants de 16 anys (node 15) i les que en presenten més de 2,5, la gran majoria de les quals són del grup de 12 anys.

Figura 114. Arbre de classificació de la variable edat amb les dades de l'estudi univariant.



Pel que fa a les segmentacions del node 2 (costat dret), aquest primer arbre mostra que la majoria de transcripcions del corpus d'anàlisi que contenen més de 4,5 heterorepeticions corresponen a l'edat de 16 anys (node 2). D'aquestes, les que presenten 1,5 adverbis acabats en *-ment* o menys tenen les mateixes probabilitats de correspondre a l'edat de 12 que a la de 16 anys (node 5) i, per tant, cal segmentar aquest node per una altra variable per discernir les dues categories. La que mostra millors resultats en aquest punt és V22, de manera que les transcripcions amb 7,5 perífrasis d'obligació o menys corresponen, en la gran majoria, a l'edat de 16 anys (node 9), mentre que les que en tenen més de 7,5 tenen més probabilitats de correspondre a l'edat de 12 anys (node 10). Quant a les transcripcions que tenen més de 1,5 adverbis acabats en *-ment* (node 6), la majoria són de parlants de 16 anys. Finalment, si tenim en compte la variable V3i, podem dividir aquestes transcripcions entre les que presenten 0,5 realitzacions de les partícules *o sea / o sigui* (node 11), força de les quals (un 60 %) corresponen a la categoria de 16 anys, i les que en contenen més de 0,5 (node 12), només un 3,7 % de les quals correspon als 12 anys.

Per tant, en definitiva, les variables lingüístiques de l'estudi univariant seleccionades pel model de classificació com a rellevants per diferenciar entre les dues edats analitzades (12 i 16 anys) són el total d'heterorepeticions (V6), de les partícules *o sea / o sigui* (V3i), d'adverbis acabats en *-ment* (V45), de les partícules *pero / però* (V3j), de les perífrasis d'obligació (V22) i de les partícules *bueno / bé* (V3c).

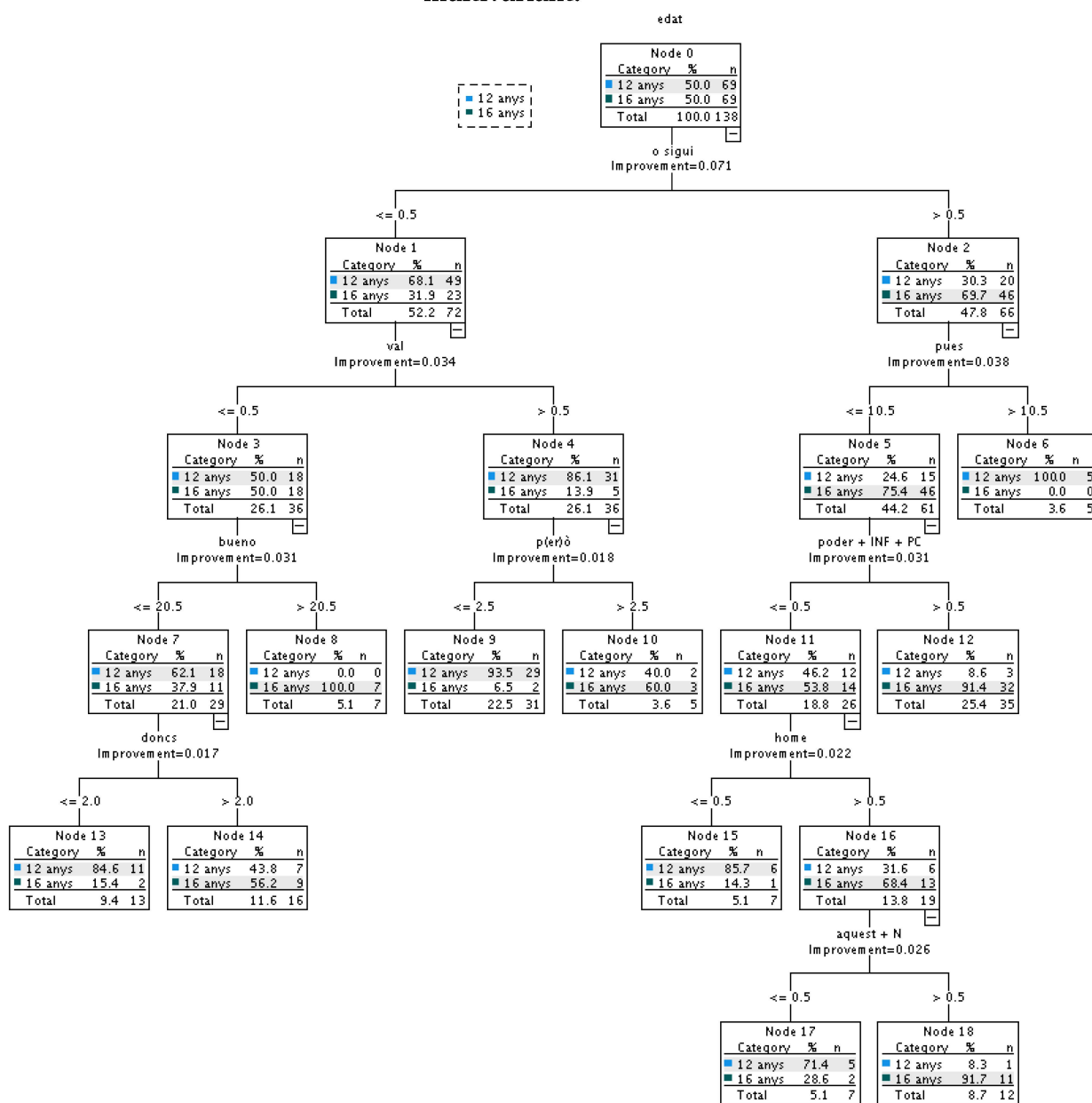
Tot i que, com hem vist, el percentatge de classificacions correctes assolit pel model generat a partir de les dades de l'estudi univariant és molt elevat, l'arbre de classificació de la variable edat obtingut a partir de les variables de l'estudi multivariant assoleix un percentatge de classificacions correctes superior, del 85,5 %, com s'observa a la Taula 115. Per tant, l'error estàndard és, també, una mica inferior que el de l'anterior model: 0,030.

Taula 115. Taula de classificació de la variable edat amb les dades de l'estudi multivariant.

Observat	Predit		
	12 anys	16 anys	Percentatge correcte
12 anys	56	13	81,2 %
16 anys	7	62	89,9 %
Percentatge global	45,7 %	54,3 %	85,5 %

Mètode de creixement: CRT. Variable dependent: edat.

Figura 115. Arbre de classificació de la variable edat amb les dades de l'estudi multivariant.



Com a l'arbre de classificació anterior, algunes de les partícules discursives juguen un paper destacat a la caracterització de la variable edat segons aquest segon arbre, generat amb les dades de l'estudi multivariant. La primera variant que s'utilitza per segmentar el node 0 és, precisament una de les que ja s'havia seleccionat a l'arbre anterior, creat a partir de les dades de l'estudi univariant, (V3i). Les transcripcions que presenten 0,5 o menys realitzacions de la variant *o sigui* formen el node 1, la majoria de casos del qual corresponen a l'edat de 12 anys. En canvi, les que en presenten més de 0,5 realitzacions es troben al node 2, gairebé un 70 % dels casos del qual pertanyen a la categoria de 16 anys. Si continuem llegint l'arbre pel costat esquerre, trobem que la segmentació següent del node 1 es fa a través de la variant *val* de V3l: de nou, es diferencia entre les transcripcions que en contenen 0,5 o menys (node 3) i les que en contenen més (node 4). La majoria (un 86,1 %) de les que es troben al segon grup corresponen a l'edat de 16 anys. A continuació, aquest node es segmenta en les transcripcions que contenen 2,5 o menys realitzacions de *p(er)ò* (V3j), que formen el node 9, i les que en tenen més realitzacions, que s'agrupen en el node 10. La gran majoria de les que es troben al node 9 pertanyen a l'edat de 12 anys, mentre que la major part de les del node 10 corresponen als 16 anys. Com que tots dos són nodes terminals, podem tornar, arribats a aquest punt, al node 3 i observar que està format per la mateixa quantitat de transcripcions corresponents a l'edat de 12 i de 16 anys. Així, es segmenta una vegada més, ara a través de *bueno* (V3c), en les que contenen 20,5 o menys ocurrencies d'aquesta partícula discursiva (node 7) i les que en tenen més (node 8). Totes les que en tenen més pertanyen al grup dels 16 anys i, per tant, el node 8 és terminal. En canvi, si bé la majoria dels casos del node 7 corresponen als 12 anys, aquest node es torna a segmentar una darrera vegada. D'una banda, podem discernir les transcripcions que, a més de totes les característiques anteriors, tenen dues o menys realitzacions de *doncs* (V3k), la majoria de les quals són de la categoria de 12 anys, i de l'altra banda, les que tenen més de dues realitzacions d'aquesta partícula, un 56 % de les quals són de la categoria de 16 anys.

Si retornem ara al node 2 i continuem llegint l'arbre pel costat dret, hi observem que la variant mitjançant la qual es segmenta per primer cop aquest node és *pues* (V3k): les transcripcions que en presenten 0,5 realitzacions o menys formen el node 5 i les que en tenen més, el node 6. Totes les que se situen en aquest últim corresponen a l'edat de 12 anys, mentre que la majoria de les que es troben al node anterior pertanyen a la de 16 anys. A continuació, es fa servir la variant *poder + INF + PC* (de la variable V14b, construccions perifràstiques que presenten els pronoms clítics en posició enclítica) per dividir les transcripcions en les que en contenen 0,5 o menys (node 11) i les que en presenten més (node 12). La gran majoria d'aquestes darreres són de l'edat de 16 anys, mentre que al node anterior, tot i que també són majoria (amb un 53,8 %), la diferència no és tan marcada entre el nombre de casos dels dos grups. Per això, el node 11 es divideix entre les transcripcions que contenen 0,5 o menys realitzacions de la variant *home* (V3h), que formen el node terminal 15, i les que en tenen més, que s'agrupen en el node 16. La majoria de les que es troben al primer d'aquests nodes corresponen a l'edat de 12 anys i la majoria de les que es troben al segon, a la de 16 anys. La darrera segmentació d'aquest arbre es fa a través del nombre de realitzacions de la variant *aquest + N* (V23): els casos que en presenten 0,5 o menys formen el node 17 i són, en la seva majoria, de l'edat de 12 anys, i els que en presenten més de 0,5 realitzacions formen el node 18 i són, en la gran majoria d'ocasions, de l'edat de 16 anys.

Per tant, totes les variants utilitzades en aquest arbre per caracteritzar els dos grups d'edat són partícules discursives (V3i2: *o sigui*, V3l2: *val*, V3k1: *pues*, V3c1: *bueno*, V3j4: *p(er)ò*, V3k3: *doncs* i V3h2: *home*) excepte *poder + INF + PC* (V14b3) i *aquest + N* (V23a).

En segon lloc, observem tot seguit els arbres de classificació que caracteritzen els grups de sexe (masculí i femení) del corpus d'anàlisi. Comencem, com en el cas anterior, per l'arbre generat a partir de les variables lingüístiques incloses a l'estudi univariant. Aquest primer, com els observats fins ara, té una profunditat de 5 nivells per sota del node arrel. Com recull la Taula 116, el percentatge de

classificacions correctes de la variable sexe gairebé arriba al 80 %. L'error estàndard se situa, així, al 0,035.

Taula 116. Taula de classificació de la variable sexe amb les dades de l'estudi univariant.

Observat	Predit		
	Masculí	Femení	Percentatge correcte
Masculí	55	9	85,9 %
Femení	21	53	71,6 %
Percentatge global	55,1 %	44,9 %	78,3 %

Mètode de creixement: CRT.

Variable dependent: sexe.

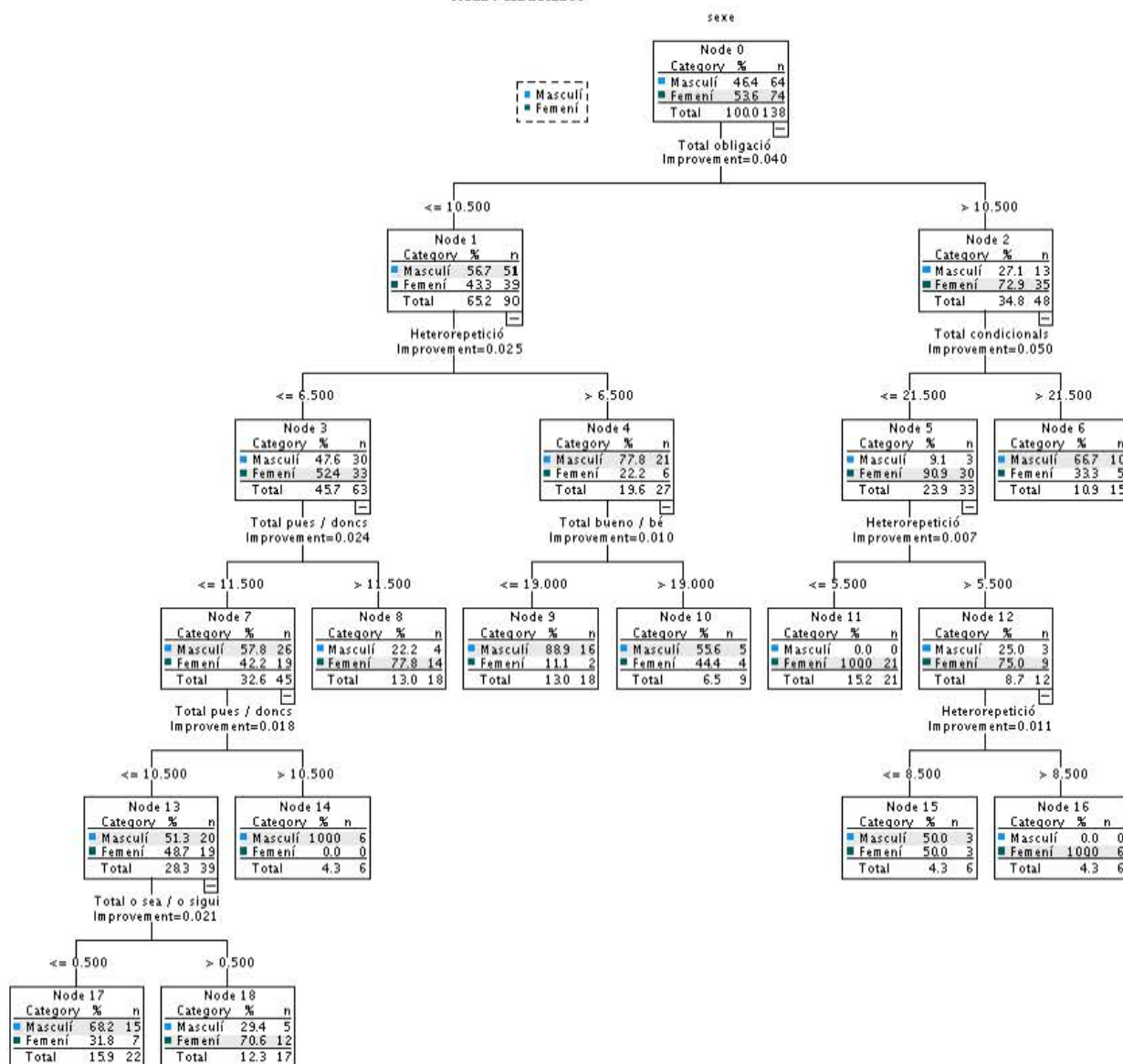
Com mostra la Figura 116, la primera segmentació del node 0 es fa entre les transcripcions que presenten 10,5 realitzacions o menys de les perífrasis verbals d'obligació (V22) i les que en presenten més. Les primeres formen el node 1 i són, en la majoria de casos, d'informants nois i les segones formen el node 2 i són, per contra, en la majoria d'ocasions, d'informants noies. Si ens fixem primer en els nodes al costat esquerre de l'arbre, veiem que el node 1 es divideix, tot seguit, entre les transcripcions que presenten 6,5 heterorepeticions (V6) o menys (node 3) i les que en contenen més (node 4). La majoria de casos del primer corresponen a informants noies i la majoria de casos del segon, a nois. El node 3 es divideix, de nou, entre el node 7, que conté els casos que presenten 11,5 o menys realitzacions de les partícules discursives *pues / doncs* (V3k), i el node 8, que agrupa els casos amb més de 11,5 realitzacions d'aquesta variable. La majoria de transcripcions del node 7 són d'informants nois, a diferència del node 8, que està format per una majoria de transcripcions d'informants noies. El node 7 es segmenta en transcripcions amb 10,5 realitzacions o menys de la variable *pues / doncs* (V3k) i aquelles amb més ocurrencies d'aquestes partícules. Les d'aquest darrer grup (node 14) són totes d'informants nois). En canvi, les del node 13 estan més repartides: un 48,7 % són de noies i un 51,3 %, de nois. Aquest node es segmenta, finalment, en els nodes 17 i 18. El node 17 conté transcripcions amb 0,5 o menys realitzacions de V3i (total del doblet *o sea / o sigui*), la majoria de les quals són de nois, i el node 18 conté aquelles que en

tenen més ocurrències, la majoria de les quals són de noies. Arribats a aquest punt al final de l'arbre, podem tornar al node 4, encara al costat esquerre del gràfic, i observar que s'utilitza la variable V3c (total de *bueno* / *bé*) per crear els nodes 9 i 10. La majoria de casos inclosos al node 9 corresponen a informants nois, mentre que els agrupats al node 10 estan més repartits (els nois hi representen el 55,6 % i les noies, el 44,4 %).

Un cop revisats tots els nodes que es desprenen del node 1, si ens situem ara al node 2 (costat dret), podem observar les segmentacions que es fan de les transcripcions que contenen més de 10,5 realitzacions de V22. La primera és mitjançant el total de construccions condicionals (V18): els casos amb 21,5 realitzacions o menys formen el node 5 (la gran majoria dels quals són d'informants noies) i els que en contenen més realitzacions formen el node 6 (la majoria dels quals corresponen a nois), que és terminal. El node 5 es segmenta a través d'una variable que ja s'havia seleccionat al costat esquerre de l'arbre, el total d'heterorepeticions (V6). Les transcripcions amb 5,5 realitzacions o menys d'aquesta variable formen el node terminal 11 i corresponen totes a informants noies. Les transcripcions amb més de 5,5 heterorepeticions, en canvi, formen el node 12, una quarta part del qual correspon a informants nois. El nombre d'heterorepeticions torna a ser, en aquest punt, el criteri seleccionat per segmentar el node 12: les transcripcions amb 8,5 realitzacions o menys s'agrupen en el node 15 i les que en presenten més realitzacions, en el node 16. Una transcripció de les incloses en el primer d'aquests nodes presenta les mateixes probabilitats de correspondre a un noi que a una noia (50 %), mentre que una de les que formen el darrer node de l'arbre serà sempre d'una noia.

En resum, de l'estudi univariant, l'algoritme en selecciona 6 variables per caracteritzar els grups de nois i noies: el total de perífrasis verbals d'obligació (V22), el d'heterorepeticions (V6), el de construccions condicionals (V18) i els dels conjunts de partícules discursives *pues* / *doncs* (V3k), *bueno* / *bé* (V3c) i *o sea* / *o sigui* (V3i).

Figura 116. Arbre de classificació de la variable sexe amb les dades de l'estudi univariant.



Quant a l'arbre de classificació de la variable sexe generat a partir de les dades de l'estudi multivariant, com s'observa a la Figura 117, té una profunditat de tres nivells per sota del node arrel. La Taula 117 mostra que el percentatge de classificacions correctes és una mica inferior que l'anterior, però tot i així supera el 70 %. L'error estàndard és del 0,038.

Taula 117. Taula de classificació de la variable sexe amb les dades de l'estudi multivariant.

Observat	Predit		
	Masculí	Femení	Percentatge correcte
Masculí	42	22	65,6 %
Femení	16	58	78,4 %
Percentatge global	42 %	58 %	72,5 %

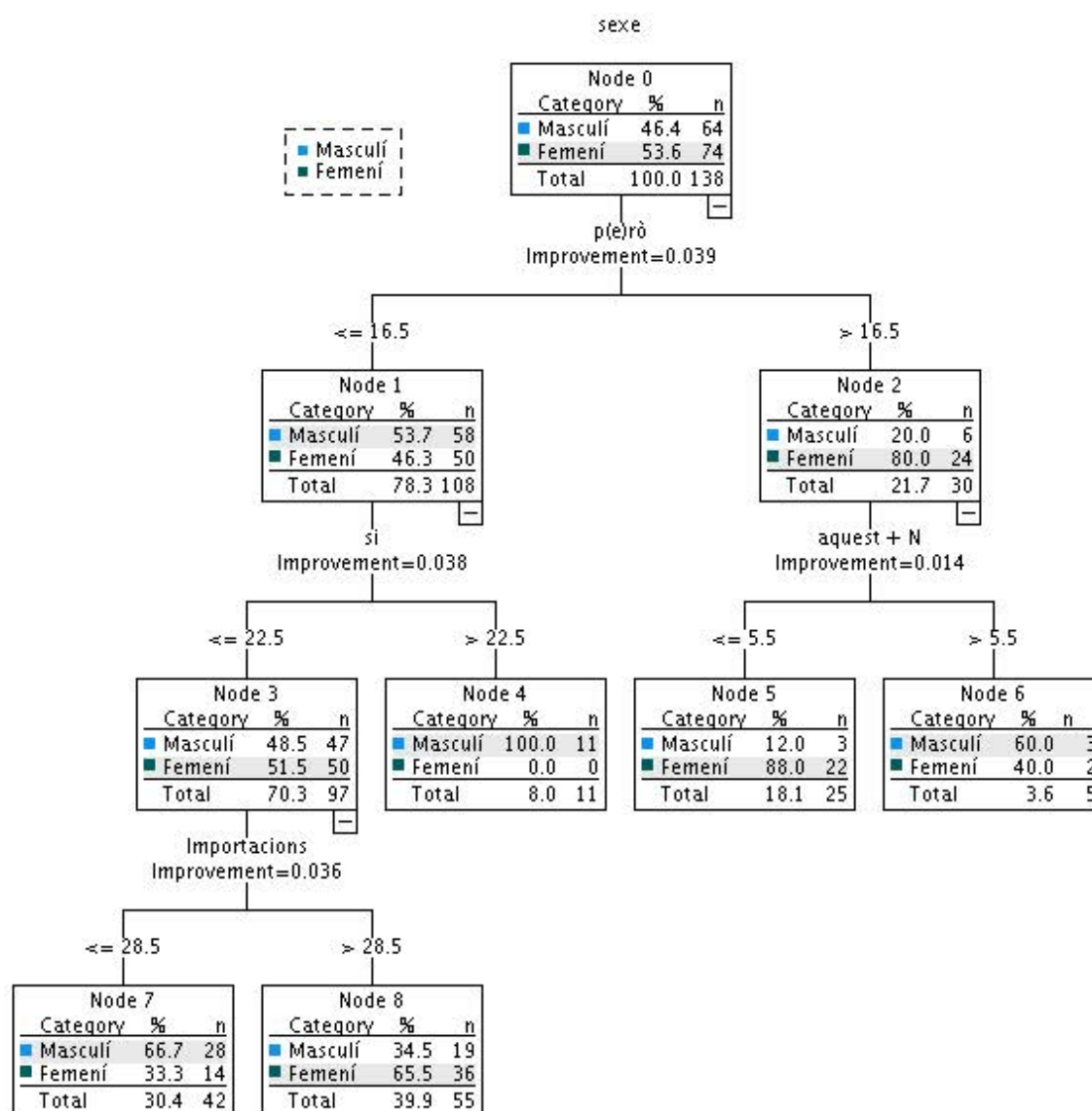
Mètode de creixement: CRT.

Variable dependent: sexe.

La primera segmentació que es fa del node 0 té en compta les realitzacions de la variant de $V3j$ $p(e)rò$: les transcripcions amb un total de 16,5 o menys ocurrencies, s'agrupen en el node 1 i les que en presenten més, en el node 2. En el primer cas, la majoria (el 53,7 %) corresponen a informants nois i, en el segon, la majoria (el 80 %) són d'informants noies. Al segon nivell hi veiem que el node 1 es segmenta a partir de la variant si de $V18$ i el node 2 a partir de la variant $aquest + N$ de $V23$. Les transcripcions del node 1 que contenen 22,5 realitzacions de la conjunció condicional si o menys conformen el node 3, la majoria de les quals corresponen a noies, mentre que les que en contenen més de 22,5 formen el node terminal 4 i corresponen totes a informants nois. En canvi, el llindar de realitzacions de la variant $aquest + N$ utilitzat per segmentar el node 2 és força més baix: les transcripcions que en presenten 5,5 ocurrencies o menys formen el node 5 i les que en contenen més de 5,5, el node 6. La distribució dels sexes en aquests dos nodes terminals és inversa, ja que al node 5 la majoria (el 88 %) de transcripcions provenen d'informants noies, però al node 6 la majoria (el 60 %) provenen d'informants nois. Retornant al node 3, l'únic del segon nivell que no és terminal, veiem que se segmenta en els nodes 7 i 8, formats respectivament per una majoria de transcripcions provinents de nois i una majoria de transcripcions d'informants noies, mitjançant la variant Importacions de $V38$. El node 7 conté les transcripcions amb 28,5 realitzacions o menys i el node 8 les que en tenen més de 28,5 ocurrencies.

En definitiva, les variants de l'estudi multivariant seleccionades per caracteritzar els grups de sexe del corpus d'anàlisi són *p(e)rò* (V3j3), *si* (V18b), *aquest + N* (V23a) i *Importacions* (V38a).

Figura 117. Arbre de classificació de la variable sexe amb les dades de l'estudi multivariant.



En últim lloc, tractem a continuació els dos arbres de classificació que caracteritzen els grups de llengua inicial. Aquests són els models amb percentatges de classificació correcta més reduïts. El primer, generat a partir de l'estudi univariant, assolix el 62,3 % de classificacions encertades, com s'observa a la Taula 118. L'error estàndard, per tant, se situa en aquest cas al 0,041. L'arbre té una profunditat de 4 nivells sota el node arrel.

Taula 118. Taula de classificació de la variable llengua inicial amb les dades de l'estudi univariànt.

Observat	Predit			
	Català	Castellà	Català i castellà	Percentatge correcte
Català	25	26	3	46,3 %
Castellà	5	55	4	85,9 %
Català i castellà	4	10	6	30 %
Percentatge global	24,6 %	65,9 %	9,4 %	62,3 %

Mètode de creixement: CRT.

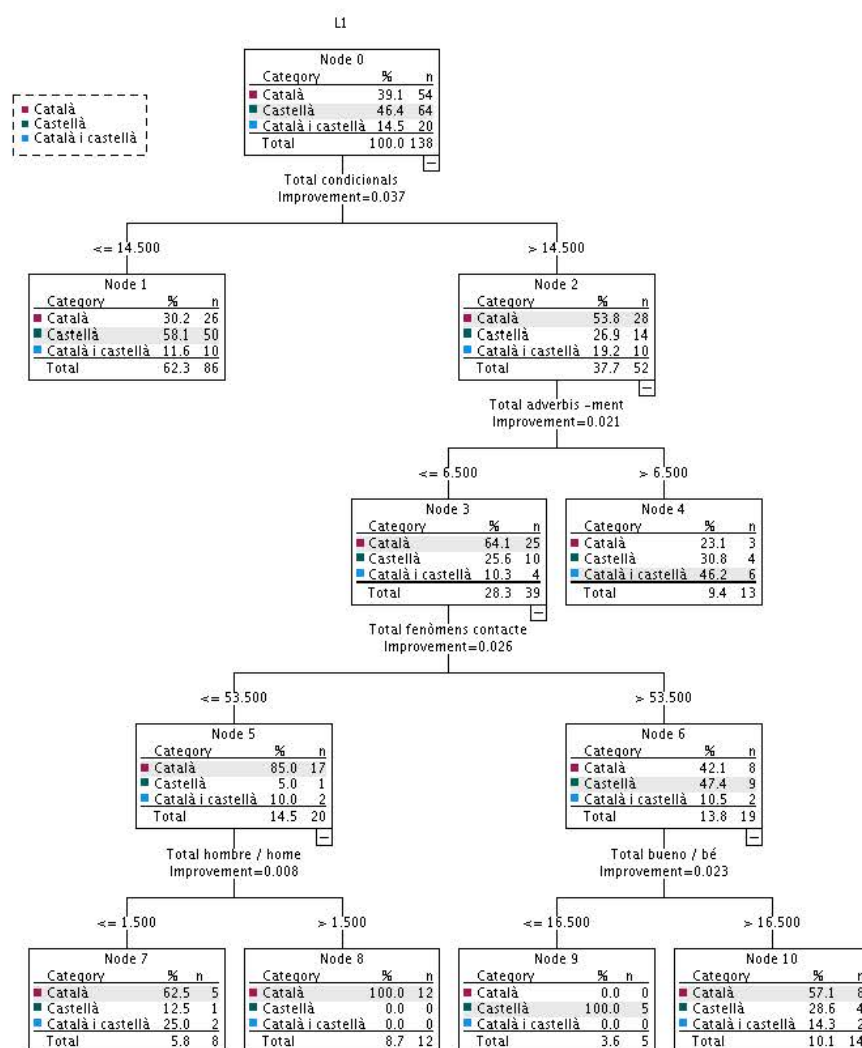
Variable dependent: L1.

En aquest arbre, el node 0, que mostra la variable dependent llengua inicial, se segmenta a través del nombre de construccions condicionals (V18) en el node 1, que conté les transcripcions amb 14,5 o menys realitzacions d'aquesta variable, i el node 2, que conté les que en tenen més de 14,5. El grup de llengua inicial més nombrós al primer node és el de castellà (amb un 58,1 %), mentre que, al segon node, el grup més nombrós és el de català (amb pràcticament un 54 %). El node 1 és terminal, però el node 2 es divideix en els nodes 3 i 4, segons si el recompte d'adverbis acabats en *-ment* (V45) és igual o inferior a 6,5 (node 3) o superior a aquesta xifra (node 4). El node 3 està format principalment per transcripcions provinents d'informants catalanoparlants inicials (en un 64 %), mentre que el node 4 conté més transcripcions d'informants bilingües inicials (46,2%) que dels altres dos grups de llengua inicial. D'aquests dos nodes, només el 3 es torna a segmentar. Aquesta vegada, mitjançant la variable fenòmens de contacte de llengües (V38): les transcripcions que en tenen 53,5 o menys (que són majoritàriament de catalanoparlants inicials) s'agrupen al node 5 i les que en presenten més de 53,5 (un 47,4 % de les quals provenen de castellanoparlants inicials) es troben al node 6. Tots dos es divideixen una darrera vegada a través de variables que recullen el total de partícules discursives. En el cas del node 5, la següent variable seleccionada és el total de *hombre / home* (V3h): les transcripcions que en presenten 1,5 realitzacions o menys formen el node 7 i les que en tenen més realitzacions, el node 8. Tant en el node 7 com en el 8 el grup més nombrós és el de catalanoparlants inicials, tot i que al node 8 també és

l'únic que hi té representació. Pel que fa al node 6, es segmenta un últim cop mitjançant la variable total de *bueno / bé* (V3c): els casos amb 16,5 realitzacions o menys formen el node 9 (i provenen tots d'informants castellanoparlants inicials), mentre que els casos amb més de 16,5 realitzacions componen el node 10 (un 57 % dels quals provenen de catalanoparlants inicials).

Així doncs, les variables de l'estudi univariant seleccionades pel model de classificació no lineal per caracteritzar la llengua inicial dels informants del corpus d'anàlisi són el total de construccions condicionals (V18), d'adverbis acabats en *-ment* (V45), de fenòmens de contacte de llengües (V38) i dels conjunts de partícules discursives *hombre / home* (V3h) i *bueno / bé* (V3c).

Figura 118. Arbre de classificació de la variable llengua inicial amb les dades de l'estudi univariant.



En el cas de la llengua inicial, el model assoleix un percentatge de classificació correcta més elevat a partir de les dades de l'estudi multivariant. Com es pot veure a la Taula 119, en aquest cas es supera el 70 % de classificacions correctes. L'error estàndard és del 0,039 i, com s'explica a continuació, l'arbre generat a partir d'aquestes dades té una profunditat de 5 nivells per sota del node arrel.

Taula 119. Taula de classificació de la variable llengua inicial amb les dades de l'estudi multivariant.

Observat	Predit			Percentatge correcte
	Català	Castellà	Català i castellà	
Català	36	17	1	66,7 %
Castellà	8	56	0	87,5 %
Català i castellà	7	8	5	25 %
Percentatge global	37 %	58,7 %	4,3 %	70,3 %

Mètode de creixement: CRT.

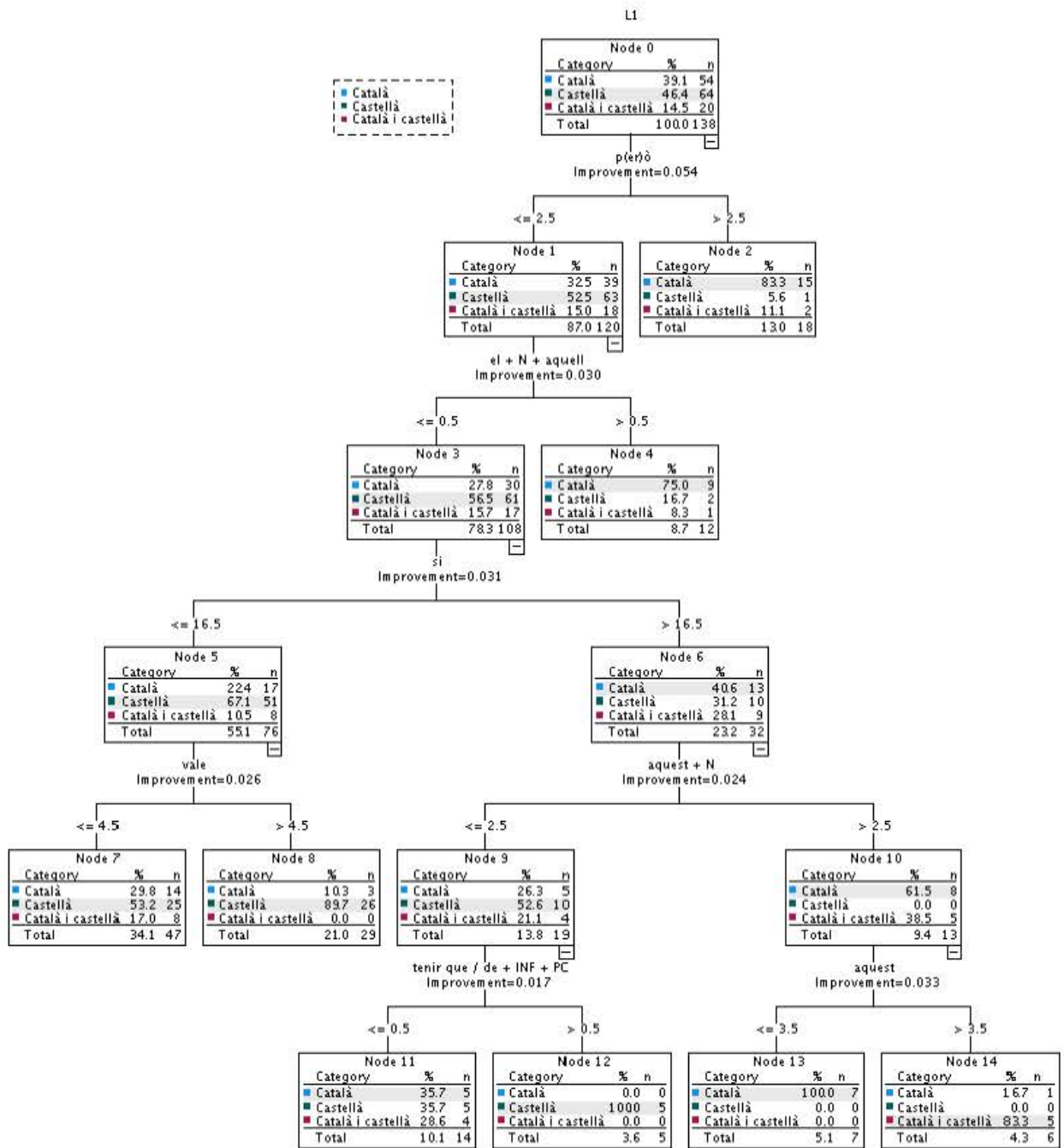
Variable dependent: L1.

Com mostra la Figura 119, la primera segmentació del node arrel en aquest arbre es fa mitjançant una de les variants de $V3j$, $p(er)ò$: les transcripcions amb 2,5 realitzacions o menys formen el node 1 i la resta es troben al node 2. Al node 1, el grup de llengua inicial més nombrós és el de castellanoparlants i, al node 2, terminal, és el de catalanoparlants. El node 1 es divideix en un nou nivell de nodes. Aquest cop, a través de la variant $el + N + aquell$ ($V23$), de forma que les transcripcions que en presenten 0,5 realitzacions o menys formen el node 3 i les que en contenen més realitzacions, el node 4, que és terminal i està format majoritàriament (en un 75 %) per transcripcions d'informants amb el català com a llengua inicial. El node 3, que conté un 56,5 % de transcripcions provinents de castellanoparlants inicials, es divideix entre les transcripcions que inclouen 16,5 realitzacions o menys de la conjunció condicional si ($V18$), que formen el node 5, i les que en tenen més realitzacions, que formen el node 6. El grup de llengua inicial més nombrós és diferent en aquests dos nodes: en el node 5, la majoria de transcripcions (un 67 %) són de castellanoparlants inicials, mentre que al node 6, la majoria (un 40,6 %) són de catalanoparlants inicials. El node 5 se segmenta en dos nodes terminals diferenciats pel total de realitzacions de $vale$ ($V31$): el

node 7 inclou transcripcions, la majoria (un 53,2 %) de castellanoparlants inicials, que en contenen 4,5 o menys i el node 8 està format per aquelles que en presenten més de 4,5, la majoria (gairebé el 90 %) de les quals són també de castellanoparlants inicials. En canvi, el node 6 se segmenta en dos nodes folla a través de la variant *aquest* + N (V23): les transcripcions del node 9 en presenten 2,5 realitzacions o menys i la majoria (el 52,6 %) són de castellanoparlants inicials i les transcripcions incloses al node 10 contenen més de 2,5 ocurrences d'aquesta variant i són majoritàriament (en un 61,5 % dels casos) de catalanoparlants inicials. El node 9 es segmenta en els nodes 11 i 12 per mitjà de la variant *tenir que / de* + INF + PC (V14b): el primer inclou les transcripcions amb 0,5 o menys realitzacions d'aquesta variant (que es distribueixen en percentatges força semblants entre els tres grups de llengua inicial), mentre que el segon conté les que mostren més realitzacions de V14b1, que, arribats a aquest punt de l'arbre, són totes de castellanoparlants inicials. Quant al node 10, es divideix en els nodes 13 i 14 per mitjà de les realitzacions de la variant *aquest* (V23). El node 13 conté les transcripcions que, a més de la resta de criteris descrits fins a arribar-hi, presenten 3,5 o menys realitzacions d'aquesta variant, que són totes de catalanoparlants inicials. En canvi, el node 14 inclou les que tenen més de 3,5 realitzacions de V23e, la majoria de les quals són de bilingües inicials (amb una única excepció d'una transcripció provinent d'un catalanoparlant inicial).

Per tant, les variants seleccionades en aquest últim arbre són *p(er)ò* (V3j4), *el* + N + *aquell* (V23d), *si* (V18b), *vale* (V3l1), *aquest* + N (V23a), *tenir que / de* + INF + PC (V14b1), *aquest* (V23e).

Figura 119. Arbre de classificació de la variable llengua inicial amb les dades de l'estudi multivariant.



En aquest apartat hem explorat les relacions entre les variables lingüístiques i sociolingüístiques de l'estudi des d'una perspectiva complementària a l'adoptada fins ara. En comptes de demanar-nos si l'edat, el sexe i la llengua inicial es correlacionen amb les variables lingüístiques, els arbres de classificació ens aporten informació sobre les característiques lingüístiques més habituals de cada grup d'informants. D'aquesta manera, llegint el recorregut que ens porta fins

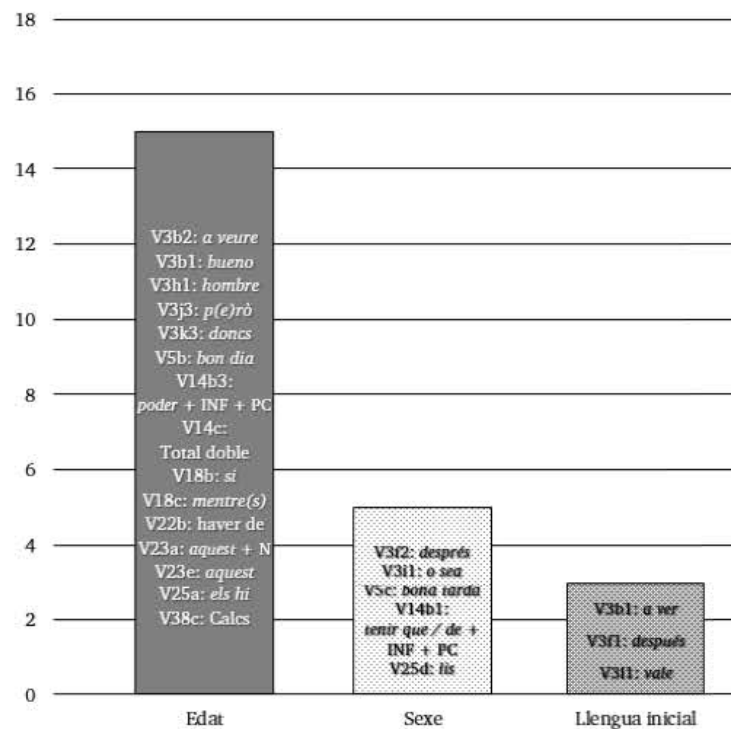
cadascun dels nodes terminals del primer arbre, una de les conclusions que en podem extreure és, per exemple, que en el cas que trobéssim a l'atzar una de les transcripcions del corpus d'anàlisi amb més de 18,5 realitzacions de V3j, més de 0,5 de V3i i 4,5 o menys realitzacions de V6, podríem estar segurs que pertany a l'edat de 16 anys, perquè seria una de les que formen el node 8 de l'arbre que caracteritza els grups d'edat a partir de les dades de l'estudi univariant. D'aquesta manera, cada node dels arbres que hem obtingut ens aporta informació sobre la probabilitat que una transcripció qualsevol de les que formen el corpus d'anàlisi provingui d'un informant de 12 o 16 anys, noi o noia i amb el català, el castellà o totes dues llengües com a inicials segons el nombre de realitzacions de les variables i variants seleccionades pels models de classificació.

6.5. Anàlisis estadístiques: síntesi

En definitiva, en aquest capítol, hem pogut observar que les variables analitzades al corpus mantenen diverses relacions entre elles. Els resultats de l'estudi estadístic descriptiu mostren, d'una banda, que la variable amb medianes de realització més elevades a tot el corpus d'anàlisi és la dels fenòmens de contacte de llengües (V38) i, especialment, les importacions (V38a), amb valors superiors a les 40 realitzacions en grups com els de les noies catalanoparlants i castellanoparlants inicials o els nois bilingües inicials als 16 anys. D'altra banda, també hem observat que alguns grups d'informants bilingües presenten medianes força diferents de les de la resta d'informants del corpus en diverses variables. Més concretament, els nois bilingües inicials mostren, als 12 anys, valors més elevats que la resta de grups en certes variables i variants pragmàtiques (com V3i1 o V3l2), morfosintàctiques (com V22a) i lèxiques (V41b). Pel que fa les noies bilingües inicials a la mateixa edat, hem comprovat que presenten medianes de realització inferiors que altres grups en variables i variants pragmàtiques com V3c1, morfosintàctiques (V23) i lèxiques (V38). Als 16 anys, tant els nois com les noies d'aquest grup de llengua inicial presenten medianes superiors a altres grups d'informants en variables i variants pragmàtiques (per exemple, V3k3), morfosintàctiques (TV14a, V22a) i també lèxiques (V45).

Quant als resultats de l'Anàlisi de Components Principals, l'edat és la variable independent que, segons els gràfics de saturacions als components rotats de l'estudi multivariant, es correlaciona clarament amb més variants lingüístiques (quinze), seguida a certa distància pel sexe (amb cinc) i la llengua inicial (amb només tres variants). Així doncs, aquests resultats indiquen que, de les 19 variables incloses a l'estudi multivariant, quinze (és a dir, un 78,95 %) estan formades per alguna variant que, aparentment, podria correlacionar-se de manera estadísticament significativa amb l'edat, el sexe i/o la llengua inicial dels informants (les que apareixen a la Figura 120).

Figura 120. Variants de l'estudi multivariant que es correlacionen amb les variables independents segons l'Anàlisi de Components Principals.



També hem observat els resultats de les proves de significació estadística. Com hem comprovat, deu de les 18 variables de l'estudi univariant (el 55,56 %) i 31 de les 64 variants (un 48,44 %) de l'estudi multivariant mostren una relació estadísticament significativa amb una o dues variables independents. Només un dels factors dependents de tot l'estudi (la variant V3j4: *p(er)ò*) ha retornat valors p per sota de 0,05 a les tres proves i manté, per tant, relacions estadísticament significatives amb totes tres variables independents alhora.

Com resumeix a continuació la Taula 120, les variables i variants per a les quals s'ha trobat una relació estadísticament significativa amb l'edat dels informants són la partícula discursiva *a ve(u)re* (V3b3) i el recompte total de V3b; *bueno* (V3c1) i el total de V3c; *home* (V3h2) i el total de V3h; *o sea* (V3i1), *o sigui* (V3i2) i el total de V3i; *p(er)ò* (V3j3), *p(er)ò* (V3j4) i el total de V3j; les variants *pos* (V3k2), *val* (V3l2) i *d'acord* (V3l3); el total d'heterorepeticions (V6); la posició proclítica dels pronoms febles que precedeixen les perífrasis verbals d'infinitiu *tenir que* i *tenir de* (V14a1) i *anar a* (V14a4), així com la posició enclítica amb *poder* + infinitiu (V14b3); la conjunció condicional *si* (V18b) i el recompte de construccions condicionals (V18); les perífrasis verbals d'obligació *tenir que* (V22a) i *haver de* (V22b) seguits d'infinitiu; el determinant demostratiu *aquest* (i formes flexionades) abans de substantiu (V23a), i el total d'adverbis acabats en *-ment* (V45). De forma semblant, les variables i variants que presenten una relació significativa amb el sexe dels informants són les partícules discursives *a ver* (V3b1) i *a ve(u)re* (V3b3), a més del recompte de V3b; les variants *després* (V3f2) i *p(er)ò* (V3j4); el recompte d'heterorepeticions (V6); el nombre de preguntes obertes (V17a); la conjunció condicional *mentre(s)* (V18c), i el total de perífrasis verbals d'obligació (V22). Finalment, les variables i variants que es mostren correlacionades amb la llengua inicial segons els resultats de les proves estadístiques són *home* (V3h2) i el total de V3h; *o sea* (V3i1) i *o sigui* (V3i2); *p(er)ò* (V3j4); *bona tarda* (V5c) i *hola bon dia* (V5d); la posició proclítica dels pronoms amb *anar a* + infinitiu (V14a4); la conjunció *si* (V18b) i el total de construccions condicionals (V18); la perífrasi verbal *caldre* + infinitiu (V22e); els sintagmes demostratius del tipus *el* (i formes flexionades) + substantiu + *aquell* (i formes flexionades) (V23d); els pronoms clítics de datiu plural *els hi* (V25a), *els* (V25b), *li* (V25c) i *lis* (V25d), independentment de la posició en què apareixen; els fenòmens de contacte de llengües que agrupa la variant *Altres* (V38d), i les referències simples a colors (V41b), així com el recompte total de mencions a colors (V41).

Taula 120. Variables i variants que mostren relacions estadísticament significatives amb les variables independents.

Variables		Variants	Edat	Sexe	L1
V3: Doblets de partícules discursives	V3b: <i>a ver</i> / <i>a veure</i>	V3b1: <i>a ver</i>	X	✓	X
		V3b3: <i>a ve(u)re</i>	✓	✓	X
	Total V3b <i>a ver</i> / <i>a veure</i>		✓	✓	X
	V3c: <i>bueno</i> / <i>bé</i>	V3c1: <i>bueno</i>	✓	X	X
	Total V3c <i>bueno</i> / <i>bé</i>		✓	X	X
	V3f: <i>después</i> / <i>després</i>	V3f2: <i>després</i>	X	✓	X
	V3h: <i>hombre</i> / <i>home</i>	V3h2: <i>home</i>	✓	X	✓
	Total V3h <i>hombre</i> / <i>home</i>		✓	X	✓
	V3i: <i>o sea</i> / <i>o sigui</i>	V3i1: <i>o sea</i>	✓	X	✓
		V3i2: <i>o sigui</i>	✓	X	✓
	Total V3i <i>o sea</i> / <i>o sigui</i>		✓	X	X
	V3j: <i>pero</i> / <i>però</i>	V3j3: <i>p(e)rò</i>	✓	X	X
		V3j4: <i>p(er)ò</i>	✓	✓	✓
	Total V3j <i>pero</i> / <i>però</i>		✓	X	X
	V3k: <i>pues</i> / <i>doncs</i>	V3k2: <i>pos</i>	✓	X	X
	V3l: <i>vale</i> / <i>d'acord</i>	V3l2: <i>val</i>	✓	X	X
V3l3: <i>d'acord</i>		✓	X	X	
V5: Obertura de la interacció		V5c: <i>bona tarda</i>	X	X	✓
		V5d: <i>hola bon dia</i>	X	X	✓
Total V6 Heterorepetició			✓	✓	X
V14: Posició dels pronoms clítics (PC) en construccions perifràstiques	V14a: Proclisi	V14a1: PC + <i>tenir que</i> / <i>de</i> + INF	✓	X	X
		V14a4: PC+ <i>anar a</i> + INF	✓	X	✓
	V14b: Enclisi	V14b3: <i>poder</i> + INF + PC	✓	X	X
V17: Tipus de preguntes		V17a: Oberta	X	✓	X

V18: Construccions condicionals	V18b: <i>si</i>	✓	✗	✓
	V18c: <i>mentre(s)</i>	✗	✓	✗
Total V18 Construccions condicionals		✓	✗	✓
V22: Perífrasis d'obligació	V22a: <i>tenir que</i> + INF	✓	✗	✗
	V22b: <i>haver de</i> + INF	✓	✗	✗
	V22e: <i>caldre</i> + INF	✗	✗	✓
Total V22 Perífrasis d'obligació		✗	✓	✗
V23: Determinants demostratius	V23a: <i>aquest</i> + N	✓	✗	✗
	V23d: <i>el</i> + N + <i>aquell</i>	✗	✗	✓
V25: Pronoms clítics de datiu plural	V25a: <i>els hi</i> + V / V + <i>-els-hi</i> / V + <i>'ls-hi</i> / V + <i>-los-hi</i>	✗	✗	✓
	V25b: <i>els</i> + V / V + <i>-los</i> / V + <i>'ls</i>	✗	✗	✓
	V25c: <i>li</i> + V / V + <i>-li</i>	✗	✗	✓
	V25d: <i>lis</i> + V / V + <i>-lis</i>	✗	✗	✓
V38: Fenòmens de contacte de llengües	V38d: Altres	✗	✗	✓
V41: Colors	V41b: <i>[blau]</i>	✗	✗	✓
Total V41 Colors		✗	✗	✓
V45 Nombre d'adverbis acabats en <i>-ment</i>		✓	✗	✗

La classificació no lineal de les dades mitjançant els arbres de classificació de les variables edat, sexe i llengua inicial han permès caracteritzar els grups d'informants a partir d'una selecció de les variables lingüístiques. Pel que fa a la classificació de les transcripcions en grups d'edat, les dades de l'estudi multivariant mostren el percentatge de classificació correcta més elevat (del 85,5 %). Les variants seleccionades en aquest model, com hem vist anteriorment, són *o sigui* (V3i2), *val* (V3l2), *pues* (V3k1), *bueno* (V3c1), *p(er)ò* (V3j4), *doncs* (V3k3), *home* (V3h2), *poder* + INF + PC (V14b3) i *aquest* + N (V23a).

En el cas dels grups de sexe, el percentatge de classificació correcta més elevat (un 78,3 %) s'ha obtingut a partir de les dades de l'estudi univariant. Així mateix, les variables utilitzades són el recompte de perífrasis verbals d'obligació

(V22), el d'heterorepeticions (V6), el de construccions condicionals (V18) i els dels conjunts de partícules discursives *pues / doncs* (V3k), *bueno / bé* (V3c) i *o sea / o sigui* (V3i).

Quant a la variable llengua inicial, el model amb un percentatge de classificació correcta més alt (del 70,3 %) correspon a l'obtingut a partir de l'estudi multivariant, del qual l'algoritme n'ha seleccionat les variants *p(er)ò* (V3j4), *el + N + aquell* (V23d), *si* (V18b), *vale* (V3l1), *aquest + N* (V23a), *tenir que / de + INF + PC* (V14b1), *aquest* (V23e) com a més rellevants per a la caracterització dels catalanoparlants, castellanoparlants i bilingües inicials.

7 DISCUSSIÓ DELS RESULTATS

El disseny d'aquest estudi, que correspon a un estudi panel, com s'ha explicat anteriorment (capítol 5), ha permès relacionar algunes de les diferències observades en el comportament lingüístic dels informants que formen el corpus d'anàlisi a la seva edat (aproximadament, 12 o 16 anys). Ara bé, cal certa cautela a l'hora d'extreure conclusions respecte a la capacitat de les dades que hem tractat al capítol anterior per descriure diferents etapes del procés d'adquisició lingüística dels informants, si bé hem pogut identificar relacions estadísticament significatives entre algunes de les variables lingüístiques analitzades i els tres grups de llengua inicial observats. Com exposen Ninio i Snow (1996: 82), els estudis basats en l'observació de l'actuació lingüística d'informants en una situació i un espai de temps determinats no poden proporcionar mai informació sobre la competència absoluta dels parlants, ja que, per exemple, la no producció d'un element lingüístic concret no respon necessàriament a una manca de competència del parlant, sinó que pot deure's al fet que, simplement, no s'hagi produït el context necessari per a la producció d'aquest element durant la interacció observada.

Pel que fa al sexe, aquest estudi demostra la necessitat d'adoptar una perspectiva empírica que tingui en compte les possibles relacions entre el comportament lingüístic i el sexe dels parlants per tal de representar la realitat de l'ús de la llengua, ja que algunes variables han mostrat correlacions estadísticament significatives amb la variable sexe. Val a dir, però, que l'operacionalització d'aquesta variable utilitzada aquí, deguda a les limitacions de la recollida de dades, és millorable en tant que es basa en una distinció binària de la corporalitat i de certes formes d'expressió de gènere i, per tant, no aconsegueix representar correctament tot el ventall d'identitats possibles. A continuació, tractem en detall les implicacions dels resultats obtinguts per a cada variable lingüística de l'estudi en relació a la bibliografia consultada i a les expectatives exposades a l'apartat **5.3.2.3** (Taula 16) pel que fa a les correlacions entre variables dependents i independents. A més, ens ocupem també de les

aplicacions principals d'aquesta investigació i dels resultats de les anàlisis que hem realitzat.

7.1. Discussió dels resultats de les variables pragmàtiques

Segons la bibliografia sobre les diferències en el comportament interaccional prototípic d'homes i dones i, donat el disseny de les variables etiquetades com a V3, l'objectiu de les quals és comptabilitzar formes equivalents de partícules discursives utilitzades en varietats col·loquials de la llengua, s'esperava que les variables que recullen els doblats de partícules discursives mostressin correlacions estadísticament significatives amb el sexe i la llengua inicial dels informants. Quant a les altres dues variables pragmàtiques, les fórmules d'obertura de la interacció (V5) i el total d'heterorepeticions (V6), s'esperava que en mostressin amb l'edat, ja que podrien reflectir habilitats pragmàtiques que es desenvolupen a l'adolescència, i, en el cas de V5, també amb la llengua inicial.

Abans d'endinsar-nos en la interpretació dels resultats de les variables pragmàtiques, però, val a dir que estudis previs sobre la influència lèxica del contacte entre el castellà i el català a la parla de la població en edat escolar també han detectat l'ús de les partícules discursives castellanes analitzades aquí com a V3. En aquest sentit, destaquen els estudis de Vila i Moreno (1996) i Galindo Solé (2006), ja que tots dos llisten gairebé totes les formes castellanes que tractem a continuació (amb l'excepció de *a ver*) com a interferències lèxiques identificades als seus corpus. El primer estudi utilitza mostres obtingudes entre 1991 i 1992 d'alumnes de 10-11 anys i 13-14 anys de Santa Coloma de Gramenet, mentre que el segon analitza les pràctiques lingüístiques d'alumnes d'11-12 anys de 52 escoles catalanes a l'hora d'esbarjo. De fet, partícules com *bueno*, *pero*, *pues* o *vale* són alguns dels fenòmens lèxics de contacte de llengües més freqüents en aquests estudis (Vila i Moreno, 1996: 419; Galindo Solé, 2006: 453-455).

7.1.1. Doblets de partícules discursives: V3b *a ver* / *a veure*

Com vèiem anteriorment a les taules 25 i 26 de l'apartat 5.6, les partícules discursives *a ver* i *a veure*, juntament amb les variants *a ve(u)re* i *aviam*, comptabilitzen un total de 250 realitzacions al corpus d'anàlisi. Ara bé, la distribució de les quatre variants és força desigual. Mentre que *a ver* i *aviam* presenten uns totals de 12 i 11 realitzacions, respectivament, *a veure* i *a ve(u)re* arriben a 45 i 182 aparicions. Tot i que caldria una anàlisi més detallada dels usos que es fan d'aquestes partícules al corpus per tal de confirmar-la, una possible explicació d'aquestes diferències en la freqüència d'aparició de les diverses variants es podria relacionar amb els graus de gramaticalització de les formes que hem observat al corpus, ja que, segons el contínuum que proposen Cuenca i Marín (2000: 235), *aviam* presentaria un grau de gramaticalització més avançat que *a veure* (o *a ve(u)re*) i *a ver*.

De les correlacions que s'esperava identificar entre les variables independents i V3b (sexe i llengua inicial), s'ha pogut corroborar que les diferències en l'ús d'aquestes partícules discursives entre els nois i les noies del corpus són estadísticament significatives. Concretament, si ens fixem en el total de V3b, són les noies les que en fan un ús més elevat, com mostra la Taula 121. Les dades de les dues variants que també mostren valors de significació menors a 0,05 en relació a la variable sexe indiquen que les noies produeixen més realitzacions de la forma *a ver* (V3b1) als 16 anys i de la variant *a ve(u)re* (V3b3) als 12 anys. En efecte, tan sols en el cas de la variant *aviam* (V3b4) trobem valors més elevats en els grups dels nois que en els de les noies.

Aquests resultats, per tant, divergeixen del que sabem sobre una construcció semànticament propera a les que recull V3b, *vamos a ver*. De fet, autores com Cuenca i Marín (200: 216), Montolío i Unamuno (2001: 195) i González-Sanz (2017) expliquen que la partícula castellana *a ver* s'hauria originat a partir de *vamos a ver*. Segons González-Sanz (2017: 234), sembla que aquesta construcció és utilitzada “mayoritariamente por participantes masculinos, lo que podría estar relacionado con un intento del hablante por alcanzar el poder en la interacción (Llorente, 1996: 128)”.

D'altra banda, malgrat que no s'ha pogut confirmar que existeixen diferències significatives quant a l'ús de les partícules discursives *a ver* / *a veure* entre els grups de sexe ni llengua inicial, sí que n'hem trobat pel que fa a les dues edats analitzades. Globalment, al corpus analitzat, els informants fan més ús d'aquesta variable als 16 anys. Quant a les variants, l'única que mostra resultats estadísticament significatius respecte l'edat, com hem vist, és *a ve(u)re* (V3b3), que és més utilitzada pels informants als 16 anys (com la resta de variants excepte *a ver*).

Taula 121. Realitzacions totals de la variable V3b i les seves variants segons el sexe i l'edat dels informants.

Estudi multivariant	Masculí		Femení	
	12 anys	16 anys	12 anys	16 anys
V3b1: <i>a ver</i>	2 (\bar{x} : 0,06; DE: 0,25)	0 (\bar{x} : 0; DE: 0)	6 (\bar{x} : 0,16; DE: 0,44)	4 (\bar{x} : 0,11; DE: 0,31)
V3b2: <i>a veure</i>	7 (\bar{x} : 0,22; DE: 0,55)	9 (\bar{x} : 0,28; DE: 0,58)	11 (\bar{x} : 0,3; DE: 0,66)	18 (\bar{x} : 0,49; DE: 1,26)
V3b3: <i>a ve(u)re</i>	5 (\bar{x} : 0,16; DE: 0,51)	21 (\bar{x} : 0,66; DE: 1,12)	36 (\bar{x} : 1; DE: 2,48)	120 (\bar{x} : 3,24; DE: 7,25)
V3b4: <i>aviam</i>	3 (\bar{x} : 0,09; DE: 0,53)	4 (\bar{x} : 0,13; DE: 0,42)	1 (\bar{x} : 0,03; DE: 0,16)	3 (\bar{x} : 0,08; DE: 0,28)
Estudi univariant				
Total V3b	17 (\bar{x} : 0,53; DE: 0,98)	34 (\bar{x} : 1,06; DE: 1,27)	54 (\bar{x} : 1,46; DE: 2,8)	145 (\bar{x} : 3,92; DE: 7,95)

Així doncs, si utilitzéssim el corpus analitzat aquí com una mostra representativa de referència per a l'estudi d'una mostra anònima, sembla que un ús elevat de la variable V3b indicaria, des d'una perspectiva estadística, una probabilitat més elevada que es tractés d'una parlant (noia) que d'un parlant (noi). Al mateix temps, segons els resultats obtinguts, també indicaria l'existència de més probabilitats que la mostra correspongués a un parlant més proper als 16 anys d'edat que no pas als 12.

Això no obstant, els resultats obtinguts no ofereixen una comprensió de l'ús que fan els informants d'aquestes partícules discursives, per la qual cosa seria interessant complementar els resultats que ens ocupen amb aproximacions qualitatives com les adoptades en estudis previs. Per exemple, Montolío i Unamuno (2001) estudien *a ver* i *a veure* (enteses com a formes “idèntiques” en ús i significat, 2001: 193) en interaccions entre alumnat i professorat i afirmen que “the use of this particle in these conversational situations not only expresses linguistic-procedural information – the natural territory of discourse markers – but also sociolinguistic information; that is to say, it specifies which speaker is controlling [sic] the conversation” (Montolío i Unamuno, 2001: 194). A més, en el seu estudi de la partícula *a ver* en un corpus de tertúlies televisades, González-Sanz (2017: 236) apunta, que *a ver* té usos diferents en contextos monologals i dialogals i, quant als darrers, distingeix entre el pla modal (en què “mitiga la fuerza ilocutiva del acto de habla y manifiesta un valor modal de deseo”) i l'interactiu (en el qual “funciona bien como marca introductora de intervenciones, bien como mecanismo de distribución del turno de habla”).

7.1.2. Doblets de partícules discursives: V3c *bueno* / *bé*

Els resultats globals de V3c demostren que *bueno* és molt més freqüent al corpus d'anàlisi que *bé*, ja que el total de realitzacions de cada variant és de 2.106 i 19, respectivament. Aquests resultats sobre l'alta freqüència de *bueno* s'alineen amb l'estudi previ d'aquestes dues partícules d'Espuny (1998), González (1998) i Vila i Moreno (1998). Si bé totes dues partícules comparteixen la majoria de funcions pragmàtiques que duen a terme, es diferencien en els contextos en què són utilitzades, ja que *bueno* “es troba, normalment, dins un discurs de to intimista, personal i familiar”, a diferència de *bé*, marcadament formal (González, 1998: 254).

L'exploració del corpus de 16 textos descrit a Tomàs i Pallejà (1998) li permet a González (1998) confirmar les tres hipòtesis de les quals parteix: que les funcions del marcador anglès *well* coincideixen, a grans trets, amb les de *bueno* i *bé*; que el propòsit comunicatiu i el grau de formalitat són els factors que expliquen la freqüència de *bueno* i *bé* als textos, i que els textos interactius i

informals presenten moltes més realitzacions de *bueno* que no de *bé*. Les dues darreres també expliquen les nostres dades. Malgrat que la situació comunicativa a la qual corresponen aquestes últimes és una prova acadèmica, la dinàmica de joc de rol proposada als informants fa que, en realitat, les converses que mantenen amb les entrevistadores conformin textos interactius i més aviat informals. Així doncs, tot i que saben que l'objectiu de l'activitat és avaluar la seva competència lingüística, els informants adopten un estil comunicatiu col·loquial que propicia l'ús de partícules discursives no normatives com *bueno* (cf. González, 1998: 249). I ho fan, a jutjar pels resultats de les proves de significació estadística, independentment dels seus sexe i llengua inicial. En aquest sentit, els resultats obtinguts per a aquesta variable reforcen l'afirmació que, almenys en aquest cas, el coneixement de les recomanacions lingüístiques no influeixen en els elements lingüístics utilitzats per molts parlants:

“La normativa de la llengua, en aquest cas la catalana, ens dóna unes pautes a seguir: *bé* és la forma «correcta», el que s'ha de dir, no *bueno*. La realitat, però, ens demostra que una quantitat considerable de parlants utilitza la segona forma en lloc de la primera, particularment en textos orals espontanis, interactius i informals. [...] Ens trobem amb casos en què el parlant és totalment conscient de la normativa de la llengua i, malgrat tot, al llarg del seu discurs utilitza molts més cops *bueno* que no pas *bé*.” (González, 1998: 247).

Com hem observat anteriorment, tan sols la prova estadística corresponent a la variable edat ha retornat valors p significatius per a *bueno* (V3c1) i per al total de V3c. La Taula 122 mostra que els parlants fan més realitzacions de V3c1 a l'edat de 16 anys que a la de 12 i que, amb el temps, les de *bé* disminueixen, si bé no de forma estadísticament significativa.

Taula 122. Realitzacions totals de la variable V3c i les seves variants segons l'edat dels informants.

Estudi multivariant	12 anys	16 anys
V3c1: <i>bueno</i>	905 (\bar{x} : 13,12; DE: 14,72)	1.201 (\bar{x} : 17,41; DE: 14,02)
V3c2: <i>bé</i>	12 (\bar{x} : 0,17; DE: 0,66)	7 (\bar{x} : 0,1; DE: 0,39)
Estudi univariant		
Total V3c	917 (\bar{x} : 13,29; DE: 14,76)	1.208 (\bar{x} : 17,51; DE: 14)

A més de reflectir el procés d'adquisició de les funcions pragmàtiques que pot assumir *bueno* en un context interaccional, l'explicació d'aquesta progressió en l'ús de *bueno* amb l'edat podria tenir diversos components. En primer lloc, com assenyalàvem a l'apartat 5.3, un dels inconvenients dels estudis panel és que l'exposició reiterada dels informants a les mateixes eines de recollida de dades (en el nostre cas, l'activitat de joc de rol) pot influenciar la seva actuació (Bretxa, 2014: 123). D'aquesta manera, i tenint en compte que, segons González (1998), *bueno* s'utilitza en interaccions de to informal i familiar, l'augment en les mostres recollides a la tercera onada del projecte podria estar relacionat amb l'increment de la familiaritat dels informants amb la situació comunicativa, tot i que la sensació de les entrevistadores és que, d'una onada a la següent, els informants no recordaven gaire el contingut de les proves. En segon lloc, podria ser que els informants mostressin una capacitat d'adequació a la situació comunicativa de compra i venda plantejada al joc de rol més desenvolupada als 16 que als 12 anys, la qual cosa repercutís en l'ús de *bueno* com a recurs pragmàtic adequat a aquest context comunicatiu, o que l'increment en l'edat també augmentés la sensació de proximitat entre els informants i l'entrevistadora. Ara bé, en tercer lloc, l'augment de realitzacions de *bueno* també podria relacionar-se amb una menor implicació en la situació comunicativa proposada, ja que, com a partícula discursiva, serveix per cohesionar el discurs oral a nivell semàntic i sintàctic, a diferència de *bé*, que “no aporta cap contingut semàntic o sintàctic al text”, sinó que serveix principalment de lligam entre emissions (González, 1998: 244-245). Una vegada més, això no obstant, aclarir l'explicació de la progressió observada

en l'ús de *bueno* queda fora de l'abast d'aquest estudi i requeriria una anàlisi qualitativa de les dades.

El que sí que permeten fer els resultats de V3c és servir de referència per a anàlisis de mostres comparables a les observades aquí. De manera que, en el cas que un parlant desconegut produís, en una interacció de característiques semblants a les analitzades, un nombre elevat de *bueno*, podria entendre's com un indicador d'una edat més pròxima als 16 que als 12 anys.

7.1.3. Doblets de partícules discursives: V3f *después* / *després*

Com vèiem a l'apartat 5.6, el total d'aparicions de V3f al corpus d'anàlisi és de 502 realitzacions, 29 de les quals corresponen a *después* (V3f1) i les altres 473 a *després* (V3f2). Ara bé, fins on arriba el nostre coneixement, no disposem de treballs previs que permetin, d'entrada, afirmar que aquests resultats són o no esperables més enllà de constatar que la variant genuïna en la llengua en què es realitzava la prova interactiva, el català, (és a dir, *després*) és més freqüent que la originada pel contacte amb el castellà (*después*). El treball d'Espuny (1998), però, mostra que el nostre no és el primer corpus oral en el qual *després* alterna amb *después*, si bé el que analitza aquesta autora prové de parlants adults.

De fet, en comparació a altres partícules discursives, el volum de treballs que s'ocupen de *després* (o *después*) és més limitat. No obstant això, destaca un primer grup d'estudis, d'orientació més aviat teòrica, que tracten aquesta partícula. En efecte, *después* s'ha inclòs com a ordenador o estructurador de la informació en diverses obres sobre partícules discursives, inclosos el capítol sobre marcadors del discurs de Martín Zorraquino i Portolés Lázaro a la *Gramática descriptiva de la lengua española* (1999: 4086) o l'obra de Landone (2009: 122 i 144) sobre aquests elements i la seva relació amb la cortesia verbal. En la mateixa línia, Borja i Sanz (2009: 121) inclou *després* a la seva proposta de classificació dels marcadors discursius en català, a la categoria d'ordenadors del discurs i, específicament, a la de marcadors de continuïtat.

Un segon grup d'estudis, que podem caracteritzar de variacionistes, tracten *después* des d'una perspectiva empírica i exploren l'ús d'aquesta forma com a sinònima de *luego* a diverses comunitats de parla. Per exemple, Santana (2016) analitza una mostra de parla culta de Sevilla provinent de 30 parlants de tres generacions diferents i hi troba que *después* és més freqüent que *luego* en les tres funcions que n'estudia (del primer en comptabilitza 138 realitzacions, mentre que del segon només 90). En aquest estudi no s'observen diferències entre homes i dones en l'ús de *después*. Al de González Marfud i Perdomo Carmona (2021: 109), dut a terme sobre 48 entrevistes de parlants de l'Havana, en canvi, sí que s'observa un ús lleugerament més elevat d'aquesta partícula per part dels homes que no pas de les dones, tant com a conector temporal (amb 63 realitzacions en els parlants home i 37 en les dones) com en la funció d'estructurador de la informació (amb 48 i 30 realitzacions, respectivament). Això no obstant, cal assenyalar que, en no haver-se realitzat proves de significació estadística a partir de les dades obtingudes en aquest darrer treball, no es pot afirmar que les diferències observades siguin més que una tendència.

En qualsevol cas, aquesta tendència sembla corroborar-se en el nostre estudi, ja que, com hem vist, la variant *després* (V3f2) mostra resultats estadísticament significatius per la variable independent sexe, corroborant així la subhipòtesi que ens plantejàvem. Concretament, les diferències significatives es troben a les dades de 12 anys. La Taula 123 recull les realitzacions totals de V3f i mostra que, efectivament, els nois produeixen més la partícula *després* que les noies als 12 anys.

Taula 123. Realitzacions totals de la variable V3f i les seves variants segons el sexe i l'edat dels informants.

Estudi multivariant	Masculí		Femení	
	12 anys	16 anys	12 anys	16 anys
V3f1: <i>después</i>	0 (\bar{x} : 0; DE: 0)	0 (\bar{x} : 0; DE: 0)	14 (\bar{x} : 0,38; DE: 1,55)	15 (\bar{x} : 0,42; DE: 1,78)
V3f2: <i>després</i>	139 (\bar{x} : 4,34; DE: 3,96)	105 (\bar{x} : 3,28; DE: 3,39)	111 (\bar{x} : 3; DE: 4,6)	118 (\bar{x} : 3,28; DE: 3,68)

Així doncs, en treballs futurs per als quals el corpus analitzat aquí pugui servir de referència, es podrà recórrer a la variable V3f (concretament, a la variant *després*) per obtenir un indicador del sexe més probable dels parlants d'aproximadament 12 anys d'edat.

7.1.4. Doblets de partícules discursives: V3h *hombre* / *home*

El total de realitzacions de V3h identificades al corpus, com vèiem anteriorment, és de 229. Aquestes, es distribueixen en el mateix sentit que el descrit per al doblet de partícules discursives immediatament anterior: la variant catalana (*home*) és més freqüent que la castellana (amb 218 i 11 realitzacions, respectivament). Per tal d'interpretar els resultats obtinguts en relació a V3h, resumim a continuació els treballs disponibles que s'han ocupat d'*home* i *hombre*.

Pel que fa a la partícula discursiva *hombre*, ha despertat l'interès de diversos pragmatistes i sociolingüistes (cf. Ruiz-González, 2019: 82). Per exemple, en un estudi recent, Ruiz-González (2019) analitza el seu ús a la secció de Granada del corpus recollit al projecte PRESEEA (*Proyecto para el Estudio Sociolingüístico del Español de España y América*). Concretament, investiga les funcions pragmàtiques d'*hombre* i la seva relació amb tres variables sociolingüístiques: l'edat, el sexe i el nivell educatiu de 54 informants. Els resultats d'aquest treball mostren que els informants més joves (independentment de la funció pragmàtica observada) i els del grup amb un nivell educatiu inferior fan més ús d'*hombre* de forma estadísticament significativa. Pel que fa al sexe, si bé els homes tendeixen a fer-ne més ús (sobretot a la generació més jove), les diferències entre els dos grups no són significatives. Tot i així, sí que s'observa una correlació estadísticament significativa entre el sexe i el nivell educatiu, de manera que

“En las mujeres el uso [de *hombre*] crece conforme desciende el nivel de instrucción, siendo especialmente destacado su empleo en el nivel educativo más bajo. Aun así, en general, son ellos, por mucho margen, quienes más lo pronunciaron, salvo en el nivel alto, donde hay más casos de mujeres.” (Ruiz-González, 2019: 87-88).

Aquests resultats contradiuen els d'estudis anteriors (com el de Martín Zorraquino, 1991, sobre Saragossa o el de Hernández Cabrera, 2015, sobre Las Palmas de Gran Canaria), en els quals aquesta partícula és més utilitzada per dones de més edat. En un treball posterior, Ruiz González (2021) se centra en una de les funcions pragmàtiques de la partícula *hombre*, la de reformulació, i constata que existeixen relacions estadísticament significatives entre el nombre de realitzacions d'aquesta funció i l'edat i el nivell d'estudis al corpus de Granada.

També comptem amb estudis centrats exclusivament en el funcionament pragmàtic d'aquesta partícula. Briz (2012) identifica una correlació entre les diverses funcions que realitza i la seva posició (inicial, final o aïllada). Cuenca i Torres (2008) també analitzen la posició i les funcions de la partícula discursiva, però consideren *hombre* juntament amb la seva contrapartida catalana *home*, així com *mujer* i *dona*. Mentre que les darreres autores argumenten que *hombre/home* i *mujer/dona* són, al capdevall, la mateixa partícula, per a Briz (2012: 53) es tracta de partícules similars, però d'ús i distribució diferents. A més dels contextuals, un dels factors que sovint s'ha utilitzat per discernir entre les nombroses funcions que pot dur a terme, és el contorn entonatiu amb què es realitza (cf. Hidalgo, 2015). En general, des d'una perspectiva teòrica, s'ha catalogat *hombre* com una partícula reactiva, de control del contacte i enfocadora de l'alteritat, ja que s'orienta cap a les relacions entre els interlocutors (Briz, 2012: 28).

Com vèiem al capítol anterior (subapartat 6.3.1), les proves de significació estadística per a la variable V3h i les seves variants mostren resultats significatius quant a l'edat i la llengua inicial (a les dades de 16 anys), tant per al total d'ocurrències de V3h com per a la variant *home* (V3h2). Així doncs, com als estudis de Ruiz González (2019, 2021) i malgrat la bibliografia sobre els rols interaccionals que típicament assumeixen homes i dones (**punt A** del subapartat 5.3.2.3.1), el sexe no resulta rellevant per explicar l'ús que fan els informants d'aquesta partícula discursiva, a diferència de l'edat. Com mostra la Taula 124, la freqüència de V3h i V3h2 incrementa entre els 12 i els 16 anys alhora que,

com esperàvem, s'observen diferències notables entre les realitzacions totals dels tres grups de llengua inicial als 16 anys.

Taula 124. Realitzacions totals de la variable V3h i les seves variants segons l'edat i la llengua inicial dels informants.

Estudi multivariant	12 anys			16 anys		
	C	B	E	C	B	E
V3h1: <i>hombre</i>	0 (\bar{x} : 0; DE: 0)	0 (\bar{x} : 0; DE: 0)	3 (\bar{x} : 0,1; DE: 0,4)	2 (\bar{x} : 0,07; DE: 0,38)	0 (\bar{x} : 0; DE: 0)	6 (\bar{x} : 0,19; DE: 1,06)
V3h2: <i>home</i>	42 (\bar{x} : 1,56; DE: 2,67)	11 (\bar{x} : 1,1; DE: 1,73)	27 (\bar{x} : 0,84; DE: 1,46)	80 (\bar{x} : 2,96; DE: 3,24)	14 (\bar{x} : 1,4; DE: 3,41)	44 (\bar{x} : 1,38; DE: 2,06)
Estudi univariant						
Total V3h	42 (\bar{x} : 1,56; DE: 2,67)	11 (\bar{x} : 1,1; DE: 1,73)	30 (\bar{x} : 0,94; DE: 1,63)	82 (\bar{x} : 3,04; DE: 3,31)	14 (\bar{x} : 1,4; DE: 3,41)	50 (\bar{x} : 1,56; DE: 2,37)

A partir de la informació anterior, doncs, una mostra anònima comparable a les interaccions analitzades aquí que presentés diverses realitzacions de les variants *hombre* i/o *home* aportaria informació sobre l'edat i, possiblement, la llengua inicial del parlant de qui provingués.

7.1.5. Doblets de partícules discursives: V3i o sea / o sigui

En total, al corpus d'anàlisi, V3i apareix en 439 ocasions, la gran majoria de les quals (393), en forma de la variant *o sigui* (V3i2) i només 46 de les quals, en forma de la variant originada pel contacte amb el castellà *o sea* (V3i1). Des d'un punt de vista qualitatiu, s'hauria de tenir en compte que l'obra de referència que s'ha utilitzat per al disseny de les variables V3 (vegeu el **punt A** del subapartat 5.3.2.3.1), el DPDE, distingeix dos significats bàsics de la partícula *o sea*, que representa amb les etiquetes *o sea 1* i *o sea 2* (Briz, 2008: 220). Ara bé, com que els criteris aplicats en aquest treball per a l'estudi de les partícules discursives són formals i no funcionals (com s'ha explicat al **punt A** del subapartat 5.3.2.3.1), les realitzacions documentades al corpus d'anàlisi no distingeixen entre els diversos significats possibles que es vehiculen mitjançant *o sea* i *o sigui*. Aquest és, tanmateix, l'objectiu d'un seguit d'estudis que s'han ocupat d'aquestes partícules des d'una perspectiva pragmàtica.

En molts d'ells, els autors assenyalen l'elevada freqüència amb què els parlants se serveixen de les partícules *o sea* i *o sigui* (per exemple, vegeu Bach, 2002: 115). Tant és així que, segons Casado Velarde (1991: 90-91), nombrosos acadèmics de la llengua castellana es van mostrar contraris a l'"abús" de *o sea* com a mot crossa o omplidor (en el sentit de González, 2001). Aquestes observacions contrasten, però, amb les més recents de Ruiz-González (2018: 194), que podrien indicar un canvi en curs cap a la reducció de l'ús d'aquesta partícula, potser en favor d'altres com *en plan*:

“Como muletilla, sin embargo, hemos podido comprobar que su uso no es tan excesivo como apuntaban algunos autores hace unas décadas, sino que, más bien, es poco o nada relevante. [...] Por lo que vemos, los jóvenes de este siglo lo usan en menor proporción, lo que podría mostrar una inversión en la tendencia a su empleo y una posible reducción de la significación de *o sea* en la oralidad dentro de unos años.”

Com passa en altres àmbits de la recerca en pragmàtica, un repàs de diversos treballs que s'ocupen de les funcions comunicatives de *o sea* i *o sigui* palesa que autors diferents han utilitzat denominacions i classificacions també diverses. A tall d'exemple, s'ha observat que Bach (2001, 2002) etiqueta *o sigui* com un *connector reformulatiu perifràstic* i proposa una tipologia en què el classifica com a *connector de reformulació expansional* (Bach, 2002), mentre que Briz (2001: 294) en parla en termes de *connector metadiscursiu* (cf. Casado Velarde, 1991: 96). Tot i així, els acadèmics coincideixen a assenyalar que *o sea* (i *o sigui*) vinculen dos elements discursius i solen establir entre ells una relació d'equivalència.

Des d'una aproximació relacionada amb la comentada fins ara, si bé contrastiva, comptem també amb alguns estudis que tracten les funcions d'aquestes partícules en diverses llengües alhora. Així doncs, Cuenca (2003) i Cuenca i Bach (2007), d'una banda, analitzen les característiques d'un seguit de marcadors reformulatius (inclòs *o sigui*) en anglès, castellà i català. De l'altra, Brumme i Schmid (2018) estudien tres partícules discursives (entre les quals es

troba *o sigui*) en català en contrast amb altres llengües romàniques (francès, italià, castellà i romanès).

Darrerament, també han aparegut investigacions més directament vinculades als nostres objectius, ja que, a més dels diversos significats que pren *o sea* en la interacció, també n'estudien la seva relació amb variables sociolingüístiques. De moment, fins on arriba el nostre coneixement, no sembla que s'hagi donat el mateix interès per la forma catalana *o sigui*. D'aquesta manera, destaquen tres treballs sobre *o sea* a ciutats castellanoparlants diferents: Hernández (2016) a la colombiana Medellín, San Martín Núñez (2017) a Santiago de Xile i Ruiz-González (2018) a la Granada peninsular.

Els resultats de l'estudi d'Hernández (2016), tot i que merament descriptius (en el sentit que no s'hi realitzen proves de significació estadística per evaluar les relacions entre les variables) s'alineen amb les tendències en relació a la mecànica conversacional indicades anteriorment (a l'apartat **2.3.1.3**), si considerem l'ús de marcadors discursius (i, específicament, de *o sea*) com a part de la «feina» per mantenir la conversa que, segons la bibliografia, tendeixen més a dur a terme les dones que els homes. La majoria (un 62,2 %) de les realitzacions de la partícula discursiva identificades al corpus (una mostra d'entrevistes semidirigides amb 101 parlants de sexe, edat, nivell educatiu i classe socioeconòmica diferents) corresponen a informants dones (Hernández, 2016: 307).

Pel que fa la resta de variables, dels tres grups de nivell educatiu que s'analitzen, el que fa més ús de la partícula *o sea* és el superior (amb un 41,7 % del total de realitzacions), el grup d'edat que més l'utilitza és el de 35 a 55 anys (amb un 35 %) i la classe social que mostra una freqüència més elevada és la mitjana (amb un 50,3 %), seguida de la popular (amb un 38,1%) i a força distància de la mitjana alta (que fa un 11,6 % de les realitzacions). En definitiva, a Hernández (2016), són les dones de 35 a 55 anys, classe socioeconòmica mitjana i nivell d'educació elevat les que fan més ús de *o sea*.

Quant a Santiago de Xile, San Martín Núñez (2017) tracta tres partícules discursives que s'hi utilitzen per a la reformulació (*o sea*, *onda* i *digamos*, per ordre de freqüència). Tots tres són més usats, de forma estadísticament significativa, per informants de nivell socioeconòmic mitjà alt (el més elevat dels considerats). En canvi, els resultats de les variables sexe i edat no mostren diferències estadísticament significatives. Tot i així, les tendències observades són que les dones en fan lleugerament més ús que els homes i que els informants que més l'utilitzen tenen entre 20 i 34 anys (seguits pels d'entre 35 i 55 i, a més distància, pels de 55 o més).

Per últim, Ruiz-González (2018) estudia els usos de *o sea* a una mostra de 54 informants de Granada obtinguda en el marc del projecte de recerca PRESEEA. Els resultats d'aquest treball són diferents als dels dos anteriors. En primer lloc, si bé els informants de nivell mitjà són els que més utilitzen aquesta partícula (amb un 41,1 % de les realitzacions), el grup de nivell educatiu que menys l'empra és el baix (amb un 18,8 % de les realitzacions). En segon lloc, la generació més jove (de 19 a 34 anys) és la que en fa un ús més baix (amb 28,5 % del total de realitzacions) i la segona generació (formada per informants d'entre 35 i 54 anys) la que l'utilitza més (amb el 40,2 % de realitzacions). En tercer lloc, són els homes els que produeixen la majoria (el 65,4 %) de les ocurrències, a diferència dels estudis anteriors. Això no obstant, no s'indica si es van dur a terme proves de significació estadística per a les correlacions bivariables entre cadascun d'aquests factors i l'ús de *o sea*, com sí que es va fer per a la relació de l'ús de la partícula amb les variables sexe i nivell educatiu (no significativa), edat i nivell d'estudis (significativa: a la segona generació, a diferència de les altres, els informants amb nivells d'estudis superior i baix fan més ús de *o sea* que els de nivell mitjà) i edat i sexe (significativa: en els homes, però no tant en les dones, l'ús de *o sea* és més gran a la segona generació).

En resum, els estudis sociolingüístics previs sobre *o sea* han correlacionat l'ús d'aquesta partícula amb els nivells socioeconòmic i educatiu, el sexe i l'edat dels parlants de diverses regions. Com hem vist al capítol anterior, els resultats de les proves estadístiques realitzades en el nostre estudi mostren diferències

significatives quant a l'edat i la llengua inicial, però no al sexe, dels informants. Concretament, com s'observa a la Taula 125, els totals de realitzacions de V3i i de les dues variants són significativament més elevats quan els informants tenien aproximadament 16 anys que no pas quan en tenien 12. A més, la subhipòtesi sobre la possible correlació entre aquest doblet de partícules discursives i la llengua inicial es confirma, ja que, als 16 anys, *o sea* és molt més freqüent entre els informants castellanoparlants inicials que entre els bilingües i castellanoparlants inicials i, als 12 anys, *o sigui* és molt més emprat pels catalanoparlants inicials que no pels altres grups de llengua inicial.

Taula 125. Realitzacions totals de la variable V3i i les seves variants segons l'edat i la llengua inicial dels informants.

Estudi multivariant	12 anys			16 anys		
	C	B	E	C	B	E
V3i1: <i>o sea</i>	3 (\bar{x} : 0,11; DE: 0,57)	0 (\bar{x} : 0; DE: 0)	7 (\bar{x} : 0,22; DE: 0,61)	1 (\bar{x} : 0,04; DE: 0,19)	1 (\bar{x} : 0,1; DE: 0,32)	34 (\bar{x} : 1,06; DE: 2,51)
V3i2: <i>o sigui</i>	72 (\bar{x} : 2,67; DE: 6,56)	43 (\bar{x} : 4,3; DE: 7,82)	17 (\bar{x} : 0,53; DE: 1,57)	131 (\bar{x} : 4,85; DE: 5,56)	45 (\bar{x} : 4,5; DE: 4,55)	85 (\bar{x} : 2,66; DE: 4,11)
Estudi univariant						
Total V3i	75 (\bar{x} : 2,78; DE: 6,54)	43 (\bar{x} : 4,3; DE: 7,82)	24 (\bar{x} : 0,75; DE: 1,68)	132 (\bar{x} : 4,89; DE: 5,56)	46 (\bar{x} : 4,6; DE: 4,6)	119 (\bar{x} : 3,72; DE: 4,8)

Per tant, els nostres resultats mostren, com els estudis previs, que l'ús de *o sea* (i *o sigui*) està condicionat per variables extralingüístiques, com l'edat dels parlants. Ara bé, a diferència d'altres treballs i de les tendències detectades a la bibliografia sobre el comportament interaccional d'homes i dones (és a dir, contràriament a la nostra subhipòtesi), els resultats obtinguts no són estadísticament significatius per sexe. Tot i això, com hem vist, l'ús d'aquestes dues partícules pot aportar informació sobre l'edat i la llengua inicial més probables dels parlants en mostres comparables al nostre corpus.

7.1.6. Doblets de partícules discursives: V3j *pero* / *però*

El recompte del total d'aparicions al corpus d'anàlisi de la variable V3j (1.791, com veïem a l'apartat 5.6) es reparteixen de manera desigual entre les quatre variants utilitzades pels parlants. La gran majoria (1.551) corresponen a *p(er)ò* (V3j3), seguida de *p(er)ò* (amb 141 realitzacions) i *pero* (amb 86) i, finalment, la variant menys freqüent és *però* (amb només 13 realitzacions). Per dotar de significat aquests resultats globals i els descrits al capítol anterior, ens ocupem a continuació dels estudis previs disponibles sobre aquestes partícules.

Com hem exposat al capítol sobre els aspectes metodològics (i concretament al subapartat 5.5.1), *pero* és una de les partícules discursives incloses en aquest estudi que no apareix al DPDE. Tot i això, sí que hi trobem la partícula *empero*, que funciona de manera semblant, i la revisió bibliogràfica mostra que, si bé tradicionalment s'ha considerat *pero* / *però* una conjunció adversativa (GEIEC, 2021c; GRAE, 2009: 2450), també se l'ha classificat com a connector (Fuentes Rodríguez, 1998; Bach Martorell, 2002: 108), marcador d'argumentació (Berenguer, 1995: 111), marcador del discurs (Borja i Sanz, 2009: 121; Perdomo Carmona, 2020: 99), partícula modal (Vercher García, 2003: 1026) i partícula discursiva (Borreguero Zuloaga, 2017). Aquests i altres autors en subratllen la diversitat de funcions pragmàtiques que realitza en contextos interaccionals, on no només enllaça fragments oracionals entre si, sinó també discursius i, fins i tot, fets extralingüístics amb torns de paraula (per exemple, un parlant pot reaccionar al fet que un altre individu es posi a menjar amb «Però si acabes de dir que no tenies gana!»; cf. Garachana Camarero, 2014). A l'estudi preliminar del corpus d'anàlisi, vam identificar usos de *pero* / *però* que aparentment compleixen els tres requisits descrits al subapartat 5.5.1 necessaris per incorporar partícules al DPDE, especialment el primer, ja que tenen clarament un valor interaccional o de connexió argumentativa, com es desprèn dels exemples següents.

Figura 121. Exemples en què V3j connecta elements discursius a nivell inferencial.

<i>Exemple 1.</i>	
ENT:	però això salta_ no/
L0001QCI:	sí però tenques això i ja està_
<i>Exemple 2.</i>	
ENT:	i no estaria ni tri:st_
L0006SCI:	bueno p(e)rò si després el nen va jugant am(b) ell jo crec que: al final s'hi acabaria acostumant_
<i>Exemple 3.</i>	
ENT:	saps si es pot es pot tocar:/
M0718QCI:	sí sí_ es pot tocar_ p(e)rò té un tacte:_ e es pot tocar p(e)rò no molt_

En definitiva, el repàs bibliogràfic mostra, sens dubte, que la classificació de *però* com a partícula discursiva és controvertida. Mostra d'aquest fet, per exemple, és que no la tracten estudis que s'ocupen de descriure la distribució i usos de diverses partícules a una comunitat de parla concreta, com Toniolo i Zurita (2014) sobre la Córdoba argentina o Santana (2015) sobre la Sevilla espanyola. Ara bé, altres autors, com Vercher García (2003: 1029) la tracten com una partícula més i fins i tot n'han afirmat que és una de les més freqüents en castellà (vegeu també Poblete, 1998).

Així doncs, sembla que aquestes dificultats a l'hora de classificar alguns usos de *pero* / *però* han derivat en una manca d'estudis variacionistes sobre aquesta partícula, ja que no hem trobat treballs que tractin la relació entre aquests elements i variables com l'edat, el sexe o el nivell d'instrucció dels parlants ni en castellà ni en català, a diferència d'altres elements que també formen part de V3. Els resultats de les anàlisis de significació estadística dutes a terme en la nostra investigació contribueixen a pal·liar aquesta manca, malgrat que en el futur caldria complementar-les amb anàlisis qualitatives dels diversos usos de *pero* / *però* que fan els informants.

D'altra banda, la revisió dels estudis previs sobre els fenòmens de contacte de llengües en mostres de parla d'informants en edat escolar a Catalunya i, especialment, dels treballs de Galindo Solé (2006: 170) i Sendra i Rovira (2012: 24) han subratllat la possibilitat que la variant *pero* estigui condicionada per una variable que no considerem en aquest estudi, la població. En efecte, la comprovació dels resultats mostra que, de les 86 realitzacions totals de *pero*, només una correspon a un informant de Manlleu, mentre que la resta són produïdes per informants de Mataró. Aquest fet s'alinea amb l'explicació que proposa Galindo Solé (2006: 170) a l'aparició de *pero* més baixa al seu corpus que al de Vila i Moreno (1996), el fet que aquesta forma seria “un tret gairebé exclusiu de Barcelona i la seva àrea d'influència”. Per tant, els resultats apunten que serà interessant considerar les possibles particularitats subdialectals relacionades amb aquesta variable en investigacions futures.

Com hem vist anteriorment, els resultats indiquen que el total d'ocurrències de V3j, la variant *p(e)rò* (V3j3) i la variant *p(er)ò* (V3j4) mostren diferències estadísticament significatives a les dues edats analitzades. Com es pot observar a la Taula 126, el total de realitzacions per a cadascun d'aquests factors augmenta amb l'edat amb l'única excepció de *p(er)ò* en el grup de noies catalanoparlants inicials. A més, aquesta darrera variant també mostra resultats estadísticament significatius per sexe als 12 anys (de manera que els nois en fan més realitzacions que les noies) i per llengua inicial a les dues edats (en què els catalanoparlants inicials són el grup que l'utilitza amb més freqüència). Així doncs, les subhipòtesis sobre les possibles correlacions entre aquest doblet de partícules discursives i el sexe i la llengua inicial dels informants es confirmen i, a més, els resultats es mostren significatius també per a la variable edat.

Taula 126. Realitzacions totals de la variable V3j i les seves variants segons l'edat, el sexe i la llengua inicial dels informants.

Estudi multivariant	12 anys						16 anys					
	Masculí			Femení			Masculí			Femení		
	C	B	E	C	B	E	C	B	E	C	B	E
V3j1: <i>pero</i>	20 (\bar{x} : 1,25; DE: 2,9)	0 (\bar{x} : 0; DE: 0)	12 (\bar{x} : 0,9; DE: 2,5)	0 (\bar{x} : 0; DE: 0)	0 (\bar{x} : 0; DE: 0)	14 (\bar{x} : 0,8; DE: 1,4)	2 (\bar{x} : 0,1; DE: 0,3)	0 (\bar{x} : 0; DE: 0)	11 (\bar{x} : 0,8; DE: 3)	7 (\bar{x} : 0,6; DE: 1,4)	1 (\bar{x} : 0,1; DE: 0,4)	19 (\bar{x} : 1,1; DE: 3)
V3j2: <i>però</i>	3 (\bar{x} : 0,2; DE: 0,5)	1 (\bar{x} : 0,5; DE: 0,7)	0 (\bar{x} : 0; DE: 0)	3 (\bar{x} : 0,3; DE: 0,9)	2 (\bar{x} : 0,3; DE: 0,7)	0 (\bar{x} : 0; DE: 0)	1 (\bar{x} : 0,1; DE: 0,3)	0 (\bar{x} : 0; DE: 0)	0 (\bar{x} : 0; DE: 0)	0 (\bar{x} : 0; DE: 0)	0 (\bar{x} : 0; DE: 0)	3 (\bar{x} : 0,2; DE: 0,5)
V3j3: <i>p(e)rò</i>	144 (\bar{x} : 9; DE: 4,9)	19 (\bar{x} : 9,5; DE: 9,2)	94 (\bar{x} : 6,7; DE: 4,1)	138 (\bar{x} : 12,5; DE: 7,5)	74 (\bar{x} : 9,3; DE: 6,8)	191 (\bar{x} : 10,6; DE: 6)	202 (\bar{x} : 12,6; DE: 8,2)	21 (\bar{x} : 10,5; DE: 2,1)	141 (\bar{x} : 10,1; DE: 6)	161 (\bar{x} : 14,6; DE: 7,3)	122 (\bar{x} : 15,3; DE: 9,2)	244 (\bar{x} : 13,6; DE: 9,1)
V3j4: <i>p(er)ò</i>	36 (\bar{x} : 2,3; DE: 3,5)	3 (\bar{x} : 1,5; DE: 0,7)	3 (\bar{x} : 0,2; DE: 0,6)	16 (\bar{x} : 1,5; DE: 3,5)	2 (\bar{x} : 0,3; DE: 0,7)	2 (\bar{x} : 0,1; DE: 0,3)	46 (\bar{x} : 2,8; DE: 4)	1 (\bar{x} : 0,5; DE: 0,7)	7 (\bar{x} : 0,5; DE: 0,8)	12 (\bar{x} : 1,1; DE: 1,2)	9 (\bar{x} : 1,1; DE: 1,2)	4 (\bar{x} : 0,2; DE: 0,5)
Estudi univariant												
Total V3j	180 (\bar{x} : 11,3; DE: 7)	22 (\bar{x} : 11; DE: 9,9)	97 (\bar{x} : 6,9; DE: 4,3)	154 (\bar{x} : 14; DE: 6,5)	76 (\bar{x} : 9,5; DE: 6,7)	193 (\bar{x} : 10,7; DE: 6)	248 (\bar{x} : 15,5; DE: 7,8)	22 (\bar{x} : 11; DE: 1,4)	148 (\bar{x} : 10,6; DE: 6,2)	173 (\bar{x} : 15,7; DE: 7,1)	131 (\bar{x} : 16,4; DE: 9,6)	248 (\bar{x} : 13,8; DE: 9,3)

Per tant, és interessant observar que a l'única variable dependent en la qual s'han hagut de considerar aspectes de la realització fonètica per distingir variants ha estat possible identificar correlacions estadísticament significatives entre les variables independents i dues realitzacions fonètiques diferents (*p(e)rò* i *p(er)ò*). Aquestes troballes reforcen la necessitat d'incorporar, en la mesura del possible, el nivell foneticofonològic a les aproximacions empíriques a la variació lingüística, com demostra també el volum de treballs que se centren precisament en fenòmens de variació d'aquest nivell (que malauradament no ha estat possible incorporar al treball que ens ocupa per limitacions pràctiques).

Finalment, en relació als objectius d'aquesta investigació, una vegada més, s'observa que els resultats obtinguts podran ser explotats per a la perfilació lingüística de mostres d'autoria desconeguda comparables a les analitzades aquí.

5.1.1. Doblets de partícules discursives: V3k *pues* / *doncs*

Els resultats globals d'aquesta variable indiquen, com ja hem vist (apartat 5.6), que V3k apareix 1.536 vegades al corpus d'anàlisi. La majoria d'aquestes realitzacions (933) corresponen a V3k3 (*doncs*). La segona variant més freqüent és *pues* (amb 477 ocurrences) i la darrera és *pos* (amb 126).

La variant castellana apareix al DPDE. De fet, a diferència de *doncs*, sobre la qual encara calen més investigacions, *pues* és una de les partícules discursives que es mencionen de manera més recurrent als treballs sobre aquests elements lingüístics que han rebut denominacions diferents (marcadors del discurs, connectors, etc.). En una de les publicacions de referència d'aquest camp, Martín Zorraquino i Portolés Lázaro (1999: 4083) el descriuen com una partícula molt freqüent al discurs oral que introdueix “un comentario nuevo e informativamente valioso con respecto del discurso que lo precede”. També va en aquest sentit la descripció que en proporciona el DPDE (Briz, Pons i Portolés, 2008). Això no obstant, una de les particularitats de *pues* és precisament el fet que pot realitzar funcions pragmàtiques diverses (per exemple, Portolés, 2001, distingeix tres *pues*: el justificador, el comentador i el connector consecutiu; però cf. Fuentes Rodríguez, Placencia i Palma Fahey, 2016: 76-77, sobre la diferència entre *pues* com a conjunció i com a marcador del discurs).

Des de la pragmàtica variacionista, *pues* és una de les partícules que ha rebut més atenció per part dels acadèmics, de manera que és comparativament fàcil trobar publicacions que n'investiguen els usos i la freqüència a mostres de parla de diversos territoris castellanoparlants. Per exemple, Grajales Alzate (2011) n'observa els usos a més de 100 entrevistes procedents de la secció del corpus PRESEEA de Medellín (Colòmbia); Vázquez Carranza (2019) analitza la variació regional mexicana en relació als usos de *pues* en posició final, i Fuentes Rodríguez, Placencia i Palma Fahey (2016) comparen el seus usos en mostres obtingudes d'estudiants universitaris a Quito, Santiago de Xile i Sevilla.

De forma complementària, també comptem amb un bon nombre de treballs que se centren en l'estudi de l'estratificació sociolingüística de *pues*. A continuació en resumim tres dels més rellevants per al nostre estudi.

En primer lloc, Manni i San Martín (2021) s'ocupen de *pues* i la variant *po* a 120 entrevistes amb parlants de més de 20 anys de Santiago de Xile, distribuïts per edat, sexe i grup socioeconòmic. Hi troben una freqüència molt més elevada de *po* que de *pues*. Pel que fa a l'edat, mentre que *po* és més utilitzat pels parlants d'entre 20 i 34 anys (amb un 42,2 % de les realitzacions), *pues* és més freqüent entre els parlants de 35 a 54 anys (amb un 63,2 % de les realitzacions). Quant al sexe, *po* es mostra força equilibrat entre homes (51,2 %) i dones (48,8 %), a diferència de *pues*, que és clarament més utilitzat per homes (71,4 %) que per dones (28,6 %). Per últim, en relació al grup socioeconòmic, *po* és més habitual als grups mitjà-baix (38,7 %) i baix (37,1 %), mentre que *pues* és utilitzat majoritàriament per informants del grup socioeconòmic baix (on es concentra el 87,2 % de realitzacions). Malgrat tot, les proves de significació estadística mostren que cap de les relacions entre *pues* i les variables independents analitzades és significativa, sinó que només ho és la relació entre *po* i el grup socioeconòmic dels parlants.

En segon lloc, Santana (2015) analitza diverses partícules discursives a dues mostres de 12 parlants amb estudis universitaris de Sevilla obtingudes amb uns 40 anys de diferència (la primera, entre 1972 i 1973 i la segona, entre 2012 i 2013). Els resultats generals d'aquest treball mostren que l'ús de partícules discursives augmenta amb el temps. Més concretament, pel que fa a *pues* (com a comentador, dins de la funció d'estructuració de la informació), és la partícula més freqüent a les dues mostres (amb un total de 510 realitzacions).

Quant al sexe dels informants, les dones el fan servir més que els homes (amb aproximadament un 55 % de les realitzacions a les dues mostres). En canvi, la distribució dels tres grups d'edat analitzats és diferent a les dues mostres: mentre que a l'obtinguda als anys 1970 els informants més joves (d'entre 21 i 32 anys) produeixen el 39 % de realitzacions, a la segona mostra és el grup d'edat mitjana (d'entre 33 i 53 anys) els que realitzen justament aquest percentatge

d'ocurrències. Tot i això, a les dues mostres els altres grups d'edat mostren percentatges similars: a la primera, el de mitjana edat realitza un 37 % dels casos i els informants de més edat, un 24 %; a la segona, els parlants de més edat en realitzen el 31 % i els més joves, el 30 %. De manera que, com afirma l'autora, “[s]e trata de una fórmula habitual en esta modalidad de habla y en este sociolecto, no solo por el índice de frecuencias registrado, elevado en ambas sub-muestras (SE 73: 58% y SE12: 68%), sino también por su extensión entre todos los encuestados” (Santana, 2015: 284).

Una altra troballa d'aquest treball que resulta interessant perquè es relaciona amb l'ús de la partícula que ens ocupa és que *pues* com a connector consecutiu només apareix, a les dues mostres analitzades, en informants del grup de més edat. Per aquest motiu, l'autora planteja que “[q]uizá estemos ante un fenómeno de cambio por el que algunas de estas partículas sean marcas específicas de generación” (Santana, 2015: 309).

En darrer lloc, Viguera Ávila (2014) adopta un enfocament molt semblant al de l'estudi anterior i investiga l'ús de diverses partícules discursives també a dues mostres obtingudes amb uns 40 anys de distància (de nou, a principi de les dècades de 1970 i de 2010). Aquesta vegada, però, els 12 parlants (la meitat de cada sexe) provenen de la ciutat de Mèxic, si bé es divideixen també en tres grups d'edat (entre 25 i 29 anys, entre 41 i 49 anys i entre 62 i 80 anys). De nou, l'ús d'aquests elements lingüístics augmenta amb el temps (Viguera Ávila, 2014: 193).

Pel que fa a *pues*, aquesta autora també distingeix entre el seu ús com a comentador (del qual n'identifica les variants *pues*, *pus* i *p's*) i com a connector consecutiu. Quant al primer, torna a resultar la partícula més freqüent, amb un total de 281 realitzacions (141 *pues*, 12 *pus* i 2 *p's*). S'utilitza més a la mostra més recent (on se n'identifiquen 155 realitzacions) que a l'obtinguda el 1971 (126 realitzacions). A la primera, són els parlants més joves (i, específicament, les dones) els que en fan més ús (65 realitzacions, 47 de les quals corresponen a les dones). En canvi, a la segona mostra, el grup de mitjana edat és el que en produeix més ocurrencies (70, 51 de les quals corresponen a homes). En general

(excepte al primer grup d'edat de la primera mostra), els homes l'utilitzen més que les dones.

En relació a *pues* com a connector consecutiu, és el segon connector d'aquest tipus en nombre de realitzacions (46), després de *entonces* (que en presenta 424). Santana (2015: 208) indica que aquestes ocurrències “se encuentran distribuid[a]s entre los grupos etarios; aparentemente sin alguna tendencia en cuanto a sexo”. Cal tenir en compte, a l'hora de considerar els resultats obtinguts en aquests estudis previs, que ni Santana (2015) ni Viguera Ávila (2014) duen a terme proves de significació estadística, ja que les seves mostres són de mides reduïdes (12 informants).

Com apuntàvem a l'inici d'aquest apartat, els estudis sobre *doncs* o *pues* en produccions lingüístiques en català són escassos. Per les similituds amb l'aproximació a aquestes partícules que prenem en el nostre estudi, destaca el treball de Vila (1998), en el qual s'analitza part del corpus descrit a Tomàs i Pallejà (1998). L'objecte d'aquest estudi són les marques transcòdiques lèxiques (MTL), enteses com unitats lèxiques les característiques fòniques, morfològiques i/o semàntiques de les quals remetent a una llengua diferent de la catalana (Vila i Moreno, 1998: 264). Tot i que aquestes unitats sumen menys d'un 2 % del total d'elements lèxics del corpus (Vila i Moreno, 1998: 269), *pues* és una de les més freqüents que hi detecta l'autor (amb 25 realitzacions, és a dir, un 4,5 % del total de MTL identificades). (La variant *pos* hi apareix una única vegada). Només dues MTL superen *pues* en aquest corpus: *lo* (amb 62 ocurrències) i *bueno* (amb 76).

Ara bé, com assenyala Vila i Moreno (1998), cal tenir en compte que les mostres que conformen el corpus no van ser seleccionades segons criteris de representativitat sociolingüística, sinó per ser considerades vàlides per estudiar la variació funcional de la llengua. Així, per exemple, “la mostra conté un nombre de parlants amb coneixements superiors de català que sens dubte ultrapassa el percentatge que una mostra sociològicament representativa els atribuiria” (Vila i Moreno, 1998: 270). Per això, “cal prendre els nostres resultats

com a estrictament orientatius i no pas com a representatius del conjunt de la societat catalana” (Vila i Moreno, 1998: 267).

En resum, doncs, aquest estudi ens informa que, així com a les mostres analitzades en altres estudis sobre la llengua castellana, també al corpus català utilitzat a les diverses publicacions de 1998 sobre la variació funcional *pues* (i, en menor mesura, la variant *pos*) hi són relativament freqüents. No obstant això, no ens aporta informació sobre la distribució de *pues* (i encara menys *doncs*) en les produccions de parlants d'edat i sexe diferents. A més, cap de les publicacions ressenyades ens n'aporta sobre la relació entre aquestes partícules i les llengües inicials dels parlants.

Els treballs consultats sobre el castellà, però, sí que ens mostren algunes tendències en relació a l'ús de *pues*, l'edat i el sexe dels parlants. A diferència d'aquests treballs, en el nostre estudi no ha estat possible observar diferències significatives en l'ús de *pues* entre els informants. Com hem vist anteriorment (subapartat 6.3.1), l'única correlació estadísticament significativa identificada s'ha establert entre *pos* (V3k2) i l'edat dels informants. La taula 127 mostra que les realitzacions d'aquesta variant incrementen entre les edats de 12 i 16 anys. D'aquesta manera, a diferència del que esperàvem, ni el sexe ni la llengua inicial es correlacionen significativament amb aquest doblet de partícules discursives.

Taula 127. Realitzacions totals de la variable V3k i les seves variants segons l'edat dels informants.

Estudi multivariant	12 anys	16 anys
V3k1: <i>pues</i>	332 (\bar{x} : 4,8; DE: 8,4)	145 (\bar{x} : 2,1; DE: 3,2)
V3k2: <i>pos</i>	44 (\bar{x} : 0,6; DE: 1,5)	82 (\bar{x} : 1,2; DE: 2,2)
V3k3: <i>doncs</i>	462 (\bar{x} : 6,7; DE: 9,3)	471 (\bar{x} : 6,8; DE: 8,3)
Estudi univariant		
Total V3k	838 (\bar{x} : 12,2; DE: 11,1)	698 (\bar{x} : 10,1; DE: 7,9)

En definitiva, per tant, de la revisió bibliogràfica sobre l'estratificació sociolingüística de *pues* i *doncs* en podem concloure, d'una banda, que, si bé sembla que la seva freqüència és elevada en produccions orals, encara manquen estudis que investiguin aquestes partícules en relació a variables com la llengua inicial dels parlants, especialment en català. De l'altra, sembla que les realitzacions d'aquestes partícules poden aportar informació sobre l'edat més probable dels parlants, ja que tant a estudis anteriors com al nostre ha estat possible identificar correlacions entre aquestes variables. Concretament, en mostres comparables a les analitzades aquí, el nombre de realitzacions de *pos* podria indicar que un parlant determinat s'acosta més probablement a l'edat de 12 o de 16 anys.

7.1.8. Doblets de partícules discursives: V3l *vale* / *d'acord*

El total de realitzacions de la variable V3l, com vèiem a l'apartat 5.6, és de 850, la gran majoria de les quals (concretament, 635) corresponen a la variant *vale* (V3l1). De la resta, n'hi ha moltes més de la variant *val* (V3l2) que de *d'acord* (172 i 43, respectivament). Així doncs, aquest doblet de partícules discursives mostra el mateix patró que V3c, a diferència de la resta de variables incloses a V3, ja que la variant originada pel contacte amb el castellà és la més freqüent al corpus d'anàlisi.

El DPDE defineix *vale* com una partícula que “[i]ndica aceptación de lo dicho o acuerdo, con frecuencia, de cierre” (Briz, Pons i Portolés, 2008). Es tracta d'una de les mencionades més habitualment a treballs i propostes de classificació d'aquests elements (per exemple, Martín Zorraquino i Portalés, 1999 o Santos Río, 2003). Aquest fet molt possiblement s'explica per l'elevada freqüència d'aparició de *vale* a les varietats col·loquials de la llengua (tant catalana com castellana). Un dels estudis que demostren aquesta alta freqüència d'ús és el de Jørgensen i Martínez López (2007) sobre converses informals i espontànies de parlants d'entre 13 i 19 anys recollides a principis del segle XXI a Madrid, en el qual *vale* és una de les partícules discursives més utilitzades, juntament amb *bueno* i *bien*.

Tot i la puntualització que es fa a la definició que en proporciona el DPDE sobre el seu ús recurrent com a indicador de tancament, *vale* també s'utilitza a l'inici de les intervencions dels parlants, com mostra Serrano Reyes (2013). Aquesta autora reflexiona sobre el tractament que rep *vale* en algunes de les primeres obres interessades per les partícules discursives en castellà i conclou que actualment ja no és un indicador de llenguatge juvenil:

“Era por aquel entonces [els anys 1990] cuando Martín Zorraquino y Portolés (1999) sostenían que el empleo de *vale* se había extendido en la conversación coloquial especialmente en las generaciones más jóvenes. Hoy en día, nos atrevemos a afirmar que estamos ante un marcador con una alta frecuencia de uso (debido en parte a su polifuncionalidad) que ya ha perdido su consideración como marca lingüística de identidad social y se encuentra ampliamente asentado en el discurso coloquial de gran parte de los hablantes.”

(Serrano Reyes, 2013: 139).

Aquesta afirmació obté un suport empíric parcial del treball de Cestero Mancera i Moreno Fernández (2008), en què combinen l'aproximació variacionista amb l'anàlisi de la conversa i la descripció pragmàtica per analitzar una mostra de 54 informants obtinguda al barri madrileny de Salamanca en el marc del projecte PRESEEA. D'una banda, els resultats d'aquest estudi indiquen que, en la mostra analitzada, la partícula *vale* no és gaire utilitzada (“pues sólo hemos encontrado 42 casos en 36 encuestas, en unas 36 horas de grabación”, Cestero Mancera i Moreno Fernández, 2008: 72). De l'altra, no obstant això, no mostra un ús més habitual en els parlants més joves, sinó que la majoria de realitzacions (el 47 %) corresponen al segon grup d'edat, parlants d'entre 35 i 54 anys, si bé els informants d'entre 20 i 34 anys també l'utilitzen força (produeixen un 40 % de les realitzacions), a diferència dels de més de 55 anys (autors tan sols de l'11 % de les ocurrencies). Pel que fa al sexe, les dones en fan un ús clarament més elevat (el 64 % de realitzacions) que els homes (que produeixen el 35 % restant). Quant a la classe social, aquests autors també troben que és més freqüent a la classe mitjana (52 % dels casos) i, finalment, en relació al nivell d'estudis, *vale* és més habitual entre els informants que han acabat l'educació primària (52 % de les ocurrencies), seguits dels que tenen formació universitària

(30 %) i, en menor mesura, dels que han acabat l'educació secundària (només el 16 %). A més, els perfils de parlants que més utilitza *vale* en aquest estudi és diferent segons el sexe. Mentre que la majoria d'usos, en el cas de les dones, corresponen a joves, de classe mitjana i nivell d'estudis baix, en el cas dels homes, són principalment d'edat i classe mitjanes amb formació universitària (Cestero Mancera i Moreno Fernández, 2008: 74).

També en relació amb els nostres objectius, trobem treballs sobre *vale* com a possible marcador de contacte interlingüístic. Per exemple, Domínguez Portela (2011) l'investiga a dos corpus de la llengua gallega. Pel que fa al català, les variants que estudiem aquí a més de la genuïna *d'acord*, és a dir, *vale* i *val* ja han estat identificades com a formes utilitzades freqüentment pels parlants en treballs com els de Vila i Moreno (1996, 1998) i Galindo Solé (2006). Encara manquen treballs, però, que en correlacionin l'ús amb característiques dels individus, de forma comparable al que fem en aquest treball.

Com hem vist al capítol anterior, a diferència del que cabia esperar, no hem trobat correlacions estadísticament significatives entre el sexe o la llengua inicial dels informants i el seu ús de V3l. Sí que n'hem trobat, en canvi, entre les variants *val* (V3l2) i *d'acord* (V3l3) i l'edat. La Taula 128 mostra el total de realitzacions d'aquesta variable per moments temporals i hi podem observar que, mentre que el total de *val* disminueix en un 62,4% entre les edats de 12 i 16 anys, el de *d'acord* gairebé es triplica.

Taula 128. Realitzacions totals de la variable V3l i les seves variants segons l'edat dels informants.

Estudi multivariant	12 anys	16 anys
V3l1: <i>vale</i>	287 (\bar{x} : 4,1; DE: 3,6)	348 (\bar{x} : 5; DE: 4,8)
V3l2: <i>val</i>	125 (\bar{x} : 1,8; DE: 2,4)	47 (\bar{x} : 0,7; DE: 1,1)
V3l3: <i>d'acord</i>	12 (\bar{x} : 0,2; DE: 0,5)	31 (\bar{x} : 0,5; DE: 1,1)
Estudi univariant		
Total V3l	424 (\bar{x} : 6,1; DE: 4,2)	426 (\bar{x} : 6,2; DE: 5,1)

Per tant, els resultats de V3l ens permeten confirmar les troballes d'estudis previs pel que fa a l'ús de *vale* i *val* i, a més, ens aporten informació de la diferència entre l'ús d'aquestes partícules i *d'acord* al corpus d'anàlisi. Quant als objectius que ens proposàvem, els resultats obtinguts confirmen que es tracta d'una variable potencialment útil per a la perfilació lingüística de mostres d'autoria desconeguda, en tant que les variants *val* i *d'acord* mostren diferències estadísticament significatives en les produccions dels mateixos parlants corresponents a edats diferents.

7.1.9. Obertura de la interacció: V5

Com hem observat al capítol 5 (apartat 5.6), el total de realitzacions de V5 identificades al corpus és de 144. Per tant, no hi ha una correspondència d'exactament una realització per transcripció (recordem que n'hi ha un total de 138), sinó que en algunes ocasions els informants no en produeixen o en produeixen més d'una (com s'ha explicat al subapartat 5.5.2). La variant que mostra més realitzacions és clarament *bon dia* (V5b), amb 94 ocurrències, mentre que *hola* (V5a) només en presenta 27, *bona tarda* (V5c) arriba a les 13 realitzacions i la composta *hola bon dia* (V5d) apareix tan sols 10 vegades.

L'ús de fórmules d'obertura de la interacció és un tret lingüístic rellevant per a la lingüística forense, ja que la variació que presenta entre parlants pot ser útil per a tasques d'anàlisi forense de textos orals o escrits. La recerca en la disciplina inclou estudis que consideren aquesta variable en produccions escrites (per exemple, Wright, 2013, i Queralt, 2015). A més, comptem amb estudis realitzats des d'altres àrees, com la pragmàtica o la fraseologia, que aborden la llibertat que tenen els parlants a l'hora de fer servir (o no) fórmules de salutació en gèneres textuais amb els quals solen treballar els lingüistes forenses, com correus electrònics (Montesinos, 2006) o converses de WhatsApp (Sampietro, 2019).

En una aproximació en certa mesura comparable a la nostra, Cantamutto (2019) analitza diversos aspectes lingüístics, incloses les obertures, en interaccions per missatges de text en relació a variables sociolingüístiques com l'edat, el sexe i el grup sociocultural dels interlocutors. Tenint en compte la rellevància de les

fórmules d'inici d'interaccions per a la lingüística forense, els estudis previs sobre el tema en textos escrits i que, sobretot des de l'expansió de les noves tecnologies i la irrupció dels serveis de missatgeria instantània, la divisió tradicional entre gèneres orals i escrits s'ha qüestionat de manera reiterada (vegi's, per exemple, López Quero, 2017), en el nostre estudi hem inclòs la variable V5 Obertura de la interacció amb l'objectiu d'obtenir una primera aproximació a aquest fenomen de variació en una mostra de parla catalana d'un nombre considerable de parlants adolescents i omplir, així, un buit existent en la recerca aplicable a la perfilació lingüística del nostre context sociolingüístic.

Certament, d'una banda, les fórmules de salutació s'han abordat des de fa dècades i des d'aproximacions diverses com l'etnografia i l'estudi de convencions culturals (per exemple, Ag Youssouf, Grimshaw i Bird, 1976; Duranti, 1997; Bibiri, Mocanu i Turculeț, 2016), l'anàlisi de la conversa (per exemple, telefònica a Ávila-Muñoz, 1998, i Coronel-Molina, 1998), la Teoria de la Rellevància (Padilla Cruz, 2003) o la pragmàtica sociocultural (Vázquez Carranza, 2020). De l'altra, comptem encara amb pocs estudis que investiguin les possibles correlacions entre aquesta variable lingüística i característiques dels parlants. Una de les poques excepcions en aquest sentit és el treball de Collí May (2009) sobre les salutacions i els comiats en relació a la cortesia verbal en una mostra d'alumnes d'un centre educatiu mexicà segons el sexe. Una altra, la conformen els estudis que s'han realitzat sobre la disponibilitat lèxica, com el de Ríos González (2007), en què compara aquesta mesura respecte a les salutacions, els temes de conversa i els comiats entre homes i dones. En català, l'escassetat de publicacions sobre l'ús de fórmules de salutació és encara més notable, tot i que ja hem observat al **punt B** del subapartat **5.3.2.3.1** que s'han fet passos en aquesta direcció des de l'estudi de les expressions fixades com a part de rutines conversacionals (Arenas Beneyto, 2005; Bladas, 2012) i des de l'estudi de la variació textual (Arenas Beneyto i Cuenca, 2007).

Així doncs, els resultats de V5 descrits al capítol anterior aporten informació sobre l'ús i la distribució de fórmules d'obertura de la interacció en una mostra de parla catalana, una aproximació que aparentment encara no s'havia

implementat en l'estudi del català actual. Per tant, en els pròxims anys, caldria replicar l'anàlisi que presentem aquí i complementar la nostra proposta amb aproximacions qualitatives per comprendre millor les implicacions i la capacitat dels resultats obtinguts de ser generalitzats a altres mostres.

Això no obstant, els resultats mostren que dues de les variants correlacionen de manera estadísticament significativa amb la llengua inicial dels informants: *bona tarda* (V5c) a les transcripcions corresponents a l'edat de 12 anys i *hola bon dia* (V5d) a les que corresponen als 16 anys. Aquests resultats podrien explicar-se a partir de l'observació que fa Arenas Beneyto (2005: 88) sobre la possible relació entre les fórmules de salutació i la competència comunicativa dels parlants, si bé caldria analitzar les interaccions que formen el corpus d'anàlisi minuciosament, el context extralingüístic en què es produeixen i les funcions pràgmatiques de les variants que hi hem quantificat per comprendre millor la correlació detectada entre la llengua inicial i dues de les variants de V5. La Taula 129 mostra que V5c als 12 anys és més utilitzada per informants bilingües inicials i V5d als 16 anys només és utilitzada per informants castellanoparlants inicials.

Taula 129. Realitzacions totals de la variable V5 i les seves variants segons l'edat i la llengua inicial dels informants.

Estudi multivariant	12 anys			16 anys		
	C	B	E	C	B	E
V5a: <i>hola</i>	7 (\bar{x} : 0,3; DE: 0,4)	0 (\bar{x} : 0; DE: 0)	10 (\bar{x} : 0,3; DE: 0,5)	3 (\bar{x} : 0,1; DE: 0,3)	2 (\bar{x} : 0,2; DE: 0,4)	5 (\bar{x} : 0,2; DE: 0,4)
V5b: <i>bon dia</i>	18 (\bar{x} : 0,7; DE: 0,5)	7 (\bar{x} : 0,7; DE: 0,5)	19 (\bar{x} : 0,6; DE: 0,5)	24 (\bar{x} : 0,9; DE: 0,3)	6 (\bar{x} : 0,6; DE: 0,5)	20 (\bar{x} : 0,6; DE: 0,5)
V5c: <i>bona tarda</i>	1 (\bar{x} : 0,04; DE: 0,2)	3 (\bar{x} : 0,3; DE: 0,5)	2 (\bar{x} : 0,06; DE: 0,2)	1 (\bar{x} : 0,04; DE: 0,2)	3 (\bar{x} : 0,3; DE: 0,5)	3 (\bar{x} : 0,1; DE: 0,3)
V5d: <i>hola bon dia</i>	1 (\bar{x} : 0,04; DE: 0,2)	0 (\bar{x} : 0; DE: 0)	2 (\bar{x} : 0,06; DE: 0,2)	0 (\bar{x} : 0; DE: 0)	0 (\bar{x} : 0; DE: 0)	7 (\bar{x} : 0,2; DE: 0,4)
Estudi univariant						
Total V5	27 (\bar{x} : 1; DE: 0,3)	10 (\bar{x} : 1; DE: 0)	33 (\bar{x} : 1,03; DE: 0,18)	28 (\bar{x} : 1,04; DE: 0,2)	11 (\bar{x} : 1,1; DE: 0,3)	35 (\bar{x} : 1,1; DE: 0,5)

En definitiva, sembla que en mostres comparables a les que conté el corpus d'anàlisi, l'ús de *bona tarda* i *hola bon dia* podria indicar la llengua inicial més probable dels parlants, tot i que és altament recomanable que per a l'avaluació d'aquesta variable es tinguin molt en compte els factors contextuals que caracteritzen la mostra en què apareix. Sobretot, l'hora del dia en què s'hauria produït i, com subratlla la bibliografia, la relació interpersonal prèvia entre els interlocutors.

7.1.10. Heterorepeticions: V6

En total, al corpus s'han detectat 712 ocurrències d'heterorepeticions (apartat 5.6). Com hem vist anteriorment (al **punt C** del subapartat 5.3.2.3.1), hem basat l'anàlisi d'aquesta variable en la tipologia de funcions de Freixa (1998: 204-206). A més, comptem amb nombroses aportacions sobre les funcions pragmàtiques d'aquest recurs lingüístic, tant en interaccions en anglès (Tannen, 1987 o Bazzanella, 1992, 2011) com en llengües romàniques (entre altres, Vigara Tauste, 1997; Blas Arroyo, 1999; Hornoïu, 2003; Bernal Linnersand, 2005), així com investigacions sobre el seu paper en l'adquisició de segones llengües i llengües estrangeres (per exemple, Svennevig, 2004; Rydland i Aukrust, 2005; Čurković Kalebić, 2016; Takeda, 2021). Recentment, també s'ha estudiat la possibilitat de detectar heterorepeticions de manera automàtica (Bigi, Bertrand i Guardiola, 2014) i s'ha investigat des de la perspectiva de comparació de llengües (Rossi, 2020). Això no obstant, no hem trobat investigacions sobre l'heterorepetició comparables amb la nostra aproximació quantitativa en relació amb variables sociolingüístiques com l'edat o el sexe dels parlants.

A partir de la bibliografia sobre l'adquisició de llengües en què s'observa l'ús d'heterorepeticions, abans de dur a terme les anàlisis estadístiques proposàvem que la variable V6 podria correlacionar-se amb l'edat dels informants. Els resultats, com hem vist al capítol anterior, corroboren aquesta correlació (com mostra la Taula 130, els totals de realitzacions augmenten amb l'edat) i, a més, també indiquen que hi ha una correlació estadísticament significativa amb el sexe (els nois fan més heterorepeticions que les noies).

Taula 130. Realitzacions totals de la variable V6 segons el sexe i l'edat dels informants.

Estudi univariant	Masculí		Femení	
	12 anys	16 anys	12 anys	16 anys
V6: heterorepeticions	148 (\bar{x} : 4,6; DE: 2,9)	238 (\bar{x} : 7,4; DE: 3,5)	138 (\bar{x} : 3,7; DE: 3,4)	188 (\bar{x} : 5,1; DE: 2,7)

Caldria investigar, des d'una perspectiva qualitativa, si l'heterorepetició forma part d'estils interaccionals diferents adoptats pels parlants. S'observa que el total de realitzacions en els informants nois és molt més gran entre les dues edats estudiades que en les noies. Això podria tenir a veure amb l'ús de les heterorepeticions com una estratègia comunicativa "còmoda" que permet als informants contribuir a la tasca interactiva amb menys esforç cognitiu en comparació a l'ús d'estratègies alternatives. Aquest fet, al seu torn, podria relacionar-se amb factors com les actituds dels informants vers la llengua de la tasca o la situació comunicativa (al capdavant, una tasca d'avaluació de la seva competència lingüística duta a terme en el context escolar)³¹. Pel que fa als nostres objectius, els resultats confirmen que l'ús d'heterorepetició pot utilitzar-se com a indicador de l'edat i el sexe més probable en mostres de la llengua d'adolescents comparables a les analitzades en aquest estudi.

7.1.11. Discussió dels resultats de les variables pragmàtiques: Síntesi

Per la seva elevada aparició al corpus d'anàlisi, és esperable que les variables analitzades a l'estudi es trobin en mostres d'interaccions en què participin adolescents, ja siguin presencials (per exemple, als centres educatius o espais d'oci) com a distància (a través de xarxes socials, jocs en línia, etc.), però també en mostres de parla monologals (com exposicions orals, publicacions o contingut audiovisual compartits a xarxes socials).

Hem observat que el nombre d'estudis que han tractat les variables pragmàtiques del nostre treball des de perspectives quantitatives o variacionistes comparables és reduït (especialment, en llengua catalana). De manera que caldria continuar aquesta línia de recerca per pal·liar el buit detectat a la bibliografia i contribuir a

³¹ Agraïxo al doctor Avel·lí Flors Mas la seva reflexió sobre la possible relació entre estils comunicatius observats en el corpus del projecte RESOL i l'ús d'heterorepeticions.

la fonamentació empírica de les tasques de perfilació lingüística de mostres en llengua catalana. A continuació, la Taula 131 resumeix els resultats de les variables pragmàtiques. Com s’hi pot observar, tot i que no sempre amb dues de les variables independents de l’estudi, com s’esperava, totes les variables pragmàtiques mostren correlacions estadísticament significatives amb l’edat, el sexe o la llengua inicial dels informants.

Taula 131. Correlacions esperades i obtingudes per a les variables pragmàtiques.

Variables pragmàtiques		S’esperava que correlacionés amb...	Es correlaciona amb...
V3: Doblets de partícules discursives	V3b: <i>a ver / a veure</i>	Sexe, llengua inicial	Edat, sexe
	V3c: <i>bueno / bé</i>	Sexe, llengua inicial	Edat
	V3f: <i>después / després</i>	Sexe, llengua inicial	Sexe
	V3h: <i>hombre / home</i>	Sexe, llengua inicial	Edat, llengua inicial
	V3i: <i>o sea / o sigui</i>	Sexe, llengua inicial	Edat, llengua inicial
	V3j: <i>pero / però</i>	Sexe, llengua inicial	Edat
	V3k: <i>pues / doncs</i>	Sexe, llengua inicial	Edat
	V3l: <i>vale / d’acord</i>	Sexe, llengua inicial	Edat
V5 Obertura de la interacció		Edat, llengua inicial	Llengua inicial
V6 Heterorepeticions		Edat	Edat, sexe

7.2. Discussió dels resultats de les variables morfosintàctiques

La informació disponible (i, en alguns casos, la manca d’aquesta informació) sobre les diverses variables morfosintàctiques que hem observat a l’estudi ens feia pensar, abans de realitzar les anàlisis de les dades recopilades, que la posició dels pronoms clítics en construccions perifràstiques (V14) i el tipus de preguntes formulades per part dels informants (V17) podrien correlacionar-se de manera estadísticament significativa amb les tres variables independents, que els determinants demostratius (V23) podrien correlacionar-se amb l’edat i la llengua inicial, que les construccions condicionals (V18) ho podrien fer amb l’edat, el nombre de preguntes no retòriques (V16), amb el sexe, i que les perífrasis verbals d’obligació (V22) i els pronoms clítics de datiu plural (V25) ho podrien fer amb la llengua inicial.

7.2.1. Posició dels pronoms clítics en construccions perifràstiques: V14

Els recomptes de les variables de l'estudi multivariant de la Taula 26 (apartat 5.6) mostren que les construccions perifràstiques amb menys realitzacions al corpus d'anàlisi són aquelles en què el pronom clític apareix en posició proclítica amb la perífrasi *anar a + infinitiu* (V14a4, 10 realitzacions), aquelles en què apareix en posició enclítica amb *haver de + infinitiu* (V14b2, 11 realitzacions) i aquelles en què apareix tant en posició proclítica com enclítica, independentment de la perífrasi verbal (V14c, 12 realitzacions). En canvi, les construccions més freqüents són, amb gran diferència, les formades per la perífrasi *poder + infinitiu*, ja sigui precedida (V14a3, 1.247 ocurrences) o seguida del pronom clític (V14b3, 150 ocurrences). En general, la proclisi (V14a, 1.864 realitzacions) és molt més freqüent que l'enclisi (V14b, 215 realitzacions). Per tant, els nostres resultats, s'alineen, pel que fa a la freqüència d'ús de la posició proclítica i enclítica, amb els obtinguts en estudis anteriors per al castellà (vegeu el **punt D** del subapartat 5.3.2.3.2).

Quant als resultats de les proves de significació entre les variables dependents i independents, les variants proclítics amb les perífrasis d'infinitiu *tenir (que/de)* i *anar a*, juntament amb la variant enclítica amb *poder* (V14a1, V14a4 i V14b3, respectivament) mostren una correlació estadísticament significativa amb l'edat dels informants. A més, la variant V14a4 també es mostra significativament correlacionada amb la llengua inicial a les dades corresponents als 12 anys. Així doncs, els nostres resultats són comparables amb els d'Illamola i Vila (2015) per al castellà en tant que el sexe tampoc es mostra estadísticament rellevant, si bé la llengua inicial en el nostre corpus sí que es correlaciona amb una de les variants. En aquest sentit, cal recordar que Illamola i Queralt (2020) també troben diferències estadísticament significatives entre els grups de llengua inicial de la seva mostra de llengua castellana.

Per tant, sembla que als factors estructurals, pragmàtics, funcionals i dialectals que influeixen en la posició dels pronoms clítics en construccions perifràstiques (Illamola i Vila, 2015: 38), podem afegir-hi característiques dels parlants com l'edat i la llengua inicial, si més no en relació amb algunes construccions. La

Taula 132 mostra el total de realitzacions per a les variants que mostren correlacions significatives amb l'edat i/o la llengua inicial. Com s'hi pot observar, tots els grups de llengua inicial fan més ús de les dues variants proclíiques als 12 que als 16 anys, mentre que la tendència per a V14b3 és inversa.

Taula 132. Realitzacions totals de les variants significatives de V14 segons l'edat i la llengua inicial dels informants.

Estudi multivariant	12 anys			16 anys		
	C	B	E	C	B	E
V14a1: PC + <i>tenir que / de</i> + INF	48 (\bar{x} : 1,78; DE: 3,15)	21 (\bar{x} : 2,1; DE: 3,41)	47 (\bar{x} : 1,47; DE: 2,54)	27 (\bar{x} : 1; DE: 2,18)	9 (\bar{x} : 0,9; DE: 1,91)	16 (\bar{x} : 0,5; DE: 1,1)
V14a4: PC + <i>anar a</i> + INF	6 (\bar{x} : 0,22; DE: 0,51)	3 (\bar{x} : 0,3; DE: 0,68)	0 (\bar{x} : 0; DE: 0)	0 (\bar{x} : 0; DE: 0)	0 (\bar{x} : 0; DE: 0)	1 (\bar{x} : 0,03; DE: 0,18)
V14b3: <i>poder</i> + INF + PC	15 (\bar{x} : 0,56; DE: 0,85)	9 (\bar{x} : 0,9; DE: 0,88)	33 (\bar{x} : 1,03; DE: 2,18)	38 (\bar{x} : 1,41; DE: 2,19)	17 (\bar{x} : 1,7; DE: 1,5)	38 (\bar{x} : 1,2; DE: 1,9)

Així doncs, en mostres comparables a l'anàlitzada les tres variants de V14 que recull la Taula 132 poden resultar útils per a la perfilació del rang d'edat més probable de parlants adolescents. No obstant això, l'aplicabilitat dels resultats de V14a4 a altres mostres és més limitada a causa del baix nombre d'ocurrències, malgrat les proves estadístiques realitzades.

7.2.2. Nombre de preguntes no retòriques: V16

En total, al corpus d'anàlisi s'han detectat 368 preguntes formulades pels informants. Com hem vist al **punt E** del subapartat **5.3.2.3.2**, a la bibliografia hi trobem tant estudis que mostren diferències en la freqüència de formulació de preguntes entre parlants homes i parlants dones (Fishman, 1983; Tuson, 1998: 216) com treballs que indiquen que aquesta variable no es correlaciona amb el sexe dels parlants (Larson i McKinley, 1998; Hawkins i Power, 1999).

Els resultats de les proves estadístiques exposats al capítol anterior (concretament, al subapartat **6.3.2**) indiquen que no hi ha diferències significatives en el nombre de preguntes no retòriques formulades pels informants del corpus d'anàlisi. La Taula 133 mostra el nombre de realitzacions totals segons l'edat, el sexe i la llengua inicial. Si comparem les dues edats, hi

podem observar que la tendència, tot i que no és estadísticament significativa, és que el nombre total de preguntes formulades pels informants augmenta amb l'edat, ja que mentre que el total de tots els grups de sexe i llengua inicial als 12 anys és de 174, als 16 arriba a les 194 realitzacions de V16.

Pel que fa al sexe, de nou, si sumem els totals de les preguntes formulades pels nois i de les formulades per les noies, podem veure que la diferència no és gaire marcada (ni, recordem, estadísticament significativa), però indica que al nostre corpus són els nois els que fan més preguntes que les noies (190 i 178, respectivament). Finalment, quant a la llengua inicial, els resultats globals semblen indicar que els nois bilingües inicials formulen força menys preguntes que els seus companys. Tot i això, tenint en compte les diferències en el nombre d'informants de cada grup de llengua inicial (hi ha força menys nois bilingües que de les altres llengües inicials, vegeu l'apartat 5.2 i la Taula 24 a l'apartat 5.6), resulta més informatiu fixar-nos en les mitjanes de realització de V16. Aquest valor mostra que no existeixen diferències gaire destacables entre els grups, ja que tots se situen entre 1,8 i 4 preguntes formulades pels informants de mitjana (amb desviacions estàndard també força reduïdes, com a màxim de 5).

Taula 133. Realitzacions totals de la variable V16 segons l'edat, el sexe i la llengua inicial dels informants.

Estudi univari- ariant	12 anys						16 anys					
	Masculí			Femení			Masculí			Femení		
	C	B	E	C	B	E	C	B	E	C	B	E
V6: pre- gunes	57 (\bar{x} : 3,6; DE: 5,05)	7 (\bar{x} : 3,5; DE: 2,12)	27 (\bar{x} : 1,9; DE: 1,75)	28 (\bar{x} : 2,5; DE: 1,92)	22 (\bar{x} : 2,8; DE: 3,15)	33 (\bar{x} : 1,8; DE: 1,54)	64 (\bar{x} : 4; DE: 4,41)	7 (\bar{x} : 3,5; DE: 2,12)	28 (\bar{x} : 2; DE: 2,11)	32 (\bar{x} : 2,91; DE: 2,91)	27 (\bar{x} : 3,38; DE: 3,16)	36 (\bar{x} : 2; DE: 1,98)

La manca de diferències estadísticament significatives en el nombre de preguntes formulades podria estar relacionada amb el disseny de la tasca interactiva que es proposava l'objectiu de la qual era vendre un producte. Tot i això, per complir aquest objectiu, els informants havien d'obtenir certa informació addicional de la seva interlocutora perquè la proporcionada a l'inici de la interacció no era suficient per identificar quin dels animals disponibles s'adequava més al destinatari. Per això, *a priori* sembla que el disseny de la tasca no constitueix una

explicació satisfactòria dels resultats. El recull bibliogràfic de Pearson, Turner i Todd-Mancillas (1991: 115) mostra que no hi ha consens entre els autors que han observat el nombre de preguntes formulades en contextos interaccionals per homes i dones sobre quin dels sexes en faria més. A més, altres estudis, com s'ha explicat anteriorment, tampoc troben una correlació entre el nombre de preguntes i el sexe dels parlants. En el cas del nostre estudi, els resultats no permeten afirmar que el nombre de preguntes pugui resultar rellevant per a la perfilació lingüística en mostres comparables a l'analitzada, si bé s'observa el total de realitzacions de V16 tendeix a augmentar amb l'edat i que els nois tendeixen a fer-ne més que les noies.

7.2.3. Tipus de preguntes: V17

El total de realitzacions de la variant V17b, que comptabilitza les preguntes tancades formulades pels informants, és de 292, un nombre considerablement més elevat que el de la variant V17a, que considera les preguntes obertes, que és de tan sols 77 a tot el corpus d'anàlisi (Taula 26 de l'apartat 5.6). Pel que fa als resultats de les proves de significació estadística, la variant V17a es mostra significativament correlacionada amb el sexe dels informants a les dades corresponents als 16 anys. A continuació, la Taula 134 mostra que les diferències entre els totals dels dos tipus de preguntes formulades pels nois i les noies són més elevades als 16 anys i que a aquesta edat els nois formulen el doble de preguntes obertes que les noies del corpus d'anàlisi.

Taula 134. Realitzacions totals de la variable V17 i les seves variants segons el sexe i l'edat dels informants.

Estudi multivariant	12 anys		16 anys	
	Masculí	Femení	Masculí	Femení
V17a: Oberta	28 (\bar{x} : 0,9; DE: 1,64)	19 (\bar{x} : 0,51; DE: 0,8)	20 (\bar{x} : 0,63; DE: 0,71)	10 (\bar{x} : 0,27; DE: 0,51)
V17b: Tancada	63 (\bar{x} : 1,9; DE: 2,63)	65 (\bar{x} : 1,76; DE: 1,72)	105 (\bar{x} : 3,28; DE: 3,39)	85 (\bar{x} : 2,3; DE: 2,27)

Com s'ha exposat al **punt F** del subapartat 5.3.2.3.2, no hem trobat estudis previs que correlacionin el nombre de preguntes obertes o tancades que formulen els parlants d'una mostra amb característiques com l'edat, el sexe o la

llengua inicial. Així doncs, els resultats obtinguts aquí constitueixen un primer pas en la direcció de pal·liar aquest buit en la bibliografia sobre l'ús de la llengua. Caldria observar si estudis amb mostres diferents extreuen resultats semblants, a més de complementar el nostre treball amb aproximacions qualitatives i anàlisis quantitatives d'altres variables que puguin interaccionar amb el tipus de preguntes formulades. Malgrat tot, sembla que el nombre de preguntes obertes podria indicar, en mostres comparables a les del nostre corpus, el sexe més probable dels adolescents de més edat. Val a dir, però, que, tot i que no és estadísticament significativa, la tendència que s'observa a les dades de 12 anys és en la mateixa direcció que en les dades de 16 anys: els nois fan més preguntes obertes.

7.2.4. Construccions condicionals: V18

A l'apartat 5.6 hem observat que la gran majoria del total de realitzacions de V18 (2.062) correspon a construccions condicionals amb la partícula *si* (V18b, 2.009) i que, d'entre la resta, la següent partícula més utilitzada és *mentre(s)* (V18c, amb 27 ocurrències), seguida de construccions amb *a no ser que* (V18f, 15 realitzacions) i de les que contenen *en tot cas* (V18g, només 11 ocurrències).

Com hem vist anteriorment (al **punt G** del subapartat 5.3.2.3.2), la bibliografia consultada conté investigacions sobre el desenvolupament de les capacitats lingüístiques a les etapes infantils i adolescents. Aquests treballs mostren que les habilitats necessàries per a la comprensió i la producció de construccions condicionals complexes s'adquireixen relativament tard (per exemple, Moreno-Ríos i García-Madruga, 2002; Mariscal i Auza, 2017; Badger i Mellanby, 2018). A més, autors com Andersen (2001: 12) assenyalen la possibilitat que a l'inici de l'adolescència encara no es dominin usos epistèmics o vinculats amb els actes de parla d'aquestes construccions (del tipus “Si el que dius és veritat, jo soc el Papa de Roma!” o “Si no et fa res que t'ho preguntí, amb qui vas anar a la festa?”). Així doncs, era esperable, segons la bibliografia disponible sobre el desenvolupament de les habilitats lingüístiques, que la variable V18 mostrés una correlació estadísticament significativa amb l'edat dels informants.

Quant a la bibliografia variacionista, no hem trobat estudis comparables amb la nostra aproximació a les construccions condicionals en català. Tot i això, sí que n'hi ha en relació al castellà. D'una banda, Santana Marrero (1997) analitza els usos de diversos nexos condicionals a una mostra de parla de Sevilla i els relaciona amb el nivell sociocultural, l'edat i el sexe dels parlants. Si bé no du a terme proves de significació estadística dels resultats, aquesta autora conclou que “las locuciones condicionales no son conmutables por sí en todos los contextos” i que “el uso de una u otra forma de expresión está condicionada [sic] por las necesidades comunicativas del discurso y que el estudio contrastivo con la variable social no arroja datos significativos para la sociolingüística” (Santana Marrero, 1997: 317).

De l'altra banda, Jiménez Fernández (1997-1998) investiga l'ús d'oracions circumstancials (incloses les condicionals) en una mostra d'informants d'11, 12 i 13 anys. Concretament, se centra en els diversos nexes utilitzats en els diferents tipus de construccions que observa i en el sexe i l'edat dels parlants. Com en el nostre cas, els resultats de Jiménez Fernández (1997-1998: 833) corroboren que *si* és el nexa condicional “de mayor presencia en la lengua”. Pel que fa a l'edat i el sexe, els que més construccions condicionals produeixen en aquest estudi són les noies d'11 anys i els nois de 13 i, mentre que els nois fan més condicionals a mesura que creixen, les noies en fan més als 11 anys que als 13. De fet, tant uns com altres disminueixen el seu ús de condicionals als 12 anys. Les diferències entre els sexes és tan sols de 1,74 % més de producció per a les noies respecte els nois.

Els resultats de les proves estadístiques del nostre estudi, com s'ha explicat al capítol anterior (subapartat 6.3.2) confirmen la correlació entre V18 i l'edat al nostre corpus d'anàlisi. Tant el total d'ocurrències de V18 com el de la variant V18b (nexa *si*) augmenten amb l'edat, ja que el primer passa de 946 realitzacions totals a les dades de 12 anys a 1.116 a les de 16 anys i el segon, de 927 als 12 anys a 1.082 quatre anys després. Això no obstant, els resultats també mostren correlacions estadísticament significatives entre els totals de V18 i de V18b i la llengua inicial. En el cas de V18, aquesta correlació només és

estadísticament significativa a les dades corresponents als 12 anys, mentre que en el de V18b ho és també a les dades dels 16 anys. En tots dos, però, el grup que fa més construccions condicionals és el de catalanoparlants inicials, seguits dels castellanoparlants i dels bilingües inicials, en aquest ordre. La variant V18c (nexa *mentre(s)*) es mostra correlacionada amb el sexe dels informants a les dades de 12 anys: els nois en fan més ús que les noies. A continuació, la Taula 135 mostra els totals i les realitzacions mitjanes de V18 i les seves variants.

Taula 135. Realitzacions totals de la variable V18 i les seves variants segons l'edat, el sexe i la llengua inicial dels informants.

Estudi multivariant	12 anys						16 anys					
	Masculí			Femení			Masculí			Femení		
	C	B	E	C	B	E	C	B	E	C	B	E
V18b: <i>si</i>	309 (\bar{x} : 19,3; DE: 11,2)	42 (\bar{x} : 21; DE: 21,2)	140 (\bar{x} : 10; DE: 6,5)	142 (\bar{x} : 12,9; DE: 8,5)	110 (\bar{x} : 13,8; DE: 6,9)	184 (\bar{x} : 10,2; DE: 5,5)	312 (\bar{x} : 19,5; DE: 11,4)	42 (\bar{x} : 21; DE: 28,3)	164 (\bar{x} : 11,7; DE: 8,8)	169 (\bar{x} : 15,4; DE: 6,4)	155 (\bar{x} : 19,4; DE: 11,7)	240 (\bar{x} : 13,3; DE: 8,7)
V18c: <i>mentre(s)</i>	2 (\bar{x} : 0,1; DE: 0,3)	0 (\bar{x} : 0; DE: 0)	4 (\bar{x} : 0,3; DE: 0,6)	0 (\bar{x} : 0; DE: 0)	0 (\bar{x} : 0; DE: 0)	1 (\bar{x} : 0,1; DE: 0,2)	1 (\bar{x} : 0,1; DE: 0,3)	0 (\bar{x} : 0; DE: 0)	17 (\bar{x} : 1,2; DE: 3,4)	0 (\bar{x} : 0; DE: 0)	0 (\bar{x} : 0; DE: 0)	2 (\bar{x} : 0,1; DE: 0,3)
V18f: <i>a no ser que</i>	4 (\bar{x} : 0,3; DE: 0,4)	0 (\bar{x} : 0; DE: 0)	2 (\bar{x} : 0,1; DE: 0,5)	0 (\bar{x} : 0; DE: 0)	1 (\bar{x} : 0,1; DE: 0,4)	1 (\bar{x} : 0,1; DE: 0,2)	4 (\bar{x} : 0,3; DE: 0,8)	0 (\bar{x} : 0; DE: 0)	0 (\bar{x} : 0; DE: 0)	0 (\bar{x} : 0; DE: 0)	3 (\bar{x} : 0,4; DE: 1,1)	0 (\bar{x} : 0; DE: 0)
V18g: <i>en tot cas</i>	1 (\bar{x} : 0,1; DE: 0,3)	0 (\bar{x} : 0; DE: 0)	0 (\bar{x} : 0; DE: 0)	1 (\bar{x} : 0,1; DE: 0,3)	0 (\bar{x} : 0; DE: 0)	2 (\bar{x} : 0,1; DE: 0,3)	0 (\bar{x} : 0; DE: 0)	6 (\bar{x} : 3; DE: 4,2)	0 (\bar{x} : 0; DE: 0)	0 (\bar{x} : 0; DE: 0)	0 (\bar{x} : 0; DE: 0)	1 (\bar{x} : 0,1; DE: 0,2)
Estudi univariant												
Total V18	316 (\bar{x} : 19,8; DE: 11,4)	42 (\bar{x} : 21; DE: 21,2)	146 (\bar{x} : 10,4; DE: 6,6)	143 (\bar{x} : 13; DE: 8,4)	111 (\bar{x} : 13,9; DE: 7)	188 (\bar{x} : 10,4; DE: 5,7)	317 (\bar{x} : 19,8; DE: 11,6)	48 (\bar{x} : 24; DE: 32,5)	181 (\bar{x} : 12,9; DE: 10,5)	169 (\bar{x} : 15,4; DE: 6,4)	158 (\bar{x} : 19,8; DE: 12,3)	243 (\bar{x} : 13,5; DE: 8,7)

Així doncs, el nombre de construccions condicionals pot resultar útil per a la perfilació lingüística de mostres comparables a l'analitzada en el nostre estudi en relació amb l'edat i la llengua inicial. En el cas del nexa *mentre(s)*, en canvi, el

reduït nombre de realitzacions complica la generalització dels resultats a mostres diferents al corpus d'anàlisi.

7.2.5. Perífrasis d'obligació: V22

De les 1.339 realitzacions totals de la variable V22 identificades al corpus d'anàlisi, 860 corresponen a la variant V22b, és a dir, a la perífrasi *haver de* seguida d'infinitiu. La següent amb un nombre més elevat d'ocurrències és V22a (*tenir que* + infinitiu), que arriba a les 360 realitzacions. Les altres dues variants, V22e (*caldre* + infinitiu) i V22c (*tenir de* + infinitiu) es mostren força igualades, amb 61 i 58 ocurrències totals, respectivament.

Podem dividir la literatura relacionada amb aquesta variable que hem consultat en dos grups: d'una banda, els treballs que documenten l'ús de la variant no normativa *tenir que* + infinitiu i, de l'altra, els que estudien les correlacions sociolingüístiques de les perífrasis verbals d'obligació. En efecte, com hem vist anteriorment (**punt H** del subapartat **5.3.2.3.2**), l'ús de *tenir que* + infinitiu és una de les marques lèxiques del contacte entre el castellà i el català més sovint mencionada als estudis sobre la interacció entre els dos sistemes lingüístics (entre altres, Hualde, 1992: 325; Rodríguez-Vida, 1997; Vila i Moreno, 1998: 266; Martínez Díaz, 2002: 87; Paloma i Rico, 2008; Sinner, 2008: 534). Un d'aquests treballs és el d'Espuny (1998), que identifica *tenir que* + infinitiu com una de les quatre formes lèxiques provinents del castellà utilitzades al corpus que analitza que alternen amb formes "pròpies del català" (en aquest cas, *haver de* + infinitiu), "sense cap explicació contextual" (Espuny, 1998: 284).

Pel que fa a les publicacions que prenen una aproximació més comparable amb la nostra, destaquen els estudis de Mier (1986) i Stokes (2015). En el primer s'analitza l'ús de diversos recursos lingüístics (entre ells, les perífrasis d'infinitiu *haver de*, *tenir de* i *tenir que*) a una mostra de 43 catalanoparlants inicials de Barcelona en relació a l'edat, el sexe, el nivell d'ocupació, d'alfabetització en català i la influència del castellà a la infantesa dels informants. L'autora mostra que l'ús de *tenir que* reflecteix l'exposició al castellà dels parlants, ja que les diferències observades entre els grups de les variables independents es poden explicar pels canvis socials esdevinguts a la societat catalana al segle XX que van

augmentar la rellevància social del castellà. Les diferències detectades en els grups d'edat i de nivells d'alfabetització en català i d'influència del castellà a la infantesa diferents estarien relacionats amb la llengua en què els parlants van ser escolaritzats; les diferències entre els sexes, amb la incorporació de les dones a l'educació i al món laboral, esferes en què haurien augmentat la seva exposició al castellà, i, finalment, les diferències detectades entre els informants de nivells d'ocupació diferents es relacionarien amb una influència més gran del castellà tant a les capes més altes (representades a l'estudi per grans empresaris) com a les més baixes (la classe treballadora), a causa de la utilitat del castellà per als negocis i de l'onada immigratòria que va portar molts treballadors provinents de la resta de l'Estat a Catalunya a la segona meitat del segle passat (tractada al subapartat 2.3.4). Respecte a les perífrasis d'obligació, Mier (1986) conclou que *tenir de* es mostra clarament en retrocés (només és emprada per sis dels seus 43 informants) i, tal com efectivament demostren treballs com els que formen el primer grup mencionat més amunt, que:

“Mentre que *tenir que* no està estigmatitzat en general, podria esdevenir-ho en el futur. Amb l'augment actual que hi ha d'interès per l'idioma català, molta gent fa curssets de català. Els mestres interessats a restablir les normes fabrianes podrien fer que *tenir que* esdevingués objecte d'una clara estigmatització. Seria més fàcil modificar l'ús d'una estructura com *tenir que* entre els adults, que no pas introduir [sic] el contrast /s/ - /z/. Per tant, es preveu una minva futura en l'ús de *tenir que*, o com a mínim, el reconeixement públic que és un castellanisme.” (Mier, 1986: 57).

L'estudi de Stokes (2015), d'altra banda, analitza l'ús de *tenir que* i *haver de* a una mostra recollida a través de Twitter el 2013 i provinent de 8 ciutats repartides per Catalunya, el País Valencià i les Illes Balears. En total, obté 9.558 realitzacions de les dues perífrasis en present d'indicatiu (un 11 % de les quals corresponen a *tenir que*). Hi troba correlacions estadísticament significatives amb la persona gramatical utilitzada, així com amb factors diatòpics (una meitat de les ciutats pertanyen a la zona dialectal occidental i l'altra a l'oriental). Quant a la primera, l'autor observa que *tenir que* mostra una forta tendència a aparèixer en primera persona, mentre que *haver de* s'utilitza principalment en tercera

persona del singular (Stokes, 2015: 453). En relació a les varietats geolèctals, les ciutats valencianes mostren un ús més igualat de les dues perífrasis que la resta de ciutats (Stokes, 2015: 454).

En definitiva, la literatura disponible sobre les perífrasis d'obligació en català col·loquial ens va conduir a pensar que hi podria haver una correlació estadísticament significativa entre la variable V22 i la llengua inicial dels informants. Ara bé, els resultats de les proves de significació estadística exposats al subapartat **6.3.2** mostren que el total de realitzacions de V22 es correlaciona de manera significativa amb el sexe dels informants a les dades corresponents als 12 anys (els nois produeixen 269 ocurrències, les noies 406), mentre que les variants V22a (*tenir que*) i V22b (*haver de*) es correlacionen amb l'edat dels informants, si bé en sentits inversos. El total d'ocurrències de V22a disminueix a mesura que els informants es fan grans (passa de 252 als 12 anys a 108 als 16 anys), alhora que el de V22b augmenta (fins a les 502 realitzacions als 16 anys, partint de 358 als 12 anys). Finalment, V22e (*caldre + infinitiu*) sí que, com es preveia, es correlaciona de manera estadísticament significativa amb la llengua inicial dels informants (malgrat que només a les dades corresponents als 12 anys). Com es pot observar a la Taula 136, els catalanoparlants inicials fan setze realitzacions d'aquesta variant, mentre que els bilingües inicials tan sols en produeixen nou i els castellanoparlants inicials, tres.

Taula 136. Realitzacions totals de la variable V22 i les seves variants segons l'edat, el sexe i la llengua inicial dels informants.

Estudi multivariant	12 anys						16 anys					
	Masculí			Femení			Masculí			Femení		
	C	B	E	C	B	E	C	B	E	C	B	E
V22a: <i>tenir que</i> + INF	26 (\bar{x} : 1,6; DE: 3,8)	24 (\bar{x} : 12; DE: 17)	37 (\bar{x} : 2,6; DE: 3)	40 (\bar{x} : 3,6; DE: 5,3)	36 (\bar{x} : 4,5; DE: 5,5)	89 (\bar{x} : 5; DE: 6,6)	13 (\bar{x} : 0,8; DE: 1,2)	5 (\bar{x} : 2,5; DE: 3,5)	24 (\bar{x} : 1,7; DE: 3,1)	27 (\bar{x} : 2,5; DE: 5,8)	17 (\bar{x} : 2,1; DE: 3,3)	22 (\bar{x} : 1,2; DE: 2)
V22b: <i>haver de</i> + INF	125 (\bar{x} : 7,8; DE: 6,1)	2 (\bar{x} : 1; DE: 0)	33 (\bar{x} : 2,4; DE: 3)	58 (\bar{x} : 5,3; DE: 5,1)	25 (\bar{x} : 3,1; DE: 3,4)	115 (\bar{x} : 6,4; DE: 6,1)	153 (\bar{x} : 9,6; DE: 9,6)	5 (\bar{x} : 2,5; DE: 2,1)	56 (\bar{x} : 4; DE: 3,4)	66 (\bar{x} : 6; DE: 5,5)	76 (\bar{x} : 9,5; DE: 6,1)	146 (\bar{x} : 8,1; DE: 8,5)
V22c: <i>tenir de</i> + INF	3 (\bar{x} : 0,2; DE: 0,5)	5 (\bar{x} : 2,5; DE: 3,5)	0 (\bar{x} : 0; DE: 0)	14 (\bar{x} : 1,3; DE: 4,2)	0 (\bar{x} : 0; DE: 0)	15 (\bar{x} : 0,8; DE: 3,1)	7 (\bar{x} : 0,4; DE: 1,5)	1 (\bar{x} : 0,5; DE: 0,7)	1 (\bar{x} : 0,1; DE: 0,3)	4 (\bar{x} : 0,4; DE: 1,2)	0 (\bar{x} : 0; DE: 0)	8 (\bar{x} : 0,4; DE: 1,7)
V22e: <i>caldre</i> + INF	8 (\bar{x} : 0,5; DE: 0,9)	5 (\bar{x} : 2,5; DE: 3,5)	1 (\bar{x} : 0,1; DE: 0,3)	8 (\bar{x} : 0,7; DE: 1,3)	4 (\bar{x} : 0,5; DE: 1,1)	2 (\bar{x} : 0,1; DE: 0,5)	7 (\bar{x} : 0,4; DE: 1,5)	4 (\bar{x} : 2; DE: 1,4)	8 (\bar{x} : 0,6; DE: 1,3)	8 (\bar{x} : 0,7; DE: 1)	3 (\bar{x} : 0,4; DE: 1,1)	3 (\bar{x} : 0,2; DE: 0,4)
Estudi univariant												
Total V22	162 (\bar{x} : 10,1; DE: 6,3)	36 (\bar{x} : 18; DE: 17)	71 (\bar{x} : 5,1; DE: 3,1)	120 (\bar{x} : 11; DE: 5,4)	65 (\bar{x} : 8,1; DE: 4,5)	221 (\bar{x} : 12,3; DE: 6,9)	180 (\bar{x} : 11,3; DE: 10,1)	15 (\bar{x} : 7,5; DE: 0,7)	89 (\bar{x} : 6,4; DE: 4)	105 (\bar{x} : 9,5; DE: 6,3)	96 (\bar{x} : 12; DE: 5,6)	179 (\bar{x} : 10; DE: 7,7)

En resum, tot i que s'esperava trobar correlacions significatives entre V22 i la llengua inicial, se n'ha identificat amb totes tres variables independents, si bé tant la detectada entre el sexe i el total de perífrasis d'obligació com la que s'ha establert entre la llengua inicial i *caldre* + infinitiu només són rellevants per a les dades recopilades quan els informants tenien 12 anys. Així doncs, la dada que possiblement resulti més interessant per a estudis posteriors amb mostres comparables a l'analitzada aquí és que el nombre d'ocurrències totals de les variants V22a i V22b poden indicar l'edat més probable dels parlants.

7.2.6. Determinants demostratius: V23

El nombre total de realitzacions de la variable V23 identificades al corpus analitzat és de 894, com hem vist a l'apartat 5.6. La variant més freqüent és V23e, és a dir, el pronom demostratiu *aquest*, amb 426 ocurrences. La segona més utilitzada és V23a, el determinant *aquest* seguit de substantiu, amb 231 realitzacions. A continuació trobem V23g, que recull sintagmes formats pel numeral indefinit, un substantiu i el sintagma preposicional *d'aquests*, que també supera el centenar d'ocurrences (116). La resta de variants mostren un ús considerablement més reduït. Tot i això, destaca V23c, és a dir, l'ús de sintagmes formats per un article definit, un substantiu i l'adjectiu de primer grau, amb 56 realitzacions. La resta de variants, amb les formes de segon grau, són menys freqüents. La variant V23b (*aquell* seguit de substantiu) comptabilitza 23 realitzacions, la variant V23d (article definit, substantiu i adjectiu *aquell*) en mostra 17, la variant V23h (article indefinit, substantiu i sintagma preposicional *d'aquells*) es fa servir 14 vegades i, per últim, el pronom *aquell* només apareix 11 vegades al corpus.

Com apunta Martín Miñana (2019: 68) en una nota al peu de pàgina, “el català general [central col·loquial, en el nostre corpus] usa habitualment la forma de primer grau tant per a proximitat com per a funcionament no marcat: *aquest*, *aquí* i *això* fan referència no només a la proximitat extrema al parlant, sinó ocasionalment també assenyala la *mitja distància* (GCC 2008: 1492) o distància neutra”. Així doncs, és molt possible que la diferència notable en el nombre de realitzacions entre les variants que inclouen formes de primer i segon grau es degui al fet que les primeres comprenen un espai referencial més ampli que les darreres, relegades a referències més allunyades dels informants. Per comprovar-ho, però, caldria una anàlisi qualitativa de les dades que queda fora de l'abast del nostre treball.

La lingüística diacrònica s'ha ocupat, en més o menys detall, del desenvolupament dels demostratius, tant en el cas del català com del castellà (per exemple, Casanova Herrero, 1993 o Enrique-Arias, 2018). A més, en castellà, és possible trobar estudis previs dedicats als usos dels demostratius. Per

exemple, Morillo-Velarde Pérez, 1992, tracta usos propis de la parla de Sevilla i Ranson (1999) estudia la variació entre les posicions anteposada i posposada dels adjectius demostratius. Serrano (2006: 125) explica que l'ús de la darrera “le indica al oyente que interprete un valor afectivo positivo”, de manera que compleix una “funció emocional”.

En canvi, a diferència del castellà oral, encara manquen treballs sobre l'ús dels determinants a les varietats orals catalanes i la seva possible interacció amb factors sociolingüístics. Ara bé, sí que comptem amb publicacions que descriuen el paradigma dels demostratius de la llengua (p. ex. GEIEC, 2021a; Brucart, 2002; cf. Saragossà, 2004), així com amb treballs que se centren en l'ús dels elements dítics (inclosos els demostratius) a la llengua oral catalana.

Entre els últims destaquen els estudis d'Albaladejo (1998), Cuenca (2010) i Martín Miñana (2019). El primer analitza la dixi d'espai al corpus descrit a Tomàs i Pallejà (1998) i en relació amb la variació textual. El segon l'analitza a una mostra de narracions orals del Corpus Audiovisual Plurilingüe (Payrató i Fitó, eds., 2008). El tercer es focalitza en les expressions demostratives en un corpus inèdit de conversa col·loquial valenciana. Per tant, ni per les seves aproximacions (que, tot i combinar en alguns casos l'anàlisi qualitativa amb la quantitativa, no es proposen correlacionar els usos amb variables sociolingüístiques, sinó descriure les funcions que realitzen els mecanismes analitzats) ni per les característiques dels corpus observats (en cap cas, mostres de parla de menors d'edat), no es tracta d'investigacions comparables amb la nostra. Malgrat tot, a Cuenca (2010) les formes més utilitzades són, com al nostre corpus, les de primer grau.

Pel que fa a les proves de significació estadística del nostre estudi, hem comprovat al capítol anterior (subapartat 6.3.2) que, mentre que el total de realitzacions de la variable V23 no es correlaciona de forma estadísticament significativa amb cap de les variables independents, la variant V23a (*aquest* + substantiu) mostra una correlació significativa amb l'edat i la variant V23d (determinant definit + substantiu + demostratiu de segon grau), amb la llengua

inicial a les dades de 12 anys. A la Taula 137 s’hi pot observar que el nombre de realitzacions (tant totals com mitjanes) de V23a augmenta entre els 12 i els 16 anys i que, a les dades de 12 anys, l’únic grup que produeix ocurrencies de V23d és el de catalanoparlants inicials.

Taula 137. Realitzacions totals de les variants significatives de V23 segons l’edat i la llengua inicial dels informants.

Estudi multivariant	12 anys			16 anys		
	C	B	E	C	B	E
V23a: <i>aquest</i> + N	55 (\bar{x} : 2,04; DE: 3,4)	8 (\bar{x} : 0,8; DE: 1,5)	26 (\bar{x} : 0,81; DE: 1,03)	68 (\bar{x} : 2,52; DE: 2,55)	19 (\bar{x} : 1,9; DE: 1,52)	55 (\bar{x} : 1,72; DE: 2,07)
V23d: <i>el</i> + N + <i>aquell</i>	8 (\bar{x} : 0,3; DE: 0,67)	0 (\bar{x} : 0; DE: 0)	0 (\bar{x} : 0; DE: 0)	5 (\bar{x} : 0,2; DE: 0,4)	1 (\bar{x} : 0,1; DE: 0,32)	3 (\bar{x} : 0,1; DE: 0,4)

Per tant, els resultats indiquen que el nombre de realitzacions de V23a pot resultar útil per a la perfilació lingüística de mostres comparables al nostre corpus. Concretament, pot constituir un indicador de l’edat més probable dels parlants. Quant a la variant V23d, en canvi, cal cautela a l’hora de generalitzar els resultats, ja que el nombre de realitzacions a les dades corresponents als 12 anys és molt limitat.

7.2.7. Pronoms clítics de datiu plural: V25

Els recomptes de les freqüències d’aparició de les variants de V25 al corpus analitzat concorden amb la variació que recullen diverses gramàtiques de la llengua (per exemple, GIEC, 2016a: 213-4), com s’ha explicat al **punt J** del subapartat **5.3.2.3.2**, ja que la variant més utilitzada és la pròpia del català central col·loquial: *els hi* (V25a), que arriba a les 150 realitzacions. La següent variable més emprada és *li* (V25c) amb 99 ocurrencies, la qual cosa corrobora la tendència que descriu Pineda (2018) segons la qual el canvi de *li* per *els* (o *els hi*) és “un cas de despronominació del pronom feble de datiu, fruit de la generalització de la duplicació de clíctic datiu” equiparable al procés que experimenten altres llengües romàniques. La tercer variant amb més ocurrencies, 37, és la forma estàndard *els* (V25b) i, finalment, trobem la forma *lis* (V25d) en un total de 16 ocasions al llarg del corpus d’anàlisi. D’aquesta manera, el

recompte final de realitzacions identificades de pronoms clítics de datiu plural és de 302.

Com afirma Massanell i Messalles (2018: 86), “[e]ls clítics constitueixen una de les àrees d’estudi més fressades de la filologia catalana”. En efecte, comptem amb nombrosos estudis que reten compte de l’ús de les formes *els hi* i *li* com a pronoms clítics de datiu plural des de la teoria sintàctica (per exemple, Todolí, 1999, Boeckx i Martín, 2013, Pineda, 2018, o Cabré i Fábregas, 2019). Ara bé, fins on arriba el nostre coneixement, el fenomen de variació que reflecteix la variable V25, l’ús alternatiu de les formes que acabem de repassar, no s’ha investigat des de la sociolingüística de la variació.

Això no obstant, els resultats de les proves de significació estadística realitzades en aquest treball mostren correlacions estadísticament significatives entre les variants de V25 i la llengua inicial dels parlants (a les dades corresponents als 12 anys i, en el cas de V25d, també als 16), fet que apunta al potencial que té aquesta aproximació per contribuir a la comprensió global del fenomen. Concretament, com es pot observar a la Taula 138, la variant V25a (*els hi*) és l’única que presenta més realitzacions en el grup de catalanoparlants inicials. La resta, V25b (*els*), V25c (*li*) i V25d (*lis*) són utilitzades, sobretot, per castellanoparlants inicials. El total d’ocurrències de V25 no mostra correlacions estadísticament significatives amb cap de les variables independents.

Taula 138. Realitzacions totals de la variable V25 i les seves variants segons l’edat i la llengua inicial dels informants.

Estudi multivariant	12 anys			16 anys		
	C	B	E	C	B	E
V25a: <i>els hi</i>	29 (\bar{x} : 1,07; DE: 1,4)	16 (\bar{x} : 1,6; DE: 2,76)	17 (\bar{x} : 0,5; DE: 1,27)	44 (\bar{x} : 1,63; DE: 3,21)	8 (\bar{x} : 0,8; DE: 1,48)	36 (\bar{x} : 1,13; DE: 2,73)
V25b: <i>els</i>	0 (\bar{x} : 0; DE: 0)	4 (\bar{x} : 0,4; DE: 0,7)	15 (\bar{x} : 0,47; DE: 1,22)	2 (\bar{x} : 0,07; DE: 0,27)	3 (\bar{x} : 0,3; DE: 0,95)	13 (\bar{x} : 0,41; DE: 0,8)
V25c: <i>li</i>	6 (\bar{x} : 0,22; DE: 0,8)	5 (\bar{x} : 0,5; DE: 1,27)	32 (\bar{x} : 1; DE: 2,2)	17 (\bar{x} : 0,63; DE: 1,08)	11 (\bar{x} : 1,1; DE: 1,45)	28 (\bar{x} : 0,88; DE: 1,5)
V25d: <i>lis</i>	0 (\bar{x} : 0; DE: 0)	3 (\bar{x} : 0,3; DE: 0,68)	8 (\bar{x} : 0,25; DE: 0,57)	0 (\bar{x} : 0; DE: 0)	0 (\bar{x} : 0; DE: 0)	5 (\bar{x} : 0,16; DE: 0,37)

Així doncs, els resultats de l'estudi mostren que, com es desprenia de la bibliografia, les variants que formen la variable V25 pronoms clítics de datiu plural es correlacionen amb la llengua inicial dels informants. Aquesta dada pot resultar rellevant per a la perfilació lingüística de mostres anònimes comparables a les transcripcions que conté el corpus d'anàlisi.

7.2.8. Discussió dels resultats de les variables morfosintàctiques: Síntesi

En aquest apartat hem observat, com mostra la Taula 139 a continuació, que només una de les variables morfosintàctiques no mostra correlacions estadísticament significatives amb cap de les variables independents de l'estudi i, per tant, no sembla especialment útil per a la perfilació lingüística de mostres comparables al corpus analitzat. En canvi, la resta de variables morfosintàctiques observades, excepte V14 i V17, es correlacionen de forma significativa amb un nombre igual o superior de variables independents al que s'esperava.

Taula 139. Correlacions esperades i obtingudes per a les variables morfosintàctiques.

Variables morfosintàctiques	S'esperava que correlacionés amb...	Es correlaciona amb...
V14: Posició dels pronoms clítics en construccions perifràstiques	Edat, sexe, llengua inicial	Edat, llengua inicial
V16: Nombre de preguntes no retòriques	Sexe	∅
V17: Tipus de preguntes	Edat, sexe, llengua inicial	Sexe
V18: Construccions condicionals	Edat	Edat, sexe, llengua inicial
V22: Perífrasis d'obligació	Llengua inicial	Edat, sexe, llengua inicial
V23: Determinants demostratius	Edat, llengua inicial	Edat, llengua inicial
V25: Pronoms clítics de datiu plural	Llengua inicial	Llengua inicial

A més, hem vist que, en general, els treballs formals sobre aquestes variables morfosintàctiques catalanes no es poden complementar amb estudis variacionistes que tractin l'ús que en fan parlants de diversos grups socials. La

nostra investigació pretén, doncs, contribuir al progrés d'aquesta línia de recerca i, en especial, a l'estudi variacionista del llenguatge adolescent. Com s'ha subratllat, també seria interessant que en un futur aquesta aproximació es complementés amb anàlisis qualitatives per tal d'obtenir un coneixement més precís sobre la relació entre les formes identificades i les diverses funcions discursives per a les quals els parlants les utilitzen sense haver de renunciar a obtenir informació sobre la seva distribució sociolingüística.

7.3. Discussió dels resultats de les variables lèxiques

A partir de les característiques de les variables lèxiques i de la bibliografia consultada, durant les primeres etapes del projecte, pensàvem que aquestes variables (V38, V41 i V45) podrien mostrar correlacions estadísticament significatives amb l'edat, el sexe i la llengua inicial dels informants.

7.3.1. Fenòmens de contacte de llengües: V38

Com s'ha exposat a l'apartat 5.6, el recompte total de realitzacions de la variable lèxica fenòmens de contacte de llengües (V38) és de 5.598, la immensa majoria de les quals (5.501) corresponen a importacions (V38a). La segona variant més freqüent al corpus d'anàlisi és V38c Calcs amb 67 ocurrències totals. Les altres dues variants, V38d Altres i V38b Híbrids, se situen a força distància, amb 19 i 11 realitzacions respectivament.

El lèxic propi de les varietats catalanes orals s'ha descrit (per exemple, Payrató, 1996) i tractat des d'aproximacions lingüístiques diferents. Tenim estudis des de la terminologia i la lexicografia (entre altres, Lorente, Cabré i de Yzaguirre, 1997; Freixa, 2015 sobre la implantació del lèxic normatiu i el lèxic implantat però no normatiu, o Freixa, 2019, sobre l'ús de manlleus a la premsa catalana), així com també des de la lingüística cognitiva (vegeu, per exemple, estudis de disponibilitat lèxica com els de Llopis Rodrigo, 2008, 2009, i Serrano Zapata, 2016, 2021).

El lèxic català s'ha abordat també des de la dialectologia (vegeu, per exemple, Veny i Massanell, 2015), si bé cal tenir en compte que “[l]es enquestes dialectològiques solen fixar-se sobretot en el lèxic més tradicional i genuí de

l'àmbit geogràfic estudiat, que sol estar directament relacionat amb el mode de vida tradicionals” (Baldaquí Escandell, 2002: 30). Relacionats amb els estudis dialectològics, destaquen treballs sobre el lèxic català que podrien situar-se entre la dialectologia i la sociolingüística de la variació, ja que combinen mètodes de recollida de dades més tradicionals (qüestionaris sobre ítems lèxics) amb la investigació de les relacions entre les formes lingüístiques i característiques dels parlants com la seva edat o llengua inicial. En aquest grup hi encabim els treballs de Baldaquí Escandell (2002) sobre la comarca de l'Alacantí i el de Bagés Fortacín (2012) sobre Flix. El primer considera com a variables extralingüístiques l'exposició al valencià com a llengua vehicular a l'escola i la llengua inicial dels informants i conclou que:

“L'efecte de la llengua dominant en els jòvens, que s'observa sobretot a través de la introducció d'innovacions no normatives i castellanitzats, afecta principalment els dos grups d'alumnes que, per raons diferents, tenen major contacte amb el castellà: els que l'han après com a primera llengua i, sobretot, els que assisteixen a programes educatius vehiculats en castellà.” (Baldaquí Escandell, 2002: 47).

Així doncs, part de la literatura disponible sobre el lèxic català identifica una certa interacció entre la llengua inicial i l'ús d'unitats lèxiques genuïnes o provinents del contacte amb el castellà. Sobre aquest tema, cal tenir presents així mateix les contribucions que se centren en les característiques del llenguatge juvenil (p. ex. Rodríguez González, coord., 2002; Stenström i Jørgensen, eds., 2009). Més concretament, els treballs que han tractat el del català ho han fet, generalment, en relació amb la seva representació als mitjans de comunicació (vegeu Wieland, 2009b i Martínez Pérez, 2020, a més dels estudis tractats a l'apartat 4.4).

Això no obstant, els estudis previs més directament relacionats amb el nostre són els que estudien els fenòmens lèxics de contacte de llengües des d'una perspectiva quantitativa (com Espuny, 1998, i Vila i Moreno, 1998). Més concretament, com s'ha explicat al capítol 4, aquesta recerca beu directament, entre altres, de la sociolingüística educativa, en el marc de la qual treballs com el de Vila i Moreno (1996), Galindo Solé (2006) o, a nivell de màster i grau,

Sendra i Rovira (2012) i Pujadas Mestre (2020) han quantificat, abans que el nostre estudi, les eleccions lèxiques que reflecteixen el contacte de llengües en mostres de la parla de menors d'edat.

Tot i provenir d'informants (pre)adolescents, les mostres d'aquests treballs previs són prou diferents per oferir resultats complementaris, tant pel que fa a les zones geogràfiques i els anys en què es va realitzar la recollida de dades com a l'edat dels informants, com es pot observar a la Taula 140. En tots els estudis, el fenomen de contacte de llengües més freqüent a les mostres analitzades és el manlleu, que es correspon a la nostra variant V38a, també la més utilitzada en el nostre estudi. A més, la llengua donant principal és sempre el castellà. Totes dues troballes són explicables a partir de la resta de la bibliografia prèvia tractada en aquest apartat. El resultat més diferent entre els estudis que ens ocupen és el pes dels fenòmens de contacte de llengües als corpus analitzats, és a dir, el percentatge de paraules que representen respecte el total de les mostres. Val a dir que el percentatge que hem calculat per al nostre estudi i que es mostra a la taula és aproximat, ja que s'ha calculat a partir, d'una banda, del total de realitzacions de V38 (i cada realització no es correspon, en tots els casos, amb una sola paraula, ja que també s'hi inclouen expressions plurilèxiques, vegeu el **punt K** del subapartat **5.3.2.3.3**) i, de l'altra, del total de paraules del corpus sense tenir en compte les intervencions de les entrevistadores. Tot i això, se situaria entre els percentatges obtinguts als estudis la mostra dels quals és més semblant a la nostra (Vila i Moreno, 1996, Galindo Solé, 2006, i Sendra i Rovira, 2012), la qual cosa resulta comprensible.

Taula 140. Comparació d'estudis previs sobre fenòmens de contacte de llengües i resultats globals de la variable V38.

Estudi	Dades	Llengua donant principal	Fenomen de contacte de llengües més freqüent	Pes dels fenòmens de contacte de llengües (ocurrències)
Vila (1996)	Santa Coloma de Gramenet (Barcelona). 1991-1992. 26 alumnes de 5è de primària (10-11 anys) i 23 alumnes de 8è d'EGB (13-14 anys)	Castellà	Manlleu	3,57 %
Galindo (2006)	52 escoles de Catalunya. Principis dels 2000. >1.000 alumnes de 6è de Primària (11-12 anys)	Castellà	Manlleu	3,55 %
Sendra (2012)	Manlleu (Barcelona). 2008-2009. 15 alumnes castellanoparlants inicials a 6è de Primària (11-12 anys) i 1r d'ESO (12-13 anys)	Castellà	Manlleu	2,45 % (6è de Primària) 3,29 % (1r d'ESO)
Pujadas (2020)	Campanet (Mallorca). 2019. 2 alumnes castellanoparlants inicials i 2 alumnes catalanoparlants inicials de 6è de Primària (11-12 anys)	Castellà	Manlleu	10,19 %
Giménez García (2022)	Manlleu i Mataró (Barcelona). 27 alumnes catalanoparlants inicials, 10 alumnes bilingües i 32 alumnes castellanoparlants inicials a 6è de Primària (11-12 anys) i 4t d'ESO (15-16 anys)	Castellà	V38a Importacions	2,75 % (aprox.)

A diferència d'aquests tres estudis previs, però, en el nostre cas es correlacionen els resultats sobre l'ús de fenòmens de contacte de llengües amb l'edat, el sexe i la llengua inicial dels informants mitjançant proves de significació estadística. Els resultats exposats al capítol anterior (subapartat **6.3.3**) mostren, malgrat tot, que l'única relació estadísticament significativa s'estableix entre la variant V38d Altres i la llengua inicial a les dades corresponents als 12 anys. A continuació, la Taula 141 mostra el total de realitzacions i les realitzacions mitjanes de V38 i les seves variants. S'hi pot veure que, a les dades de 12 anys, l'únic grup que produeix la variant V38d és el de castellanoparlants inicials. Pel que fa a la variant V38a Importacions i als resultats globals de V38, hi podem observar la tendència (no significativa des del punt de vista estadístic) que els catalanoparlants inicials en fan lleugerament més ús que els castellanoparlants i els bilingües inicials.

Taula 141. Realitzacions totals (ocurrències) de la variable V38 i les seves variants segons l'edat i la llengua inicial dels informants.

Estudi multivariant	12 anys			16 anys		
	C	B	E	C	B	E
V38a: Importacions	1.120 (\bar{x} : 41,5; DE: 32)	341 (\bar{x} : 34,1; DE: 20,3)	1.216 (\bar{x} : 38; DE: 18,2)	1.125 (\bar{x} : 41,7; DE: 24,3)	387 (\bar{x} : 38,7; DE: 15,7)	1.312 (\bar{x} : 41; DE: 24,5)
V38b: Híbrids	0 (\bar{x} : 0; DE: 0)	1 (\bar{x} : 0,1; DE: 0,3)	7 (\bar{x} : 0,2; DE: 0,6)	1 (\bar{x} : 0; DE: 0,2)	0 (\bar{x} : 0; DE: 0)	2 (\bar{x} : 0; DE: 0,2)
V38c: Calcs	8 (\bar{x} : 0,3; DE: 0,5)	7 (\bar{x} : 0,7; DE: 0,9)	17 (\bar{x} : 0,5; DE: 0,8)	16 (\bar{x} : 0,6; DE: 1)	2 (\bar{x} : 0,2; DE: 0,4)	17 (\bar{x} : 0,5; DE: 0,8)
V38d: Altres	0 (\bar{x} : 0; DE: 0)	0 (\bar{x} : 0; DE: 0)	11 (\bar{x} : 0,3; DE: 1,3)	4 (\bar{x} : 0,2; DE: 0,6)	0 (\bar{x} : 0; DE: 0)	4 (\bar{x} : 0,1; DE: 0,3)
Estudi univariant						
Total V38	1.128 (\bar{x} : 41,8; DE: 32)	349 (\bar{x} : 34,9; DE: 20,2)	1.251 (\bar{x} : 39,1; DE: 18,5)	1.146 (\bar{x} : 42,4; DE: 24,2)	389 (\bar{x} : 38,9; DE: 15,8)	1.335 (\bar{x} : 41,7; DE: 24,7)

D'aquesta manera, per tant, els resultats obtinguts per a la variable V38 s'alineen amb i són explicables a partir dels estudis previs disponibles sobre els fenòmens de contacte de llengües, si bé només s'ha pogut identificar una correlació estadísticament significativa (entre V38d i la llengua inicial als 12 anys). Quant a la possible aplicació d'aquests resultats en la perfilació lingüística de mostres

d'autoria desconeguda comparables al corpus d'anàlisi, caldrà certa cautela a l'hora d'extreure conclusions sobre la llengua inicial més probable dels parlants a partir del seu ús de fenòmens de contacte de llengües, si bé el nombre d'ocurrències detectades permet identificar certes tendències.

7.3.2. Colors: V41

Al corpus d'anàlisi s'hi ha detectat un total de 115 realitzacions de la variable V41. En la majoria d'ocasions (77), la variant utilitzada pels informants és V41b, és a dir, el nom del color, mentre que la variant V41a, *color* seguit de l'adjectiu referent al color en qüestió, s'hi produeix un total de 38 vegades.

L'estudi de l'expressió lingüística dels colors s'ha tractat des de nombroses disciplines i perspectives. Per exemple, la filologia s'ha ocupat de les referències a colors en obres medievals (Martinell Gifre, 1986, o Montero Curiel, 1994), mentre que l'antropologia lingüística ha contribuït al coneixement general sobre el tema mitjançant la teoria dels termes bàsics de color (*basic color terms*) de Berlin i Kay (1969/1991) i diversos treballs que l'apliquen a cultures i comunitats de parla diferents (per exemple, Pawłowski, 2006; sobre el català, vegeu Davies, Corbett i Bayo Margalef, 1995). A més, comptem amb treballs que comparen la semàntica dels termes utilitzats en llengües (i cultures) diferents per fer referència als colors (com el de Zhang i Pengcuo, 2020, sobre el xinès i l'anglès). La semàntica és, precisament, una de les àrees de la lingüística que més s'ha ocupat dels colors (Biggam, 2012). Des de la psicolingüística i la lingüística cognitiva, s'ha recorregut en un bon nombre d'ocasions a l'estudi de la disponibilitat lèxica per accedir als termes utilitzats pels parlants per anomenar els colors. A continuació, tractem succintament tres exemples d'aquest últim grup de treballs previs, ja que indiquen una possible correlació entre les formes lingüístiques que fan referència a colors i el sexe dels parlants.

A l'estudi d'Al-Harashsheh i Obeidat (2017), la mostra utilitzada es compon de 40 homes i 40 dones universitaris d'Irbid, Jordània. Els investigadors els van proporcionar una roda amb l'espectre cromàtic i demanar que escriguissin en àrab el nom dels colors que hi percebessin. Les dones van anomenar un total de

111 termes referents als colors, mentre que els homes tan sols 63. Els autors expliquen que “Jordanian women associate colors with the names of socially familiar objects in order to distinguish between them” (Al-Harashseh i Obeidat, 2017: 248).

Espinosa-Zaragoza (2021) utilitza igualment una mostra d'estudiants universitaris. En el seu cas, 39 dones i 24 homes amb el castellà o el català i el català com a llengües inicials. L'autora se serveix d'un formulari en línia per recollir les dades que, a més de la prova de disponibilitat lèxica, inclouen informació sobre el nivell educatiu, l'edat, el gènere, l'orientació sexual i les aficions relacionades amb colors (per exemple, arts plàstiques, interiorisme, moda, indústria automobilística, etc.). Els resultats no mostren diferències gaire marcades entre els termes bàsics de color produïts per homes i dones, però sí certa “predilection for qualified colour terms” (Espinosa-Zaragoza, 2017: 134) en els homes i el fet que “females feel more open to using MCTs [metonymic colour terms] than males”, tendència ja observada en estudis anteriors (Espinosa-Zaragoza, 2021: 139).

El darrer exemple de l'ús de les tasques de disponibilitat lèxica en aproximacions empíriques a les formes referents als colors és el treball de Paredes García (2006), en el qual es combina amb dades del projecte PRESEEA. La mostra de les enquestes de disponibilitat lèxica la formen 484 alumnes de 3r i 4t de l'ESO i 1r i 2n de batxillerat, mentre que del projecte PRESEEA se n'utilitzen 36 gravacions obtingudes al barri madrileny de Salamanca. Aquesta vegada, a la prova d'enumeració de termes, els resultats mostren diferències entre els sexes “no solo en cuanto al conocimiento de este campo léxico, sino que también parecen indicar que es diferente tanto la percepción del color como, incluso, la discriminación de la gama”, una vegada més, en el sentit que les dones produeixen més termes que els homes, si bé les diferències no arriben a ser significatives (enumeren 313 i 294 paraules diferents, respectivament) “y pueden deberse a que la muestra estaba compuesta por un número mayor de mujeres” (Paredes García, 2006: 134). Això no obstant l'autor observa diferències més aviat qualitatives: “Según mis datos, la diferencia entre mujeres

y hombres tiene que ver no sólo con la cantidad de términos metonímicos que emplean los dos sexos (más las mujeres que los hombres), sino también con el valor semántico de dichos sustantivos (los de las mujeres establecen una escala cromática mucho más amplia que los de los hombres)” (Paredes García, 2006: 134).

Pel que fa a l'anàlisi del corpus provinent de PRESEEA, el nombre de referències cromàtiques identificades és relativament baix: “sólo aparecen 116 palabras en un corpus de aproximadamente 300000, lo que representa un exiguo 0,039% con relación al total de unidades” (Paredes García, 2006: 135). Si fem el càlcul equivalent al nostre corpus, podem observar que el resultat és força semblant, ja que, del total de 333.612 paraules transcrites del corpus, només s'han detectat 115 realitzacions de V41, la qual cosa equival al 0,035 %. Quant a les diferències detectades per Paredes García (2006) entre els sexes a les gravacions que estudia, mentre que el nombre total de substantius diferents utilitzats per homes i dones és molt semblant (10 i 11, respectivament), la clau sembla trobar-se en el “número de veces que [se] emplea las palabras referidas al color: en las conversaciones de las mujeres aparecen 82 referencias al color por tan solo 36 de los hombres” (Paredes García, 2006: 137).

En aquest sentit, les reflexions finals que fa l'autor podrien, aparentment, aplicar-se també al nostre corpus d'anàlisi, ja que, tot i que tant les converses de PRESEEA com la prova interactiva de RESOL contenen fragments descriptius en els quals podria esperar-se que els informants fessin referències als colors, “su aparición en el discurso es muy reducida” (Paredes García, 2006: 138). És a dir, la comparació de la prova de disponibilitat lèxica i dels resultats de l'anàlisi de les interaccions indica que no hi ha una relació directa entre el coneixement de formes referents a colors per part dels parlants i l'ús que en fan a les entrevistes. Segons l'autor, això es deu a diverses raons: el tema de conversa, i el fet que el color es percebria com un element irrellevant excepte en situacions determinades. Així doncs, Paredes García (2006: 138) proposa que els colors s'utilitzin en tres circumstàncies:

- “cuando el término del color es imprescindible como elemento discriminador de unidades,
- cuando el contexto no resulta suficientemente explícito,
- cuando el término de color pertenece a una unidad fraseológica o se usa con valor metafórico o metonímico”.

Seria interessant observar si aquestes observacions podrien explicar també la freqüència d'ús de la variable V41 al nostre corpus. A més de les aproximacions que hem tractat fins ara, les referències a colors també han estat objecte d'estudi de la sociolingüística variacionista, si bé menys sovint. Al-Jehani (1990), per exemple, correlaciona les dades obtingudes mitjançant una prova d'elicitació amb el sexe (homes o dones), el grup ètnic (nòmades o urbanites) i l'edat (15 anys o menys, entre 20 i 30 anys, 45 anys o més) en una mostra de 36 informants de la Meca, a l'Aràbia Saudita. Els resultats mostren diferències per totes tres variables. Les dones produeixen més termes diferents que els homes, els informants d'entre 20 i 30 en produeixen més que els altres grups d'edat i els urbanites, més que els nòmades.

Quant a la llengua catalana, no hem pogut identificar estudis previs sobre les referències als colors des de la sociolingüística de la variació, però sí els treballs lexicogràfics de Grossmann (1980) i Bozzo Duran (1986), així com l'anàlisi semàntica des de la lingüística cognitiva de Domènech Bagaria i Estopà Bagot (2015). En definitiva, doncs, la literatura prèvia sobre la variació en les referències lingüístiques a colors en relació amb característiques dels parlants apunta a una possible correlació amb el sexe (a més dels tractats fins ara, cf. López García i Morant, 1991; García Mouton, 2003 ; Espinosa-Zaragoza, 2021: 120-126), de manera que esperàvem identificar correlacions estadísticament significatives entre la variable V41 Colors i aquesta variable. A més, però, la manca d'estudis amb objectius comparables als nostres per al català ens va conduir a mantenir la porta oberta a identificar-ne també amb les altres dues variables independents, l'edat i la llengua inicial dels informants.

Malgrat tot, els resultats exposats al capítol anterior (subapartat **6.3.3**) indiquen que les úniques relacions estadísticament significatives s'estableixen entre el total de realitzacions de V41 i de la variant V41b (absència del substantiu *color*) i la llengua inicial a les dades corresponents als 16 anys. Com mostra la Taula 142, als 16 anys, són els catalanoparlants inicials els que fan més ús de V41b i V41, a una distància notable de la resta d'informants, mentre que les distàncies entre els grups a les dades de 12 anys són molt menors.

Taula 142. Realitzacions totals de la variable V41 i les seves variants segons l'edat i la llengua inicial dels informants.

Estudi multivariant	12 anys			16 anys		
	C	B	E	C	B	E
V41a: <i>color</i> [blau]	12 (\bar{x} : 0,4; DE: 0,8)	3 (\bar{x} : 0,3; DE: 0,7)	11 (\bar{x} : 0,3; DE: 0,8)	4 (\bar{x} : 0,2; DE: 0,5)	2 (\bar{x} : 0,2; DE: 0,4)	6 (\bar{x} : 0,2; DE: 0,5)
V41b: [blau]	14 (\bar{x} : 0,5; DE: 1)	1 (\bar{x} : 0,1; DE: 0,3)	18 (\bar{x} : 0,6; DE: 1,4)	35 (\bar{x} : 1,3; DE: 2)	4 (\bar{x} : 0,4; DE: 1)	5 (\bar{x} : 0,2; DE: 0,4)
Estudi univariant						
Total V41	26 (\bar{x} : 1; DE: 1,7)	4 (\bar{x} : 0,4; DE: 0,7)	29 (\bar{x} : 0,9; DE: 2,1)	39 (\bar{x} : 1,4; DE: 2,1)	6 (\bar{x} : 0,6; DE: 1,3)	11 (\bar{x} : 0,3; DE: 0,8)

En resum, els resultats obtinguts podrien resultar rellevants per a la perfilació lingüística de mostres dubitades comparables a les del corpus analitzat, però no respecte la variable que més probable semblava a partir de la revisió bibliogràfica, el sexe, sinó pel que fa a la llengua inicial dels parlants adolescents de més edat.

7.3.3. Nombre d'adverbis acabats en *-ment*: V45

El nombre d'ocurrències totals de l'última variable de l'estudi, els adverbis acabats en *-ment* produïts pels informants, és de 391, com hem vist al capítol 5 (apartat **5.6**). La revisió bibliogràfica, com també s'ha exposat anteriorment (al punt **M** del subapartat **5.3.2.3.3**) ha permès identificar treballs sintàctics i diacrònics sobre aquests adverbis (per exemple, Marlangeon, 2002; Torner, 2007; Hummel, 2013), a més de publicacions que apunten a una possible relació dels adverbis en *-ment* amb l'edat (Nippold, 1998) i el sexe (Tuson, 1998) dels parlants.

Sobre l'ús d'aquests adverbis en edats diferents, cal destacar, a més, l'estudi de Barriga Villanueva (1994), en el qual analitza converses amb 24 nens i nenes de sis anys i nois i noies de dotze anys. Els seus resultats mostren moltes menys realitzacions en les dades provinents d'infants de sis anys que en les d'infants de dotze:

“Como se desprende de estas tablas, los adverbios en *mente* son escasos en el habla de niños de seis años, frente a una relativa abundancia en el habla de niños de doce (relación 1 a 10). Estos resultados no parecen fortuitos sino que obedecen a ciertos rasgos del habla vinculados directamente tanto con el desarrollo cognoscitivo y lingüístico, así como con el conocimiento del mundo de los niños en momentos clave de su formación. Como ya se había mencionado, los conceptos de modalidad, espacialidad, temporalidad o cantidad que subyacen a estos adverbios, resultan difíciles de comprender sobre todo para niños que inician los años escolares.” (Barriga Villanueva, 1994: 566-567).

L'observació d'aquesta variable en produccions lingüístiques infantils, per tant, reflecteix part del procés d'adquisició lingüística que té lloc a la infantesa. En aquest sentit, l'autora proposa que l'ús d'adverbis en *-ment* “supone un gradual dominio que atraviesa varios niveles: sintáctico, semántico y discursivo y que trasciende las fronteras de la mera competencia lingüística para llegar a la competencia comunicativa” (Barriga Villanueva, 1994: 565).

A més del sexe i l'edat, però, també s'ha estudiat la relació de l'ús d'aquests adverbis amb altres variables sociolingüístiques. Sobre la llengua anglesa, trobem estudis com el de Macaulay (2002), que analitza dues mostres de parla d'Escòcia (una d'Ayr i una altra de Glasgow) i corrobora els resultats d'un treball anterior (Bernstein, 1971) pel que fa a la correlació entre la producció d'adverbis acabats en *-ly* i la classe social. Els parlants de classe social mitjana els utilitzen, en general, més de dues vegades més que els parlants de classe treballadora i aquest patró s'observa tant entre els adults com entre els adolescents.

En un altre treball variacionista, Tagliamonte i Ito (2002) analitzen la variació entre els adverbis acabats en *-ly* i *zero adverbs*, és a dir, l'ús adverbial d'adjectius (per exemple, *it was {really/real} small*) en una mostra de 91 parlants (37 homes

i 54 dones d'entre 15 i 91 anys d'edat) de York, Anglaterra, en relació amb l'edat, el sexe i el nivell educatiu dels informants, però també amb el significat i la funció dels adverbis. En general, identifiquen un canvi dràstic en temps aparent: mentre que els parlants de més edat prefereixen construccions amb adjectius en funció adverbial (del tipus *real good*), els més joves utilitzen majoritàriament adverbis acabats en *-ly* (com *really good*; Tagliamonte i Ito, 2002: 252). Les autores observen, no obstant això, que la majoria de realitzacions d'adverbis en *-ly* corresponen a *really* (un 66 %), de manera que aïllen aquest adverbi en la seva anàlisi, ja que es comporta de forma diferent a la resta d'adverbis:

“However, the frequency of use of *really* increases markedly from oldest to youngest, with the youngest speakers hardly ever using the zero-marked form – only 7 percent. In contrast, the frequency of the zero form amongst all the other adverbs decreases only marginally across the generations” (Tagliamonte i Ito, 2002: 251).

Les proves de significació estadística mostren que la variable de més pes és el significat de l'adverbi, que pot ser concret o abstracte: “adverbs which are associated with concrete meaning highly favour the zero-marked form at .69, while those associated with abstract meaning highly disfavour at .22, making them much more likely to be marked with *-ly*” (2002: 256). D'entre les variables socials, la correlació més forta és amb la variable nivell educatiu, tant per a *really* com per a la resta d'adverbis, seguida a poca distància per l'edat. En canvi, el sexe dels informants no mostra significació estadística en el cas de *really*, però sí per als altres adverbis: “males favour the zero form at .62; females disfavour the form at .44” (2002: 254).

Pel que fa al català i al castellà, comptem amb estudis correlacionals que s'ocupen de la relació entre l'ús d'adverbis en *-ment* i nivells de formalitat diferents. En català, destaca l'estudi de Gelpí (1998) i, en castellà, el de Kortschak i Hummel (2018). El primer, recordem, es planteja la hipòtesi que el to de les mostres (formal o informal) està relacionat amb l'aparició d'aquests adverbis i els resultats quantitius obtinguts li permeten a l'autora confirmar-la:

“si el text té to formal, la freqüència d'adverbis serà més alta que si el text té to informal” (Gelpí, 1998: 140).

A l'estudi de Kortschak i Hummel (2018) s'analitza un subcorpus de C-Oral-Rom (Cresti i Moneglia, 2005) de més de 400 parlants. La majoria (188) són de la ciutat de Madrid i voltants, 61 són d'altres regions castellanoparlants i 161 són d'origen desconegut. Aquesta mostra es divideix en un subcorpus formal i un d'informal i els autors comparen els resultats que n'obtenen amb dades de la parla de Mèxic explotades prèviament per Arjona Iglesias (1991). En aquesta investigació, com a la de Gelpí (1998), els textos formals presenten més adverbis en *-ment* que els informals. Concretament, els autors afirmen que “los adverbios en *-mente* ocurren tres veces más en los textos formales” (Kortschak i Hummel, 2018: 12), tant al seu corpus com a les dades mexicanes. A més, si bé no observen que el nivell d'estudis mostri cap tendència aparent respecte l'ús d'adverbis en *-ment*, sí que troben que “todos los hablantes que muestran una predilección por este tipo de adverbios pertenecen al estrato más cultivado” (2018: 17).

Altres factors que es desprèn de la bibliografia que caldria tenir en compte en estudis futurs sobre els adverbis acabats en *-ment* són les diferències individuals entre els parlants (segons Kortschak i Hummel, 2018: 17, “[l]a preferència por usar o no usar los adverbios en *-mente* es una característica básicamente idiolectal o individual situacional”), la influència del cotext (ja que De Smet, 2018, demostra, amb dos corpus diferents, que l'aparició d'un adverbi acabat en *-ly* propicia, de manera estadísticament significativa, una nova aparició d'un altre adverbi d'aquest tipus amb una arrel diferent a les línies següents) i, finalment, possibles processos de canvi en marxa, com la pluralització d'adverbis en *-mente* que mostren Felú Arquiola i Pato (2019).

Així doncs, malgrat que, com es després d'aquest breu repàs bibliogràfic, els adverbis acabats en *-ment* han rebut força atenció de lingüistes de tradicions diferents, fins on arriba el nostre coneixement no s'ha abordat la possible correlació entre el seu ús per part d'informants adolescents i les variables que tractem en el nostre estudi. Si bé, com hem vist, podíem esperar que les proves

de significació estadística mostressin resultats positius per a les variables edat i sexe, com s'ha explicat al capítol anterior (subapartat **6.3.3**), només ha estat així en el primer cas. La Taula 143 recull el total de realitzacions de V45 i s'hi pot observar que a les dades de 16 anys el total i la realització mitjana per part dels informants és més del doble que a les de 12 anys.

Taula 143. Realitzacions totals de la variable V45 i les seves variants segons l'edat dels informants.

Estudi univariant	12 anys	16 anys
Total V45	124 (\bar{x} : 1,80; DE: 3,03)	267 (\bar{x} : 3,87; DE: 4,47)

D'aquesta manera, per tant, la variable lèxica V45, és a dir, el recompte d'adverbis acabats en *-ment* pot resultar útil per a la identificació de l'edat més probable de parlants adolescents en mostres comparables pel que fa a les característiques lingüístiques (català central col·loquial) i textuais (mostra conversacional, no planificada, etc.) amb les que formen el corpus d'anàlisi.

7.3.4. Discussió dels resultats de les variables lèxiques: Síntesi

A partir dels resultats obtinguts a les proves de significació estadística hem pogut comprovar que cap de les tres variables lèxiques d'aquesta investigació mostren resultats rellevants per a la identificació del sexe més probable de parlants adolescents. Ara bé, tant V38 com V41 poden aportar informació útil per a la identificació de la llengua inicial dels parlants a través de l'anàlisi de les seves produccions lingüístiques i V45 pot situar els parlants adolescents en el rang d'edat que s'estén dels 12 als 16 anys.

Els repàs dels estudis previs sobre cadascuna de les variables lèxiques observades ens ha permès explicar no sols el potencial de les variables sinó també, en el cas de V38 i V45, els resultats obtinguts i, en tots els casos, l'abast de les contribucions que fa l'estudi proposat aquí al coneixement previ d'aquests fenòmens de variació identificats al corpus. L'aproximació realitzada aquí a les fórmules de referència als colors (V41) obre un camí altrament inexplorat en llengua catalana, mentre que l'estudi proposat de V38 i de V45 amplien les dades obtingudes en treballs variacionistes previs sobre els fenòmens de contacte

de llengües i els adverbis acabats en *-ment*, tant pel que fa a la nostra llengua com als estudis de la parla dels adolescents. La Taula 144 mostra les variables independents amb les quals s'esperava que es correlacionessin de forma estadísticament significativa les variables lingüístiques i les variables amb les quals l'anàlisi ha permès confirmar que mantenen una relació significativa.

Taula 144. Correlacions esperades i obtingudes per a les variables lèxiques.

Variables lèxiques	S'esperava que correlacionés amb...	Es correlaciona amb...
V38: Fenòmens de contacte de llengües	Edat, sexe, llengua inicial	Llengua inicial
V41: Colors	Edat, sexe, llengua inicial	Llengua inicial
V45: Nombre d'adverbis acabats en <i>-ment</i>	Edat, sexe	Edat

7.4. Discussió dels resultats: síntesi

En aquest capítol, hem pogut dotar de significat els resultats quantitatius exposats al capítol anterior mitjançant la contextualització bibliogràfica de l'estudi de cadascuna de les variables pragmàtiques, morfosintàctiques i lèxiques que hem analitzat. A més, hem reflexionat sobre les conseqüències que poden tenir els resultats obtinguts per a la fonamentació empírica de la perfilació sociolingüística de mostres de parla adolescent en català.

Si considerem ara els resultats globals que s'esperaven en comparació amb els obtinguts, podem constatar que tres de les vint variables (és a dir, el 15 %) mostren correlacions estadísticament significatives amb més variables independents de les que esperàvem. Aquestes variables són el nombre d'heterorepeticions (V6), les construccions condicionals (V18) i les perífrasis d'obligació (V22). Unes altres tres variables, els doblats de partícules *a ver / a veure* (V3b), *hombre / home* (V3h) i *o sea / o sigui* (V3i), mostren correlacions significatives amb el mateix nombre de variables que s'esperava (dues), però una de les identificades és diferent a la que s'havia proposat abans de les anàlisis. A continuació, trobem dues variables (un 10 % de les incloses a l'estudi) que mostren les mateixes correlacions que s'esperaven: els determinants

demonstratius (V23) i els pronoms clítics de datiu plural (V25). Les altres dotze variables (el 60 %) mostren menys correlacions estadísticament significatives de les esperades inicialment.

Com hem comprovat en aquest capítol, el corpus d'anàlisi presenta característiques valuoses per a la fonamentació empírica de la perfilació lingüística. Es tracta d'un corpus format per material audiovisual transcrit i corresponent al mateix grup d'individus en dos moments temporals diferents (estudi panel). La situació comunicativa enregistrada, a més a més, és controlada, ja que es tracta de la mateixa prova interactiva (tasca de joc de rol), però, alhora, permet que els informants s'expressin oralment amb més llibertat i espontaneïtat que altres mètodes de recollida de dades (com proves de disponibilitat lèxica, elicitació de sinònims, qüestionaris o entrevistes semidirigides), la qual cosa es fa palesa en la creativitat demostrada per alguns dels informants en el control temàtic de la conversa amb l'entrevistadora. Per tant, els resultats obtinguts de l'anàlisi quantitativa d'aquest corpus constitueix una aportació rellevant per a la pràctica de la perfilació sociolingüística general, però també, a nivell local, per a la pràctica de perfilació sociolingüística de mostres en llengua catalana i, encara més específicament, de mostres que hagin pogut estar produïdes per parlants adolescents.

8 CONCLUSIONS

Aquesta tesi sorgeix d'una necessitat específica, la de comptar amb estudis descriptius de les varietats col·loquials de català utilitzades per parlants adolescents que correlacionin trets lingüístics amb variables socials, com l'edat, el sexe i la llengua inicial dels parlants, i que, d'aquesta manera, siguin directament aplicables a tasques de perfilació sociolingüística de mostres d'autoria desconeguda sol·licitades en el marc de processos judicials o d'investigacions privades o policials. La revisió bibliogràfica ha demostrat que, en el cas de la llengua catalana, encara no existeixen estudis previs que cobreixin aquesta necessitat, ja que tant els estudis que s'ocupen de la variació (és a dir, que analitzen trets lingüístics en relació a característiques dels parlants) com els que tracten l'ús de la llengua que fan els parlants adolescents presenten diverses limitacions importants per a la seva aplicació en la perfilació lingüística.

En primer lloc, la majoria dels treballs que tracten la variació sociolingüística del català es van realitzar fa vint anys o més i, per tant, difícilment són aplicables a mostres actuals, ja que la llengua (i, especialment, les varietats utilitzades per adolescents) evoluciona constantment. En segon, aquests estudis solen centrar-se en la parla pròpia d'una localitat o d'una regió específica que resulta especialment interessant als lingüistes a càrrec de la recerca perquè pot presentar canvis lingüístics en marxa o perquè es troba en una zona de transició entre varietats o per altres raons relacionades amb la intersecció entre la llengua i els sectors de població (per exemple, és una zona on s'han produït canvis demogràfics destacats). Malgrat l'innegable interès acadèmic d'aquests treballs, doncs, la generalització dels resultats a mostres potencialment provinents d'altres àrees geogràfiques on es parla la mateixa llengua queda seriosament compromesa.

Un tercer grup de limitacions que presenten diversos estudis variacionistes sobre la llengua catalana per a la seva aplicació a la perfilació sociolingüística de textos possiblement produïts per adolescents són les relacionades amb la mostra que utilitzen. En molts casos, la mida d'aquestes mostres es limita a la inclusió d'un nombre concret de representants dels grups etaris i socials que es pretén

estudiar. Per exemple, si partim de tres variables socials binàries (considerem, posem per cas, el sexe, dos grups d'edat, i dues classes socials) i utilitzem dos representants de cada grup, amb 16 individus podem dur a terme un estudi que ens permeti obtenir informació sobre la relació entre aquestes variables i els trets lingüístics que analitzem. El problema que presenta aquest procediment de mostreig estratificat no proporcional per a l'aplicació dels resultats a la perfilació lingüística, però, és que, en termes demogràfics, no podem afirmar que una mostra com aquesta de 16 persones sigui representativa d'una comunitat, ja que no té en compte la distribució real de cada grup a la societat de partida (si hi ha, per exemple, molts més individus d'una de les dues edats analitzades).

Pel que fa als treballs sobre la llengua catalana dels adolescents, provenen de diverses aproximacions a la lingüística: l'adquisició del llenguatge, la sociolingüística educativa o la didàctica de les llengües. Hem observat que comptem, d'una banda, amb pocs estudis, en comparació amb els que descriuen els usos lingüístics d'altres grups etaris i, de l'altra, que molts dels que s'han publicat se centren en la qüestió de la representació de les varietats juvenils als mitjans de comunicació, especialment pel que fa la tensió entre la voluntat de respectar la normativa i la de representar de forma realista el llenguatge juvenil. La resta, és a dir, les publicacions que descriuen trets lingüístics concrets que caracteritzen la parla dels adolescents catalanòfons i que aporten informació directament rellevant per a la perfilació lingüística se centren, majoritàriament, en els fenòmens de contacte entre el català i el castellà i, una vegada més, utilitzen mostres de població que presenten serioses limitacions per a la generalització dels resultats.

Així doncs, l'angle adoptat en aquesta tesi, que combina un corpus de parla adolescent de mida considerable (69 informants) obtingut com a part d'un projecte d'investigació en sociolingüística educativa, el plantejament metodològic propi de la sociolingüística de la variació i l'enfocament pràctic de la lingüística forense constitueix una contribució innovadora a l'àmbit acadèmic català que resulta prometedora, en tant que obre un gran ventall de possibilitats d'ampliació de la recerca en aquesta llengua en el marc de la lingüística

aplicada. En aquesta tesi, partim, doncs, de la pregunta de recerca “hi ha variables lingüístiques que permeten identificar l’edat, el sexe i la llengua inicial més probable en el discurs d’un parlant adolescent en català?”. Per respondre-la, l’objectiu que ens hem plantejat és el d’analitzar un seguit de variables lingüístiques aplicables a la perfilació sociolingüística de produccions anònimes en català produïdes per (pre)adolescents. El sisè capítol recull els resultats de les nostres anàlisis i, com s’exposa al setè capítol, només una de les variables lingüístiques que hem examinat no mostra cap correlació estadísticament significativa ni amb l’edat, ni amb el sexe ni amb la llengua inicial dels informants. Quant a la resta de variables, com a mínim una de les variants que les formen mostra correlacions estadísticament significatives amb una o més variables independents. Per tant, la resposta a la nostra pregunta de recerca ha de ser afirmativa.

Per tal d’assolir el nostre objectiu principal, hem dividit la recerca en quatre objectius específics. El primer consistia a identificar variables susceptibles de ser realitzades pels informants que formen el corpus d’anàlisi. Per assolir-lo, ens hem servit, d’una banda, d’una revisió bibliogràfica de la literatura descriptiva i variacionista de la llengua catalana (i, en alguns casos, de la llengua castellana), així com de la dedicada a la lingüística forense en català i castellà. De l’altra banda, hem dut a terme una primera exploració d’una selecció del corpus d’anàlisi que ens ha permès identificar, de les possibles variables lingüístiques utilitzades en estudis previs, les més susceptibles d’aparèixer a la nostra mostra en un nombre de realitzacions suficient per a l’anàlisi estadística, descartar aquelles que no eren aplicables a la varietat de llengua del corpus i proposar algunes variables lingüístiques que encara no s’havien utilitzat en estudis de la variació en català. Finalment, doncs, hem inclòs a l’estudi un total de 21 variables lingüístiques de tres nivells diferents (deu de pragmàtiques, vuit de morfosintàctiques i tres de lèxiques).

El segon objectiu específic que hem assolit és el d’identificar al corpus les variants que realitzen els informants. Concretament, a causa de les restriccions imposades pel tractament estadístic de les dades, hem considerat tan sols

aquelles variants i variables que apareixen un mínim de deu vegades en tot el corpus. D'aquesta manera, a tall d'exemple, hem descartat per falta d'ocurrències variables com les fórmules de comiat o l'ús de verbs acabats en *-guer* (*volguer*, *poguer*, etc.) i variants com estructures condicionals amb els nexes *amb/en que* (V18d), *con que* (V18e) o *només que* (V18h). En relació a les variants que hem analitzat, cal subratllar també que, des d'una aproximació descriptivista a l'estudi de la llengua coherent amb la metodologia variacionista en la qual se sustenta el treball, hem inclòs variants no normatives però pròpies de les varietats orals col·loquials (per exemple, V22a *tenir que*).

El tercer objectiu específic ha consistit en la quantificació de totes les ocurrències de les variants i variables de l'estudi presents al corpus. Per a aquesta tasca, es va recórrer tant al procediment "manual" de marcatge de realitzacions (diverses iteracions del cicle d'operacions: lectura atenta, identificació i codificació d'ocurrències i revisió) com a l'ús d'un conjunt de comandaments executats amb la interfície de consola Windows Powershell per tal d'agilitzar el recompte automàtic i evitar errors humans (com s'ha explicat al subapartat 5.3.2.4).

El darrer objectiu específic d'aquest treball ha consistit en l'aplicació de tècniques estadístiques sobre les dades obtingudes amb la finalitat de provar o refutar les tres hipòtesis de les quals parteix l'estudi, que es relacionen amb l'existència de diferències estadísticament significatives en l'ús de les variables lingüístiques analitzades per part dels informants segons la seva edat (hipòtesi 1), sexe (hipòtesi 2) i llengua inicial (hipòtesi 3). El procediment d'anàlisi estadística aplicat, això no obstant, proporciona més informació sobre les relacions entre les variables dependents i independents. A banda de l'estudi estadístic exploratori inicial, que ha permès observar valors utilitzats habitualment en avaluacions quantitatives d'un conjunt de dades com la mitjana o la mediana, ens hem servit de l'Anàlisi de Components Principals per explorar l'existència i la intensitat de relacions entre cadascuna de les variables lingüístiques i les tres variables independents. Per corroborar el pes estadístic de les relacions observades mitjançant aquesta tècnica, però, hem aplicat al conjunt de dades quantitatives tres proves de significació estadística no paramètriques. Per comprovar la

significació de les relacions entre l'edat i les variables dependents, s'ha utilitzat la prova dels rangs amb signe de Wilcoxon, tenint en compte la dependència de les dades derivada del fet que provenen de mesures dels mateixos informants fetes en dos temps diferents. Per comprovar la significació de les relacions entre el sexe dels informants i les variables dependents s'ha utilitzat la prova U de Mann-Whitney i, finalment, per comprovar la significació de les relacions entre la llengua inicial i les variables lingüístiques, s'ha aplicat el test de Kruskal-Wallis.

Quant a les hipòtesis de què partia l'estudi, en general podem afirmar que els resultats les confirmen totes tres, si bé cal precisar aquesta afirmació quant a les relacions estadísticament significatives que, segons la bibliografia prèvia, es podia esperar obtenir amb cadascuna de les variables lingüístiques i les que hem identificat mitjançant les proves de significació que acabem de descriure. Sobre els resultats específics en relació amb les subhipòtesis de les quals partia l'estudi de cadascuna de les variables lingüístiques, recordem que es recullen a les taules 130, 138 i 143 incloses als subapartats de síntesi del setè capítol.

Pel que fa a la primera hipòtesi, de totes les variables dependents de l'estudi, s'esperava que nou mostressin diferències estadísticament significatives segons l'edat. D'aquestes nou correlacions esperables, l'estudi realitzat n'ha corroborat l'existència de cinc i, a més, s'ha trobat que set dels vuit conjunts de partícules discursives analitzats com a V3 i la variable V22 Perífrasis d'obligació també mostren diferències significatives segons l'edat dels informants. En total, doncs, si comptabilitzem cadascun dels dobles de partícules discursives que ha mostrat una correlació significativa amb l'edat com una variable "de ple dret", tenim que tretze variables lingüístiques es correlacionen de forma estadísticament significativa amb l'edat dels informants i, per tant, els resultats obtinguts poden deure's, segons la sociolingüística de la variació, a la gradació per edats.

En canvi, les quatre variables que esperàvem que mostressin diferències per edat i per a les quals hem obtingut valors p superiors al nivell de significació podrien o bé reflectir habilitats lingüístiques que els informants arriben a dominar abans de les edats analitzades (per tant, abans dels 12 anys) o bé aquelles que, tot i

que encara no s'han adquirit fins a nivells adults de competència lingüística, no mostren diferències prou acusades entre les edats analitzades com perquè siguin estadísticament significatives. El primer cas podria ser el de la variable V5 Obertura de la interacció (recordem que només en una transcripció de tot el corpus no es realitza cap variant), mentre que el segon podria ser el de la variable V38 Fenòmens de contacte de llengües. Ara bé, també cal considerar la possibilitat que simplement les variables analitzades no estiguin relacionades amb l'edat dels parlants (per exemple, sembla que l'ús d'una o altra variant de V41 Colors podria correspondre's a l'estil idiolectal dels informants) o que el plantejament de les variables no sigui l'adequat per trobar les relacions que mantenen amb l'edat (estudis futurs podrien, en aquest sentit, explorar amb més deteniment l'ús dels fenòmens que nosaltres hem agrupat en la variant V38a Importacions o V38d Altres, per exemple).

En relació amb la segona hipòtesi, esperàvem identificar diferències estadísticament significatives entre els nois i les noies en set variables lingüístiques. Ara bé, els resultats mostren que existeixen en el cas de dues d'aquestes variables, V3 Doblets de partícules discursives (concretament, V3b *a ver / a veure* i V3f *después / després*) i V17 Tipus de preguntes, i de tres de les altres variables: V6 Heterorepeticions, V18 Construccions condicionals i V22 Perífrasis d'obligació.

Els motius pels quals els resultats no han mostrat les correlacions que s'esperaven a partir de la bibliografia per a la variable sexe són diversos. Primer, cal considerar que la major part dels estudis sobre les diferències entre els estils interaccionals entre homes i dones es van dur a terme fa més de vint anys, en comunitats de parla diferents a la nostra i utilitzant com a informants parlants adults, no adolescents. Totes aquestes diferències amb el nostre corpus podrien explicar les diferències en els resultats dels estudis previs i del nostre. Cal tenir en compte que les societats i els seus usos lingüístics canvien constantment i que els adolescents es distancien d'altres grups etaris mitjançant, entre altres recursos al seu abast, el seu comportament lingüístic (Tuson, 1998).

Pel que fa a la tercera hipòtesi, de les nou variables dependents que suggeríem *a priori*, a partir de la revisió bibliogràfica, que podien mostrar diferències significatives entre els grups de llengua inicial, vuit en mostren, segons els resultats de la prova de significació. A més, cal afegir-hi la variable V18 Construccions condicionals, que també en presenta. Així doncs, si considerem el conjunt de doblats de partícules discursives com una sola característica lingüística, la variable llengua inicial és la variable independent que es correlaciona de forma estadísticament significativa amb més trets lingüístics diferents. En aquest sentit, els nostres resultats s'alineen amb les observacions d'autors com Sorolla i Flors-Mas (2020) i Flors-Mas, Sorolla, Pradilla i Vila (2021) sobre el paper destacat que juga la llengua inicial en l'explicació dels usos lingüístics dels parlants al context sociolingüístic català.

A l'hora de considerar les possibles raons per les quals la llengua inicial es mostra o no correlacionada amb les variables lingüístiques cal també recordar que la distribució d'informants al corpus no és igual a tots els grups. Concretament, ens mostrem reticents a extreure conclusions sobre el comportament lingüístic dels parlants que declaren tenir tant el català com el castellà com a llengües inicials perquè només n'hi ha deu al corpus d'anàlisi, a diferència dels altres dos grups, que estan formats per més de 25 informants.

Una de les conclusions metodològiques més clares que es desprèn dels resultats obtinguts és la importància d'aplicar, en la mesura del possible, les proves de significació estadística d'un estudi a totes les variables lingüístiques que examini, tant a aquelles que la revisió de literatura especialitzada ja ha demostrat que solen mostrar-se correlacionades amb les variables independents com aquelles per a les quals encara no s'han obtingut resultats rellevants. En el nostre treball, aquesta decisió metodològica ha permès identificar sis correlacions que no es preveien abans de l'estudi.

Cal destacar igualment altres aspectes metodològics d'aquest treball que afegeixen valor a l'estudi. Com hem explicat, els estudis panel com el que es presenta aquí són escassos a la bibliografia sobre la variació i, encara més específicament, en la que tracta la llengua catalana. A més, tampoc és habitual

trobar estudis centrats en parlants preadolescents i adolescents ni que tinguin un corpus de més de 65 informants. Així mateix, la varietat de tècniques estadístiques utilitzades també constitueix un pas endavant en la recerca sobre la variació, ja que no es limita a l'estadística descriptiva, com els primers estudis correlacionals, ni a les tècniques de significació estadística, com els que podríem encabir en una segona onada metodològica, sinó que l'Anàlisi de Components Principals ens permet aprofundir en les relacions que indiquen les proves de significació i, mitjançant els arbres de classificació, podem identificar quines variables, d'entre totes les dependents, caracteritzen de manera més efectiva els informants de les edats, els sexes i les llengües inicials que investiguem.

Aquest estudi de variables pragmàtiques, morfosintàctiques i lèxiques utilitzades per 69 informants (pre)adolescents de tres grups de llengua inicial en dos moments temporals (a les edats de 12-13 anys i 15-16 anys) contribueix de manera directa a la fonamentació empírica i la precisió de les tasques de perfilació sociolingüística de mostres dubitades en català que puguin haver estat produïdes per joves menors d'edat. D'aquesta manera, representa un exemple de com la recerca sobre la variació lingüística present en una mostra determinada pot nodrir una àrea concreta de la lingüística aplicada, en el nostre cas, una de les tasques que formen la lingüística forense, la perfilació sociolingüística. A més, posa de manifest que l'estudi de la variació, no només fa valioses contribucions a la descripció de les llengües (com el fet mateix de demostrar que la variació lingüística és sistemàtica i no caòtica), sinó que també pot fer aportacions destacades a diversos àmbits, com l'estudi del desenvolupament i l'adquisició lingüístics, la sociolingüística educativa i la política lingüística o la lingüística forense. Aquesta última és l'àrea que, com hem pogut comprovar amb la revisió bibliogràfica, necessita més impuls acadèmic en llengua catalana. Mostra d'això és el fet que aquest estudi sigui, fins on arriba el nostre coneixement, el primer a enfocar-se cap a la perfilació lingüística no només pel que fa a parlants de català menors d'edat, sinó també, en efecte, de qualsevol edat. En altres paraules, comptem, actualment, amb molt pocs estudis empírics sobre la llengua catalana

dedicats a la millora de pràctiques en lingüística forense, a diferència d'altres llengües i contextos acadèmics.

Així doncs, les aportacions principals d'aquesta tesi a la lingüística catalana són dues. D'una banda, l'anàlisi quantitativa realitzada sobre l'ús de 21 variables lingüístiques en una mostra de mida considerable (69 informants) de parla (pre)adolescent, ja que representa una clara contribució a la descripció de les varietats catalanes emprades per usuaris de la llengua que es troben en l'últim tram de l'etapa escolar (entre 12 i 16 anys). D'una altra, l'anàlisi de les relacions entre aquestes variables lingüístiques i les tres variables independents, ja que, com s'ha explicat, a més de contribuir a la comprensió global de la interacció entre la llengua i la societat en un context demolingüístic complex com el català, permeten identificar l'edat, el sexe i la llengua inicial més probables dels autors de mostres lingüístiques comparables al nostre corpus pel que fa a la varietat i la modalitat de llengua. Aquesta darrera aportació obre una possible línia d'investigacions futures centrades en l'impuls de la lingüística forense en català a partir de l'aplicació de la metodologia pròpia dels estudis sobre la variació lingüística, que compten amb una àmplia tradició en els territoris de parla catalana. Els fruits d'aquesta línia d'investigació podrien ser de tots colors.

Una de les limitacions més evidents d'aquesta tesi és el nombre de variables independents que hem pogut incloure a l'estudi, de manera que una de les possibles recerques futures és precisament la de realitzar un estudi semblant que consideri més variables sociolingüístiques, com la localitat o àrea geogràfica (recordem que hem trobat indicis que algunes de les variables que apareixen al corpus es veuen influenciades per la població dels informants, Manlleu o Mataró). També seria interessant tenir estudis sobre altres dialectes, ja que les dues localitats de les quals prové el nostre corpus pertanyen al català central. Com hem vist, altres variables habituals a la sociolingüística de la variació són la classe social o l'origen ètnic dels parlants (alguns exemples de l'aplicació d'aquests factors a l'estudi de la variació lingüística al context català i espanyol són Marzo i Turell, 2007 o Ballart Macabich, 2013).

La següent limitació de la nostra investigació que es pot convertir en una oportunitat d'ampliar la recerca sobre el català aplicable a la lingüística forense és el fet que el corpus utilitzat es va recollir entre el 2006 i el 2012, és a dir, fa més de deu anys. Per continuar incrementant el volum de dades sobre la llengua catalana dels adolescents, seria de gran interès dur a terme un estudi comparable al que presentem aquí amb una mostra més recent i, fins i tot (si fos possible), amb una mostra encara més antiga. D'aquesta manera, es podria observar si els resultats obtinguts aquí caracteritzen tan sols els adolescents catalans nascuts a final dels anys 1990 o també els d'altres generacions.

Una altra qüestió relativa a la mostra que hem analitzat és el fet que hi ha un desequilibri pel que fa al nombre d'integrants del grup de bilingües inicials, com ja hem indicat. Seria necessari que estudis futurs comptessin amb un nombre més semblant d'informants de cada grup de llengua inicial. Pel que fa al contingut del corpus, caldria que nous treballs els objectius dels quals estiguessin relacionats, com el nostre, amb la lingüística forense poguessin analitzar mostres més comparables, quant al tema tractat pels informants, als assumptes que poden trobar-se en textos rellevants per a processos judicials i investigacions. Per exemple, en el cas de voler contribuir a l'anàlisi de mostres provinents de les xarxes socials relacionades amb casos de (ciber)assetjament de menors, l'ideal seria comptar amb mostres en què els informants valoren negativament algú.

En definitiva, però, com hem vist al llarg de la discussió dels resultats, en català encara no disposem d'un nombre gaire elevat de treballs sobre la distribució sociolingüística actual de les variables lingüístiques que hem analitzat (o d'altres). La recerca de la variació lingüística contribueix, com hem observat, al coneixement general sobre les llengües i pot servir, per tant, per a finalitats molt diverses. En aquesta tesi, ens hem centrat en la llengua dels (pre)adolescents amb la finalitat de contribuir a la perfilació sociolingüística en contextos forenses. Part dels actes delictius amb els quals poden veure's involucrats els menors d'edat i que, per tant, poden fer que es trobin immersos en processos investigatius o judicials tenen lloc a internet. En aquest sentit, cal tenir en compte que la facilitat d'accés a aquest mitjà i la universalització dels dispositius

mòbils permeten que els adolescents hi passin molt de temps (el 31,6 % dels menors d'entre 11 i 18 anys passa més de 5 hores al dia connectat a internet, segons Andrade, Guadix, Rial i Suárez 2021: 25). Així, per exemple, Andrade, Guadix, Rial i Suárez (2021: 73) estimen que un 22,5 % dels menors que van enquestar podrien estar patint ciberassetjament i que un 19,7 % podrien estar actuant com a ciberassetjadors. L'assetjament no és l'única activitat delictiva que pot realitzar-se majoritàriament a través de la llengua, sinó que també hem de tenir en compte delictes com el frau o la captació de menors, per nombrar-ne d'altres. Si bé els riscos associats a l'ús d'internet no són els únics que poden fer que els adolescents arribin als àmbits judicials, on poden jugar un paper destacat les proves lingüístiques, aquestes dades són un indicador de la vulnerabilitat dels joves a internet, que s'ha de sumar a la que poden patir al món no virtual. Els usos lingüístics dels adolescents, com hem observat, poden resultar clau en contextos forenses, de manera que cal continuar investigant-los.

Finalment, subratllem la necessitat d'adoptar una perspectiva holística en la recerca lingüística. En aquest treball, ens hem aproximat a les dades des d'una metodologia quantitativa, però, com s'ha indicat al llarg del treball, caldria que investigacions futures ampliessin els nostres resultats mitjançant anàlisis qualitatives. A més, però, també és necessari que nous treballs tinguin en compte altres eixos de la variació lingüística dels quals no hem pogut donar compte aquí. Parafrasejant Lorente Casafont (1998: 87-88):

“Hi ha una relació directa entre qualsevol canvi [lingüístic] i causes de caràcter social o funcional; i quan un parlant selecciona una variant lèxica ho fa seguint motivacions contextuals (variació funcional), dialectals (variació geogràfica) o personals (variació idiolectal o estilística). Si els models teòrics han d'explicitar els criteris que condicionen les opcions dels parlants, sembla plausible que les motivacions externes han de poder ser-hi codificades”.

En el cas del llenguatge adolescent, a més de factors sociolingüístics com els que hem tractat aquí, cal que tinguem presents que el complex procés d'adquisició lingüística que presentàvem de forma simplificada a la introducció encara es troba en marxa (o, considerant que no acaba mai, més efervescent que a

l'adultesa). En definitiva, l'aproximació variacionista posa a l'abast del lingüista un marc metodològic que permet contribuir “a l'estudi de la llengua des de la perspectiva polinivellada” (Pradilla 2011: 137), ja que permet considerar la interacció d'un gran nombre de factors que condicionen l'actuació lingüística dels parlants. D'aquesta manera, si entenem la sociolingüística variacionista com un calidoscopi (Pradilla, 2011), per comprendre les complexitats de l'univers de la llengua en ús necessitem treballs que aconseguixin combinar de manera satisfactòria aquesta eina amb el microscopi de la teoria lingüística.

9 BIBLIOGRAFIA

- Ag Youssouf, I., Grimshaw, A. D. i Bird, C. S. (1976). greetings in the desert. *American ethnologist*, vol. 3 (4), p. 797-824. <https://doi.org/10.1525/ae.1976.3.4.02a00140>
- Aijón Oliva, M. Á. i Borrego Nieto, J. (2013). La variación gramatical como forma y significado. El uso de los clíticos verbales en el español peninsular. *Lingüística*, vol. 29 (2), p. 93-126.
- Ainsworth, J. (2010). Curtailing coercion in police interrogation: the failed promise of Miranda v. Arizona. A: Coulthard, M. i Johnson, A. (Eds.). *The Routledge Handbook of Forensic Linguistics*. Londres/Nova York: Routledge, p. 111-125.
- Albaladejo, M. (1998). La dixi d'espai: entre el discurs planificat i l'espontani. A: Payrató, L. (ed.). *Oralment. Estudis de variació funcional*. Barcelona: Publicacions de l'Abadia de Montserrat, p. 145-165.
- Alcover, A. M. i Moll, F. B. (1926-1968). *Diccionari Català-Valencià-Balear* [DCVB]. Recuperat a través de: <https://dcvb.iec.cat/>
- Alea Riera, V.; Guillén Estany, M.; Muñoz Vaquer, C.; Maqueda de Anta, I.; Torrelles Puig, E., i Viladomiu Canela, N. (2005). Pruebas para K muestras independientes. *Uso del SPSS*. Recuperat a través de: http://www.ub.edu/aplica_infor/spss/cap6-4.htm
- Al-Harashseh, A. M. i Obeidat, M. M. (2017). A Sociolinguistics Study of Colors Used in Colloquial Jordanian Arabic. *Dirsat for Social Sciences and Humanities*, vol. 44 (4), p. 241-250.
- Al-Jehani, N. M. (1990). Color Terms in Mecca: A Sociolinguistic Perspective. *Anthropological Linguistics*, vol. 32 (1-2), p. 163-174. <https://www.jstor.org/stable/30028145>
- Alonso, H. i Suïls, J. (1993). La morfologia verbal de subjuntiu al Segrià: estudi prospectiu. *Sintagma*, vol. 5, p. 5-17.
- Alter, S. G. (2005). *William Dwight Whitney and the Science of Language*. Baltimore/Londres: The Johns Hopkins University Press.

- Alturo, N. (1995). La variació d'haver auxiliar al català nord-occidental. A: Turell, M. T. (ed.). *La sociolingüística de la variació*. Barcelona: PPU, p. 221-255.
- Alturo, N. i Turell, M. T. (1990). Linguistic change in el Pont de Suert: The study of variation of /ʒ/. *Language Variation and Change*, vol. 2 (1), p. 19-30. <https://doi.org/10.1017/S0954394500000247>
- Alturo, N. i Turell, M. T. (1991). Per una sociolingüística de la variació: la lingüística correlacional i la llengua catalana. *Treballs de Sociolingüística Catalana*, vol. 9, p. 128-131.
- Amat Rodrigo, J. (2016). Prueba de los rangos con signo de Wilcoxon. *Ciencia de Datos*. Recuperat a través de: [https://www.cienciadedatos.net/documentos/18 prueba de los rangos c on signo de wilcoxon](https://www.cienciadedatos.net/documentos/18_prueba_de_los_rangos_c_on_signo_de_wilcoxon)
- Amat Rodrigo, J. (2017). Mann Whitney Wilcoxon. Test de Wilcoxon Mann Whitney como alternativa al t-test. *Ciencia de Datos*. Recuperat a través de: [https://www.cienciadedatos.net/documentos/17 mann%E2%80%93whitney u test#Mann Whitney Wilcoxon](https://www.cienciadedatos.net/documentos/17_mann%E2%80%93whitney_u_test#Mann_Whitney_Wilcoxon)
- Andersen, G. (2001). *Pragmatic markers and sociolinguistic variation. A relevance-theoretic approach to the language of adolescents*. Amsterdam/Filadèlfia: John Benjamins Publishing Company.
- Andrade, B.; Guadix, I.; Rial, A., i Suárez, F. (2021). *Impacto de la tecnología en la adolescencia. Relaciones, riesgos y oportunidades*. Madrid: UNICEF España.
- Angouri, J. (2010). Quantitative, Qualitative or Both? Combinng Methods in Linguistic Research. A: Litosseliti, L. (ed.). *Research Methods in Linguistics*. London/New York: Continuum, p. 29-45.
- Anònim. (1979). The Origin of 'Sociolinguistics'. *Language in Society*, vol. 8 (1), p. 141-141. <https://doi.org/10.1017/S0047404500006102>
- Aparici Aznar, M. i Noguera Llopart, E. (2012). *Desenvolupament i avaluació del llenguatge oral*. Barcelona: UOC.

- Areiza Londoño, R., Cisneros Estupiñán, M. i Tabares Idárraga, L. E. (2012). *Sociolingüística. Enfoques pragmático y variacionista*. Bogotá: ECOE Ediciones.
- Arenas Beneyto, L. (2005). Les salutacions com a construccions idiomàtiques. *Estudios Catalanes. Revista internacional de lengua, literatura y cultura catalanas*, vol. 3 (3), p. 79-89.
- Arenas Beneyto, L. i Cuenca, M. J. (2007). Salutacions i variació textual. *Quaderns de Filologia. Estudis Lingüístics*. vol. XII, p. 253-269.
- Argenter, J. A. (2020). Languages in Contact: A Sociocultural Approach. A: Argenter, J. A. i Lüdtke, J. (eds.). *Manual of Catalan Linguistics*. Berlín/Boston: De Gruyter, p. 597-628.
- Arias Flores, E. (2012). “En aquella época el pobre era pobre”: La repetición, figura poética del habla oral. *Lengua y Habla*, vol. 16, p. 16-31.
- Arjona Iglesias, M. (1991). Los adverbios terminados en *-mente*. A: Arjona, Iglesias, M. *Estudios sintácticos sobre el habla popular mexicana*. México, D. F.: UNAM, p. 25-63.
- Aschenberg, H. i Loureda Lamas, Ó. (2011). Introducción. Marcadores del discurso: descripción, definición, contraste. A: Aschenberg, H i Loureda Lamas, Ó. (ed.). *Marcadores del discurso: de la descripción a la definición*. Madrid/Frankfurt: Iberoamericana/Vervuert, p. 9-34.
- Ash, S. (2013). Social class. A: Chambers, J. K. i Schilling, N. (eds.). *The Handbook of Language Variation and Change*. Malden/Oxford/Chichester: Wiley-Blackwell, p. 350-367.
- Austin, J. (1962). *How to do things with words*. Oxford: Oxford University Press.
- Ávila-Muñoz, A. M. (1998). Aproximación a la estructura de las secuencias de apertura y cierre en las conversaciones telefónicas en español. *ELUA. Estudios de Lingüística*, vol. 12, p. 45-68. <https://doi.org/10.14198/ELUA1998.12.03>
- Bach Martorell, C. (2001). Coherencia tipológica en los conectores reformulativos del catalán. A: de Bustos Tovar, et al. (Eds.). *Lengua, discurso, texto: (I Simposio Internacional de Análisis del Discurso)*, vol. 1, p. 523-538. Madrid: Visor Libros.

- Bach Martorell, C. (2002). Els Connectors Reformulatius Catalans Anàlisi i Proposta d'Aplicació Lexicogràfica. Tesi doctoral inèdita. Universitat Pompeu Fabra.
- Badger; J. R. i Mellanby, J. (2018). Producing and understanding conditionals: When does it happen and why does it matter? *Journal of Child Language Acquisition and Development*, vol. 6 (1), p. 21-41.
- Badia i Margarit, A. M. (1994). *Gramàtica de la llengua catalana*. Barcelona: Enciclopèdia Catalana.
- Bagés Fortacín, N. (2012). Vitalitat i ús del lèxic del parlar de Flix. Variació lingüística generacional. *Miscel·lània del CERE*, vol. 23, p. 411-432.
- Bailey, C. J. N. (1973). *Variation and Linguistic Theory*. Arlington: Center for Applied Linguistics.
- Baker, S. C., Watson, B. M., i Gallois, C. (2020). Social psychology and language. A: Braddick, O. (Ed.). *Oxford Research Encyclopedia of Psychology*. Oxford: Oxford University Press.
- Baldaquí Escandell, J. M. (2002). El model lèxic dels estudiants dels programes educatius bilingües. *Catalan Review*, vol. 16 (1-2), p. 29-49.
- Ball, M. J. (2010). (Ed.). *The Routledge Handbook of Sociolinguistics Around the World*. Londres/Nova York: Routledge.
- Ball, M. J. (2010). Introduction. A: (Ed.). *The Routledge Handbook of Sociolinguistics Around the World*. Londres/Nova York: Routledge, p. 1-3.
- Ballart Macabich, J. (2013). Variació fònica al català de Barcelona: l'accent xava. *Treballs de Sociolingüística Catalana*, vol. 23, p. 133-151. <https://doi.org/10.2436/20.2504.01.55>
- Baran, M. (2017). Fórmulas de saludo y de despedida como relacionemas. Apuntes desde la diacronía y la sincronía de la lengua. *Studia Iberystyczne* vol. 16, p. 7-20. <https://doi.org/10.12797/SI.16.2017.16.01>
- Barcelona School of Management [UPF-BSM]. (2008). Els perits judicials lingüistes podran obtenir una titulació universitària oficial. *Notícies*. Recuperat a través de: <https://www.bsm.upf.edu/ca/noticies/els-perits-judicials-linguistes-podran-obtenir-una-titulacio-universitaria-oficial>

- Baron, N. S. i Ling, R. (2011). Necessary smileys & useless periods. Redefining punctuation in electronically-mediated communication. *Visible Language*, vol. 45 (1-2), p 45-67.
- Barriga Villanueva, R. (1994). Algunas notas sobre el uso de los adverbios en mente en el habla infantil. *Nueva Revista de Filología Hispánica*, vol. 42 (2), p. 563-572. <https://www.jstor.org/stable/40299616>
- Bastardas i Boada, A. (2012). El català i els joves: propostes de política lingüística del Consell Social de la Llengua Catalana. *Treballs de Sociolingüística Catalana*, vol. 22, p. 77-92. <https://doi.org/10.2436/20.2504.01.43>
- Bastida Serra, C. (2017). Emocions i comunicació electrònica textual en català. Un estudi multimetodològic dels recursos emprats per adolescents i adults joves. Tesis doctoral inèdita. Universitat d'Andorra.
- Bazzanella, C. (1992). Aspetti pragmatici della ripetizione dialogica. A: Gobber, G. (Ed.). *Linguistica pragmatica*. Roma: Bulzoni, p. 433-454.
- Bazzanella, C. (2011). Redundancy, repetition, and intensity in discourse. *Language Sciences*, vol. 33, p. 243-254. <https://doi.org/10.1016/j.langsci.2010.10.002>
- Berenguer, J. A. (1995). Marcadores discursivos y relato conversacional. *Caplletra*, vol. 18, p. 109-120.
- Berlanga Silvente, V., Rubio Hurtado, M. J. i Vilà Baños, R. (2013). Com aplicar arbres de decisió en SPSS. *REIRE Revista d'Innovació i Recerca en Educació*, vol. 6 (1), p. 65-79. <https://doi.org/10.1344/reire2013.6.1615>
- Berlín, B. i Kay, P. (1991). *Basic color terms: their universality and evolution*. Berkeley: University of California Press.
- Bernal Linnarsand, M. (2005). La alo-repetición como estrategia de cortesía y descortesía en la conversación conflictiva. A: Murillo Medrano, J. (Ed.). *II Coloquio Internacional del Programa EDICE*. Estocolm i Costa Rica: Programa EDICE, p. 55-79.
- Bernal, E. i Sinner, C. (2009). *Al seu rotllo: aproximació al llenguatge juvenil català*. *Zeitschrift für Katalanistik*, vol. 22, p. 7-36.

- Bernstein, B. (1971). *Theoretical Studies towards a Sociology of Language*. Vol. 1, Class, Codes and Control. London: Routledge and Kegan Paul.
- Bibiri, A. D., Mocanu, M., y Turculeț, A. (2016). The Intonation of Romanian Greetings: A Sociolinguistics Approach. *International Journal of Cognitive and Language Sciences*, vol. 10 (4), p. 1123-1131.
- Biggam, C. P. (2012). *The Semantics of Colour. A Historical Approach*. Cambridge: Cambridge University Press.
- Bigi, B., Bertrand, R. i Guardiola, M. (2014). Automatic detection of other-repetition occurrences: application to French conversational speech. A: Calzolari, N. et al. (Eds.). *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, p. 836-842. Reykjavík: European Language Resources Association.
- Blackwell, S. (2013). History of forensic linguistics. A: Chappelle, C. A. (Ed.). *The Encyclopedia of Applied Linguistics*. Oxford: Blackwell Publishing, p. 1-7.
- Bladas, Ò. (2012). Conversational routines, formulaic language and subjectification. *Journal of Pragmatics*, vol. 44, p. 929-957. <https://doi.org/10.1016/j.pragma.2012.04.009>
- Blanco Canales, A. (2000). Redes sociales y variación lingüística. *REIS: Revista Española de Investigaciones Sociológicas*, vol. 91, p. 115-136.
- Blas Arroyo, J. L. (1994). *Sociolingüística del español. Desarrollos y perspectivas en el estudio de la lengua española en contexto social*. Madrid: Cátedra.
- Blas Arroyo, J. L. (1998). Un caso de variación pragmática sobre la ampliación significativa de un marcador discursivo en el español actual. Aspectos estructurales y sociolingüísticos. *Analecta Malacitana*, vol. 21 (2), p. 543-571.
- Blas Arroyo, J. L. (1999). “Diga por qué, diga por qué...”. La repetición en el debate político-electoral. *Revista de Investigación Lingüística*, núm. 1 (2), p. 5-42.
- Bloomfield, L. (1933/1984). *Language*. Chicago/Londres: The University of Chicago Press.

- Bobadilla, F. i Bobadilla, G. (1980-1981). El estudio de tres variables sociolingüísticas en Rancagua: problemas preliminares. *Boletín de Filología*, vol. 31 (2), p. 721-741.
- Boeckx, C. i Martín, T. (2013). *El clíctic datiu és més que un clíctic*. Lleida: Pagès editors.
- Boix i Fuster, E. i Vila i Moreno, V. (1998). *Sociolingüística de la llengua catalana*. Barcelona: Ariel.
- Boix-Fuster, E. i Cicres, J. (2013). Article *in memoriam* sobre el llegat de M. Teresa Turell. *Treballs de Sociolingüística Catalana*, vol. 23, p. 13-18. <https://doi.org/10.2436/20.2504.01.48>
- Bonet i Alsina, E. (1992). Sobre un dels usos de *hi* en lloc de *li*. *Els Marges*, vol. 46, p. 101-110.
- Bonet, E. (2002). Cliticització. A: Solà, J.; Lloret, M. R.; Mascaró, J., i Pérez Saldanya, M. (Dir.). *Gramàtica del català contemporani*, vol. 1. Barcelona: Empúries, p. 933-989.
- Borja i Sanz, J. (2009). Fraseologia i cohesió textual: els marcadors discursius en català. A: Faluba K. (Ed.). *Actes del Catorzè Col·loqui Internacional de Llengua i Literatura Catalanes*, vol. 2. Barcelona: Publicacions de l'Abadia de Montserrat, p. 111-124.
- Borràs-Comes, J. (2013). Català nord-occidental. A: Cabré, T. i Prieto, P. (Eds.). *L'entonació dels dialectes Catalans*. Barcelona: Publicacions de l'Abadia de Montserrat, p. 101-126.
- Borreguero Zuloaga, M. (2017). Los relatos coloquiales: partículas discursivas y polifonía. *Pragmalingüística*, vol. 25, p. 62-88.
- Bozzo Duran, M. (1986). Els noms dels colors al Diccionari General de la Llengua Catalana. *Llengua i Literatura*, 1, p. 237-251.
- Brasseur, J. i Jimenez, B. C. (1989). Performance of university students on the Fullerton subtest of idioms. *Journal of Communication Disorders*, vol. 22, 351-359. [https://doi.org/10.1016/0021-9924\(89\)90010-5](https://doi.org/10.1016/0021-9924(89)90010-5)
- Bretxa i Riera, V. (2014). El salt a secundària. Els canvis en les tries lingüístiques i culturals dels preadolescents mataronins en la transició educativa. Tesis doctoral inèdita. Universitat de Barcelona.

- Bretxa, V. i Vila, F. X. (2014). L'evolució dels usos lingüístics dins l'aula des de sisè de primària fins a quart d'ESO. *Revista de Llengua i Dret*, vol. 62, p. 106-123. <https://doi.org/10.2436/20.8030.02.81>
- Bretxa, V.; Comajoan, L.; Ubalde, J., i Vila, F. X. (2016). Changes in the linguistic confidence of primary and secondary students in Catalonia: a longitudinal study. *Language, Culture and Curriculum*, vol. 29 (1), p. 56-72. <https://doi.org/10.1080/07908318.2016.1132657>
- Breu, Marta. (2013). *Vocabulari terminològic LGBT (lèsbic, gai, bisexual i transgènere)*. Barcelona: TERMCAT, Centre de Terminologia. [http://www.termcat.cat/ca/Diccionaris En Linia/164/](http://www.termcat.cat/ca/Diccionaris%20En%20Linia/164/)
- Bright, W. (1966). *Sociolinguistics: proceedings of the UCLC Sociolinguistics Conference, 1964*. The Hague: Mouton.
- Bright, W. (1997). Social factors in language change. A: Coulmas, F. (Ed.). *The Handbook of Sociolinguistics*. Malden/Oxford/Victoria: Blackwell Publishing, p. 81-91.
- Britain, D. (1991). *Dialect and Space: A Geolinguistic Analysis of Speech Variables in the Fens*. Tesi doctoral inèdita. University of Essex.
- Briz Gómez, A. (2008). Diccionario de partículas discursivas del español. A: Azorín Fernández, D. et al. (Coords.). *Actas del II Congreso Internacional de Lexicografía Hispánica. El diccionario como puente entre las lenguas y culturas del mundo*. Biblioteca Virtual Miguel de Cervantes, p 217-227.
- Briz, A., Pons, S. i Portolés, J. (Coords.). (2008). *Diccionario de partículas discursivas del español*. www.dpde.es
- Briz, Antonio. (2001). El uso de *o sea* en la conversación. A: De Kock, J. (Ed.). *Lingüística con Corpus. Catorce aplicaciones sobre el español*. Salamanca: Ediciones de la Universidad de Salamanca, p. 287-318.
- Briz, Gómez, A. (2012). La definición de las partículas discursivas *hombre y mujer*. *Anuario de Lingüística Hispánica*, vol. 28, p. 27-55.
- Brown, P. i Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.

- Brucart, J. (2002). Els determinants. A: Solà, J.; Lloret, M. R.; Mascaró, J., i Pérez Saldanya, M. (Dir.). *Gramàtica del català contemporani*, vol. 2. Barcelona: Empúries, p. 1437-1516.
- Brumme, J. i Schmid, B. (2018). Convergències i divergències en l'ús de les partícules discursives. A: Balaş, O. D. i Montoliu Pauli, X. (Eds.). *Actes del XVIIIè Col·loqui de l'AILLC (Bucarest, 2018)*, p. 96-108. <https://doi.org/10.2436/15.8090.02.6>
- Busquet Duran, J.; Medina Cambrón, A., i Sort i Jané, J. (2006). *La recerca en comunicació. Què hem de saber? Quins passos hem de seguir?* Barcelona: Editorial UOC.
- Busquet, N. (2010). El parlar salat: descripció, àmbit geogràfic i ús. Tesi doctoral inèdita. Universitat de Barcelona.
- Cabanellas Torres, J. (2020). Anàlisi del sistema universitari espanyol en relació a la classe social dels estudiants i les seves oportunitats. Treball de fi de grau inèdit. Universitat de les Illes Balears.
- Cabanes Fitor, V. (1996). Les perífrasis modals de necessitat-obligació i probabilitat en català. Seguiment diacrònic. Segles XIII al XX. *Caplletra*, vol. 20, p. 129-164.
- Cabré, T. i Fàbregas, A. (2019). 3rd person clitic combinations across Catalan varieties: Consequences of the nature of the dative clitic. *The Linguistic Review*, vol. 36 (2), p. 151-190. <https://doi.org/10.1515/tlr-2018-2010>
- Calderón Noguera, D. F. i Durán Mendivelso, B. N. (2009). Caracterización sociolingüística de la comunidad de habla de Tunja. *Cuadernos de Lingüística Hispánica*, vol. 14, p. 139-158.
- Calvet, L. J. (2003). Reflections on the Origins of Sociolinguistics in Europe. A: Paulston, C. B. i Tucker, G. R. (Eds.). *Sociolinguistics: The Essential Readings*. Oxford: Blackwell, p. 17-24.
- Cantamutto, L. (2019). Aperturas y cierres en la interacción digital escrita: fórmulas de saludo en español bonaerense. *Tonos Digital*, vol. 37.

- Carbonero Cano, P. i Santana Marrero, J. (2010). Marcadores del discurso, variación dialectal y variación social. A: Loureda Lamas, Ó. i Acín-Villa, E. (Coords.). *Los estudios sobre marcadores del discurso en español, hoy*. Madrid: Arco Libros, p. 497-522.
- Carrera-Sabaté, J. (2006). Some connections between linguistic change and the written language: the behavior of speakers aged 3 to 20. *Language Variation and Change*, vol. 18 (1), p. 15-34.
<https://doi.org/10.1017/S0954394506060017>
- Casado Velarde, M. (1991). Los operadores discursivos *es decir, esto es, o sea, y a saber* en español actual: valores de lengua y funciones textuales. *Revista Lingüística Española Actual (LEA)*, vol. 13 (1), p. 87-116.
- Casanova Herrero, E. (1993). Evolució i interferència en el sistema demostratiu català. Una explicació. A: Ferrando Francés, A., Meseguer, L. i Alemany Ferrer, R. (coords.). *Actes del novè Col·loqui Internacional de Llengua i Literatura Catalanes: Alacant-Elx, 9-14 de setembre de 1991*, vol. 3. Barcelona: Publicacions de l'Abadia de Montserrat, p. 161-195.
- Casanova Seuma, L. (1998). Un estudi tipològic del català col·loquial. *Sintagma: revista de lingüística*, vol. 10, p. 5-25.
- Casesnoves Ferrer, R. (2012). La sociolingüística de la variació. Material docent en línia. Universitat Oberta de Catalunya.
- Casesnoves Ferrer, R. i Gavaldà Ferré, N. (2014). Maria Teresa Turell (1949-2013). *Estudis Romànics*, vol. 36, p. 711-712.
- Casesnoves-Ferrer, R.; Forcadell, M. i Gavaldà Ferré, N. (Eds.). (2013). *Ens queda la paraula: estudis de lingüística aplicada en honor a M. Teresa Turell*. Barcelona: Institut Universitari de Lingüística Aplicada. Universitat Pompeu Fabra.
- CCMA. (s. d.). ésAdir. Corporació Catalana de Mitjans Audiovisuals. Generalitat de Catalunya. Recuperat a través de: <https://esadir.cat/>
- Cestero Mancera, A. M. i Moreno Fernández, F. (2008). Usos y funciones de *vale* y *¡venga!* en el habla de Madrid. *Boletín de Lingüística*, vol. XX (29), p. 65-84.

- Chacón Beltrán, R. (2004). *An Introduction to Sociolinguistics*. Madrid: Universidad Nacional de Educación a Distancia.
- Chambers, J. K. (2003). *Sociolinguistic Theory*. Oxford: Wiley-Blackwell.
- Chambers, J. K. i Schilling, N. (2013). *The Handbook of Language Variation and Change*. Malden/Oxford/Chichester: Wiley-Blackwell.
- Cheshire, J. (1987). Syntactic variation, the linguistic variable and sociolinguistic theory. *Linguistics*, vol. 25 (2), p. 257-282. <https://doi.org/10.1515/ling.1987.25.2.257>
- Cheshire, J. (2004). Sex and Gender in Variationist Research. A: Chambers i Schilling, N. (Eds.). *The Handbook of Language Variation and Change*. Malden/Oxford/Chichester: Wiley-Blackwell, p. 423-443. <https://doi.org/10.1002/9780470756591.ch17>
- Chevrot, J. P. i Foulkes, P. (2013). Introduction: Language acquisition and sociolinguistic variation. *Linguistics*, vol. 51 (2), p. 251-254. <https://doi.org/10.1515/ling-2013-0010>
- Cicres i Bosch, J. (2004). Anàlisi espectral de [f 0 s x] en espanyol i la seva utilitat en la identificació de parlants. *Interlingüística*, vol. 15 (1), p. 291-298.
- Cicres i Bosch, J. (2007). Aplicació de l'Anàlisi de l'entonació i de l'alienació tonal a la identificació de parlants en fonètica forense. Tesi doctoral inèdita. Universitat Pompeu Fabra.
- Cicres i Bosch, J. (2011b). Transcripció i autenticació de gravacions en contextos judicials. *Llengua, Societat i Comunicació*, vol. 9, p. 26-32.
- Cicres i Bosch, J. i Turell, M. T. (2004). El análisis multidimensional de la voz como herramienta para la identificación del hablante en fonética forense. A: Cano López, P. (Coord.). *Actas del VI Congreso de Lingüística General: Santiago de Compostela, 3-7 de mayo de 2004*, vol. 1, p. 803-812.
- Cicres, J. (2011a). Los sonidos fricativos sordos y sus implicaciones forenses. *Estudios filológicos*, vol. 48, p. 33-48. <https://doi.org/10.4067/S0071-17132011000200003>

- Cicres, J. i Gavaldà, N. (2014). La lingüística forense: la llengua com a evidència. *Revista de Llengua i Dret*, vol. 61, p. 60-71.
<https://doi.org/10.2436/20.8030.02.54>
- Cicres, J. i Queralt, S. (2019). An n-gram based approach to the automatic classification of schoolchildren's writing. *Vigo International Journal of Applied Linguistics*, vol. 16, p. 53-80.
<https://doi.org/10.35869/vial.v0i16.93>
- Clark, E. V. (2003). Honing conversational skills. A: *First Language Acquisition*. Cambridge: Cambridge University Press.
- Clua Julve, E. (1999). Distància lingüística i classificació de varietats dialectals. *Caplletra*, vol. 26, p. 11-26.
- Coates, J. (2004). *Women, Men and Language*. Harlow: Pearson Longman.
- Cohen, M. (1956). *Pour une sociologie du langage*. Paris: Albin Michel.
- Collí May, J. (2009). Greetings and farewells between women and men at UQROO. A: Marín Marín. (Ed.). *Memoria del V foro de estudios en lenguas*. Chetumal: Universidad de Quintana Roo, p. 67-81.
- Collins, P. (2010). Sociolinguistics in Australia. A: Ball, M. J. (Ed.). *The Routledge Handbook of Sociolinguistics Around the World*. Londres/Nova York: Routledge, p. 151-158.
- Colomina, J. (2002). Paradigmes flectius de les altres classes nominals. A: Solà, J.; Lloret, M. R.; Mascaró, J., i Pérez Saldanya, M. (Dirs.). *Gramàtica del català contemporani*, vol. 1. Barcelona: Empúries, p. 535-582.
- Comajoan, L. (2003). Adquisició de primeres i segones llengües: perspectiva històrica i qüestions actuals. *Caplletra*, vol. 35, p. 15-46.
- Comajoan, L.; Vila, F. X.; Bretxa, V.; Sorolla, N.; Tenorio, X., i Melià, J. (2013). Els usos lingüístics en família i amb amics de l'alumnat autòcton i al·lòcton de sisè de primària a Catalunya, Mallorca i la Franja. A: Vila, F. X. i Salvat, E. (Eds.). *Noves immigracions i llengües*. Barcelona: MRR.
- Cornips, L. i Gregersen, F. (2016). The impact of Labov's contribution to general linguistic theory. *Journal of Sociolinguistics*, vol. 20 (4), p. 498-524.

- Coronel-Molina, S. M. (1998). Openings and closings in telephone conversations between native Spanish speakers. *Working Papers in Educational Linguistics*, vol. 14 (1), p. 49-68.
- Coulmas, F. (Ed.). (1997). *The Handbook of Sociolinguistics*. Malden/Oxford/Victoria: Blackwell.
- Coulthard, M. i Johnson, A. (2007). *An Introduction to Forensic Linguistics: Language in Evidence*. Londres/Nova York: Routledge.
- Coulthard, M., May, A., i Sousa-Silva, R. (Eds.). (2020). *The Routledge handbook of forensic linguistics*. Londres/Nova York: Routledge.
- Coulthard, M.; Grant, T., i Kredens, K. (2011). Forensic Linguistics. A: Wodak, R.; Johnstone, B., i Kerswill, P. (Eds.). *The SAGE Handbook of Sociolinguistics*. Thousand Oaks: SAGE Publications, p. 529-544.
- Cresti, E. i Moneglia, M. (2005). (Eds.). *C-ORAL-ROM: Integrated reference corpora for spoken Romance language*. *Studies in Corpus Linguistics* 15. Amsterdam/Filadèlfia: Benjamins.
- Crystal, D. (1987). *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press.
- CSASE. (2014). *Estudi sociodemogràfic i lingüístic de l'alumnat de 4t d'ESO de Catalunya. 2013*. Barcelona: Consell Superior d'Avaluació del Sistema Educatiu. Departament d'Educació. Generalitat de Catalunya.
- Cuenca, M. J. (2003). Two ways to reformulate: a contrastive analysis of reformulation markers. *Journal of Pragmatics*, vol. 35, p. 1069-1093. [https://doi.org/10.1016/S0378-2166\(03\)00004-3](https://doi.org/10.1016/S0378-2166(03)00004-3)
- Cuenca, M. J. (2010). Díctics espacials i gramàtica en narracions orals. *Estudis Romànics*, vol. 32, p. 101-123. <https://doi.org/10.2436/20.2500.01.49>
- Cuenca, M. J. i Bach, C. (2007). Contrasting the form and use of reformulation markers. *Discourse Studies*, vol. 9 (2), p. 149-175. <https://doi.org/10.1177/1461445607075347>
- Cuenca, M. J. i Marín, M. J. (2000). Verbos de percepció gramaticalizados como conectores. Análisis contrastivo español-catalán. *Revista Española de Lingüística Aplicada*, vol. extra 1, p. 215-238.

- Cuenca, M. J. i Marín, M. J. (2012). Discourse markers and modality in spoken Catalan: The case of (*és*) *clar*. *Journal of Pragmatics*, vol. 44, p. 2211-2225. <https://doi.org/10.1016/j.pragma.2012.09.006>
- Cuenca, M. J. i Torres Vilatarsana, M. J. (2008). Usos de *hombre/home* y *mujer/dona* como marcadores del discurso en la conversación coloquial. *Verba: Anuario galego de filoloxía*, vol. 35, p. 235-256.
- Čurković Kalebić, S. (2016). Repetitions in L2 and L1 Conversations. A: Akbarov, A. (Ed.). *Current Research on Language Learning and Teaching. Case Study of Bosnia and Herzegovina*. Newcastle upon Tyne: Cambridge Scholars Publishing, p 233-244.
- D. A. (2008a). Metodología cualitativa. *Diccionario de términos clave de ELE*. Recuperat a través de: https://cvc.cervantes.es/ensenanza/biblioteca_ele/diccio_ele/diccionario/metodologiacualitativa.htm
- D. A. (2008b). Metodología cuantitativa. *Diccionario de términos clave de ELE*. Recuperat a través de: https://cvc.cervantes.es/ensenanza/biblioteca_ele/diccio_ele/diccionario/metodologiacuantitativa.htm
- Davies, I.; Corbett, G.; Bayo Margalef, J. (1995). Colour terms in Catalan: An investigation of eighty informants, concentrating on the purple and blue regions. *Transactions of the Philological Society*, vol. 93 (1), p. 17-49. <https://doi.org/10.1111/j.1467-968X.1995.tb00435.x>
- De Marlangeon, Silvia Beatriz Kaul. (2002). *Los adverbios en -mente del español de hoy y su función semántica de cuantificación*. Frankfurt am Main/Madrid: Vervuert Verlagsgesellschaft.
- De Rosselló i Peralta, C. (2014). Les alternances de codi: un recurs comunicatiu. *Comunicació Educativa*, vol. 27, p. 27-33.
- De Smet, H. (2018). Unwitting Inventors: Speakers Use -ly- Adverbs More Creatively when Primed. *Zeitschrift für Anglistik und Amerikanistik*, vol. 66 (3), p. 329-340. <https://doi.org/10.1515/zaa-2018-0028>
- Degand, L.; Cornillie, B., i Pietrandrea, P. (2013). Discourse markers and modal particles: two sides of the same coin? A: Degand, L.; Cornillie, B.;

- Pietrandrea, P. (Eds.). *Discourse markers and modal particles. Categorization and Description*. Amsterdam/Filadèlfia: John Benjamins, p. 1-18.
- DGPL. (2015). *Anàlisi de l'Enquesta d'usos lingüístics de la població 2013. Resum dels factors clau*. Barcelona: Direcció General de Política Lingüística. Departament de Cultura. Generalitat de Catalunya.
- Di Vito, N. (2010). Sociolinguistics in France. A: (Ed.). *The Routledge Handbook of Sociolinguistics Around the World*. Londres/Nova York: Routledge, p. 310-326.
- Díaz Monroy, L. G. (2007). *Estadística multivariada: Inferencia y métodos*. Bogotá: Departamento de Estadística. Facultad de Ciencias. Universidad Nacional de Colombia.
- Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics*, vol. 17 (4), p. 463-489. <https://doi.org/10.1515/COG.2006.015>
- Domènech Bagaria, O. i Estopà Bagot, R. (2015). La resemantització com a recurs creatiu: la neologia cromàtica. *Caplletra*, vol. 59, p. 159-183.
- Domènech Bagaria, O. i Gelpí Arroyo, C. (2017). Eines i recursos per a la traducció jurídica cap al català: dels tipus de recursos a les necessitats dels usuaris. A: Jiménez Salcedo, J. i Monzó-Nebot, E. (Eds.). *Les llengües minoritzades en l'ordre postmonolingüe/Minoritized languages under a postmonolingual order/Las lenguas minorizadas en el orden posmonolingüe*. Castelló de la Plana: Universitat Jaume I, p. 221-234.
- Domingo, A.; Vidal-Coso, E., i Serra, J. (2012). Apunts sobre sistemes migratoris i llengua als territoris de parla catalana. *Treballs de la Societat Catalana de Geografia*, vol. 73, p. 145-166. <https://doi.org/10.2436/20.3002.01.8>
- Domínguez Portela, S. (2011). O marcador discursivo vale: usos ou abusos? *Estudos de lingüística galega*, vol. 3, p. 211-226. <https://doi.org/10.3309/1989-578X-11-10>
- Du Plessis, N. M. (2017). A Rule for Naming Objects. *Biological Theory*, vol. 12, p. 39-49. <https://doi.org/10.1007/s13752-016-0256-0>
- Duranti, A. (1997). Universal and Culture-Specific Properties of Greetings. *Journal of Linguistic Anthropology*, vol. 7 (1), p. 63-97.

- Durrell, M. (2004). Sociolect/Soziolekt. A: Wiegand, H. E. (Ed.). *Sociolinguistics: An International Handbook of the Science of Language and Society/Soziolinguistik: Ein internationales Handbuch zur Wissenschaft von Sprache un Gesellschaft*. Volume 1/1. Teilband. Berlín/Nova York: Walter de Gruyter, p. 200-205.
- Eades, D.; Fraser, H.; Siegel, J.; McNamara, T., i Baker, B. (2003). Linguistic identification in the determination of nationality: A preliminary report. *Language Policy*, vol. 2, p. 179-199.
<https://doi.org/10.1023/A:1024640612273>
- Eckert, P. (1997). Gender and sociolinguistic variation. A: Coates, J. (Ed.). *Language and Gender. A Reader*. Oxford: Blackwell.
- Eckert, P. (2000). *Linguistic Variation as Social Practice: The Linguistic Construction Of Identity In Belten High*. Malden: Blackwell.
- Eckert, P. (2005). Variation, convention, and social meaning. Paper presented at the annual meeting of the Linguistic Society of America. Oakland, CA, January 7. Recuperat a través de:
<https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.553.1313&rep=rep1&type=pdf>
- Eckert, P. (2012). Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology*, vol. 41, p. 78-100.
- Eckert, P. i McConnell-Ginet, S. (2003). *Language and Gender*. Cambridge: Cambridge University Press.
- Ellis, E. (2006). Monolingualism: The unmarked case. *Sociolinguistic Studies/Estudios de Sociolingüística*, vol. 7 (2), p. 173-196.
<https://doi.org/10.1558/sols.v7i2.173>
- Enrique-Arias, A. (2018). Factores diatópicos en la variación entre “este” y “aqueste” en la historia del español. A: Arnal Purroy, M. L.; Castañer Martín, R. M.; Enguita Utrilla, J. M.; Lagüéns Gracia, V., i Martín Zorraquino, M. A. (coords.). *Actas del X Congreso Internacional de Historia de la Lengua Española: Zaragoza, 7-11 de septiembre de 2015*, vol. 2. Saragossa: Diputación de Zaragoza, p. 1553-1569.

- Espinosa-Zaragoza, I. (2021). Colour and gender: language nuances. *Feminismo/s*, vol. 38, p. 115-147. <https://doi.org/10.14198/fem.2021.38.05>
- Espuny, J. (1998). Aspectes de la interferència lèxica castellana en el discurs oral català. A: Payrató, L. (ed.). *Oralment. Estudis de variació funcional*. Barcelona: Publicacions de l'Abadia de Montserrat, p. 275-290.
- Ettxebarria Arostegui, M. (2013). La variación lingüística: precisiones en torno a la noción en diversas teorías lingüísticas. *Oihenart, cuadernos de lengua y literatura*, vol. 28, p. 207-239.
- Ettxebarria, J.; Joaristi, L. i Lizasoain, L. (1990). Programación y análisis estadísticos básicos con SPSS/PC+. Madrid: Paraninfo.
- Fabà Prats, A. i Torres-Pla, J. (2020). De la llengua inicial a la llengua d'identificació a Catalunya: qui adopta el català i qui se n'allunya. *Treballs de Sociolingüística Catalana*, vol. 30, p. 213-232. <https://doi.org/10.2436/20.2504.01.168>
- Fasold, R. W. (1990). *The Sociolinguistics of Language*. Oxford: Blackwell.
- Fasold, R. W. (1991). The quiet demise of variable rules. *American Speech*, vol. 66, p. 3-21. <https://doi.org/10.2307/455431>
- Feagin, C. (2013). Entering the community. Fieldwork. A: Chambers, J. K. i Schilling, N. (Eds.). *The Handbook of Language Variation and Change*. Malden/Oxford/Chichester: Wiley-Blackwell, p. 19-37.
- Felíu Arquiola, E. i Pato, E. (2019). ¿Realmentes existen?: la “pluralización” de los adverbios en *-mente* en español actual. *Onomázein, revista de lingüística, filología y traducción*, vol. 44, p. 166-190. <https://doi.org/10.7764/onomazein.44.08>
- Ferguson, C. A. (1959). Diglossia. *Word*, vol. 15 (2), p. 325-340. <https://doi.org/10.1080/00437956.1959.11659702>
- Ferguson, C. A. (1997). History of sociolinguistics. A: Paulston, C. B. i Tucker, G. R. (Eds.). *The early days of sociolinguistics: memories and reflections*. Dallas: The Summer Institute of Linguistics, p. 77-95.

- Fernández Planas, A. M., *et al.* (2011). *Llengua, Societat i Comunicació*, vol. 9. Lingüística forense. Recuperat a través de: <https://revistes.ub.edu/index.php/LSC/issue/view/266>
- Fernández Planas, A. M.; Roseano, P.; García, W. E.; Cerdà Massó, R.; Romera Barrios, L.; Carrera i Sabaté, J.; Szmidt, D.; Labraña, S., i Martínez Celdrán, E. (2015). Cap a un nou mapa dialectal del català? Consideracions a partir de dades prosòdiques tractades dialectomètricament. *Estudios de fonética experimental*, vol. 24, p. 257-86.
- Figuerola, E. (1994). *Sociolinguistic Metatheory*. Oxford/Nova York/Tokyo: Elsevier.
- Fischer, K. (Ed.). (2006). *Approaches to discourse particles*. Amsterdam: Elsevier.
- Fischer, K. i Alm, M. (2013). A radical construction grammar perspective on the modal particle-discourse particle distinction. A: Degand, L.; Cornillie, B., i Pietrandrea, P. (Eds.). *Discourse Markers and Modal Particles. Categorization and Description*. Amsterdam/Filadèlfia: John Benjamins, p. 47-88.
- Fischer, S. (2002). *The Catalan Clitic System. A Diachronic Perspective on its Syntax and Phonology*. Berlín: Mouton de Gruyter.
- Fishman, P. (1983). Interaction: The Work Women Do. A: Thorne, B.; Kramarae, C., i Henley, N. (Eds.). *Language, Gender and Society*. Rowley: Newbury House, p. 89-101.
- Fitó Pardo, J. (2009). Multimodalitat i dixi d'espai en joves catalans multilingües. Tesi doctoral inèdita. Universitat de Barcelona.
- Flors, A. (2015). Young People and Languages in Catalonia: The state of the Question. *Linguapax Review*. A: Cru, J. (Ed.). *The role of youth in language revitalisation/El paper dels joves en la revitalització lingüística*. Linguapax International, p. 30-42.
- Flors-Mas, A. (2017). Usos lingüístics i identitats socials entre adolescents catalans i valencians. Tesi doctoral inèdita. Universitat de Barcelona.
- Flors-Mas, A. (2018). Language use in secondary school classrooms in Catalonia and the Valencia Region: Students and teachers as actors in language-in-education policy. *Revista de Llengua i Dret*, vol. 70, p. 103-116. <https://doi.org/10.2436/rld.i70.2018.3175>

- Flors-Mas, A.; Sorolla, N; Pradilla, M. À., i Vila, F. X. (2021). The recent evolution of first languages in Catalonia. Between minoritization and language maintenance. *Language Problems and Language Planning*, vol. 45 (1), p. 31-55. <https://doi.org/10.1075/lplp.20028.flo>
- Floyd, M. B. (1990). Development of subjunctive mood in children's Spanish: a review. *Confluencia*, vol. 5 (2), p. 93-104. <http://www.jstor.org/stable/27921930>
- Fogarty, K. (2010). "Just say it in your own words": The social interactional nature of investigative interviews into child sexual abuse. Tesi doctoral inèdita. The University of Adelaide.
- Fogarty, K; Augoustinos, M., i Kettler, L. (2013). Re-thinking rapport through the lens of progressivity in investigative interviews into child sexual abuse. *Discourse Studies*, vol. 15 (4), p. 395-420. <https://doi.org/10.1177/1461445613482429>
- Font Rotchés, D. (2008). Els patrons entonatius de les interrogatives absolutes del català central. *Llengua & Literatura*, vol. 19, p. 299-329.
- Forensiclàb. (2013). *Idiolectometria Forense i Índex de Similitud Idiolectal*. Unitat de Variació Lingüística. Institut Universitari de Lingüística Aplicada. Universitat Pompeu Fabra. Recuperat a través de: http://www.iula.upf.edu/rec/forensic_isi/docums/forensic_isi_ca.pdf
- Fought, C. (Ed.). (2004). *Sociolinguistic Variation: Critical Reflections*. Oxford: Oxford University Press.
- Freed, A. F. (1994). The form and function of questions in informal dyadic conversation. *Journal of Pragmatics*, vol. 21, p. 621-644. [https://doi.org/10.1016/0378-2166\(94\)90101-5](https://doi.org/10.1016/0378-2166(94)90101-5)
- Freixa Aymerich, J. (2019). Augment constant de manlleus en la premsa catalana: una conseqüència de la mundialització? *Treballs de Sociolingüística Catalana*, vol. 29, p. 43-64. <https://doi.org/10.2436/20.2504.01.150>
- Freixa, J. (1998). A propòsit de la repetició en el discurs. A: Payrató, L. (ed.). *Oralment. Estudis de variació funcional*. Barcelona: Publicacions de l'Abadia de Montserrat, p. 195-222.

- Freixa, J. (2015). La implantació del lèxic normatiu: estudi complementari del lèxic implantat no normatiu. *Treballs de Sociolingüística Catalana*, vol. 25, p. 65-80. <https://doi.org/10.2436/20.2504.01.91>
- Fuentes Rodríguez, C. (1998). *Pero, ¿cuál es su valor?* *Philologia hispalensis*, vol. 12 (1), p. 123-145.
- Fuentes Rodríguez, C.; Placencia, M. E., i Palma Fahey, M. (2016). Regional pragmatic variation in the use of the discourse marker *pues* in informal talk among university students in Quito (Ecuador), Santiago (Chile) and Seville (Spain). *Journal of pragmatics*, vol. 97, p. 74-92. <https://doi.org/10.1016/j.pragma.2016.03.006>
- Galián Conesa, M. D.; Ato Lozano, E., i Carranza Carnicero, J. A. (2010). Explosión del vocabulario infantil y primer lenguaje: una revisión. *Anales de Psicología*, vol. 26 (2), p. 341-347.
- Galindo Solé, M. (2005). Les Funcions discursives de l'alternança de codis als patis de les escoles de Catalunya. *Treballs de sociolingüística catalana*, vol. 18, p. 249-72.
- Galindo Solé, M. (2006). Les llengües a l'hora del pati. Usos lingüístics en les converses dels infants de primària a Catalunya. Tesi doctoral inèdita. Universitat de Barcelona.
- Garachana Camarero, M. (2012). Discourse Marker. *Diccionari de lingüística on line*. Recuperat a través de: <http://www.ub.edu/diccionarilinguistica/print/309>
- Garachana Camarero, M. (2014). Marcador discursivo. *Diccionari de Lingüística en línia*. Recuperat a través de: <http://www.ub.edu/diccionarilinguistica/node/144>
- Garayzábal Heinze, E.; Queralt Estevez, S., Reigosa Riveiros, M. (2019). *Fundamentos de la lingüística forense*. Madrid: Síntesis.
- Garayzábal, E.; Jiménez Bernal, M., i Reigosa Riveiros, M. (2012). *Lingüística forense la lingüística en el ámbito legal y policial*. Madrid: Euphonia Ediciones.
- Garcés Gómez, M. P. (2002-2004). La repetición: formas y funciones en el discurso oral. *Archivo de filología aragonesa*, vol. 59-60 (1), p. 437-456.

- García Mouton, P. (2003). *Así hablan las mujeres. Curiosidades y tópicos del uso femenino del lenguaje*. Madrid: La esfera de los libros.
- Gargallo Gil, J. E. (2002). Dues menes de frontera (lingüística i administrativa), diversos models de parlars (i de parlants) xurros fronterers. *Caplletra*, vol. 32, p. 127-141.
- Gauthier, C. (2016). ¿Lo tengo que hacer o tengo que hacerlo?: Diferencias semánticas entre la posición proclítica y enclítica del pronombre clítico *lo* en la perífrasis verbal *tener que* + infinitivo. Treball de fi de grau inèdit. Uppsala University.
- Gavaldà Ferré, N. (2011). Sociolingüística de la variació i lingüística forense. *Llengua, Societat i Comunicació*, vol. 9, p. 49-58. <https://doi.org/10.1344/LSC-2011.9.7>
- Gavaldà Ferré, N. (2013). Index of idiolectal similitude for the phonological module of English applied to forensic speech comparison. Tesi doctoral inèdita. Universitat Pompeu Fabra.
- Gelpí, C. (1998). Anàlisi de freqüències lèxiques: noms, verbs, adjectius i adverbis en *-ment*. A: Payrató, L. (ed.). *Oralmènt. Estudis de variació funcional*. Barcelona: Publicacions de l'Abadia de Montserrat, p. 129-141.
- Gelpí, C. (2013). Errores frecuentes en el lenguaje jurídico: usos incorrectos de las preposiciones. *Iuris: Actualidad y práctica del derecho*, vol. 198, p. 22-27.
- Generalitat de Catalunya. (s. d.). *Optimot*. Consultes lingüístiques. Generalitat de Catalunya. Recuperat a través de: <https://aplicacions.llengua.gencat.cat/llc/AppJava/index.html>
- Gibbons, J. (2013). Maite Turell's Contribution to Attribution Analysis in Forensic Linguistics. A: Casesnoves-Ferrer, R.; Forcadell Guinjoan, M., i Gavaldà Ferré, N. (Eds.). *Ens queda la paraula: Estudis de lingüística aplicada en honor de M. T. Turell*. Barcelona: IULA, p. 368-372.
- Gibbons, J. i Turell, M. T. (Eds.). (2008). *Dimensions of Forensic Linguistics*. Amsterdam/Filadèlfia: John Benjamins Publishing Company.
- Giménez, R. i Queralt, S. (2021). Marcas ideològiques en el discurso de los terroristas del 17-A. *Al-Ghurabá*, vol. 46, p. 19-25.

- Giménez, R.; Elstein, S., i Queralt, S. (2020). The pandemic and the forensic linguistics caseworker's wellbeing: effects and recommendations. *The International Journal of Speech, Language and the Law*, vol. 27 (2), p. 233-254. <https://doi.org/10.1558/ijssl.19548>
- Gimeno Menéndez, F. (1990). *Dialectología y sociolingüística españolas*. Alacant: Universidad de Alicante.
- Gimeno Menéndez, F. (2003). Historia de la dialectología y sociolingüística españolas. A: Alemany Bay, C.; Aracil Varón, B.; Mataix Azuar, R.; Mendiola Oñate, P.; Valero Juan, E. M., i Villaverde Pérez, A. (Coords.). *Con Alonso Zamora Vicente Actas del Congreso Internacional La Lengua, la Academia, lo Popular, los Clásicos, los Contemporáneos*. Alacant: Universidad de Alicante, p. 67-84.
- Goldshmidt, O. T. i Weller, L. (2000). "Talking emotions": Gender differences in a variety of conversational contexts. *Symbolic Interaction*, vol. 23 (2), p. 117-134. <https://doi.org/10.1525/si.2000.23.2.117>
- Gómez Duran, G. (2011). Gramàtica del català rossellonès. Tesi doctoral inèdita. Universitat Autònoma de Barcelona.
- Gómez Lázaro, M. (2018). La subida de clítico en las perífrasis verbales con más de un verbo auxiliar. *Estudios Interlingüísticos*, vol. 6, p. 63-75.
- González Balletbò, I.; Pujolar Cos, J.; Font Tanyà, A., i Martínez Sanmartí, R. (2014). Llengua i joves: Usos i percepcions lingüístics de la joventut catalana. Barcelona: Generalitat de Catalunya. Recuperat a través de: https://treballiaferssocials.gencat.cat/web/.content/JOVENTUT_document_s/arxiu/publicacions/col_estudis/ESTUDIS30.pdf
- González Marfud, A. M. i Perdomo Carmona, M. (2021). Para *luego*, es tarde. Marcadores discursivos *luego* y *después* en el habla culta de La Habana. *Cuadernos de la ALFAL*, vol. 13 (1), p. 88-111.
- González Salgado, J. A. (2011). L'elecció lingüística com a font de problemes jurídics. *Revista de Llengua i Dret*, vol. 55, p. 57-79.
- González, M. (1998). *Bé i bueno*. Apunts sobre l'ús dels marcadors discursius. A: Payrató, L. (ed.). *Oralment. Estudis de variació funcional*. Barcelona: Publicacions de l'Abadia de Montserrat, p. 241-257.

- González, M. (2001). Els marcadors pragmàtics compostos en el relat oral anglès i català. *Caplletra*, vol. 30, p. 73-94.
- González, M. (2004). *Pragmatic Markers in Oral Narrative. The Case of English and Catalan*. Amsterdam/Filadèlfia: John Benjamins Publishing Company.
- González-Sanz, M. (2017). Una aproximación pragmática y sintáctica al marcador conversacional *a ver*. *Pragmalingüística*, vol. 25, p. 232-248.
- Goodman, G. S. i Melinder, A. (2007). Child witness research and forensic interviews of young children: A review. *Legal and Criminological Psychology*, vol. 12 (1), p. 1-19.
<https://doi.org/10.1348/135532506X156620>
- Goodman, M.; Adams, N.; Corneil, T.; Kreukels, B.; Motmans, J., i Coleman, E. (2019). Size and distribution of transgender and gender nonconforming populations: A narrative review. *Endocrinology and Metabolism Clinics of North America*, vol. 48, p. 303-321.
<https://doi.org/10.1016/j.ecl.2019.01.001>
- Grajales Alzate, R. (2011). Funciones del marcador discursivo *pues* en el habla de Medellín. *Forma y Función*, vol. 24 (1), p. 25-45.
- Grant Anthony P. (2020). Contact-Induced Linguistic Change: An Introduction. A: Grant, A. P. (Ed.). *The Oxford Handbook of Language Contact*. Oxford: Oxford University Press.
<https://doi.org/10.1093/oxfordhb/9780199945092.013.1>
- Grant, T. (2010). Text messaging forensics. Txt 4n6: Idiolect free authorship analysis? A: Coulthard, M. i Johnson, A. (Eds.). *The Routledge Handbook of Forensic Linguistics*. Nova York: Routledge, p. 508-522.
- Gregersen, F. i Pedersen, I. L. (1991). Copenhagen as a Speech Community. A: Berge, K. L. i Kotsinas, U. B. (Eds.). *Storstadsspråk och storstadskultur i Norden. Föredrag från ett forskarsymposium*. Estocolm: Institutionen för nordiska språk, p. 57-69.
- Grossmann, M. (1980). Sistema lexical dels noms de color en català. *Estudis Romànics*, p. 99-118.
- Grup Enciclopèdia Catalana. (s. d.). *Diccionari.cat*. Recuperat a través de: <http://diccionari.cat/>

- Güell, S.; Buldó, M.; Pradilla, M. À.; Saura, A., i Verge, J. A. (2006). Pervivència de la realització fricativa labiodental a Valls. A: Blas, J. L.; Casanova, M.; Velando, M. (Eds.). *Discurso y sociedad: Contribuciones al estudio de la lengua en contexto social*. Castelló de la Plana: Publicacions de la Universitat Jaume I, p. 391-400.
- Guevara, G. (2015). Funciones del marcador discursivo *pues*: en un corpus oral. *Dialógica, revista multidisciplinaria*, vol 12 (1), p. 294-323.
- Gumperz, J. J. i Hymes, D. (Eds.). (1964). The Ethnography of Communication. *American Anthropologist*, vol. 66 (6), part 2.
- Guy, G. (1980). Variation in the group and in the individual. A: Labov, W. (Ed.). *Locating language in time and space*. Nova York: Academic Press, p. 1-36.
- Guy, G. (1993). The quantitative análisis of linguistic variation. A: Preston, D. R. (Ed.). *American Dialect Research*. Amsterdam/Filadèlfia: John Benjamins, p. 223-249.
- Halliday, M. A. K. i Hasan, R. (1976). *Cohesion in English*. Nova York: Longman.
- Haugen, E. (1950). The Analysis of Linguistic Borrowing. *Language*, vol. 26, p. 210-231.
- Hawkins, K. i Power, C. B. (1999). Gender differences in questions asked during small decision-making group discussions. *Small Group Research*, vol. 30 (2), p. 235-256.
- Hazen, K. (2011). Labov: language variation and change. A: Wodak, R.; Johnstone, B., i Kerswill, P. (Eds.). *The SAGE Handbook of Sociolinguistics*. Thousand Oaks: SAGE Publications, p. 24-39.
- Hernández Cabrera, C. (2015). Marcadores del discurso de Las Palmas de Gran Canaria. A: Valencia Espinoza, A. i Viguera Ávila, A. (Eds.). *Marcadores hispánicos. Usos de España y América en el corpus de estudio de la norma culta*. Ciudad de México: UNAM, p. 257-302.
- Hernández Campoy, J. M. (1993). Dialectología tradicional, sociolingüística laboviana y geolingüística trudgilliana: tres aproximaciones al estudio de la variación. *ELUA, Estudios de Lingüística Universidad de Alicante*, vol. 9, p. 151-181. <http://doi.org/10.14198/ELUA1993.9.08>

- Hernández, M. A. (2016). El marcador discursivo o sea en el español hablado de Medellín. *Lingüística y Literatura*, vol. 69, p. 295-314. <https://doi.org/10.17533/udea.lyl.n69a13>
- Heydon, G. (2019). *Researching Forensic Linguistics: Approaches and Applications*. Londres/Nova York: Routledge.
- Hidalgo Navarro, A. (2015). Prosodia y partículas discursivas: sobre las funciones de atenuación, intensificación como valores (des)cortesés en los marcadores conversacionales. *Círculo de Lingüística Aplicada a la Comunicación*, vol. 62, p. 76-104. http://dx.doi.org/10.5209/rev_CLAC.2015.v62.49499
- Hodson, T. C. (1939). Sociolinguistics in India. *Man in India*, vol. 19, p. 94-98.
- Holmes, J. (2001). *An Introduction to Sociolinguistics*. Harlow: Pearson Education Limited.
- Hornoiu, D. (2003). Repetition and gender in Romanian conversational discourse. *Ovidius University Annals of Philology*, vol. 14, p. 121-131.
- Hualde, J. I. (1992). *Catalan: descriptive grammar*. London: Routledge.
- Hudson, R. A. (2000). *La sociolingüística*. Barcelona: Editorial Anagrama.
- Hummel, M. (2013). Sincronía y diacronía de los llamados adjetivos adverbializados y de los adverbios en *-mente*. *Anuario de Letras. Lingüística y Filología*, vol. 1 (2), p. 215-281.
- Hyland, K. (2000). *Disciplinary Discourses: Social Interactions in Academic Writing*. Harlow: Pearson Education.
- Hymes, D. (2000). The emergence of sociolinguistics: A response to Samarin. *Journal of Sociolinguistics*, vol. 4 (2), p. 312-315. <https://doi.org/10.1111/1467-9481.00117>
- IBM. (2021). Creación de árboles de decisión. *SPSS Statistics*. Recuperat a través de: <https://www.ibm.com/docs/es/spss-statistics/27.0.0?topic=trees-creating-decision>
- IEC. (2021a). 11.2. Els demostratius. 11. Els demostratius i els possessius. *Gramàtica essencial de la llengua catalana*. [GEIEC]. Barcelona: IEC. Recuperat a través de: https://geiec.iec.cat/capitol_veure.asp?id_gelc=217&capitol=11

- IEC. (2021b). 13.5.3 Combinacions de dos pronoms. 13. Els pronoms personals i els pronoms *ho*, *en* i *hi*. *Gramàtica essencial de la llengua catalana*. [GEIEC]. Barcelona: IEC. Recuperat a través de: https://geiec.iec.cat/capitol_veure.asp?id_gelc=265&capitol=13
- IEC. (2021c). 22.2.4. La coordinació adversativa. 22. La coordinació. *Gramàtica essencial de la llengua catalana*. [GEIEC]. Barcelona: IEC. Recuperat a través de: https://geiec.iec.cat/capitol_veure.asp?id_gelc=163&capitol=22#
- Illamola i Gómez, C. (2015). Contacto de lenguas y la expresión de la posterioridad temporal en el español de Cataluña. Tesis doctoral inèdita. Universitat de Barcelona.
- Illamola, C. i Queralt, S. (2020). Alternancia en la posición de los pronombres átonos en perífrasis verbales como variable para construir perfiles lingüísticos forenses. *Tonos Digital*, vol. 39 (2), p. 1-24.
- Illamola, C. i Vila, F. X. (2015). Análisis de la posición de los pronombres átonos en construcciones perifrásticas entre escolares de la Región Metropolitana de Barcelona. *Diálogo de la Lengua*, vol. 7, p. 36-57.
- Institut d'Estadística de Catalunya [Idescat] i Direcció General de Política Lingüística del Departament de Cultura [DGPL]. (2015). *Enquesta d'usos lingüístics de la població 2013*. Barcelona: Generalitat de Catalunya, Departament de Cultura, Direcció General de Política Lingüística i Institut d'Estadística de Catalunya. Recuperat a través de: <https://www.idescat.cat/cat/idescat/publicacions/cataleg/pdfdocs/eulp2013.pdf>
- Institut d'Estadística de Catalunya [Idescat] i Direcció General de Política Lingüística [DGPL]. (2014a). Població de Catalunya segons coneixement del català i grups d'edat. Taxes. Dades enllaçades 2003-2008-2013. *Institut d'Estadística de Catalunya*. Recuperat a través de: <https://www.idescat.cat/pub/?id=eulp&n=7289>

- Institut d'Estadística de Catalunya [Idescat] i Direcció General de Política Lingüística [DGPL]. (2014b). Població de Catalunya segons coneixement del castellà i edat quinquennal. Taxes. *Institut d'Estadística de Catalunya*. Recuperat a través de: <https://www.idescat.cat/pub/?id=eulp&n=4985>
- Institut d'Estadística de Catalunya [Idescat] i Direcció General de Política Lingüística [DGPL]. (2014c). Població segons llengua inicial. %. *Institut d'Estadística de Catalunya*. Recuperat a través de: <https://www.idescat.cat/pub/?id=eulp&n=3566&t=201300&by=at>
- Institut d'Estadística de Catalunya [Idescat] i Direcció General de Política Lingüística [DGPL]. (2018). Usos lingüístics de la població. Llengua inicial, d'identificació i habitual. *Institut d'Estadística de Catalunya*. Recuperat a través de: <https://www.idescat.cat/indicadors/?id=anuals&n=10364>
- Institut d'Estadística de Catalunya [Idescat] i Secretaria de Política Lingüística del Departament de la Vicepresidència [SPL]. (2009). Enquesta d'usos lingüístics de la població 2008. Barcelona: Institut d'Estadística de Catalunya. Recuperat a través de: <https://www.idescat.cat/cat/idescat/publicacions/cataleg/pdfdocs/eulp2008.pdf>
- Institut d'Estadística de Catalunya [Idescat]. (2018a). Població de Manlleu a 1 de gener. Per sexe i edat quinquennal. *Institut d'Estadística de Catalunya*. Recuperat a través de: <https://www.idescat.cat/pub/?id=pmh&n=9548&geo=mun%3A081120&t=200700>
- Institut d'Estadística de Catalunya [Idescat]. (2018b). Població de Mataró a 1 de gener. Per sexe i edat quinquennal. *Institut d'Estadística de Catalunya*. Recuperat a través de: <https://www.idescat.cat/pub/?id=pmh&n=9548&geo=mun%3A081213&t=200700>
- Institut d'Estadística de Catalunya [Idescat]. (2019a). El municipi en xifres. Manlleu (Osona). *Institut d'Estadística de Catalunya*. Recuperat a través de: <https://www.idescat.cat/emex/?id=081120>

- Institut d'Estadística de Catalunya [Idescat]. (2019b). El municipi en xifres. Mataró (Maresme). *Institut d'Estadística de Catalunya*. Recuperat a través de: <https://www.idescat.cat/emex/?id=081213>
- Institut d'Estadística de Catalunya [Idescat]. (2021). Justícia juvenil. Menors i joves atesos. Comarques i Aran. *Institut d'Estadística de Catalunya*. Recuperat a través de: <https://www.idescat.cat/pub/?id=aec&n=893>
- Institut d'Estudis Catalans. (2009a). *Diccionari de la llengua catalana*. [DIEC]. 2a ed. Recuperat a través de: <http://dlc.iec.cat/>
- Institut d'Estudis Catalans. (2009b). Malament. *Diccionari de la llengua catalana*. 2a ed. Recuperat a través de: <http://dlc.iec.cat/>
- Institut d'Estudis Catalans. (2016a). 8. La flexió dels pronoms personals. La forma dels pronoms febles. *Gramàtica de la llengua catalana*. [GIEC]. Barcelona: IEC.
- Institut d'Estudis Catalans. (2016b). 24. El temps, l'aspecte i el mode. Les perífrasis verbals. *Gramàtica de la llengua catalana*. [GIEC]. Barcelona: IEC.
- Institut d'Estudis Catalans. (2021d). 29.3.1. Les perífrasis modals. 29. Les formes verbals no finites: infinitiu, gerundi i participi. *Gramàtica essencial de la llengua catalana*. [GEIEC]. Barcelona: IEC. Recuperat a través de: https://geiec.iec.cat/capitol_veure.asp?id_gelc=350&capitol=29
- Institut Universitari de Lingüística Aplicada [IULA]. (2013). *Idiolectometria aplicada a la Lingüística Forense*. Recuperat a través de: http://www.iula.upf.edu/rec/forensic_isi/index.htm
- Instituto Nacional de Estadística. (2021). Estadística de condenados: Menores. Últimos datos. *Instituto Nacional de Estadística*. Recuperat a través de: https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=estadistica_C&cid=1254736176795&menu=ultiDatos&idp=1254735573206
- Jiménez Fernández, R. (1997-1998). Sobre el uso de las oraciones circunstanciales en niños de 11 a 13 años. *CAUCE, Revista de Filología y su Didáctica*, vol. 20-21, p. 813-844.
- Jiménez Rodríguez, J. (2010). Adquisición y desarrollo del lenguaje. A: Muñoz García, A. (Coord.). *Psicología del desarrollo en la etapa de educación infantil*. Madrid: Ediciones Pirámide, p 101-120.

- Johnstone, B. (Ed.). (1994). *Repetition in Discourse. Interdisciplinary Perspectives*. New Jersey: Ablex Publishing Corporation.
- Jol, G. i Stommel, W. (2021). The Interactional Costs of “Neutrality” in Police Interviews with Child Witnesses. *Research on Language and Social Interaction*, vol. 54 (3), p. 299-318. <https://doi.org/10.1080/08351813.2021.1939532>
- Jørgensen, A. M. i Martínez López, J. A. (2007). Los marcadores del discurso del lenguaje juvenil de Madrid. *Revista Virtual de Estudos da Linguagem*, vol. 5 (9), p. 1-19.
- Kerbrat-Orecchioni, C. (1980). *L'énonciation. De la subjectivité dans le langage*. Paris: Armand Colin.
- Khorounjaia, E. i Tolchinsky, L. (2004). Discursive constraints on the lexical realization of arguments in Spanish. A: Berman, R. A. (Ed.). *Language Development across Childhood and Adolescence*. Amsterdam: John Benjamins, p. 83-110.
- Koerner, E. F. K. (2002). William Labov and the origins of sociolinguistics in America. A: *Toward a History of American Linguistics*. Londres/Nova York: Routledge, p. 253-284.
- Koerner, K. (1991). Toward a History of Modern Sociolinguistics. *American Speech*, vol. 66 (1), p. 57-70. <http://www.jstor.org/stable/455434>
- Koppel, M.; Akiva, N., i Dagan, I. (2006). Feature instability as a criterion for selecting potential style markers. *Journal of the American Society for Information and Science Technology*, vol. 57 (11), p. 1519-1525.
- Kortmann, B. i Szmrecsanyi, B. (2004). Global synopsis: morphological and syntactic variation in English. A: Kortmann, B. i Schneider, E. W. (Eds.). *A Handbook of Varieties of English A Multimedia Reference Tool*. Berlín/Boston: De Gruyter Mouton, p. 2334-2394. <https://doi.org.sire.ub.edu/10.1515/9783110197181-142>
- Kortschak, C. i Hummel, M. (2018). Los adverbios en *-mente* en los registros formal e informal del español hablado peninsular y mexicano. *Nueva Revista de Filología Hispánica*, vol. 66 (1), 3-43. <https://doi.org/10.24201/nrfh.v66i1.3391>

- Kredens, K. i Góralewska- Łach, G. (1998). Language as sole incriminating evidence: the Augustynek case. *Journal of Speech Language and the Law*, vol. 5 (2), p. 193-202. <https://doi.org/10.1558/sll.1998.5.2.193>
- Kring, A. M., i Gordon, A. H. (1998). Sex differences in emotion: expression, experience, and physiology. *Journal of personality and social psychology*, vol. 74 (3), p. 686-703. <https://doi.org/10.1037//0022-3514.74.3.686>
- Labov, W. (1963). The social motivation of language change. *Word*, vol. 19, p. 273-309.
- Labov, W. (1966). *The Social Stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.
- Labov, W. (1969). Contraction, deletion, and inherent variability of the English copula. *Language*, vol. 45, p. 715-762.
- Labov, W. (1972). *Sociolinguistic patterns*. Oxford: Basil Blackwell.
- Labov, W. (1973). *Language in the Inner City. Studies in the Black English Vernacular*. Filadelfia: University of Pennsylvania Press.
- Labov, W. (1982). Building on Empirical Foundations. A: Lehmann, W. P.; Malkiel, Y. (Eds.). *Perspectives on historical linguistics*. Amsterdam/Filadelfia: John Benjamins, p. 17-92.
- Labov, W. (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change*, vol. 2 (2), p. 205-254. <https://doi.org/10.1017/S0954394500000338>
- Labov, W. (1991). The three dialects of English. A: Eckert, P. (Ed.). *New Ways of Analyzing Sound Change*. Nova York: Academic Press, p. 1-44.
- Labov, W. (2001) *Principles of Linguistic Change*, vol 2., External factors. Oxford: Blackwell.
- Labov, W. (2004) Quantitative Analysis of Linguistic Variation/Quantitative Analyse sprachlicher Variation. A: Wiegand, H. E. (Ed.). *Sociolinguistics: An International Handbook of the Science of Language and Society/Soziolinguistik: Ein internationales Handbuch zur Wissenschaft von Sprache un Gesellschaft*. Volume 1/1. Teilband. Berlín/Nova York: Walter de Gruyter, p. 6-21.

- Labov, W. (2006). *The Social Stratification of English in New York City*. (2a ed.). Nova York: Cambridge University Press.
- Landone, E. (2009). *Los marcadores del discurso y la cortesía verbal en español*. Bern: Peter Lang.
- Larrea Mendizabal, I. (2016). Les actituds lingüístiques dels immigrants panjabis adults a Catalunya. Tesi doctoral inèdita. Universitat Pompeu Fabra.
- Larrosa Barbero, M. (2003-2004). Metodología sociolingüística. *Anuario de Lingüística Hispánica*, vol. 19-20, p. 141-178.
- Larson, V. L. i McKinley, N. L. (1998). Characteristics of adolescents' conversations: A longitudinal study. *Clinical Linguistics & Phonetics*, vol. 12 (3), p. 183-203. <https://doi.org/10.3109/02699209808985221>
- León, G.; Peñalba, J. A., i Escudero, I. (2010). Profe, ¿puedo preguntar? Una breve introducción a la interacción de preguntas y respuestas entre profesor y alumno (2002). *Revista Nebrija de Lingüística Aplicada*, vol. 7, p. 1-14.
- Li, W. (2009). Different Communication Rules between the English and Chinese Greetings. *Asian Culture and History*, vol. 1 (2), p. 72-74. <https://doi.org/10.5539/ach.v1n2p72>
- Llamas Saiz, C. (2010). Los marcadores del discurso y su sintaxis. A Loureda Lamas, Ó. i Acín-Villa, E. (Coords.). *Los estudios sobre marcadores del discurso en español, hoy*. Madrid: Arco Libros, p. 183-240.
- Llopis Rodrigo, F. (2008). Disponibilitat léxica del valencià: classificació dels fenòmens de contacte lingüístic. A: Blas Arroyo, J. L.; Casanova Ávalos, M.; Velando Casanova, M.; Vellón Lahoz, J. (coords.). *Discurso y sociedad II: Nuevas contribuciones al estudio de la lengua en contexto social*, Castellón de la Plana: Universitat Jaume I, p. 571-582.
- Llopis Rodrigo, F. (2009). *El lèxic del valencià central: un estudi de disponibilitat léxica*. València: Servei de Publicacions de la Universitat de València.
- López Escobedo, F. (2011). El análisis de las características dinámicas de la señal de habla como posible marca para la contratación e identificación forense de voz: un estudio para el español de México. Tesi doctoral inèdita. Universitat Pompeu Fabra.

- López García, Á. i Morant, R. (1991). *Gramática femenina*. Madrid: Cátedra.
- López Morales, H. (1990). La sociolingüística actual. A: Moreno Fernández, F. (Rec.). *Estudios sobre variación lingüística*. Alcalá de Henares: Servicio de publicaciones de la Universidad de Alcalá de Henares, p. 79-87.
- López Quero, S. (2017). Conversación escrita y conversación oral: análisis comparativo. *Tonos Digital*, vol. 32, p. 1-33.
- López Serena, A. i Borreguero Zuloaga, M. (2010). Los marcadores del discurso y la variación lengua hablada vs. lengua escrita. A: Loureda Lamas, Ó. i Acín-Villa, E. (Coords.). *Los estudios sobre marcadores del discurso en español, hoy*. Madrid: Arco Libros, p. 415-496.
- Lorente Casafont M. (2016). La fraseología interferida: primera aproximació als factors interns de la interferència. A: Pradilla, M. A. (Ed.). *Miscel·lània d'homenatge a Joan Martí i Castell*, vol. 2. Tarragona: Publicacions Universitat Rovira i Virgili, p. 111-120.
- Lorente Casafont, M. (1998). Variació lèxica. De la idiosincràsia al tractament sistemàtic de la diversitat. *Caplletra*, vol. 25, p. 83-98.
- Lorente, M.; Cabré, M.T., i de Yzaguirre, L. (1997). Lèxic i variació. A: Boix, E; Lorente, M.; Lloret, M. R.; Payrató, L., i Perea, M. P. (eds.). *Anàlisi de la variació lingüística. Actes de la II Jornada sobre la variació lingüística i del III Col·loqui Lingüístic de la Universitat de Barcelona (CLUB-3)*. Barcelona: PPU, p. 121-147.
- Loureda Lamas, Ó. i Acín-Villa, E. (Coords.). (2010). *Los estudios sobre marcadores del discurso en español, hoy*. Madrid: Arco Libros.
- Lüdi, G. (Ed.). (1987). *Devenir bilingue – parler bilingue. Actes du 2e colloque su le bilinguisme*. Tübingen: Max Niemeyer.
- Macaulay, R. (2002). Extremely interesting, very interesting, or only quite interesting? Adverbs and social class. *Journal of Sociolinguistics*, vol. 6 (3), p. 398-417.
- Macaulay, R. K. S. (1978). Variation and consistency in Glaswegian English. A: Trudgill, P. (Ed.). *Sociolinguistic Patterns in British English*. Londres: Edward Arnold, p. 132-143.

- MacLeod, N. (2013). Forensic Linguistics. A: Chappelle, C. A. (Ed.). *The Blackwell Encyclopedia of Applied Linguistics*. Oxford: Wiley-Blackwell, p. 2159-2166.
- MacLeod, N. (2016). "I thought I'd be safe there": Pre-empting blame in the talk of women reporting rape. *Journal of Pragmatics*, vol. 96, p. 96-109.
<https://doi.org/10.1016/j.pragma.2016.03.008>
- Mallinson, C. (2009). Sociolinguistics and sociology: current directions, future partnerships. *Language and Linguistics Compass*, vol. 3 (4), p. 1034-1051.
<https://doi.org/10.1111/j.1749-818x.2009.00144.x>
- Manni, J. i San Martín, A. (2021). Las funciones pragmáticas y la estratificación sociolingüística de las partículas discursivas *po* y *pues* en el habla de Santiago de Chile. *Nueva Revista del Pacífico*, vol. 74, p. 259-289.
- Marí, I. (2011). De la normalització a la sostenibilitat: els límits de la planificació de l'estatus. *Treballs de Sociolingüística Catalana*, vol. 21, p. 83-94 <https://doi.org/10.2436/20.2504.01>
- Marín Jordà, M. J. (2005). Gramaticalització i funció discursiva dels verbs de percepció. *Caplletra*, vol. 38, p. 47-71.
- Marín, M. J. i Cuenca, M. J. (2012). De l'atribució a la modalitat: construccions amb *és que* en català oral. *Caplletra*, vol. 52, p. 65-94.
- Mariscal, S. i Auza Benavides, A. (2017). Typical language development of monolingual spanish-speaking children. A: Auza Benavides A., Schwartz R. (Eds.). *Language Development and Disorders in Spanish-speaking Children*. Cham: Springer. https://doi.org/10.1007/978-3-319-53646-0_1
- Marquina Zarauza, M. (2016). Estudio fonético-acústico de la variación inter e intrahablante de hablantes bilingües de catalán y de castellano. Tesis doctoral inédita. Universitat Pompeu Fabra.
- Marquina, M. i Queralt, S. (2014). Similarity threshold to detect plagiarism in Spanish. *RAEL: Revista Electrónica de Lingüística Aplicada*, vol. 13 (1), p. 79-95.
- Martín Miñana, P. (2019). La referència del demostratiu en la conversa col·loquial. A: Robles i Sabater, F. i Bertomeu-Pi, P. (eds.). *La construcción del discurso en español y catalán/La construcció del discurs en espanyol i català*. Alemanya: Buske, p. 65-86.

- Martín Zorraquino, M. A. (1991). Elementos de cohesión en el habla de Zaragoza. A: Buesa, T.; Enguita, J. M. (Coords.). *I Curso de Geografía Lingüística de Aragón*. Zaragoza: Institución Fernando el Católico, p. 253-286.
- Martín Zorraquino, M. A. (2010). Los marcadores del discurso y su morfología. A Loureda Lamas, Ó. i Acín-Villa, E. (Coords.). *Los estudios sobre marcadores del discurso en español, hoy*. Madrid: Arco Libros, p. 93-182.
- Martín Zorraquino, M. y Portolés, J. (1999). Los marcadores del discurso. A: Bosque, I. i Demonte, V. (Eds.). *Gramática descriptiva de la lengua española*. Madrid: Espasa Calpe, p. 4051-4207.
- Martin, J. R. (1992). *English text. System and Structure*. Amsterdam/Filadèlfia: John Benjamins.
- Martinell Gifre, E. (1986). Expresión lingüística del color en el “Lapidario” de Alfonso X. *Cahiers de linguistique hispanique médiévale*, vol. 11, p. 133-149.
- Martínez Díaz, E. (2002). Las perífrasis modales de obligación “tener que + infinitivo” y “haber de + infinitivo”: Variación e interferencia en el español de Barcelona. Tesis doctoral inédita. Universitat de Barcelona.
- Martínez Díaz, E. (2003). La frecuencia de uso de “haber” y “tener” en las estructuras perifrásticas de obligación. Algún fenómeno de variación en el español de Cataluña. *Interlingüística*, vol. 14, p. 681-694.
- Martínez Pérez, A. V. (2020). El model de llengua de Merlí i els peripatètics. Variació i canvi lingüístic als mitjans de comunicació. A: Radatz, H. I. (ed.). *Canvi lingüístic, estandardització i identitat en català/Linguistic Change, Standardization and Identity in Catalan*. Amsterdam/Philadelphia: John Benjamins, p. 25-49.
- Martínez Pons, M.; Llos Casadellà, B.; Bridgewater Mateu, M., i Querol Puig, E. (2021). *Estudi sociodemogràfic i lingüístic de l'alumnat de 4t d'ESO a Catalunya 2006-2013-2021*. Barcelona: Consell Superior d'Avaluació del Sistema Educatiu. Departament d'Educació. Generalitat de Catalunya.
- Martínez Ruiz, R. (2000). La deixis. A: Briz, A. i Grupo Val.Es.Co. *¿Cómo se comenta un texto coloquial?* Barcelona: Ariel, p. 243-262.

- Marzo, Á. i Turell i Julià, M. T. (2007). Las comunidades gitanas. A: Turell i Julià, M. T. (Ed.). *El plurilingüismo en España*. Barcelona: Universitat Pompeu Fabra, p. 265-285.
- Mas, A. i Montoya, B. (2004). La sociolingüística de la variació als Països Catalans: estat de la qüestió. *Caplletra*, vol. 37, p. 243-266.
- Massanell i Messalles, M. (2018). Avançant en el coneixement de la morfologia històrica de la llengua catalana. *Anuari de filologia. Estudis de lingüística*, vol. 8, p. 85-102. <https://doi.org/10.1344/AFEL2018.8.4>
- Massanell, M. (2020). Dialects. A: Argenter, J. A. i Lüdtke, J. (Coords.). *Manual of Catalan Linguistics*. Berlín/Boston: Walter de Gruyter, p. 371-396.
- McElhinny, B. (2008). Theorizing gender in sociolinguistics and linguistic anthropology. A: Holmes, J. i Meyerhoff, M. (Eds.). *The Handbook of Language and Gender*. Malden/Oxford/Melbourne: Blackwell Publishing, p. 21-42.
- McGhee-Bidlack, B. (1991). The development of noun definitions: A metalinguistic analysis. *Journal of Child Language*, vol. 18, p. 417-434.
- McLaughlin (1978: 55 i s., citat a Singleton i Ryan, 2004
- McLaughlin, B. (1978/1984). *Second Language Acquisition in Childhood. Vol. I. Preschool Children*. Hillsdale: Lawrence Erlbaum.
- Meyerhoff, M. (2002). Social psychology of language and language variation. *University of Pennsylvania Working Papers in Linguistics*, vol. 8 (3), Article 12.
- Meyerhoff, M. (2014). Variation and gender. A: Ehrlich, S.; Meyerhoff, M., i Holmes, J. (Eds.). *The Handbook of Language, Gender, and Sexuality*. Malden/Oxford/Chichester: Wiley Blackwell, p. 87-102.
- Mier, J. Z. (1982). A Sociolinguistic Study of Selected Aspects of Catalan. Tesi doctoral. The University of Michigan.
- Mier, J. Z. (1986). Estudi sociolingüístic de certs aspectes de la llengua catalana. *Treballs de Sociolingüística Catalana*, vol. 6, p. 33-112.
- Milian i Massana, A. (2009). Dels drets lingüístics al dret lingüístic: vint-i-cinc anys d'evolució a Catalunya. *Revista de Llengua i Dret*, vol. 51, p. 63-93.

- Milroy, J. i Milroy, L. (1997). Varieties and variation. A: Coulmas, F. (Ed.). *The Handbook of Sociolinguistics*. Malden/Oxford/Victoria: Blackwell Publishing, p. 47-64.
- Milroy, L. (1987). *Language and Social Networks*. Oxford: Blackwell.
- Milroy, L. i Gordon, M. (2003). *Sociolinguistics. Method and Interpretation*. Malden/Oxford/Melbourne/Berlín: Blackwell Publishing.
- Montero Curiel, M. L. (1994). La expresión del color en un bestiario medieval. *Anuario de estudios filológicos*, vol. 17, p. 369-384.
- Montesinos, A. I. (2006). Unitats fraseològiques i internet. A: Salvador Liern, V. i Climent Raga, L. (Eds.). *El Discurs prefabricat II: fraseologia i comunicació social*. Castelló de la Plana: Servei de Comunicació i Publicacions de la Universitat Jaume I, p. 197-212.
- Montolío Durán, E. i Unamuno, V. (2001). The discourse marker *a ver* (Catalan, *a veure*) in teacher-student interaction. *Journal of Pragmatics*, vol. 33, p. 193-208.
- Montoya i Abat, B. (2002). L'estudi del canvi lingüístic segons la sociolingüística (aplicació a la morfologia del català balear). A: Cano Ivorra, M. A.; Martines Peres, J.; Martines Peres, V., i Ponsoda Sanmartín, J. J. (Coords.). *Les claus del canvi lingüístic*. València: Institut Interuniversitari de Filologia Valenciana, p. 275-300.
- Morales, A. (1999). Adquisición y aprendizaje de la lengua materna: algunas precisiones. *REALE: Revista de Estudios de Adquisición de la Lengua Española*, vol. 12, p. 45-58.
- Moreno Fernández, F. (1998). *Principios de sociolingüística y sociología del lenguaje*. Barcelona: Editorial Ariel.
- Moreno-Ríos, S. i García-Madruga, J. A. (2002). El desarrollo del razonamiento sobre lo que podría haber ocurrido: condicionales indicativos y subjuntivos. *Infancia y Aprendizaje*, vol. 25 (4), p. 485-498. <https://doi.org/10.1174/021037002762064055>
- Morillo-Velarde Pérez, R. (1992). Un modelo de variación sintáctica dialectal. El demostrativo de realce en el andaluz. A: Ariza Viguera, M.; Cano-Aguilar, R.; Mendoza, J. i Narbona Jiménez, A. (eds.). *Actas del II Congreso*

Internacional de Historia de la Lengua española, vol. 2. Madrid: Pabellón de España, p. 219-228.

- Mufwene, S. S.; Rickford, J. R.; Bailey, G., i Baugh, J. (1998). *African-American English. Structure, history and use*. Abingdon/Nova York: Routledge.
- Mugler, F. i Benton, R. (2010). Sociolinguistics in the South Pacific. A: (Ed.). *The Routledge Handbook of Sociolinguistics Around the World*. Londres/Nova York: Routledge, p. 170-178.
- Müller, U.; Overton, W. F., i Reene, K. (2001). Development of Conditional Reasoning: A Longitudinal Study. *Journal of Cognition and Development*, vol. 2 (1), p. 27-49. https://doi.org/10.1207/S15327647JCD0201_2
- Murillo Ornat, S. (2010). Los marcadores del discurso y su semántica. A: Loureda Lamas, Ó. i Acín-Villa, E. (Coords.). *Los estudios sobre marcadores del discurso en español, hoy*. Madrid: Arco Libros, p. 241-280.
- Myers-Scotton, C. (1992). Comparing codeswitching and borrowing. *Journal of Multilingual and Multicultural Development*, vol. 13 (1 i 2), p. 19-39.
- Nagy, M. i Meyerhoff, N. (2008). Introduction – Social lives in Language. A: Nagy, M. i Meyerhoff, N. (Eds.). *Social Lives in Language – Sociolinguistics and multilingual speech communities. Celebrating the work of Gillian Sankoff*. Amsterdam/Filadèlfia: John Benjamins Publishing Company, p. 1-16.
- Navarro, P. Els parlars de transició entre el català occidental i l'oriental. *Caplletra*, vol. 26, p. 91-105.
- Neubert, A. (1976). What is sociolinguistics? Three postulates for sociolinguistic research. *Archivum Linguisticum*, vol. 7 (2), p. 152-160.
- Newbrook, M. (1982). Sociolinguistic reflexes of dialect interference in West Wirral. Tesi doctoral inèdita. Reading University.
- Nichols P. C. (1986). Prepositions in Black and White English of South Carolina. A: Montgomery, M. i Bailey, G. H. (Eds.). *Language Variety in the South: Perspectives in Black and White*. University: University of Alabama Press, p. 73-84.
- Nini, A. (2014). Authorship profiling in a forensic context. Tesi doctoral inèdita. Aston University.

- Ninio, A. i Snow, C. E. (1996). *Pragmatic Development*. Boulder/Oxford: Westview Press.
- Ninio, A. i Snow, C. E. (1999). The development of pragmatics: Learning to use language appropriately. A: Bhatia, T. K. i Ritchie, W. C. (Eds.). *Handbook of Language Acquisition*. Nova York: Academic Press, p. 347-383.
- Nippold, M. A. (1998). *Later language development: the school-age and adolescent years*. Austin: PRO-ED.
- Nippold, M. A. (2004). Research on later language development. International perspectives. A: Berman, R. A. (Ed.). *Language Development across Childhood and Adolescence*. Amsterdam/Filadèlfia: John Benjamins Publishing Company, p. 1-8.
- Nippold, M. A. i Rudzinski, M. (1993). Familiarity and transparency in idiom explanation: A developmental study of children and adolescents. *Journal of Speech, Language and Hearing Research*, vol. 40, p. 245-253.
- Nippold, M. A.; Frantz-Kaspar, M. W., i Vigeland, L. M. (2017). Spoken language production in young adults: examining syntactic complexity. *Journal of Speech, Language, and Hearing Research*, vol. 60, p. 1339-1347.
- Nippold, M.; Uhden, L., Schwarz, I. (1997). Proverb explanation through the lifespan: A developmental study of adolescence and adults. *Journal of Speech, Language and Hearing Research*, vol. 36, p. 728-737.
- Nolan, F. (1983). *The Phonetic Bases of Speaker Recognition*. Cambridge: Cambridge University Press.
- Nolan, F. i Grigoras, C. (2005). A case for formant analysis in forensic speaker identification. *International Journal of Speech Language and the Law*, vol. 12 (2), p. 143-173. <https://doi.org/10.1558/sll.2005.12.2.143>
- Norricks, N. R. (1987). Functions of repetition in conversation. *Text-Interdisciplinary Journal for the Study of Discourse*, vol. 7 (3), p. 245-264.
- Nussbaum, L. i Unamuno, V. (2006). (Eds.). *Usos i competències multilingües entre escolars d'origen immigrant*. Bellaterra: Servei de Publicacions de la Universitat Autònoma de Barcelona.

- Nussbaum, L. i Unamuno, V. (Eds.). (2006). *Usos i competències multilingües entre escolars d'origen immigrant*. Bellaterra: Universitat Autònoma de Barcelona. Servei de Publicacions.
- Olsson, J. (2008). *Forensic Linguistics: Second Edition*. Londres/Nova York: Continuum International Publishing Group.
- Ortega-Rivera, E. i Solana, M. (2015). Migracions a Catalunya: cinc dècades de canvis i continuïtats. A: Domingo, A. (Coord.). *Recerca i immigració VII. Migracions dels segles XX i XXI: una mirada candeliana*. Barcelona: Generalitat de Catalunya. Recuperat a través de: https://ddd.uab.cat/pub/caplli/2015/168378/recerca_immigracioVII_a2015p43.pdf
- Oxburgh, G. E.; Myklebust, T., i Grant, T. (2010). The question of question types in police interviews: a review of the literature from a psychological and linguistic perspective. *International Journal of Speech, Language and the Law*, vol. 17 (1), p. 45-66. <https://doi.org/10.1558/ijssl.v17i1.45>
- Padilla Cruz, M. (2003). Perspectivas pragmáticas sobre los saludos. *Interlingüística*, vol. 14, p. 815-828.
- Paloma, D. i Rico, A. (2008). *Diccionari de dubtes i barbarismes*. Barcelona: Edicions 62.
- Pàmies, X. (2012). Com tradueixo. *Els Marges*, vol. 96, p. 106-116.
- Paredes García, F. (2006). El campo léxico de los colores: convergencias y divergencias en grupos sociales de Madrid. A: Alba Valencia, A. (Ed.). XIV Congreso de la ALFAL. Actas (ed. en CD-ROM). Santiago de Chile. p 131-140.
- Paulston, C. B. i Tucker, G. R. (Eds.). *Sociolinguistics: The Essential Readings*. Oxford: Blackwell.
- Paunonen, H. (1994). The Finnish language in Helsinki. A: Nordberg, B. (Ed.). *The Sociolinguistics of Urbanization: The Case of the Nordic Countries*. Berlín: Walter de Gruyter, p. 223-245.
- Pawłowski, A. (2006). Quantitative linguistics in the study of colour terminology. A research report. A: Biggam, C. P. i Kay, C. J. (eds.). *Progress*

- in Colour Studies. Volume I. Language and Culture.* Amsterdam/Philadelphia: John Benjamins, p. 37-55.
- Payrató, L. (1996a). Transcripció del discurs oral. A: Payrató, L.; Boix, E.; Lloret, M. R. i Lorente, M. (Eds.). *Corpus, corpora*. Barcelona: PPU, p. 181-216.
- Payrató, L. (1996b). *Català col·loquial. Aspectes d'ús corrent de la llengua catalana*. València: Universitat de València. 3a. ed.
- Payrató, L. (1998). Preàmbul. A: Payrató, L. (ed.). *Oralment. Estudis de variació funcional*. Barcelona: Publicacions de l'Abadia de Montserrat, p. 5-33.
- Payrató, L. (2002). L'enunciació i la modalitat oracional. A: Solà, J.; Lloret, M. R.; Mascaró, J., i Pérez Saldanya, M. (Dir.). *Gramàtica del català contemporani*, vol 2. Barcelona: Empúries, p. 1149-1220.
- Payrató, L. (2010). *Pragmàtica, discurs i llengua oral: introducció a l'anàlisi funcional de textos*. Barcelona: UOC.
- Payrató, L. (Ed.). (1998). *Oralment. Estudis de variació funcional*. Barcelona: Publicacions de l'Abadia de Montserrat.
- Payrató, L. i Fitó, J. (Eds.). (2008). *Corpus audiovisual plurilingüe*. Barcelona: Publicacions i Edicions de la Universitat de Barcelona.
- Pearson, J. C.; Turner, L. H., i Todd-Mancillas, W. (1991). *Gender and Communication*. Dubuque: Wm. C. Brown.
- Perdomo Carmona, M. (2020). Contribució al estudio de los marcadores discursivos en un corpus del español actual de Cuba. Tesis doctoral inèdita. Universidad de Zaragoza.
- Perea, M. P. (2012). La dialectometría y su aplicación en el estudio de las variedades del catalán. *Revista De Filología Asturiana*, vol. 9 (9-10), 109-129.
- Pereira, D. I. (2011). Análisis acústico de los marcadores discursivos *a ver, bueno, claro, vale, ¿cómo?* y *ya*. *Onomázein, revista de lingüística, filología y traducción*, vol. 24 (2), p. 85-100.
- Perkins, R. (2015). Native Language Identification (NLID) for Forensic Authorship Analysis of Weblogs. A: Dawson, M. i Omar, M. (Eds.). *New Threats and Countermeasures in Digital Crime and Cyber Terrorism*.

- Hershey: Information Science Reference, p. 213-234.
<https://doi.org/10.4018/978-1-4666-8345-7.ch012>
- Perrin, L.; Deshaies, D., i Paradis, C.. (2003). Pragmatic functions of local diaphonic repetitions in conversation. *Journal of Pragmatics*, vol. 35, p. 1843-1860. [https://doi.org/10.1016/S0378-2166\(03\)00117-6](https://doi.org/10.1016/S0378-2166(03)00117-6)
- Peterson, C.; Dowden, C., i Tobin, J. (1999). Interviewing preschoolers: Comparisons of yes/no and wh- questions. *Law and Human Behavior*, vol. 23 (5), p. 539-555. <https://doi.org/10.1023/A:1022396112719>
- Pineda, A. (2018). *Li-per-els[i]* o la despronominalització del clíctic datiu en català: un fenomen incipient. *Quaderns de Filologia: Estudis Lingüístics*, vol. 23, p. 247-281. <https://doi.org/10.7203/qf.23.13530>
- Plaza, C. (1995). Lleialtat lingüística, edat i nivell educatiu: la *e* posttònica a la Conca de Barberà. A: Turell, M. T. (Ed.). *La sociolingüística de la variació*. Barcelona: PPU, p. 117-138.
- Poblete, M. T. (1998). Los marcadores discursivo-conversacionales de más alta frecuencia en el español de Valdivia (Chile). *Estudios Filológicos*, vol. 33, p. 93-103. <https://doi.org/10.4067/S0071-17131998003300007>
- Poblete, M. T. (2002). La elipsis en la interacción comunicativa y variación sociolingüística. *Estudios Filológicos*, vol. 37, p. 117-131. <https://doi.org/10.4067/S0071-17132002003700007>
- Pons Bordería, S. (2006). A functional approach to discourse markers. A: Fischer, K. (Ed.). *Approaches to Discourse Particles*. Amsterdam: Elsevier, p. 77-99.
- Pons Parera, E. (2011). El dret lingüístic en l'àmbit català. *Treballs de Sociolingüística Catalana*, vol. 21, p. 115-124
<https://doi.org/10.2436/20.2504.01/30>
- Pons-Rodríguez, L. (2010). Los marcadores del discurso en la Historia del español. A: Loureda Lamas, Ó. i Acín-Villa, E. (Coords.). *Los estudios sobre marcadores del discurso en español, hoy*. Madrid: Arco Libros. 523-616.
- Poole, D. A. (2016). *Interviewing Children. The Science of Conversation in Forensic Contexts*. Washington: American Psychological Association.

- Poplack, S. (2012). What does the Nonce Borrowing Hypothesis hypothesize? *Bilingualism: Language and Cognition*, vol. 15 (3), p. 644-648. <https://doi.org/10.1017/S1366728911000496>
- Poplack, S. i Sankoff, D. (1988). Code-Switching. A: Ammon, U.; Dittmar, N., i Mattheier, K. J. (Eds.). *Sociolinguistics: an international handbook of the science of language and society*, vol. 2. Berlín: Walter de Gruyter, p. 1174-1180.
- Porroche Ballesteros, M. (2014). Sobre el marcador discursivo *a ver*. *Español actual: Revista de español vivo*, vol. 102, p. 91-110.
- Portolés, J. (2001). *Marcadores del discurso*. Barcelona: Ariel.
- Pradilla, M. À. (1995). El desafricament prepalatal intervocàlic al català de transició nord-occidental/Valencià. A: Turell, M. T. (Ed.). *La sociolingüística de la variació*. Barcelona: PPU, p. 53-117.
- Pradilla, M. À. (2001). La sociolingüística de la variació: aproximació metodològica (i II). *Noves SL.: Revista de sociolingüística*, vol. 2 (estiu). Recuperat a través de: https://www.gencat.cat/llengua/noves/noves/hm01estiu/metodologia/pradilla1_9.htm
- Pradilla, M. À. (2003). Les representacions del variacionisme en l'àmbit de la llengua catalana. Transferències i transaccions. *Noves SL. Revista de Sociolingüística*, vol. 12 (tardor). Recuperat a través de: <https://www.gencat.cat/llengua/noves/noves/hm03tardor/docs/pradilla.pdf>
- Pradilla, M. À. (2008). *Sociolingüística de la variació i llengua catalana*. Barcelona: IEC.
- Pradilla, M. À. (2011). L'univers calidoscòpic de la sociolingüística de la variació en l'àmbit de la llengua catalana. *Treballs de Sociolingüística Catalana*, vol. 21, p. 125-140.
- Pradilla, M. À. (2020). Social and Functional Variation in Catalan. A: Argenter, J. A. i Lüdtke, J. (eds.). *Manual of Catalan Linguistics*. Berlín/Boston: De Gruyter, p. 397-420.

- Precht, K. (2008). Sex similarities and differences in stance in informal American conversation. *Journal of Sociolinguistics*, vol. 12 (1), p. 89-111.
- Preston (2004). Three kinds of sociolinguistics: A psycholinguistic perspective. A: Fought, C. (Ed.). (2004). *Sociolinguistic Variation: Critical Reflections*. Oxford: Oxford University Press, p. 140-158.
- Pujadas Mestre, F. A. (2020). Estudi dels fenòmens de contacte de llengües en el lèxic d'estudiants de sisè de primària. Treball de final de grau inèdit. Universitat de les Illes Balears.
- Purtí, E. (2006). La incorporació tardana de l'alumnat al·loglot: l'experiència dels primers programes adreçats als estudiants extracomunitaris a Catalunya. A: Vila, F. X.; Boix-Fuster, E., i Alturo, N. (Eds.). *Integrar, des de la fragilitat? Societats plurilingües davant els reptes de les immigracions multilingües: Suïssa, Brussel·les, Luxemburg, Quebec i Catalunya*. Barcelona: IEC, p. 83-87.
- Pusch, C. D. (Ed.). (2006). *La gramàtica pronominal del català: variació – evolució – funció/The grammar of Catalan pronouns: variation – evolution – function*. Aquisgrà: Shaker Verlag.
- Putnam, G. N. i O'Hern, E. M. (1955). The Status Significance of an Isolated Urban Dialect. *Language*, vol. 31 (4), p. 1-32.
- Queen, R. (2013). Gender, Sex, Sexuality, and Sexual Identities. A: Chambers, J. K. i Schilling, N. (eds.). *The Handbook of Language Variation and Change*. Malden/Oxford/Chichester: Wiley-Blackwell, p. 368-387.
- Queralt, S. (2014). Acerca de la prueba lingüística en atribución de autoría hoy. *Revista de Llengua i Dret*, vol. 62, p. 35-48.
- Queralt, S. (2015). Estudio piloto para la evaluación de evidencias lingüísticas en la comparación forense de textos mediante distribuciones poblacionales y relaciones de verosimilitudes. Tesis doctoral inèdita. Universitat Pompeu Fabra.
- Queralt, S. (2022). Cuando el amor es una estafa. A: Galindo, M. M. i Méndez, M. C. (Eds.). *La lingüística del amor: de la pasión a la palabra*. Madrid: Editorial Pie de Página, p. 229-246.

- Queralt, S. (2022). Linguistic profiling. A: Perkins, R.; Picornell, I., i Coulthard, M. (Eds.). *Methodologies and Challenges in Forensic Linguistic Casework*. Wiley, p. 42-60.
- Queralt, S. i Cicres, J. (En premsa). ¿Niños y niñas escriben igual? Análisis de n-gramas de categorías morfológicas. *Culture and Education*.
- Queralt, S. i Giménez, R. (2019). Podem elaborar perfils sociolingüístics forenses a partir de corpus discursius d'àmbit escolar? *Apunts de sociolingüística i política lingüística*, vol. 5, p. 1-5. Recuperat a través de: <https://cuscub.files.wordpress.com/2019/01/apunts-cusc-num-5-gener-2019-linguistica-forense.pdf>
- Queralt, S. i Giménez, R. (2021). Challenges in cases of authorship attribution and possible manipulation of minors. A: *15th Biennial Conference of the International Association of Forensic Linguists. Book of Abstracts*. Birmingham: IAFL.
- Queralt, S. i Giménez, R. (En premsa, 2022). Análisis lingüístico para establecer la existencia de coacción en el discurso de un acusado por los atentados del 17A. *VII Anuario Internacional de Criminología y Ciencias Forenses*, vol. 7.
- Queralt, S.; Marquina, M., i Giménez, R. (2018). Evidencias lingüísticas del plagio en el periodismo español. *Estudios sobre el Mensaje Periodístico*, vol. 24 (2), p. 1559-1578. <https://doi.org/10.5209/ESMP.62234>
- Queralt, S.; Spassova, M. S., i Turell, M. T. (2011). L'ús de les combinacions de seqüències de categories gramaticals com a nova tècnica de comparació forense de textos escrits. *Llengua, Societat i Comunicació*, vol. 9, p. 60-68.
- Rafanell, A. (1991). Un recorregut per la prehistòria de la dialectologia catalana. *Estudi general: Revista de la Facultat de Lletres de la Universitat de Girona*, vol. 11, p. 13-31.
- Ramírez Salado, M. (2017). Antecedentes de la lingüística forense: ¿desde cuándo se estudia el lenguaje como evidencia? *Pragmalingüística*, vol. 25, p. 525-539.
- Ranson, D. (1999). Variación sintáctica del adjetivo demostrativo en español. A: Serrano, M. J. (ed.). *Estudios de variación sintáctica*. Madrid/Frankfurt: Iberoamericana/Vervuert, p. 121-142.

- Rash, F. (2004). Linguistic Politeness and Greeting Rituals in German-speaking Switzerland. *Linguistik Online*, vol. 20 (3), p. 47-72. <https://doi.org/10.13092/lo.20.1063>
- Raymond, G. (2003). Grammar and Social Organization: Yes/No Interrogatives and the Structure of Responding. *American Sociological Review*, vol. 68, p. 939-967. <http://dx.doi.org/10.2307/1519752>
- Real Academia Española. (2009). 31.10 La coordinación adversativa. *Nueva gramática de la lengua española*, vol. II. Sintaxis II. [GRAE]Madrid: Espasa, p 2450-2459.
- Reguera, A. (2010). Metodología de la investigación lingüística. Prácticas de escritura. *Boletín de Filología*, vol. 45 (2), p. 319-323. <https://doi.org/10.4067/S0718-93032010000200012>
- Ribera, J. E. (2019). “En el sur de esta sucursal del infierno que llamamos Ciudad de México”. La deixis espacial y la deixis discursiva. Instituto de Investigaciones Filológicas (IIF). Centro de Lingüística Hispánica (CLH). Universidad Nacional Autónoma de México (UNAM). Recuperat a través de: <https://www.amla.org.mx/wp-content/uploads/2019/03/Conferencia-Deixis-espacial-y-deixis-discursiva-BIS.pdf>
- Rigau, G. (1998). La variació sintàctica: uniformitat en la diversitat. *Caplletra*, vol. 25, p. 63-82.
- Ríos González, G. (2007). Diferencias léxicas entre el hombre y la mujer en tres centros de interés: *Saludos, Temas de conversación y Despedidas*. *Filología y Lingüística*, vol. 33 (1), p. 151-166.
- Rodríguez González, F. (Coord.). (2002). *El lenguaje de los jóvenes*. Barcelona: Ariel.
- Rodríguez-Vida, S. (1997). *Catalán-castellano: frente a frente*. Barcelona: Inforbooks.
- Rokach, L. i Maimon, O. (2015). *Data Mining with Decision Trees. Theory and Applications*. Singapore/Hackensack/Londres: World Scientific.

- Romero, S. (2001). Una aproximació a l'estudi dels processos de canvi lingüístic en varietats geogràfiques no prestigioses: el cas de la Conca de Tremp. *Noves SL: Revista de Sociolingüística*, (hivern). Recuperat a través de: https://www.gencat.cat/llengua/noves/noves/hm01hivern/catalana/rome-ro1_7.htm
- Romero, S. (2003). La diversificació estilística com a factor de variació i canvi lingüístics. *Noves SL: Revista de Sociolingüística*, (tardor). Recuperat a través de: <https://www.gencat.cat/llengua/noves/noves/hm03tardor/docs/romero.pdf>
- Rose, P. (2002). *Forensic speaker identification*. Londres/Nova York: Taylor & Francis.
- Rossi, G. (2020). Other-repetition in conversation across languages: Bringing prosody into pragmatic typology. *Language in Society*, vol. 49, p. 495-520. <https://doi.org/10.1017=S0047404520000251>
- Rudman, J. (1998). The State of Authorship Attribution Studies: Some Problems and Solutions. *Computers and the Humanities*, vol. 31 (4), p. 351-365. <https://www.jstor.org/stable/30200436>
- Rudman, J. (2012). The State of Non-Traditional Authorship Attribution Studies—2012: Some Problems and Solutions. *English Studies*, vol. 93 (3), p. 259-274. <https://doi.org/10.1080/0013838X.2012.668785>
- Ruiz González, N. (2021). *Hombre, quiero decir que...* Análisis sociolingüístico de la partícula *hombre* como reformulador en el corpus PRESEEA de Granada. *Tonos Digital*, vol. 40 (1), p. 1-25.
- Ruiz-González, N. (2018). La reformulación discursiva en el español de Granada: el caso de *o sea*. *Itinerarios: Revista de estudios lingüísticos, literarios, históricos y antropológicos*, vol. 28, p. 177-199.
- Ruiz-González, N. (2019). Marcadores discursivos en andaluz/Discourse Markers in Andalusian. *Revista Internacional de Lingüística Iberoamericana*, vol. 17 (2-34), p. 79-100. <https://www.jstor.org/stable/10.2307/48587506>

- Rydland, V. i Aukrust, V. G. (2005). Lexical Repetition in Second Language Learners' Peer Play Interaction. *Language Learning*, vol. 55 (2), p. 229-274.
- Salvador, G. (1952). Fonética masculina y fonética femenina en el habla de Vertientes y Tarifa (Granada). *Orbis*, vol. 1, p. 19-24.
- Sampietro, A. (2019). Emoji and rapport management in Spanish WhatsApp chats. *Journal of Pragmatics*, vol. 143, p. 109-120.
<https://doi.org/10.1016/j.pragma.2019.02.009>
- Samuelson, L. K. i McMurray, B. (2017). What does it take to learn a word? *WIREs Cognitive Science*, vol. 8 (1-2), p. 1-10.
<https://doi.org/10.1002/wcs.1421>
- San Martín Núñez, A. (2017). Análisis sociolingüístico de los reformuladores de explicación en el español hablado de Santiago de Chile. *Revista Signos. Estudios de Lingüística*, vol. 50 (93), p. 124-147.
<https://doi.org/10.4067/S0718-09342017000100006>
- San Martín Núñez, A. (2017). Análisis sociolingüístico de los reformuladores de explicación en el español hablado de Santiago de Chile. *Revista Signos. Estudios de lingüística*, vol. 50 (93), p. 124-147.
<https://doi.org/10.4067/S0718-09342017000100006>
- Santamaria, L. (2010). El col·loquial de la ficció. La caracterització de la parla dels joves. *Llengua i Ús*, vol 49, p. 78-89.
- Santana (2016). *Luego y después: uso discursivo y variación/Luego and después: discursive use and variation. Rilce. Revista de Filología Hispánica*, vol. 32 (2), p. 513-535. <https://doi.org/10.15581/008.32.2.513-35>
- Santana Marrero, J. (1997). Las locuciones condicionales. ¿Un caso de variación sintáctica? *Interlingüística*, vol. 8, p. 311-318.
- Santana Marrero, J. (2015). Marcadores discursivos de Sevilla. A: Vigueras Ávila, A. i Valencia, A. (Coords.). *Más sobre marcadores hispánicos. Usos de España y América en el corpus de estudio de la norma culta*. Mèxic D. F.: UNAM, p. 443-490.
- Santos Río, L. (2003). *Diccionario de partículas*. Salamanca: Luso-Española de Ediciones.

- Saragossà, A. (2004). El valor dels adjectius demostratius segons les llengües catalana i castellana: problemes i solucions. *Quaderns de Filologia. Estudis Lingüístics*, vol. 9, p. 199-243.
- Segura Llopes, C. (2003). *Variació dialectal i estandardització al Baix Vinalopó*. Alacant/Barcelona: Institut Universitari de Filologia Valenciana i Publicacions de l'Abadia de Montserrat.
- Sendra i Rovira, M. (2012). Estudi dels fenòmens de contacte de llengües en el lèxic d'escolars de primera llengua castellana i la seva evolució lingüística (de 6è de primària a 1r d'ESO). Treball de final de màster inèdit. Universitat Autònoma de Barcelona, Universitat de Barcelona.
- Serrano Reyes, P. (2013). Análisis socio-pragmático del marcador conversacional vale como iniciador de intervención/Sociopragmatic analysis of conversational marker vale at initial context of intervention. *Estudios interlingüísticos*, vol. 1, p. 135-147.
- Serrano Zapata, M. (2016). Fenòmens de contacte lingüístic en el lèxic disponible català de Lleida. *Studia Romanica Posnaniensia*, vol. 43 (2), p. 51-64. <https://doi.org/10.14746/strop.2016.432.003>
- Serrano Zapata, M. (2021). Transferències lingüístiques castellà-català: un estudi de disponibilitat lèxica a Lleida. *Treballs de Sociolingüística Catalana*, vol. 31, p. 149-164. <https://doi.org/10.2436/20.2504.01.183>
- Serrano, M. J. (2006). Acción e interacción social en variación sintáctica y discursivo-pragmática. A: Blas Arroyo, J. L.; Casanova Ávalos, M.; Velando Casanova, M. (eds.). *Discurso y sociedad: contribuciones al estudio de la lengua en contexto social*. Castelló de la Plana: Publicacions de la Universitat Jaume I, p. 121-140.
- Serrano, M. J. (2008). El rol de la variable *sexo* o *género* en sociolingüística: ¿diferencia, dominio o interacción? *Boletín de Filología*, vol. 43, p. 175-192.
- Shuy (1969). Sex as a factor in sociolinguistic research. Presentat a Anthropological Society of Washington. Recuperat a través de: <https://files.eric.ed.gov/fulltext/ED027522.pdf>
- Shuy, R. W. (1993). *Language Crimes: The Use and Abuse of Language Evidence in the Courtroom*. Cambridge: Blackwell.

- Shuy, R. W. (2003). Children, Language, and the Law. *Montana Law Review*, vol. 66 (1), p. 61-79.
- Shuy, R. W. (2008). *Fighting Over Words: Language and Civil Law Cases*. Oxford: Oxford University Press.
- Shuy, R. W. (Ed.). (1965). *Social Dialects and Language Learning: Proceedings of the Bloomington, Indiana, Conference, 1964*. Champaign, Ill.: National Council of Teachers of English.
- Silva-Corvalán, C. (2001). *Sociolingüística y pragmática del español*. Washington: Georgetown University Press.
- Singleton, D. i Ryan, L. (2004). *Language Acquisition: The Age Factor*. Clevedon/Buffalo/Toronto: Multilingual Matters Ltd.
- Sinner, C. (2003). Valores y usos de *haber (de) + infinitivo* y *tener que + infinitivo* en diferentes variedades del castellano. A: Pusch, C. D. i Wesch, A. (Eds.). *Verbalperiphrasen in den (ibero-)romanischen Sprachen*. Hamburg: Helmut Buske, p. 193-206.
- Sinner, C. (2008). Castellano y catalán en contacto: oralidad y contextos informales/Spanish and Catalan in contact: orality and informal contexts. *Oihenart*, vol. 23, p. 521-543.
- Sinner, C. i Wieland, K. (Eds.). (2015). *Norma, ús i actituds lingüístiques. El paper del català en la vida quotidiana*. Leipzig: Leipziger Universitätsverlag.
- Snell, J. (2010). From sociolinguistic variation to socially strategic stylisation. *Journal of Sociolinguistics*, vol. 14 (5), p. 630-656.
- Solan, L. M. i Tiersma, P. T. (2005). *Speaking of Crime: The Language of Criminal Justice*. Chicago: University of Chicago Press.
- Solé Solé, E. (2002). Els noms col·lectius catalans descripció i reconeixement. Tesi doctoral inèdita. Universitari Pompeu Fabra.
- Soler Manyer, C.; Mas Craviotto, X., i Viladomat Monterde, G. (s.d.). Comhoderia.cat. Recuperat a través de: <http://www.ub.edu/comhoderia/>
- Sorolla Vidal, N. (2016). Tria de llengües i rols sociolingüístics a la Franja des de la perspectiva de l'anàlisi de xarxes socials. Tesi doctoral inèdita. Universitat de Barcelona.

- Sorolla, N. i Flors-Mas, A. (2020). The use of Catalan by millennials in Catalonia: less influence of linguistic origin. *Revista de Llengua i Dret*, vol. 73, p. 50-68. <https://doi.org/10.2436/rld.i73.2020.3448>
- Spang, K. (1979). *Fundamentos de retórica*. Pamplona: Universidad de Navarra.
- Spassova, M. S. (2009). El potencial discriminatorio de las secuencias de categorías gramaticales en la atribución forense de autoría de textos en español. Tesis doctoral inèdita. Universitat Pompeu Fabra.
- Spassova, M.; Turell, M. T., i Cicres, J. (Eds.). (2007). *Proceedings of the 2nd European IAFL Conference on Forensic Linguistics/Language and the Law*. Barcelona: Documenta Universitaria.
- Spolsky, B. (2011). Ferguson and Fishman: Sociolinguistics and the Sociology of Language. A: Wodak, R.; Johnstone, B., i Kerswill, P. (Eds.). *The SAGE Handbook of Sociolinguistics*. Thousand Oaks: SAGE Publications, p. 3-15.
- Stanford, J. N. (2016). A call for more diverse sources of data: Variationist approaches in non-English contexts. *Journal of Sociolinguistics*, vol. 20 (4), p. 525-541.
- Stenström, A. B. i Jørgensen, A. M. (Eds.). (2009). *Youngspeak in a Multilingual Perspective*. Amsterdam/Filadèlfia: John Benjamins.
- Stivers, T. (2010). An overview of the question–response system in American English conversation. *Journal of Pragmatics*, vol. 42, p. 2772-2781.
- Stokes, C. R. (2015). The use of Catalan verbal periphrases *haver de* and *tenir que* on Twitter. *Sociolinguistic Studies*, vol. 9 (4), p. 445-466.
- Stuart-Smith, J. i Haddican, B. (2010). Sociolinguistics in the British Isles. A: (Ed.). *The Routledge Handbook of Sociolinguistics Around the World*. Londres/Nova York: Routledge, p. 296-309.
- Suïls, J.; Ariño, S.; Alturo, N., i Turell, M. T. (2009). El canvi lingüístic al Pont de Suert, vint anys després (estudi en temps real a partir de l'estudi en temps aparent de 1986). A: Creus, I.; Puig, M., i Veny, J. R. (Ed.). *Actes del XV Col·loqui de l'Associació Internacional de Llengua i Literatura Catalanes (Lleida, 7-11 setembre 2009)*, vol. 1. Associació Internacional de Llengua i Literatura Catalanes: Publicacions de l'Abadia de Montserrat, p. 297-314.

- Svennevig, J. (2004). Other-repetition as display of hearing, understanding and emotional stance. *Discourse Studies*, vol. 6 (4), p. 489-516. <http://www.jstor.org/stable/24048492>
- Tagliamonte, S. A. (2006). *Analysing Sociolinguistic Variation*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511801624>
- Tagliamonte, S. A. (2012). *Variationist Sociolinguistics: Change, Observation, Interpretation*. Oxford: Wiley-Blackwell.
- Tagliamonte, S. i Ito, R. (2002). Think *really different*: Continuity and specialization in the English dual form adverbs. *Journal of Sociolinguistics*, vol. 6 (2), p. 236-266.
- Takeda, L. U. (2021). Elicitation of mutual understanding and achievement of coherence. Allo-repetition in Japanese EFL speaking and chat-style writing interactions. A: Lee, C. (Ed.). *Second Language Pragmatics and English Language Education in East Asia*. Londres/Nova York: Routledge, p. 41-58.
- Tannen, D. (1987). Repetition in Conversation: Toward a Poetics of Talk. *Language*, vol. 63 (3), p. 574-605. <https://www.jstor.org/stable/415006>
- Tannen, D. (1989). *Talking voices: Repetition, dialogue, and imagery in conversational discourse*. Cambridge: Cambridge University Press.
- Taplin, J. E.; Staudenmayer, H., i Taddonio, J. L. (1974). Developmental changes in conditional reasoning: Linguistic or logical? *Journal of Experimental Child Psychology*, vol. 17, p. 360-373.
- Tarabini, A. (2021). Abordant la desigualtat educativa: el rol de la pedagogia, el currículum i les creences docents. *Anuari de l'Educació de les Illes Balears*, vol. 2021, p. 16-31.
- Tenorio Segarra, X. (2013). Llengua inicial i llengua amb progenitors. Una mateixa realitat o dues variables diferents? *Treballs de Sociolingüística Catalana*, vol. 23, p. 407-426. <https://doi.org/10.2436/20.2504.01.66>
- Thomas, E. R. (2007). Phonological and phonetic characteristics of African American Vernacular English. *Language and Linguistics Compass*, vol. 1, p. 450-475.
- Todolí i Cervera, J. (1998). *Els pronoms personals*. València: Publicacions de la Universitat de València.

- Todolí, J. (1999). Els pronoms clítics del valencià i el barceloní contrastats. A: Kailuweit, R. i Radatz, H. I. (eds.). *Katalanisch: Sprachwissenschaft und Sprachkultur. Akten des 14. Deutschen Katalanistentags im Rahmen von Romania I. Jena, 28.9.-2.10.1997*. Frankfurt am Main: Vervuert, p. 253-265.
- Todolí, J. (2002). Els pronoms. A: Solà, J.; Lloret, M. R.; Mascaró, J., i Pérez Saldanya, M. (Dir.). *Gramàtica del català contemporani*, vol. 2. Barcelona: Empúries, p. 1337-1433.
- Tolchinsky, L.; Martí, M. A., i Llauradó, A. (2010). The growth of the written lexicon in Catalan. From childhood to adolescence. A: Vernon Carter, S. A. i Alvarado, M. (Eds.). *Developmental aspects of written language*. Amsterdam/Filadèlfia: John Benjamins Publishing Company, pp. 206-235.
<https://doi.org/10.1075/wll.13.2.02tol>
- Tomàs i Pallejà, R. (1998). Constitució i característiques del corpus utilitzat. A: Payrató, L. (ed.). *Oralment. Estudis de variació funcional*. Barcelona: Publicacions de l'Abadia de Montserrat, p. 35-54.
- Toniolo, M. T. i Zurita, M. E. (2014). Marcadores discursivos en Córdoba, Argentina. *Cuadernos de la ALFAL*, vol. 5, p. 69-106.
- Torner Castells, S. (2007). *De los adjetivos calificativos a los adverbios en “-mente” semántica y gramática*. Madrid: Visor Libros.
- Torres Cacoullous, R. (1999). Construction frequency and reductive change: Diachronic and register variation in Spanish clitic climbing. *Language Variation and Change*, vol. 11, p. 143-170.
- Trousdale, G. (2010). *An Introduction to English Sociolinguistics*. Edimburg: Edinburgh University Press.
- Troya Déniz, M. i Pérez Martín, A. M. (2011). Distribución de clítics con perífrasis verbales en hablantes universitarios de Las Palmas de Gran Canaria. *Lingüística*, vol. 26, p. 9-25.
- Trudgill, P. (1974). *The Social Differentiation of English in Norwich*. Cambridge: Cambridge University Press.
- Trudgill, P. (2000). *Sociolinguistics. An introduction to language and society*. (4a ed.). Londres: Penguin Books.

- Trudgill, P. (2006). Social Differentiation. A: Chambers, J. K, Trudgill, P. y Schilling-Estes, N. (Eds.). (2006). *The Handbook of Language Variation and Change*. Malden: Blackwell Publishing, p. 373-374.
- Tucker, G. R. (1997). The development of sociolinguistics as a field of study: Concluding observations. A: Paulston, C. B. i Tucker, G. R. (Eds.). *The early days of sociolinguistics: memories and reflections*. Dallas: The Summer Institute of Linguistics, p. 317-324.
- Turell, M. T. (1995a). L'alternança de llengües i el préstec en una comunitat inter-ètnica. A: Turell, M. T. (Ed.). (1995). *La sociolingüística de la variació*. Barcelona: PPU. p. 259-293.
- Turell, M. T. (1995b). The variationist view of variation: evidence from Catalan-speaking communities. *Catalan Review: International Journal of Catalan Culture*, vol. 12, p. 275-290.
- Turell, M. T. (2003). El temps aparent i el temps real en estudis de variació i canvi lingüístic. *Noves SL.: Revista de sociolingüística*, vol. 3, (tardor), p. 1-11. Recuperat a través de: <https://www.gencat.cat/llengua/noves/noves/hm03tardor/docs/turell.pdf>
- Turell, M. T. (2004). Textual kidnapping revisited: the case of plagiarism in literary translation. *The International Journal of Speech, Language and the Law*, vol. 11 (1), p. 1-26. <https://doi.org/10.1558/ijssl.v11i1.1>
- Turell, M. T. (2006). Aplicaciones forenses de la lingüística descriptiva y de corpus. A: Juan Garau, M.; Amengual Pizarro, M., i Salazar Noguera, J. (Eds.). *Lingüística aplicada en la sociedad de la información y la comunicación*. Edicions UIB, p. 43-64.
- Turell, M. T. (2007). Plagio y traducción literaria. *Vasos Comunicantes*, vol. 37 (1), p. 43-54.
- Turell, M. T. (2008). Plagiarism. A: Gibbons, J. i Turell, M. T. (Eds.). *Dimensions of Forensic Linguistics*. Amsterdam/Filadèlfia: John Benjamins, p. 265-299.
- Turell, M. T. (2009). The task and role of the forensic linguist in giving evidence. A: Amengual Pizarro, M.; Juan Garau, M., i Salazar Noguera, J. (Eds.). *New perspectives on English studies*. Edicions UIB.

- Turell, M. T. (2010). The use of textual, grammatical and sociolinguistic evidence in forensic text comparison. *The International Journal of Speech, Language and the Law*, vol. 17 (2), p. 211-250.
<https://doi.org/10.1558/ijssl.v17i2.211>
- Turell, M. T. (2011). La tasca del lingüista detectiu en casos de detecció de plagi i determinació d'autoria de textos escrits. *Llengua, Societat i Comunicació*, vol. 9, p. 69-85.
- Turell, M. T. (2012). Detection approaches to plagiarism in literary translation. A: Lanero Fernández, J. J. i Chamosa, J. L. (Coords.). *Lengua, traducción, recepción: en honor de Julio César Santoyo*, vol. 2. Lleó: Universidad de León, p. 535-559.
- Turell, M. T. (2013). Retrospectiva del model labovià de variació lingüística des del català. *Treballs de Sociolingüística Catalana*, vol. 23, p. 25-56.
<https://doi.org/10.2436/20.2504.01.50>
- Turell, M. T. (Ed.). (1995). *La sociolingüística de la variació*. Barcelona: PPU.
- Turell, M. T. (Ed.). (2005). *Lingüística forense, lengua y derecho. Conceptos, métodos y aplicaciones*. Girona: Documenta Universitaria.
- Turell, M. T. i Gavaldà, N. (2013). Towards an index of idiolectal similitude (or distance) in forensic authorship analysis. *Journal of Law and Policy*, vol. 21 (2), p. 495-514.
- Tuson Valls, A. (1998). Diferència sexual i variació lingüística. *Caplletra*, vol. 25, p. 205-220.
- Universitat Pompeu Fabra [UPF]. (s. d.). Servei gratuït d'assessorament lingüístic. UPFSolidària. Recuperat a través de:
<https://www.upf.edu/web/upfsolidaria/servei-gratuit-d-assessorament-linguistic>
- Valls Alecha, E. (2010). Cap a una diversificació metodològica de la dialectometria catalana: primers resultats d'aplicar la distància de Levenshtein al corpus oral dialectal. *Estudios interlingüísticos*, vol. 20, p. 1-10.

- Valls Alecha, E. (2013). Direccionalitat, ritme, abast i naturalesa del canvi lingüístic en curs en català nord-occidental. De l'anàlisi dialectomètrica a la reflexió sociolingüística. Tesi doctoral inèdita. Universitat de Barcelona.
- Valls, E. (2013). Canvi morfològic vs. canvi fonològic en català nord-occidental. *Treballs de Sociolingüística Catalana*, vol. 23, p. 57-79.
- Valls, E. i Clua, E. (2013). Distància de Levenshtein vs. "mètode COD". Dos sistemes de mesura de la distància fonètica aplicats al Corpus Oral Dialectal. A: Casanova Herrero, E. i Calvo Rigual, C. (Coord.). *Actas del XXVI Congreso Internacional de Lingüística y de Filología Románicas*, vol. 6. Berlin: De Gruyter, p. 255-267.
- Valls, E. i González, M. (2016). La subagrupació romànica de la llengua catalana: una aproximació dialectomètrica de base fonètica a l'Atlas Linguistique Roman. *Studia Romanica Posnaniensia*, vol. 43 (2), p. 5-27. <https://doi.org/10.14746/strop.2016.432.001>
- Van Lacker Sidtis, D. i Wolf, R. (2015). Pragmatic verbal repetition: Review and application of a new method of quantification. *Text and Talk*, vol. 35 (2), p. 263-267.
- Vanrell Bosch, M. M.; Mascaró, I.; Prieto, P.; Torres-Tamarit, F. (2010). Preguntar per saber i preguntar per confirmar. L'entonació de les interrogatives absolutes informatives i confirmatòries en català central i balear. *Randa*, vol. 64, p. 77-95.
- Vázquez Carranza, A. (2019). "Chilapa pues": variación regional en el uso de pues en posición final en el español mexicano. *Revista internacional de lingüística iberoamericana*, vol. 33, p. 147-165.
- Vázquez Carranza, A. (2020). Saludos y despedidas: tipología y contraste entre datos intuitivos y observacionales. *Pragmática Sociocultural/Sociocultural Pragmatics*, vol. 8 (2), p. 182-203. <https://doi.org/10.1515/soprag-2020-0009>
- Veny, J. i Massanell, M. (2015). *Dialectologia catalana. Aproximació pràctica als parlars catalans*. Universitat de València, Universitat d'Alacant, Universitat de Barcelona.

- Veny, J. i Pons Griera, L. (1964-1978). *Atlas Lingüístic del Domini Català*. Recuperat a través de: <https://aldc.espais.iec.cat/>
- Vercher García, E. J. (2003). Problemas de definición y clasificación de las partículas modales: Estudio comparado entre el ruso y el español. *Interlingüística*, vol. 14, p. 1017-1032.
- Verdoodt, A. F. (1997). The demography of language. A: Coulmas, F. (Ed.). *The Handbook of Sociolinguistics*. Malden/Oxford/Victoria: Blackwell Publishing, p. 33-43.
- Viaplana, J. (2002). *Dialectologia*. València. Universitat de València.
- Vigara Tauste, A. M. (1997). Comodidad y recurrencia en la organización del discurso coloquial. *Espéculo, Revista de Estudios Literarios*, vol. 7. Recuperat a través de: <https://webs.ucm.es/info/especulo/numero7/index.htm>
- Vigueras Ávila, A. (2014). Marcadores del discurso de la ciudad de México. *Cuadernos de la ALFAL*, vol. 5, p. 191-245
- Vila i Moreno, F. X. (1996). When Classes Are Over. Language Choice and Language Contact in Bilingual Education in Catalonia. Tesi doctoral inèdita. Vrije Universiteit Brussel.
- Vila i Moreno, F. X. (1998). *Bueno, vale ja de criticar, no?* Marques transcòdiques lèxiques i variació funcional en català. A: Payrató, L. (ed.). *Oralmnt. Estudis de variació funcional*. Barcelona: Publicacions de l'Abadia de Montserrat, p. 259-274.
- Vila i Moreno, F. X. (2004). Introducció. El model lingüístic escolar a Catalunya: balanç i reptes. *Llengua, Societat i Comunicació*, vol. 1, p. 1-4. <https://doi.org/10.1344/LSC-2004.1.1>
- Vila, F. X. (2013). De l'«etapa autonòmica» a la «sociolingüística del multilingüisme»: una dècada de recerca sociolingüística sobre les noves immigracions. A: Vila, F. X. i Salva, E. (Eds.). *Noves immigracions i llengües*. Barcelona: MRR, p. 11-28.
- Vila, F. X. (2020). Language Demography A: Argenter, J. A. i Lüdtke, J. (eds.). *Manual of Catalan Linguistics*. Berlín/Boston: De Gruyter, p. 629-648.
- Vila, F. X. (Ed.) (2012). *Survival and Development of Language Communities. Prospects and Challenges*. Bristol/Buffalo/Toronto: Multilingual Matters.

- Vila, F. X.; Ubalde, J.; Bretxa, V., i Comajoan-Colomé, L. (2020) Changes in language use with peers during adolescence: a longitudinal study in Catalonia. *International Journal of Bilingual Education and Bilingualism*, vol. 23 (9), p. 1158-1173.
<https://doi.org/10.1080/13670050.2018.1436517>
- Wardhaugh, R. (2006). *Language Variation. A: An Introduction to Sociolinguistics*. Malden/Oxford/Victoria: Blackwell Publishing, p. 135-161.
- Waterman, A. H., Blades, M., i Spencer, C. (2001). Interviewing children and adults: The effect of question format on the tendency to speculate. *Applied Cognitive Psychology*, vol. 15 (5), p. 521-531.
<https://doi.org/10.1002/acp.741>
- Weinreich, U. (1953). *Languages in Contact. Findings and Problems*. La Haia/París: Mouton.
- Weinreich, U.; Labov, W. i Herzog, M. (1968). Empirical Foundations for a Theory of Language Change. A: Lehmann, W. i Malkiel, Y. (Eds.). *Directions for Historical Linguistics*. Austin: University of Texas Press, p. 95-188.
- Wesch, A. (2008). Mujeres de hacer faenas que limpian racholas: sobre algunos catalanismos léxicos. A: Sinner, C. i Wesch, A. (Eds.). *El castellano en las tierras de habla catalana*. Madrid/Frankfurt: Iberoamericana/Vervuert, p. 57-64.
- Wieland (2009a). El llenguatge juvenil a Barcelona i la seva representació als mitjans de comunicació. *Noves SL.: Revista de sociolingüística*, (hivern), p. 1-7.
Recuperat a través de:
<http://www.gencat.cat/llengua/noves/noves/hm09hivern/docs/wieland.pdf>
- Wieland, K. (2009b). El llenguatge juvenil català: un refugi per als barbarismes i els estrangerismes? Convergències i divergències en l'ús de préstecs en la varietat diafàsica juvenil i la seva representació als mitjans de comunicació catalans. A: Faluba, K. i Szijj, I. (eds.) *Actes del Catorzè Col·loqui Internacional de Llengua i Literatura Catalanes, Budapest, 2006*, vol. III. Barcelona: Publicacions de l'Abadia de Montserrat, p. 399-412.

- Winford, Donald. *An Introduction to contact linguistics*. Malden (Mass.): Blackwell, 2003.
- Wolfram, W. (2006). Variation and Language: Overview. A: Brown, K. (Ed.). *Encyclopedia of Language & Linguistics*, p. 333-340. <https://doi.org/10.1016/B0-08-044854-2/04256-5>
- Wright, D. (2013). Stylistic variation within genre conventions in the Enron email corpus: developing a text-sensitive methodology for authorship research. *International Journal of Speech, Language and the Law*, vol. 20 (1), p. 45-75. <https://doi.org/10.1558/ijssl.v20i1.45>
- Wright, D. (2014). *Stylistics versus Statistics: A corpus linguistic approach to combining techniques in forensic authorship analysis using Enron emails*. Tesi doctoral inédita. The University of Leeds.
- Zabalegui, N. (2008). La posición de los pronombres átonos en construcciones con verbos no conjugados en el español actual de Caracas. *Akademos*, vol. 10 (2), p. 83-107.
- Zhang, F. i Pengcuo, D. (2020). Comparison of Inner Meanings of Color Terms Between Chinese and English from the Perspective of Sociolinguistics. A: *Advances in Social Science, Education and Humanities Research*, vol. 507, p. 73-79. <https://doi.org/10.2991/assehr.k.201215.307>
- Zuljan Kumar, D. (2019). Pragmatic functions of repetitions in spontaneous spoken dialect discourse. *Dialectologia et Geolinguistica*, vol. 27 (1), p. 123-141. <https://doi.org/10.1515/dialect-2019-0007>

10 ANNEXOS

Annex 1. Informants del corpus d'anàlisi per població, sexe i llengua inicial.

Població	Informant	Sexe	Llengua inicial
Manlleu	L0001	Masculí	Català i castellà
Manlleu	L0002	Femení	Castellà
Manlleu	L0003	Masculí	Castellà
Manlleu	L0004	Femení	Català
Manlleu	L0005	Femení	Català i castellà
Manlleu	L0006	Masculí	Castellà
Manlleu	L0007	Femení	Castellà
Manlleu	L0008	Femení	Castellà
Manlleu	L0009	Masculí	Castellà
Manlleu	L0012	Masculí	Castellà
Manlleu	L0016	Masculí	Castellà
Manlleu	L0018	Femení	Castellà
Manlleu	L0019	Masculí	Castellà
Manlleu	L0020	Masculí	Català
Manlleu	L0021	Masculí	Català
Manlleu	L0025	Masculí	Català
Manlleu	L0027	Masculí	Català
Manlleu	L0031	Femení	Castellà
Manlleu	L0032	Femení	Català i castellà
Manlleu	L0035	Masculí	Català
Manlleu	L0037	Femení	Català
Manlleu	L0038	Masculí	Català
Manlleu	L0043	Masculí	Català
Manlleu	L0046	Femení	Català i castellà
Manlleu	L0047	Masculí	Castellà
Manlleu	L0048	Femení	Català i castellà
Manlleu	L0049	Femení	Català i castellà
Manlleu	L0050	Femení	Català
Manlleu	L0051	Masculí	Català
Manlleu	L0052	Femení	Català
Manlleu	L0057	Masculí	Català
Manlleu	L0058	Masculí	Català
Manlleu	L0060	Femení	Català
Mataró	M0043	Femení	Castellà
Mataró	M0045	Masculí	Català
Mataró	M0048	Masculí	Català
Mataró	M0049	Femení	Castellà
Mataró	M0051	Masculí	Català
Mataró	M0054	Femení	Català
Mataró	M0089	Masculí	Castellà
Mataró	M0090	Masculí	Castellà
Mataró	M0094	Femení	Català
Mataró	M0096	Femení	Castellà
Mataró	M0105	Femení	Castellà
Mataró	M0289	Femení	Castellà

Annexos

Mataró	M0290	Masculí	Castellà
Mataró	M0293	Masculí	Castellà
Mataró	M0295	Femení	Castellà
Mataró	M0297	Femení	Català i castellà
Mataró	M0453	Masculí	Català
Mataró	M0456	Masculí	Castellà
Mataró	M0462	Femení	Català
Mataró	M0469	Femení	Català
Mataró	M0470	Masculí	Català
Mataró	M0471	Femení	Castellà
Mataró	M0472	Femení	Castellà
Mataró	M0718	Masculí	Castellà
Mataró	M0719	Femení	Castellà
Mataró	M0722	Masculí	Castellà
Mataró	M0723	Femení	Català i castellà
Mataró	M0725	Femení	Castellà
Mataró	M0728	Femení	Castellà
Mataró	M0732	Femení	Castellà
Mataró	M0734	Femení	Castellà
Mataró	M0737	Femení	Català i castellà
Mataró	M0914	Femení	Català
Mataró	M0917	Femení	Català
Mataró	M0918	Masculí	Català
Mataró	M0925	Masculí	Català i castellà

Annex 2. Pautes de transcripció.

Tipus de fenomen	Fenomen		Transcripció
Aspectes prosòdics	Allargament	Breu	:
		Mitjà	::
		Llarg	::: (o més)
	Truncament	Mot o grup tonal	-
	Altres fenòmens prosòdics		{(fenomen) text afectat}
	Multiplicitat d'altres fenòmens prosòdics		{(fenomen, fenomen) text afectat}
Aspectes vocals	Riure no simultani amb el text (un símbol per síl·laba)		@
	Riure simultani amb el text		{(@) text afectat}
	Sons paralingüístics	Assentiment	mhm / ahà
		Dubte	m:: / e::
		Desacord	ntx / pf
		Valoració	pse
		Èmfasi	buff
Altres		Formes escrites convencionals	

	Altres fenòmens vocals (somniaient, cantant, tossint...)		{(fenomen) text afectat}
	Multiplicitat d'altres fenòmens vocals		{(fenomen, fenomen) text afectat}
Pauses i encavalcaments	Pausa	Breu	(Pausa curta)
		Mitjana	(Pausa mitjana)
		Llarga	(Pausa llarga)
	Encavalcament		A: [text afectat] B: [text afectat]
Aspectes gestuals o externs a la conversa	Fenòmens no simultanis		((descripció del fenomen))
	Fenòmens simultanis		{(fenomen) text afectat}
Regulacions i comentaris	Aclariments fonètics	Elisió	(so elidit)
		Transcripció fonètica	Text (/text transcrit/)
	Comentaris		((comentari))
	Fragment dirigit a un participant determinat		((a PARTICIPANT))
	Fragments intel·ligibles (una lletra per síl·laba)		((x/xx/xxx))
Fragments incerts		{(??) fragment incert}	
Altres aspectes	Alternances de codi		{(L2) text afectat}

Font: adaptat de Flors-Mas (2017: 610-611).

Annex 3. Realitzacions de V38 al corpus d'anàlisi.

	Freq.	Nr. parlants	Transcripcions	V38
<i>a boleo</i>	1	1	M0045SCI	V38a
<i>acudeix</i>	1	1	M0090SCI	V38d
<i>a la seva bola</i>	5	3	L0027QCI M0054QCI M0054SCI M0089QCI	V38c
<i>a lo bestia</i>	1	1	M0045SCI	V38a
<i>a lo millor</i>	68	22	L0002QCI L0002SCI L0019QCI L0020SCI L0027QCI L0035QCI L0035SCI L0052QCI M0048QCI M0293QCI M0293SCI M0295SCI M0453SCI M0462QCI M0718QCI M0725SCI M0728QCI M0728SCI M0734QCI M0734SCI M0737QCI	V38a
<i>a más</i>	2	2	M0289QCI M0293SCI	V38d
<i>a mí no me gustan</i>	1	1	M0725SCI	V38a
<i>a mi prima le pasó</i>	1	1	M0725SCI	V38a

Annexos

<i>a saco</i>	1	1	L0047QCI	V38a
<i>a ver</i>	13	11	L0003SCI L005QCI M0043SCI M0096QCI M0105QCI M0293SCI M0470SCI M0719SCI M0725QCI M0728SCI M0737SCI	V38a
<i>abultar</i>	1	1	M0290QCI	V38a
<i>acaso</i>	1	1	M0728SCI	V38a
<i>acondicionat -da</i>	2	1	M0297QCI M0297SCI	V38a
<i>aconsejable</i>	1	1	M0737QCI	V38a
<i>acoplar-se</i>	1	1	M0054SCI	V38a
<i>acostumbrar(-se)</i>	8	5	L0035QCI L0047QCI L0047SCI M0289SCI M0290QCI M0293SCI	V38a
<i>aquario</i>	2	2	L0002QCI M0914SCI	V38a
<i>acurrucar-se</i>	3	2	L0048QCI M0295SCI	V38a
<i>además</i>	1	1	M0734SCI	V38a
<i>ademés</i>	6	4	L0046QCI M0051QCI M0453QCI M0723SCI	V38a
<i>adiestrar</i>	8	4	L0008QCI M0290QCI M0453QCI M0734QCI	V38a
<i>adrestrat</i>	1	1	M0734QCI	V38a
<i>agarrar-se</i>	1	1	M0090SCI	V38a
<i>agobiar(-se)</i>	8	5	L0007QCI L0008QCI M0096SCI M0722QCI M0925QCI	V38a
<i>ah no eso ya lo he dicho</i>	1	1	M0456SCI	V38a
<i>ahí</i>	5	4	M0718QCI M0728SCI M0732SCI M0734QCI M0734SCI	V38a
<i>ahorrar</i>	1	1	M0734SCI	V38a
<i>al seu aire</i>	2	2	L005SCI L0019QCI	V38c
<i>al seu rotllo</i>	1	1	M0725QCI	V38c
<i>al tanto</i>	2	1	L0003QCI	V38a
<i>alcance</i>	1	1	M0918QCI	V38a
<i>alfombra</i>	1	1	L0019QCI	V38a
<i>algo</i>	242	84	L0001QCI L0001SCI L0002QCI L0003QCI L0003SCI L0004QCI L0004SCI L0006QCI L0006SCI L0007QCI L0007SCI L0008QCI L0008SCI L0009QCI L0012SCI L0016QCI L0018SCI L0020QCI L0020SCI L0021SCI L0027QCI L0031SCI L0035QCI L0037QCI L0038QCI L0038SCI L0046QCI L0046SCI L0047QCI L0047SCI L0048QCI L0048SCI L0049QCI L0049SCI L0050SCI L0051SCI L0052QCI L0052SCI L0057QCI L0057SCI L0058QCI L0058SCI L0060SCI M0043SCI M0045QCI M0045SCI M0048QCI M0051QCI	V38a

			M0054QCI M0054SCI M0089SCI M0090SCI M0094QCI M0096QCI M0096SCI M0105SCI M0289QCI M0290QCI M0290SCI M0293QCI M0293SCI M0295QCI M0453QCI M0453SCI M0456QCI M0456SCI M0462QCI M0462SCI M0469QCI M0471SCI M0472QCI M0472SCI M0719SCI M0723QCI M0723SCI M0728SCI M0737SCI M0914QCI M0914SCI M0917QCI M0917SCI M0918QCI M0925QCI M0925SCI	
<i>almohada</i>	1	1	L0006SCI	V38a
<i>amaestrat</i>	1	1	L0006SCI	V38a
<i>amastrat</i>	1	1	M0049SCI	V38a
<i>anchura</i>	1	1	M0472SCI	V38a
<i>apalancat</i>	1	1	L0021QCI	V38a
<i>apanyar-se(-les)</i>	2	2	L0048QCI L0058QCI	V38a
<i>aplastar(-se)</i>	2	2	L0058QCI M0089QCI	V38a
<i>apoiar-se</i>	4	4	L0008QCI L0052QCI M0293SCI M0914QCI	V38a
<i>apreci</i>	1	1	L0009QCI	V38a
<i>aranyar</i>	25	13	L0006QCI L0018QCI M0096QCI M0096SCI M0290SCI M0295SCI M0470QCI M0470SCI M0472QCI M0472SCI M0719QCI M0723QCI M0728SCI	V38a
<i>arisc -a</i>	4	3	M0456QCI M0725SCI M0734SCI	V38a
<i>arrampar</i>	1	1	M0089SCI	V38d
<i>así</i>	10	8	M0096SCI M0289QCI M0290SCI M0295QCI M0295SCI M0456SCI M0722SCI M0725QCI	V38a
<i>asins</i>	1	1	M0732SCI	V38a
<i>assentar-se</i>	3	3	L0051SCI M0090SCI M0094SCI	V38a
<i>atar</i>	1	1	M0456QCI	V38a
<i>aunque</i>	3	2	L0003SCI M0295SCI	V38a
<i>balsa</i>	2	2	L0006SCI L0009SCI	V38a
<i>barato</i>	18	14	L0003QCI L0020SCI L0025QCI L0027QCI L0038SCI M0049SCI M0054QCI M0054SCI M0094QCI M0094SCI M0293QCI M0462QCI M0918QCI M0918SCI	V38a
<i>barreny</i>	1	1	M0725SCI	V38a
<i>basura</i>	3	2	L0008QCI L0048QCI	V38a
<i>biberón</i>	1	1	M0089SCI	V38a
<i>bitxo</i>	23	21	L0003SCI L005SCI L0018SCI L0021SCI L0027SCI L0037QCI L0037SCI L0048SCI L0058QCI M049QCI M0049SCI M0051SCI	V38a

Annexos

			M0094QCI M0293SCI M0472QCI M0472SCI M0723QCI M0914QCI M0914SCI M0925SCI	
<i>bombilla</i>	1	1	M0290SCI	V38a
<i>bones</i>	2	2	L0003QCI L0046QCI	V38c
<i>bou</i>	3	1	L0020QCI L0020SCI	V38c
<i>bueno</i>	2106	69	L0001QCI L0001SCI L0002QCI L0002SCI L0003QCI L0003SCI L0004QCI L0004SCI L0005QCI L0005SCI L0006QCI L0006SCI L0007QCI L0007SCI L0008QCI L0008SCI L0009QCI L0009SCI L0012QCI L0012SCI L0016QCI L0016SCI L0018QCI L0018SCI L0019QCI L0019SCI L0020QCI L0020SCI L0021QCI L0021SCI L0025QCI L0025SCI L0027QCI L0027SCI L0031QCI L0031SCI L0032QCI L0032SCI L0035QCI L0035SCI L0037QCI L0038QCI L0038SCI L0043QCI L0043SCI L0046QCI L0046SCI L0047QCI L0047SCI L0048QCI L0048SCI L0049QCI L0049SCI L0050QCI L0050SCI L0051QCI L0051SCI L0052QCI L0052SCI L0057QCI L0057SCI L0058QCI L0058SCI L0060QCI L0060SCI M0043QCI M0043SCI M0045QCI M0045SCI M0048QCI M0048SCI M0049QCI M0051QCI M0051SCI M0054QCI M0054SCI M0089QCI M0089SCI M0090QCI M0094QCI M0094SCI M0096QCI M0096SCI M0105QCI M0289QCI M0289SCI M0290QCI M0293QCI M0293SCI M0295QCI M0295SCI M0297QCI M0297SCI M0453QCI M0453SCI M0456QCI M0456SCI M0462QCI M0462SCI M0469QCI M0469SCI M0470QCI M0470SCI M0471QCI M0471SCI M0472QCI M0472SCI M0718QCI M0719QCI M0722QCI M0722SCI M0723SCI M0725QCI M0725SCI M0728QCI M0732QCI M0732SCI M0734QCI M0734SCI M0737QCI M0914QCI M0914SCI M0917QCI M0917SCI M0918QCI M0918SCI M0925QCI M0925SCI	V38a
<i>bufar-se</i>	2	1	M0719SCI	V38c
<i>búho</i>	13	5	L0049QCI M0054QCI M0290QCI M0722SCI M0728QCI	V38a
<i>calentar</i>	1	1	M0456SCI	V38a

<i>cama</i>	1	1	M0289SCI	V38a
<i>camada</i>	1	1	M049QCI	V38a
<i>cànvio, etc.</i>	24	18	L0003SCI L0007QCI L0009SCI L0025QCI L0032SCI L0048QCI L0051QCI L0052QCI L0052SCI L0057QCI L0058SCI L0060QCI M0045QCI M0045SCI M0048QCI M0089QCI M0718QCI M0723SCI M0737SCI M0925SCI	V38c
<i>capcera</i>	1	1	M0096SCI	V38b
<i>carantonyes</i>	2	2	M0051QCI M0722SCI	V38a
<i>cargar-se-les</i>	1	1	L0058SCI	V38a
<i>cariciar</i>	1	1	M0732SCI	V38a
<i>cariny</i>	1	1	M0096SCI	V38a
<i>carinyo</i>	28	17	L0004SCI L0007QCI L0027QCI L0046SCI L0049QCI L0049SCI L0057QCI L0060QCI M0049SCI M0051QCI M0051SCI M0054QCI M0094QCI M0295SCI M0462QCI M0723SCI M0723QCI M0737SCI	V38a
<i>carinyós -a</i>	59	29	L0001SCI L0003SCI L0004QCI L0004SCI L0008QCI L0008SCI L0009SCI L0018QCI L0021QCI L0025QCI L0027QCI L0046QCI L0046SCI L0047QCI L0048QCI L0049QCI L0049SCI L0050QCI L0050SCI L0051SCI L0052SCI L0057SCI L0058QCI L0058SCI M0048SCI M0049QCI M0049SCI M0051QCI M0051SCI M0054QCI M0295SCI M0472QCI M0725SCI M0725QCI M0728SCI M0732SCI M0734QCI M0737SCI	V38a
<i>casi</i>	13	11	L0027QCI L0032QCI L0035QCI L0047SCI L0049QCI L0049SCI L0050SCI L0051SCI M0094SCI M0105SCI M0925QCI	V38a
<i>casibé</i>	1	1	M0918QCI	V38b
<i>celos</i>	2	1	M0719SCI	V38a
<i>celosos</i>	1	1	M0290SCI	V38a
<i>centímetros</i>	9	4	L0020SCI L0020QCI L0047SCI L0057QCI M0723SCI	V38a
<i>cerrat</i>	1	1	M0293SCI	V38a
<i>claro</i>	1	1	M0456QCI	V38a
<i>comeder</i>	1	1	M0734SCI	V38a
<i>comedor</i>	1	1	M0722SCI	V38a
<i>comida</i>	2	2	M0290QCI M0722SCI	V38a
<i>como</i>	1	1	M0734QCI	V38a
<i>¿cómo cómo?</i>	1	1	M0290SCI	V38a

Annexos

<i>como que no</i>	1	1	L005QCI	V38a
<i>cómo se dice</i>	1	1	M0105SCI	V38a
<i>cómo se llama esto</i>	1	1	L0006QCI	V38a
<i>còmodo</i>	2	2	L0004SCI L0018QCI	V38a
<i>contra</i>	2	1	L0031QCI	V38c
<i>convenir (convendria, convengui)</i>	2	2	L0047SCI M0925SCI	V38b
<i>correa</i>	1	1	M0719SCI	V38a
<i>corredera</i>	1	1	L0047QCI	V38a
<i>corretejar</i>	2	1	L0047SCI	V38b
<i>cortar</i>	1	1	M0734SCI	V38a
<i>creo</i>	1	1	M0469QCI	V38a
<i>cresqui</i>	1	1	M0043QCI	V38b
<i>criadero</i>	1	1	M0051QCI	V38a
<i>crystal</i>	5	3	M0293QCI M0472SCI M0734SCI	V38a
<i>cualquier</i>	1	1	M0723QCI	V38a
<i>cuarenta y dos</i>	1	1	M0722SCI	V38a
<i>cuarenta-dos</i>	1	1	M0725SCI	V38b
<i>cucaracha</i>	22	20	L0001SCI L0003QCI L0003SCI L0006SCI L0007SCI L0020SCI L0046QCI L0049QCI M0045SCI M0048QCI M0051SCI M0089QCI M0096QCI M0105QCI M0289SCI M0456QCI M0456SCI M0472SCI M0732QCI M0918QCI	V38a
<i>cuelga (subst.)</i>	1	1	M0090SCI	V38a
<i>cuida(d)o</i>	23	15	L0001QCI L0027QCI L0031QCI L0049QCI L0058QCI M0045SCI M0049SCI M0051QCI M0051SCI M0105QCI M0456SCI M0462QCI M0471SCI M0472SCI M0728QCI M0737QCI M0918QCI	V38a
<i>cuidadós -a</i>	1	1	L0007SCI	V38b
<i>cuidats (subst.)</i>	5	3	L0046SCI M0289QCI M0295QCI M0295SCI	V38a
<i>dale</i>	1	1	M0918QCI	V38a
<i>dar-lo mimos</i>	1	1	M0290SCI	V38a
<i>de esto -a</i>	4	2	M0096SCI M0734QCI	V38a
<i>demás</i>	1	1	M0718QCI	V38a
<i>demés</i>	6	6	L0052QCI M0094QCI M0734SCI M0737QCI M0918QCI M0925SCI	V38a
<i>desmontar</i>	3	3	L0031SCI L0037SCI M0914SCI	V38a
<i>después</i>	29	6	L005QCI L005SCI L0046SCI M0462QCI M0462SCI M0470SCI	V38a
<i>desvantatge</i>	3	2	L0025QCI M0054SCI	V38a
<i>disfrutar</i>	1	1	L0003QCI	V38a
<i>donar igual</i>	4	3	M0456SCI M0722QCI M0732QCI	V38c

<i>donar-se compte</i>	1	1	M0737SCI	V38c
<i>durader</i>	1	1	M0051SCI	V38a
<i>en castellano</i>	1	1	M0722SCI	V38a
<i>en sèrio</i>	2	2	L0019QCI M0725SCI	V38a
<i>en ves de</i>	2	2	L0047QCI L0058SCI	V38a
<i>encarinyar-se</i>	3	2	L005QCI M0718QCI	V38a
<i>encuentro</i>	1	1	M0725SCI	V38a
<i>enfermetat</i>	10	8	L0009QCI L0018SCI L0027QCI M0048QCI M0048SCI M0051SCI M0462QCI M0718SCI M0719QCI	V38a
<i>enfrentar-se</i>	1	1	M0051QCI	V38a
<i>enganche (subst.)</i>	1	1	L0047QCI	V38a
<i>en la jaula</i>	2	1	M105SCI	V38a
<i>entonces</i>	1	1	M0734QCI	V38a
<i>equis</i>	2	1	L0007QCI	V38a
<i>escarabaix</i>	2	1	M0105SCI	V38a
<i>escarabajo</i>	1	1	M0043QCI	V38a
<i>escondre's</i>	1	1	M0290SCI	V38a
<i>escurridís -a</i>	2	1	M0295SCI	V38a
<i>eso</i>	1	1	L0048QCI	V38a
<i>espaciós -a</i>	1	1	L0047SCI	V38a
<i>esta</i>	2	2	M0295SCI M0456QCI	V38a
<i>esté</i>	2	2	M0290SCI M0732SCI	V38a
<i>esto</i>	2	2	M0722SCI M0734QCI	V38a
<i>esto es de verdad, ¿eh?</i>	1	1	M0725SCI	V38a
<i>expositori</i>	2	1	L0025QCI	V38d
<i>extrany</i>	7	7	L0049SCI M0089QCI M0295SCI M0723QCI M0734SCI M0925QCI M0925SCI	V38a
<i>fiel</i>	1	1	L0004SCI	V38a
<i>financiació</i>	1	1	L0019QCI	V38a
<i>foco</i>	1	1	M0737SCI	V38a
<i>fruts</i>	6	6	L0001SCI L0006QCI L0006SCI M0290QCI M0295SCI M0719SCI	V38a
<i>gasto (subst.)</i>	2	1	M0925QCI	V38a
<i>gordo -a</i>	4	4	L0001SCI M0048QCI M0048SCI M0096SCI	V38a
<i>grillo</i>	2	2	M0290QCI M0723SCI	V38a
<i>guai</i>	1	1	M049QCI	V38a
<i>a gusta</i>	1	1	M0090SCI	V38a
<i>hasta</i>	5	3	L0027QCI L0051SCI M0293QCI	V38a
<i>hembra</i>	1	1	M0925SCI	V38a
<i>hinbernació</i>	1	1	M0215SCI	V38a
<i>hinbernader</i>	1	1	L0051QCI	V38a
<i>hinbernar</i>	1	1	M0295SCI	V38a

Annexos

<i>hobby</i>	1	1	M0462QCI	V38a
<i>hombre</i>	11	4	L0019QCI M0456SCI M0469QCI M0734SCI	V38a
<i>hombro</i>	1	1	M0728QCI	V38a
<i>hormiga</i>	1	1	M0456QCI	V38a
<i>incluït -da</i>	11	11	L0006SCI L0007QCI L0012SCI L0019QCI L0032SCI L0035QCI M0045SCI M0051SCI M0470SCI M0734SCI M0918SCI	V38a
<i>independent</i>	2	2	L0008QCI M0723QCI	V38a
<i>jalar</i>	1	1	M0045SCI	V38a
<i>jaula</i>	51	18	L0002QCI L0003SCI L0006SCI L0007QCI L0009QCI L0009SCI L0020SCI M0051SCI M0090SCI M0105SCI M0290QCI M0290SCI M0293SCI M0456QCI M0456SCI M0471SCI M0722SCI M0728SCI M0732SCI M0734QCI M0734SCI M0918QCI	V38a
<i>jo(d)er</i>	1	1	M0917QCI	V38a
<i>ju(e)guetó</i>	4	3	L0007QCI M0289SCI M0718SCI	V38a
<i>justar</i>	1	1	M0293SCI	V38a
<i>lagart</i>	1	1	M0295SCI	V38a
<i>lagartija</i>	4	2	L0020QCI M0290QCI M0290SCI	V38a
<i>lagarto</i>	1	1	L0031SCI	V38a
<i>làmpara</i>	2	2	M0295QCI M0295SCI	V38a
<i>lana</i>	2	1	M0290SCI	V38a
<i>lata</i>	1	1	M0725SCI	V38a
<i>lavar</i>	1	1	M0719SCI	V38a
<i>lechuga</i>	2	1	L005QCI L005SCI	V38a
<i>lechuza</i>	1	1	M0049SCI	V38a
<i>limar</i>	1	1	M0722SCI	V38a
<i>llagart</i>	1	1	M0293QCI	V38a
<i>llagartana</i>	1	1	M0090SCI	V38b
<i>llamatiu/-va</i>	1	1	M0295SCI	V38a
<i>llechuza</i>	1	1	L0018QCI	V38a
<i>llimpiar</i>	2	2	M0289SCI M0293QCI	V38a
<i>lo</i>	303	57	L0003QCI L0003SCI L0005QCI L0006QCI L0007SCI L0008QCI L0009QCI L0012QCI L0016QCI L0018QCI L0018SCI L0019QCI L0019SCI L0020QCI L0020SCI L0021QCI L0021SCI L0025QCI L0025SCI L0027QCI L0031QCI L0031SCI L0032QCI L0032SCI L0035SCI L0038QCI L0038SCI L0046QCI L0047QCI L0047SCI L0048QCI L0049QCI L0050QCI L0050SCI L0051QCI L0051SCI	V38a

			L0052QCI L0052SCI L0057QCI L0057SCI L0058QCI L0058SCI L0060QCI L0060SCI M0043QCI M0045QCI M0045SCI M0048QCI M0048SCI M0049SCI M0051QCI M0051SCI M0054QCI M0054SCI M0090SCI M0094QCI M0094SCI M0096QCI M0096SCI M0289QCI M0293QCI M0293SCI M0295SCI M0297SCI M0453QCI M0453SCI M0456QCI M0456SCI M0462QCI M0469QCI M0469SCI M0470SCI M0471SCI M0472QCI M0472SCI M0723QCI M0725SCI M0728QCI M0728SCI M0734QCI M0734SCI M0737SCI M0914SCI M0917QCI M0918QCI M0918SCI M0925QCI M0925SCI	
<i>lòros</i>	1	1	L0018SCI	V38a
<i>los grillos</i>	1	1	M0725QCI	V38a
<i>maltratar</i>	1	1	M0728QCI	V38a
<i>manguera</i>	1	1	L0032SCI	V38a
<i>manso</i>	2	2	M0045SCI M0297QCI	V38a
<i>mariposa</i>	1	1	M0293SCI	V38a
<i>más</i>	1	1	M0295QCI	V38a
<i>mascota</i>	17	13	L0025QCI L0027QCI L0032QCI L0038QCI L0043SCI L0052QCI L0058QCI M0090SCI M0105SCI M0456QCI M0722QCI M0722SCI M0917QCI	V38c
<i>maullar</i>	3	1	M0453SCI	V38a
<i>medir</i>	91	27	L0001SCI L0002QCI L0004QCI L0008SCI L0016QCI L0016SCI L0018SCI L0019SCI L0020QCI L0020SCI L0021QCI L0021SCI L0027QCI L0037SCI L0043QCI L0043SCI L0047SCI L0048SCI M0090QCI M0090SCI M0094QCI M0094SCI M0290SCI M0295SCI M0453QCI M0453SCI M0462SCI M0470QCI M0470SCI M0471SCI M0472SCI M0719SCI M0725SCI M0732SCI M0734QCI	V38a
<i>mellor</i>	1	1	M0295SCI	V38a
<i>menos</i>	14	11	L0005SCI L0006SCI L0020SCI L0027SCI L0048QCI L0049QCI L0058SCI M0045QCI M0045SCI M0054SCI M0094SCI	V38a
<i>mentres</i>	37	17	L0007SCI L0021QCI L0025SCI L0035QCI L0047QCI L0047SCI L0051SCI L0058QCI M0043QCI	V38a

Annexos

			M0045SCI M0048QCI M0089QCI M0096QCI M0096SCI M0105QCI M0290QCI M0456QCI M0456SCI M0462QCI M0470SCI	
<i>mesa</i>	1	1	M0090SCI	V38a
<i>metro</i>	3	2	L0020SCI L0058SCI	V38a
<i>mideix(en)</i>	24	6	M0293QCI M0295SCI M0719QCI M0719SCI M0728QCI M0728SCI M0732SCI M0734SCI	V38a
<i>mimar</i>	5	4	L0005QCI L0005SCI M0722SCI M0728SCI M0734SCI	V38a
<i>mimat</i>	1	1	M0051SCI	V38a
<i>mimos</i>	4	4	L0057QCI M0469SCI M0719SCI M0734SCI	V38a
<i>mims</i>	1	1	M0734SCI	V38a
<i>modo</i>	1	1	M0295SCI	V38a
<i>mono</i>	9	8	L0048QCI L0050QCI L0051SCI M049QCI M0453SCI M0470SCI M0472SCI M0737QCI	V38a
<i>montar</i>	1	1	M0914QCI	V38a
<i>morder (mordi, morden)</i>	2	2	M0456QCI M0734QCI	V38a
<i>murciélago</i>	1	1	M0090SCI	V38a
<i>ningun -a</i>	8	6	M0090SCI M0293QCI M0293SCI M0456SCI M0719QCI M0734SCI M0737SCI	V38a
<i>no sé cómo se dice</i>	3	1	M0722SCI	V38a
<i>no sé cómo se llama</i>	4	3	M0105QCI M0290SCI M0722SCI	V38a
<i>no sé qué no sé cuántos</i>	1	1	M0045SCI	V38a
<i>nombre</i>	1	1	M0290SCI	V38a
<i>nòria</i>	2	2	L0004SCI M0914QCI	V38a
<i>o sea</i>	48	17	M0045SCI M0096SCI M0105QCI M0289QCI M0289SCI M0290QCI M0295QCI M0456QCI M0462QCI M0719SCI M0722QCI M0725SCI M0728QCI M0728SCI M0734QCI M0914QCI M0925QCI	V38a
<i>okay</i>	1	1	L0035QCI	V38a
<i>olorar</i>	4	3	M0456QCI M0456SCI M0725SCI M0295QCI	V38c
<i>oruga</i>	1	1	M0453QCI	V38a
<i>oscuritat</i>	1	1	M0096SCI	V38a
<i>pa(ra) abajo</i>	1	1	M0456QCI	V38a
<i>pack</i>	5	5	L0006QCI L0012QCI L0058QCI M0090SCI M0925SCI	V38a
<i>palomitas</i>	1	1	M0105QCI	V38a
<i>par</i>	2	2	M0456QCI M0718QCI	V38c

<i>para</i>	1	1	M0049SCI	V38a
<i>patada</i>	1	1	M0295QCI	V38a
<i>pecera</i>	2	2	M0105QCI M0725QCI	V38a
<i>peinar</i>	1	1	M0722SCI	V38a
<i>peligrós -a</i>	1	1	M0719SCI	V38a
<i>peluche</i>	1	1	M0472SCI	
<i>peluquera</i>	2	1	M0462QCI M0462SCI	V38a
<i>peluqueria</i>	3	2	L0008SCI M0462QCI M0462SCI	V38a
<i>peluix</i>	1	1	M0094SCI	V38a
<i>pepites</i>	1	1	M0054QCI	V38a
<i>per lo tant</i>	7	1	M0295QCI M0295SCI	
<i>pescaderia</i>	2	2	M0453SCI M0456QCI	V38a
<i>pescat</i>	1	1	M0456QCI	V38a
<i>pessunyes</i>	1	1	M0728SCI	V38a
<i>pico</i>	1	1	L0016SCI	V38a
<i>pienso (subst.)</i>	10	6	M0051QCI M0289SCI M0293QCI M0295SCI M0456SCI M0737QCI	V38a
<i>pins</i>	1	1	L0006QCI	V38d
<i>pipes</i>	81	50	L0002QCI L0004SCI L0007QCI L0007SCI L0008SCI L0008QCI L0009SCI L0016QCI L0018QCI L0018SCI L0019SCI L0020QCI L0020SCI L0021QCI L0021SCI L0025QCI L0025SCI L0027QCI L0032QCI L0032SCI L0035QCI L0037QCI L0037SCI L0038QCI L0038SCI L0043QCI L0043SCI L0046QCI L0046SCI L0048SCI L0050SCI L0051QCI L0051SCI L0052QCI L0057QCI L0058QCI L0060QCI M0043QCI M0045SCI M0051QCI M0051SCI M0054QCI M0094QCI M0094SCI M0105QCI M0105SCI M0290SCI M0293QCI M0293SCI M0295SCI M0453SCI M0462SCI M0469QCI M0470QCI M0470SCI M0471SCI M0472QCI M0472SCI M0718QCI M0719QCI M0722QCI M0722SCI M0723SCI M0728QCI M0728SCI M0732QCI M0732SCI M0734QCI M0734SCI M0737SCI M0918QCI M0925QCI M0925SCI	V38a
<i>pisto</i>	1	1	M0734SCI	V38a
<i>plan</i>	6	3	L0046QCI L0057QCI L0058QCI L0058SCI M0456QCI	V38a
<i>pluma</i>	2	2	L0058QCI M0094SCI	V38a
<i>popó</i>	1	1	L0027SCI	V38a
<i>pos</i>	122	46	L0001QCI L0003QCI L0004SCI	V38a

Annexos

			L0005QCI L0006QCI L0008QCI L0009SCI L0019SCI L0020QCI L0031QCI L0037QCI L0046QCI L0049QCI L0052QCI M0045QCI M0045SCI M0048QCI M0048SCI M0049QCI M0051QCI M0054QCI M0094QCI M0096QCI M0096SCI M0105QCI M0105SCI M0290SCI M0293QCI M0293SCI M0295QCI M0295SCI M0297QCI M0456QCI M0456SCI M0462QCI M0470QCI M0470SCI M0472SCI M0719QCI M0722SCI M0723QCI M0723SCI M0728SCI M0734SCI M0737SCI M0914SCI	
<i>propio</i>	1	1	M0719SCI	V38a
<i>pues</i>	486	52	L0003QCI L0003SCI L0004SCI L0005QCI L0006SCI L0007QCI L0008QCI L0009QCI L0009SCI L0012SCI L0016QCI L0019QCI L0020SCI L0021QCI L0021SCI L0027QCI L0027SCI L0031QCI L0031SCI L0032QCI L0032SCI L0035QCI L0046SCI L0048QCI L0048SCI L0049SCI L0050QCI L0051QCI L0057QCI L0057SCI L0058QCI M0043QCI M0043SCI M0045QCI M045SCI M048QCI M0048SCI M0049SCI M0051SCI M0054QCI M0054SCI M0089QCI M0089SCI M0090SCI M0094QCI M0094SCI M0096QCI M0096SCI M0105QCI M0105SCI M0290QCI M0293QCI M0293SCI M0297QCI M0453SCI M0456QCI M0456SCI M0462QCI M0462SCI M0469SCI M0470QCI M0470SCI M0471QCI M0471SCI M0472SCI M0719QCI M0723QCI M0723SCI M0725SCI M0728SCI M0732QCI M0732SCI M0914SCI M0917SCI M0925QCI M0925SCI	V38a
<i>puesto (subst.)</i>	2	1	M0051SCI	V38a
<i>qué dice</i>	1	1	M0297QCI	V38a
<i>que pillaría a malas no sé</i>	1	1	M0725SCI	V38a
<i>qué tonto</i>	1	1	L0012SCI	V38a
<i>què va</i>	4	4	M0725SCI L0047QCI L0060QCI M0734QCI	V38d
<i>quicos</i>	16	14	L0008QCI L0012SCI L0016SCI L0018SCI L0019SCI L0020QCI L0043SCI L0048SCI M0045QCI M0045SCI M0462QCI M0462SCI	V38a

			M0469QCI M0470SCI M0472QCI M0917SCI	
<i>quitar</i>	1	1	M0293SCI	V38a
<i>rajar</i>	1	1	M0048QCI	V38a
<i>rama</i>	1	1	M0718QCI	V38a
<i>rana</i>	8	5	M0718QCI L0009SCI M0048QCI M0456QCI M0728QCI	V38a
<i>ranconet</i>	1	1	M0734QCI	V38b
<i>raro -s</i>	11	7	L0016QCI L0018QCI L0031SCI L0058QCI M0054SCI M0472SCI M0723SCI	V38a
<i>rasgat -da</i>	1	1	M0737QCI	V38a
<i>raspa</i>	1	1	M0290SCI	V38a
<i>rato</i>	20	15	L0005SCI L0020SCI L0027SCI L0035SCI L0037QCI L0052QCI L0057QCI L0058QCI M0453QCI M0453SCI M0456QCI M0462SCI M0725SCI M0918SCI M0925SCI	V38a
<i>ratonet</i>	1	1	M0734QCI	V38b
<i>raza</i>	1	1	M0723SCI	V38a
<i>reclosit</i>	1	1	M0051SCI	V38c
<i>reganyar</i>	6	3	M0293SCI M0719QCI M0725QCI	V38a
<i>resto (subst.)</i>	1	1	M0453SCI	V38a
<i>retxontxón -a</i>	1	1	M0048SCI	V38a
<i>rincó</i>	1	1	M0090SCI	V38a
<i>rociar</i>	1	1	M0734SCI	V38a
<i>roedor</i>	1	1	L0027SCI	V38a
<i>ruidós -a</i>	1	1	M049QCI	V38a
<i>s'entreteniria</i>	1	1	M0295SCI	V38d
<i>sacar</i>	12	6	M0090SCI M0289SCI M0293QCI M0728QCI M0732QCI M0734QCI	V38a
<i>salita</i>	1	1	M0728SCI	V38a
<i>saltamontens</i>	1	1	M0725SCI	V38a
<i>saltamontes</i>	28	19	L0009SCI L0016QCI L0018QCI L0018SCI L0020QCI L0020SCI L0031SCI L0043QCI L0052QCI M0043QCI M0045SCI M0048QCI M0054SCI M0290SCI M0462SCI M0718QCI M0723QCI M0732SCI M0914SCI M0917QCI M0918SCI	V38a
<i>saltamuntanyes</i>	1	1	L0025QCI	V38d
<i>sanguiuella</i>	2	1	M0456QCI	V38a
<i>sardinita</i>	1	1	M0295SCI	V38a
<i>secar</i>	1	1	M0043SCI	V38a
<i>sempre i quan</i>	1	1	M0295QCI	V38c
<i>sentar-se</i>	1	1	L0021QCI	V38a
<i>serrín</i>	3	3	L0005QCI L0049QCI M0722SCI	V38a

Annexos

<i>sesenta</i>	2	2	M0293QCI M0722SCI	V38a
<i>sillita</i>	1	1	L0009SCI	V38a
<i>silló</i>	1	1	M0719SCI	V38a
<i>sobat -da</i>	1	1	L0027QCI	V38a
<i>sobrí</i>	2	1	M0090SCI	V38a
<i>soletat</i>	1	1	L0057QCI	V38a
<i>soltar</i>	3	2	M0295SCI M0734SCI	V38a
<i>su</i>	2	1	M0719SCI	V38a
<i>subir-se</i>	2	1	M0734SCI	V38a
<i>suelto -a</i>	21	11	L0016SCI L0019SCI L0027QCI L0027SCI L0047QCI M0048QCI M0054QCI M0462QCI M0723QCI M0723SCI M0732QCI M0737QCI M0914QCI	V38a
<i>tamany</i>	18	10	L0001QCI L0001SCI L0047QCI L0057QCI L0058SCI L0060QCI M0049QCI M0049SCI M0295QCI M0456SCI M0723QCI M0918QCI	V38a
<i>tela</i>	1	1	L0060QCI	V38c
<i>tendria, tendries, tendrà, tendràs</i>	13	9	M0048SCI M0290SCI M0293SCI M0295SCI M0722SCI M0732QCI M0734SCI M0737SCI M0918SCI	V38c
<i>tetabrick</i>	2	1	M0293SCI	V38a
<i>tiburó</i>	1	1	M0462QCI	V38a
<i>tipo</i>	49	22	L0006SCI L0007QCI L0009QCI L0027QCI L0031QCI L0035QCI L0035SCI L0037SCI L0038SCI L0046QCI L0049QCI L0051QCI L0052QCI M0045QCI M0048QCI M0048SCI M0049QCI M0049SCI M0054SCI M0089QCI M0293QCI M0295QCI M0462SCI M0914QCI M0918SCI	V38a
<i>tíque</i>	1	1	M0925QCI	V38a
<i>tíquet</i>	2	2	L0007QCI M0925QCI	V38a
<i>toallita</i>	1	1	M0453SCI	V38a
<i>tomar</i>	1	1	M0293SCI	V38a
<i>tono</i>	1	1	L0003QCI	V38a
<i>tonteria</i>	3	3	L0046QCI L0047QCI M0295SCI	V38a
<i>tonto -a</i>	1	1	L0058SCI	V38a
<i>transportín</i>	2	2	M0051QCI M0725QCI	V38a
<i>tratar</i>	1	1	M0096SCI	V38a
<i>trepar</i>	1	1	M0295SCI	V38a
<i>tubo</i>	4	4	L0021SCI L0025SCI L0048SCI M0723SCI	V38a
<i>tupper</i>	2	2	M0293SCI M0469SCI	V38a
<i>xulo</i>	1	1	L0025SCI	V38a
<i>txutxe</i>	11	5	L0004SCI L0006QCI L0016SCI	V38a

			M0051SCI M0472SCI	
<i>vacil·lar</i>	1	1	L0019SCI	V38c
<i>vago -a</i>	1	1	L0038QCI	V38a
<i>val</i>	172	50	L0001SCI L0002QCI L0004SCI L0005QCI L0005SCI L0006SCI L0008SCI L0012SCI L0018SCI L0019SCI L0025QCI L0027SCI L0031QCI L0031SCI L0032QCI L0035QCI L0035SCI L0037SCI L0038SCI L0043QCI L0043SCI L0046QCI L0046SCI L0047SCI L0048SCI L0049QCI L0049SCI L0050SCI L0051QCI L0051SCI L0052QCI L0052SCI L0058QCI L0058SCI L0060SCI M0043SCI M0045SCI M0049QCI M0049SCI M0051QCI M0051SCI M0089SCI M0090QCI M0096QCI M0096SCI M0289SCI M0290QCI M0293QCI M0293SCI M0295SCI M0297SCI M0453SCI M0456QCI M0456SCI M0462QCI M0469QCI M0472SCI M0722QCI M0722SCI M0728SCI M0734QCI M0734SCI M0737QCI M0917QCI M0917SCI M0925QCI M0925SCI	V38a
<i>vale</i>	635	66	L0001QCI L0002QCI L0002SCI L0003QCI L0003SCI L0004QCI L0004SCI L0005QCI L0005SCI L0006QCI L0006SCI L0007QCI L0007SCI L0008QCI L0008SCI L0009QCI L0009SCI L0012QCI L0012SCI L0016QCI L0016SCI L0018QCI L0018SCI L0019QCI L0019SCI L0020SCI L0021QCI L0021SCI L0027QCI L0027SCI L0031QCI L0032QCI L0032SCI L0035QCI L0035SCI L0038QCI L0038SCI L0043QCI L0043SCI L0046QCI L0046SCI L0047SCI L0048SCI L0049QCI L0049SCI L0050QCI L0050SCI L0051QCI L0051SCI L0052QCI L0052SCI L0057QCI L0057SCI L0058QCI L0058SCI L0060QCI L0060SCI M0043QCI M0043SCI M0045QCI M0045SCI M0048SCI M0049QCI M0049SCI M0051QCI M0051SCI M0054QCI M0054SCI M0089QCI M0089SCI M0090QCI M0090SCI M0094QCI M0094SCI M0096QCI M0096SCI M0105QCI M0105SCI M0289QCI M0289SCI M0290QCI	V38a

Annexos

			M0290SCI M0293QCI M0293SCI M0295QCI M0295SCI M0297QCI M0297SCI M0453QCI M0453SCI M0456QCI M0456SCI M0471QCI M0471SCI M0462QCI M0462SCI M0469QCI M0470QCI M0470SCI M0472QCI M0472SCI M0719QCI M0722SCI M0723QCI M0723SCI M0725QCI M0725SCI M0728QCI M0728SCI M0732QCI M0732SCI M0734QCI M0734SCI M0737QCI M0737SCI M0914QCI M0914SCI M0917QCI M0917SCI M0918QCI M0918SCI M0925QCI M0925SCI	
<i>valen (d'acord')</i>	1	1	L0043QCI	V38a
<i>valla</i>	2	2	L0027SCI M0918SCI	V38a
<i>vamos allá</i>	1	1	M0049QCI	V38a
<i>vamos</i>	3	1	M0472SCI	V38a
<i>vantatge</i>	3	3	L0025QCI M0054SCI M0737SCI	V38a
<i>vàrios</i>	2	2	L0016SCI L0031QCI	V38a
<i>vaya</i>	7	6	L0003SCI L0047SCI M0051QCI M0289SCI M0453SCI M0734QCI	V38a
<i>venden</i>	1	1	M0725SCI	V38a
<i>venga</i>	2	2	L0051SCI M0456QCI	V38a
<i>venta</i>	1	1	M0054SCI	V38a
<i>verges (reixes')</i>	1	1	M0734SCI	V38a
<i>vidriera</i>	3	3	L0007QCI M0737QCI M0925SCI	V38a
<i>viruta</i>	2	1	L0008QCI	V38a
<i>vistobueno</i>	1	1	M0054QCI	V38a
<i>vive en una jaula</i>	1	1	M0105SCI	V38a
<i>vivir (viveix, veveix)</i>	4	1	M0090SCI	V38d
<i>voltereta</i>	2	1	M0045SCI	V38a
<i>volvería</i>	1	1	M0293SCI	V38a
<i>xarca</i>	3	3	M0295QCI M0289SCI M0718QCI	V38a
<i>xupar</i>	1	1	L0005SCI	V38a
<i>y esas cosas</i>	1	1	M0290QCI	V38a
<i>y eso</i>	9	4	L0001SCI L0048QCI L0049QCI M0105QCI	V38a

Annex 4. Estudi estadístic exploratori.

Annex 4.1. Estudi estadístic exploratori de V3B (sexe=masculí, edat=12 anys).

	Edat 12 anys Sexe Masculí L1 Català						Edat 12 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>a ver</i>	,0	,00	.	.	,00	,00	,1	,00	.	.	,00	1,00
<i>a veure</i>	,3	,00	.	.	,00	2,00	,1	,00	.	.	,00	2,00
<i>a ve(u)re</i>	,3	,00	.	.	,00	2,00	,1	,00	.	.	,00	1,00
<i>aviam</i>	,3	,00	.	.	,00	3,00	,0	,00	.	.	,00	,00
	Edat 12 anys Sexe Masculí L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>a ver</i>	,0	,00	.	.	,00	,00						
<i>a veure</i>	,0	,00	.	.	,00	,00						
<i>a ve(u)re</i>	,0	,00	.	.	,00	,00						
<i>aviam</i>	,0	,00	.	.	,00	,00						

Annexos

Annex 4.2. Estudi estadístic exploratori de V3B (sexe=femení, edat=12 anys).

	Edat 12 anys Sexe Femení L1 Català						Edat 12 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>a ver</i>	,1	,00	.	.	,00	1,00	,2	,00	.	.	,00	2,00
<i>a veure</i>	,5	,00	.	.	,00	2,00	,3	,00	.	.	,00	2,00
<i>a ve(u)re</i>	,7	,00	.	.	,00	5,00	1,2	,00	.	.	,00	14,00
<i>aviam</i>	,0	,00	.	.	,00	,00	,0	,00	.	.	,00	,00

	Edat 12 anys Sexe Femení L1 Català i castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>a ver</i>	,1	,00	.	.	,00	1,00
<i>a veure</i>	,0	,00	.	.	,00	,00
<i>a ve(u)re</i>	,9	,50	,00	2,00	,00	3,00
<i>aviam</i>	,1	,00	.	.	,00	1,00

Annex 4.3. Estudi estadístic exploratori de V3B (sexe=masculí, edat=16 anys).

	Edat 16 anys Sexe Masculí L1 Català						Edat 16 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>a ver</i>	,0	,00	.	.	,00	,00	,0	,00	.	.	,00	,00
<i>a veure</i>	,3	,00	.	.	,00	2,00	,2	,00	.	.	,00	1,00
<i>a ve(u)re</i>	,9	,50	,00	1,00	,00	5,00	,4	,00	.	.	,00	3,00
<i>aviam</i>	,1	,00	.	.	,00	1,00	,2	,00	.	.	,00	2,00

	Edat 16 anys Sexe Masculí L1 Català i castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>a ver</i>	,0	,00	.	.	,00	,00
<i>a veure</i>	1,0	1,00	,00	2,00	,00	2,00
<i>a ve(u)re</i>	,0	,00	.	.	,00	,00
<i>aviam</i>	,0	,00	.	.	,00	,00

Annexos

Annex 4.4. Estudi estadístic exploratori de V3B (sexe=femení, edat=16 anys).

	Edat 16 anys Sexe Femení L1 Català						Edat 16 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>a ver</i>	,0	,00	.	.	,00	,00	,2	,00	.	.	,00	1,00
<i>a veure</i>	,6	,00	.	.	,00	6,00	,6	,00	.	.	,00	4,00
<i>a ve(u)re</i>	3,5	,00	.	.	,00	27,00	1,9	,00	.	.	,00	22,00
<i>aviam</i>	,1	,00	.	.	,00	1,00	,0	,00	.	.	,00	,00

	Edat 16 anys Sexe Femení L1 Català i castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>a ver</i>	,1	,00	.	.	,00	1,00
<i>a veure</i>	,0	,00	.	.	,00	,00
<i>a ve(u)re</i>	5,9	1,00	,00	14,00	,00	27,00
<i>aviam</i>	,3	,00	.	.	,00	1,00

Annex 4.5. Estudi estadístic exploratori de V3C (sexe=masculí, edat=12 anys).

	Edat 12 anys Sexe Masculí L1 Català						Edat 12 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>bueno</i>	19,7	12,50	4,00	25,00	2,00	76,00	10,9	9,00	4,00	14,00	,00	34,00
<i>bé</i>	,00	,00	.	.	,00	,00	,1	,00	.	.	,00	1,00

	Edat 12 anys Sexe Masculí L1 Català i castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>bueno</i>	14,0	14,00	11,00	17,00	11,00	17,00
<i>bé</i>	,00	,00	.	.	,00	,00

Annexos

Annex 4.6. Estudi estadístic exploratori de V3C (sexe=femení, edat=12 anys).

	Edat 12 anys Sexe Femení L1 Català						Edat 12 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>bueno</i>	16,1	11,00	7,00	17,00	,00	65,00	9,5	8,00	3,00	17,00	,00	25,00
<i>bé</i>	,5	,00	.	.	,00	4,00	,1	,00	.	.	,00	1,00
	Edat 12 anys Sexe Femení L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>bueno</i>	7,6	3,00	3,00	20,00	,00	20,00						
<i>bé</i>	,6	,00	.	.	,00	3,00						

Annex 4.7. Estudi estadístic exploratori de V3C (sexe=masculí, edat=16 anys).

	Edat 16 anys Sexe Masculí L1 Català						Edat 16 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>bueno</i>	14,1	10,00	7,00	15,00	3,00	49,00	12,5	9,00	4,00	19,00	1,00	40,00
<i>bé</i>	,2	,00	.	.	,00	2,00	,2	,00	.	.	,00	2,00

	Edat 16 anys Sexe Masculí L1 Català i castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>bueno</i>	22,5	22,50	12,00	33,00	12,00	33,00
<i>bé</i>	,0	,00	.	.	,00	,00

Annexos

Annex 4.8. Estudi estadístic exploratori de V3C (sexe=femení, edat=16 anys).

	Edat 16 anys Sexe Femení L1 Català						Edat 16 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>bueno</i>	21,5	25,00	3,00	34,00	2,00	44,00	22,3	24,00	9,00	35,00	1,00	48,00
<i>bé</i>	,0	,00	.	.	,00	,00	,0	,00	.	.	,00	,00
	Edat 16 anys Sexe Femení L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>bueno</i>	14,6	11,00	3,00	36,00	,00	36,00						
<i>bé</i>	,1	,00	.	.	,00	1,00						

Annex 4.9. Estudi estadístic exploratori de V3F (sexe=masculí, edat=12 anys).

	Edat 12 anys Sexe Masculí L1 Català						Edat 12 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>después</i>	,0	,00	.	.	,00	,00	,0	,00	.	.	,00	,00
<i>després</i>	5,9	5,00	2,00	8,00	1,00	13,00	2,6	1,50	1,00	3,00	,00	9,00

	Edat 12 anys Sexe Masculí L1 Català i castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>después</i>	,0	,00	.	.	,00	,00
<i>després</i>	4,0	4,00	1,00	7,00	1,00	7,00

Annexos

Annex 4.10. Estudi estadístic exploratori de V3F (sexe=femení, edat=12 anys).

	Edat 12 anys Sexe Femení L1 Català						Edat 12 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>después</i>	,1	,00	.	.	,00	1,00	,5	,00	.	.	,00	9,00
<i>després</i>	1,1	,00	.	.	,00	4,00	4,1	3,00	2,00	7,00	,00	22,00
	Edat 12 anys Sexe Femení L1 Català i castellà											
							Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>después</i>							,5	,00	.	.	,00	3,00
<i>després</i>							3,3	1,00	1,00	11,00	,00	11,00

Annex 4.11. Estudi estadístic exploratori de V3F (sexe=masculí, edat=16 anys).

	Edat 16 anys Sexe Masculí L1 Català						Edat 16 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>después</i>	,0	,00	.	.	,00	,00	,0	,00	.	.	,00	,00
<i>després</i>	4,4	3,00	2,00	7,00	,00	13,00	2,4	2,00	1,00	5,00	,00	6,00

	Edat 16 anys Sexe Masculí L1 Català i castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>después</i>	,0	,00	.	.	,00	,00
<i>després</i>	,0	,00	.	.	,00	,00

Annexos

Annex 4.12. Estudi estadístic exploratori de V3F (sexe=femení, edat=16 anys).

	Edat 16 anys Sexe Femení L1 Català						Edat 16 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>después</i>	,0	,00	.	.	,00	,00	,5	,00	.	.	,00	9,00
<i>després</i>	4,0	2,00	,00	6,00	,00	15,00	2,6	2,00	1,00	5,00	,00	10,00
	Edat 16 anys Sexe Femení L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>después</i>	,8	,00	.	.	,00	6,00						
<i>després</i>	3,4	1,50	1,00	7,00	,00	10,00						

Annex 4.13. Estudi estadístic exploratori de V3H (sexe=masculí, edat=12 anys).

	Edat 12 anys Sexe Masculí L1 Català						Edat 12 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>hombre</i>	,0	,00	.	.	,00	,00	,1	,00	.	.	,00	2,00
<i>home</i>	1,1	,00	.	.	,00	7,00	1,5	,50	,00	2,00	,00	5,00

	Edat 12 anys Sexe Masculí L1 Català i castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>hombre</i>	,0	,00	.	.	,00	,00
<i>home</i>	1,5	1,50	1,00	2,00	1,00	2,00

Annexos

Annex 4.14. Estudi estadístic exploratori de V3H (sexe=femení, edat=12 anys).

	Edat 12 anys Sexe Femení L1 Català						Edat 12 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>hombre</i>	,0	,00	.	.	,00	,00	,1	,00	.	.	,00	1,00
<i>home</i>	2,2	1,00	,00	8,00	,00	10,00	,3	,00	.	.	,00	2,00
	Edat 12 anys Sexe Femení L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>hombre</i>	,0	,00	.	.	,00	,00						
<i>home</i>	1,0	,00	.	.	,00	5,00						

Annex 4.15. Estudi estadístic exploratori de V3H (sexe=masculí, edat=16 anys).

	Edat 16 anys Sexe Masculí L1 Català						Edat 16 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>hombre</i>	,0	,00	.	.	,00	,00	,4	,00	.	.	,00	6,00
<i>home</i>	3,4	2,00	1,00	6,00	,00	12,00	1,5	1,00	1,00	4,00	,00	8,00
	Edat 16 anys Sexe Masculí L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>hombre</i>	,0	,00	.	.	,00	,00						
<i>home</i>	,0	,00	.	.	,00	,00						

Annexos

Annex 4.16. Estudi estadístic exploratori de V3H (sexe=femení, edat=16 anys).

	Edat 16 anys Sexe Femení L1 Català						Edat 16 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>hombre</i>	,2	,00	.	.	,00	2,00	,0	,00	.	.	,00	,00
<i>home</i>	2,4	1,00	1,00	4,00	,00	8,00	1,3	,50	,00	1,00	,00	6,00
	Edat 16 anys Sexe Femení L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>hombre</i>	,0	,00	.	.	,00	,00						
<i>home</i>	1,8	,50	,00	1,00	,00	11,00						

Annex 4.17. Estudi estadístic exploratori de V3I (sexe=masculí, edat=12 anys).

	Edat 12 anys Sexe Masculí L1 Català						Edat 12 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>o sea</i>	,2	,00	.	.	,00	3,00	,0	,00	.	.	,00	,00
<i>o sigui</i>	1,8	,00	.	.	,00	12,00	,6	,00	.	.	,00	7,00
	Edat 12 anys Sexe Masculí L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>o sea</i>	,0	,00	.	.	,00	,00						
<i>o sigui</i>	6,0	6,00	,00	12,00	,00	12,00						

Annexos

Annex 4.18. Estudi estadístic exploratori de V3I (sexe=femení, edat=12 anys).

	Edat 12 anys Sexe Femení L1 Català						Edat 12 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>o sea</i>	,0	,00	.	.	,00	,00	,4	,00	.	.	,00	3,00
<i>o sigui</i>	4,0	,00	.	.	,00	32,00	,4	,00	.	.	,00	4,00
	Edat 12 anys Sexe Femení L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>o sea</i>	,0	,00	.	.	,00	,00						
<i>o sigui</i>	3,9	1,00	,00	24,00	,00	24,00						

Annex 4.19. Estudi estadístic exploratori de V3I (sexe=masculí, edat=16 anys).

	Edat 16 anys Sexe Masculí L1 Català						Edat 16 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>o sea</i>	,0	,00	.	.	,00	,00	,9	,00	.	.	,00	6,00
<i>o sigui</i>	5,1	4,00	3,00	10,00	,00	20,00	,9	,00	.	.	,00	5,00
	Edat 16 anys Sexe Masculí L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>o sea</i>	,5	,50	,00	1,00	,00	1,00						
<i>o sigui</i>	3,0	3,00	,00	6,00	,00	6,00						

Annexos

Annex 4.20. Estudi estadístic exploratori de V3I (sexe=femení, edat=16 anys).

	Edat 16 anys Sexe Femení L1 Català						Edat 16 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>o sea</i>	,1	,00	.	.	,00	1,00	1,2	,00	.	.	,00	12,00
<i>o sigui</i>	4,5	4,00	4,00	20,00	,00	20,00	4,0	1,50	,00	6,00	,00	14,00
	Edat 16 anys Sexe Femení L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>o sea</i>	,0	,00	.	.	,00	,00						
<i>o sigui</i>	4,9	3,00	3,00	16,00	1,00	16,00						

Annex 4.21. Estudi estadístic exploratori de V3J (sexe=masculí, edat=12 anys).

	Edat 12 anys Sexe Masculí L1 Català						Edat 12 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>pero</i>	1,3	,00	.	.	,00	10,00	,9	,00	.	.	,00	9,00
<i>però</i>	,2	,00	.	.	,00	2,00	,0	,00	.	.	,00	,00
<i>p(e)rò</i>	9,0	7,00	7,00	14,00	1,00	17,00	6,7	7,00	3,00	10,00	,00	13,00
<i>p(er)ò</i>	2,3	1,00	1,00	4,00	,00	13,00	,2	,00	.	.	,00	2,00

	Edat 12 anys Sexe Masculí L1 Català i castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>pero</i>	,0	,00	.	.	,00	,00
<i>però</i>	,5	,50	,00	1,00	,00	1,00
<i>p(e)rò</i>	9,5	9,50	3,00	16,00	3,00	16,00
<i>p(er)ò</i>	1,5	1,50	1,00	2,00	1,00	2,00

Annexos

Annex 4.22. Estudi estadístic exploratori de V3J (sexe=femení, edat=12 anys).

	Edat 12 anys Sexe Femení L1 Català						Edat 12 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>pero</i>	,0	,00	.	.	,00	,00	,8	,00	.	.	,00	5,00
<i>però</i>	,3	,00	.	.	,00	3,00	,0	,00	.	.	,00	,00
<i>p(e)rò</i>	12,5	13,00	8,00	24,00	2,00	25,00	10,6	11,00	8,00	14,00	,00	19,00
<i>p(er)ò</i>	1,5	,00	.	.	,00	12,00	,1	,00	.	.	,00	1,00

	Edat 12 anys Sexe Femení L1 Català i castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>pero</i>	,0	,00	.	.	,00	,00
<i>però</i>	,3	,00	.	.	,00	2,00
<i>p(e)rò</i>	9,3	6,50	3,00	18,00	3,00	20,00
<i>p(er)ò</i>	,3	,00	.	.	,00	2,00

Annex 4.23. Estudi estadístic exploratori de V3J (sexe=masculí, edat=16 anys).

	Edat 16 anys Sexe Masculí L1 Català						Edat 16 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>pero</i>	,1	,00	.	.	,00	1,00	,8	,00	.	.	,00	11,00
<i>però</i>	,1	,00	.	.	,00	1,00	,0	,00	.	.	,00	,00
<i>p(e)rò</i>	12,6	11,50	7,00	15,00	2,00	35,00	10,1	11,00	6,00	16,00	,00	20,00
<i>p(er)ò</i>	2,9	1,00	,00	4,00	,00	13,00	,5	,00	.	.	,00	3,00

	Edat 16 anys Sexe Masculí L1 Català i castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>pero</i>	,0	,00	.	.	,00	,00
<i>però</i>	,0	,00	.	.	,00	,00
<i>p(e)rò</i>	10,5	10,50	9,00	12,00	9,00	12,00
<i>p(er)ò</i>	,5	,50	,00	1,00	,00	1,00

Annexos

Annex 4.24. Estudi estadístic exploratori de V3J (sexe=femení, edat=16 anys).

	Edat 16 anys Sexe Femení L1 Català						Edat 16 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>pero</i>	,6	,00	.	.	,00	4,00	1,1	,00	.	.	,00	13,00
<i>però</i>	,0	,00	.	.	,00	,00	,2	,00	.	.	,00	2,00
<i>p(e)rò</i>	14,6	13,00	8,00	21,00	7,00	28,00	13,6	14,50	4,00	21,00	2,00	32,00
<i>p(er)ò</i>	1,1	1,00	,00	3,00	,00	3,00	,2	,00	.	.	,00	2,00

	Edat 16 anys Sexe Femení L1 Català i castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>pero</i>	,1	,00	.	.	,00	1,00
<i>però</i>	,0	,00	.	.	,00	,00
<i>p(e)rò</i>	15,3	14,00	6,00	28,00	6,00	29,00
<i>p(er)ò</i>	1,1	1,00	1,00	3,00	,00	3,00

Annex 4.25. Estudi estadístic exploratori de V3K (sexe=masculí, edat=12 anys).

	Edat 12 anys Sexe Masculí L1 Català						Edat 12 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>pues</i>	5,3	1,00	,00	7,00	,00	43,00	3,0	1,00	,00	6,00	,00	17,00
<i>pos</i>	,3	,00	.	.	,00	2,00	1,2	,50	,00	1,00	,00	9,00
<i>doncs</i>	5,5	4,00	1,00	9,00	,00	24,00	5,5	1,50	,00	7,00	,00	21,00
	Edat 12 anys Sexe Masculí L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>pues</i>	6,5	6,50	,00	13,00	,00	13,00						
<i>pos</i>	,0	,00	.	.	,00	,00						
<i>doncs</i>	3,5	3,50	,00	7,00	,00	7,00						

Annexos

Annex 4.26. Estudi estadístic exploratori de V3K (sexe=femení, edat=12 anys).

	Edat 12 anys Sexe Femení L1 Català						Edat 12 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>pues</i>	4,7	2,00	,00	5,00	,00	19,00	4,3	1,00	,00	12,00	,00	25,00
<i>pos</i>	,8	,00	.	.	,00	4,00	,4	,00	.	.	,00	2,00
<i>doncs</i>	6,8	2,00	,00	16,00	,00	23,00	9,3	2,00	,00	14,00	,00	46,00
	Edat 12 anys Sexe Femení L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>pues</i>	7,7	2,00	,00	28,00	,00	28,00						
<i>pos</i>	,8	,00	.	.	,00	5,00						
<i>doncs</i>	5,9	1,50	,00	15,00	,00	24,00						

Annex 4.27. Estudi estadístic exploratori de V3K (sexe=masculí, edat=16 anys).

	Edat 16 anys Sexe Masculí L1 Català						Edat 16 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>pues</i>	2,1	1,00	,00	4,00	,00	8,00	2,1	1,00	,00	5,00	,00	10,00
<i>pos</i>	,4	,00	.	.	,00	3,00	,9	,00	.	.	,00	8,00
<i>doncs</i>	6,9	6,50	1,00	11,00	,00	23,00	3,6	2,00	,00	5,00	,00	12,00
	Edat 16 anys Sexe Masculí L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>pues</i>	,5	,50	,00	1,00	,00	1,00						
<i>pos</i>	1,0	1,00	.	.	1,00	1,00						
<i>doncs</i>	8,5	8,50	7,00	10,00	7,00	10,00						

Annexos

Annex 4.28. Estudi estadístic exploratori de V3K (sexe=femení, edat=16 anys).

	Edat 16 anys Sexe Femení L1 Català						Edat 16 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>pues</i>	1,2	,00	.	.	,00	6,00	2,7	,50	,00	1,00	,00	17,00
<i>pos</i>	2,1	,00	.	.	,00	9,00	1,5	,00	.	.	,00	10,00
<i>doncs</i>	6,0	3,00	1,00	10,00	1,00	27,00	9,4	5,00	,00	13,00	,00	42,00
	Edat 16 anys Sexe Femení L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>pues</i>	2,6	2,50	,00	6,00	,00	7,00						
<i>pos</i>	1,3	1,00	1,00	4,00	,00	4,00						
<i>doncs</i>	7,1	5,50	1,00	10,00	,00	23,00						

Annex 4.29. Estudi estadístic exploratori de V3L (sexe=masculí, edat=12 anys).

	Edat 12 anys Sexe Masculí L1 Català						Edat 12 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>vale</i>	4,5	3,4
<i>val</i>	1,5	1,3
<i>d'acord</i>	,4	,00	.	.	,00	3,00	,1	,00	.	.	,00	2,00

	Edat 12 anys Sexe Masculí L1 Català i castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>vale</i>	4,0
<i>val</i>	4,5
<i>d'acord</i>	,0	,00	.	.	,00	,00

Annexos

Annex 4.30. Estudi estadístic exploratori de V3L (sexe=femení, edat=12 anys).

	Edat 12 anys Sexe Femení L1 Català						Edat 12 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>vale</i>	3,7	4,7
<i>val</i>	2,5	2,0
<i>d'acord</i>	,0	,00	.	.	,00	,00	,1	,00	.	.	,00	1,00

	Edat 12 anys Sexe Femení L1 Català i castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>vale</i>	4,2
<i>val</i>	1,1
<i>d'acord</i>	,4	,00	.	.	,00	2,00

Annex 4.31. Estudi estadístic exploratori de V3L (sexe=masculí, edat=16 anys).

	Edat 16 anys Sexe Masculí L1 Català						Edat 16 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>vale</i>	6,7	3,4
<i>val</i>	0,6	,6
<i>d'acord</i>	,7	,00	.	.	,00	5,00	,7	,00	.	.	,00	5,00
	Edat 16 anys Sexe Masculí L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>vale</i>	2,0						
<i>val</i>	,5						
<i>d'acord</i>	,5	,50	,00	1,00	,00	1,00						

Annexos

Annex 4.32. Estudi estadístic exploratori de V3L (sexe=femení, edat=16 anys).

	Edat 16 anys Sexe Femení L1 Català						Edat 16 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>vale</i>	4,3	5,8
<i>val</i>	,8	,5
<i>d'acord</i>	,0	,00	.	.	,00	,00	,2	,00	.	.	,00	1,00

	Edat 16 anys Sexe Femení L1 Català i castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>vale</i>	4,7
<i>val</i>	1,5
<i>d'acord</i>	,5	,00	.	.	,00	4,00

Annex 4.33. Estudi estadístic exploratori de V5 (sexe=masculí, edat=12 anys).

	Edat 12 anys Sexe Masculí L1 Català						Edat 12 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>hola</i>	,3	,00	.	.	,00	1,00	,3	,00	.	.	,00	1,00
<i>bon dia</i>	,6	1,00	1,00	1,00	,00	1,00	,6	1,00	1,00	1,00	,00	1,00
<i>bona tarda</i>	,1	,00	.	.	,00	1,00	,1	,00	.	.	,00	1,00
<i>hola bon dia</i>	.0	,00	.	.	,00	,00	,1	,00	.	.	,00	1,00

	Edat 12 anys Sexe Masculí L1 Català i castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>hola</i>	,0	,00	.	.	,00	,00
<i>bon dia</i>	1,0	1,00	.	.	1,00	1,00
<i>bona tarda</i>	,0	,00	.	.	,00	,00
<i>hola bon dia</i>	,0	,00	.	.	,00	,00

Annexos

Annex 4.34. Estudi estadístic exploratori de V5 (sexe=femení, edat=12 anys).

	Edat 12 anys Sexe Femení L1 Català						Edat 12 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>hola</i>	,2	,00	.	.	,00	1,00	,3	,00	.	.	,00	1,00
<i>bon dia</i>	,8	1,00	1,00	1,00	,00	1,00	,6	1,00	1,00	1,00	,00	1,00
<i>bona tarda</i>	,0	,00	.	.	,00	,00	,1	,00	.	.	,00	1,00
<i>hola bon dia</i>	,1	,00	.	.	,00	1,00	,1	,00	.	.	,00	1,00
	Edat 12 anys Sexe Femení L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>hola</i>	,0	,00	.	.	,00	,00						
<i>bon dia</i>	,6	1,00	1,00	1,00	,00	1,00						
<i>bona tarda</i>	,4	,00	.	.	,00	1,00						
<i>hola bon dia</i>	,0	,00	.	.	,00	,00						

Annex 4.35. Estudi estadístic exploratori de V5 (sexe=masculí, edat=16 anys).

	Edat 16 anys Sexe Masculí L1 Català						Edat 16 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>hola</i>	,1	,00	.	.	,00	1,00	,0	,00	.	.	,00	,00
<i>bon dia</i>	,9	1,00	1,00	1,00	,00	1,00	,6	1,00	1,00	1,00	,00	1,00
<i>bona tarda</i>	,1	,00	.	.	,00	1,00	,1	,00	.	.	,00	1,00
<i>hola bon dia</i>	,0	,00	.	.	,00	,00	,1	,00	.	.	,00	1,00

Edat 16 anys Sexe Masculí L1 Català i castellà						
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>hola</i>	,5	,50	,00	1,00	,00	1,00
<i>bon dia</i>	,5	,50	,00	1,00	,00	1,00
<i>bona tarda</i>	,5	,50	,00	1,00	,00	1,00
<i>hola bon dia</i>	,0	,00	.	.	,00	,00

Annexos

Annex 4.36. Estudi estadístic exploratori de V5 (sexe=femení, edat=16 anys).

	Edat 16 anys Sexe Femení L1 Català						Edat 16 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>hola</i>	,2	,00	.	.	,00	1,00	,3	,00	.	.	,00	1,00
<i>bon dia</i>	,8	1,00	1,00	1,00	,00	1,00	,6	1,00	1,00	2,00	,00	2,00
<i>bona tarda</i>	,0	,00	.	.	,00	,00	,1	,00	.	.	,00	1,00
<i>hola bon dia</i>	,0	,00	.	.	,00	,00	,3	,00	.	.	,00	1,00

	Edat 16 anys Sexe Femení L1 Català i castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>hola</i>	,1	,00	.	.	,00	1,00
<i>bon dia</i>	,6	1,00	1,00	1,00	,00	1,00
<i>bona tarda</i>	,3	,00	.	.	,00	1,00
<i>hola bon dia</i>	,0	,00	.	.	,00	,00

Annex 4.37. Estudi estadístic exploratori de V6 (sexe=masculí, edat=12 anys).

	Edat 12 anys Sexe Masculí L1 Català						Edat 12 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>Heterorepetició</i>	4,8	4,00	3,00	6,00	2,00	14,00	4,3	3,50	3,00	5,00	1,00	12,00
	Edat 12 anys Sexe Masculí L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>Heterorepetició</i>	5,5	5,50	4,00	7,00	4,00	7,00						

Annexos

Annex 4.38. Estudi estadístic exploratori de V6 (sexe=femení, edat=12 anys).

	Edat 12 anys Sexe Femení L1 Català						Edat 12 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>Heterorepetició</i>	5,0	3,00	2,00	7,00	1,00	15,00	3,2	2,00	1,00	4,00	,00	11,00
	Edat 12 anys Sexe Femení L1 Català i castellà											
							Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>Heterorepetició</i>							3,3	3,00	2,00	6,00	1,00	6,00

Annex 4.39. Estudi estadístic exploratori de V6 (sexe=masculí, edat=16 anys).

	Edat 16 anys Sexe Masculí L1 Català						Edat 16 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>Heterorepetició</i>	7,4	8,00	7,00	11,00	1,00	12,00	7,9	7,00	6,00	10,00	3,00	16,00
	Edat 16 anys Sexe Masculí L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>Heterorepetició</i>	4,0	4,00	1,00	7,00	1,00	7,00						

Annexos

Annex 4.40. Estudi estadístic exploratori de V6 (sexe=femení, edat=16 anys).

	Edat 16 anys Sexe Femení L1 Català						Edat 16 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>Heterorepetició</i>	4,1	4,00	3,00	7,00	1,00	8,00	5,2	5,00	3,00	8,00	1,00	10,00
	Edat 16 anys Sexe Femení L1 Català i castellà											
							Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>Heterorepetició</i>							6,1	5,00	5,00	13,00	4,00	13,00

Annex 4.41. Estudi estadístic exploratori de V14a (sexe=masculí, edat=12 anys).

	Edat 12 anys Sexe Masculí L1 Català						Edat 12 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
PC + tenir que / de + INF	1,1	,00	.	.	,00	7,00	1,6	1,00	,00	3,00	,00	5,00
PC + haver de + INF	4,3	3,00	2,00	8,00	,00	18,00	1,4	1,00	,00	3,00	,00	5,00
PC + poder + INF	9,6	8,50	5,00	14,00	1,00	25,00	6,1	5,00	3,00	10,00	1,00	17,00
PC+ anar a + INF	,4	,00	.	.	,00	2,00	,0	,00	.	.	,00	,00
	Edat 12 anys Sexe Masculí L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
PC + tenir que / de + INF	6,0	6,00	1,00	11,00	1,00	11,00						
PC + haver de + INF	,5	,50	,00	1,00	,00	1,00						
PC + poder + INF	19,0	19,00	11,00	27,00	11,00	27,00						
PC+ anar a + INF	,0	,00	.	.	,00	,00						

Annexos

Annex 4.42. Estudi estadístic exploratori de V14a (sexe=femení, edat=12 anys).

	Edat 12 anys Sexe Femení L1 Català						Edat 12 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
PC + tenir que / de + INF	2,8	,00	.	.	,00	12,00	1,3	,00	.	.	,00	13,00
PC + haver de + INF	2,7	2,00	1,00	4,00	,00	10,00	3,5	2,50	1,00	5,00	,00	11,00
PC + poder + INF	7,7	7,00	4,00	12,00	2,00	14,00	8,9	7,50	6,00	10,00	,00	27,00
PC+ anar a + INF	,0	,00	.	.	,00	,00	,0	,00	.	.	,00	,00
	Edat 12 anys Sexe Femení L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
PC + tenir que / de + INF	1,1	,50	,00	3,00	,00	4,00						
PC + haver de + INF	2,3	1,50	,00	5,00	,00	6,00						
PC + poder + INF	7,0	5,00	1,00	17,00	1,00	18,00						
PC+ anar a + INF	,4	,00	.	.	,00	2,00						

Annex 4.43. Estudi estadístic exploratori de V14a (sexe=masculí, edat=16 anys).

	Edat 16 anys Sexe Masculí L1 Català						Edat 16 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
PC + tenir que / de + INF	,6	,00	.	.	,00	6,00	,3	,00	.	.	,00	1,00
PC + haver de + INF	3,8	3,00	3,00	7,00	,00	17,00	1,9	1,50	,00	4,00	,00	5,00
PC + poder + INF	9,8	9,00	8,00	14,00	4,00	20,00	10,1	7,00	7,00	11,00	2,00	42,00
PC+ anar a + INF	,0	,00	.	.	,00	,00	,0	,00	.	.	,00	,00
	Edat 16 anys Sexe Masculí L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
PC + tenir que / de + INF	,5	,50	,00	1,00	,00	1,00						
PC + haver de + INF	1,5	1,50	1,00	2,00	1,00	2,00						
PC + poder + INF	11,0	11,00	.	.	11,00	11,00						
PC+ anar a + INF	,0	,00	.	.	,00	,00						

Annexos

Annex 4.44. Estudi estadístic exploratori de V14a (sexe=femení, edat=16 anys).

	Edat 16 anys Sexe Femení L1 Català						Edat 16 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
PC + tenir que / de + INF	1,5	,00	.	.	,00	8,00	,7	,00	.	.	,00	5,00
PC + haver de + INF	2,9	2,00	1,00	7,00	,00	11,00	3,8	2,50	,00	4,00	,00	13,00
PC + poder + INF	8,3	8,00	5,00	13,00	4,00	15,00	9,3	9,00	8,00	12,00	3,00	18,00
PC+ anar a + INF	,0	,00	.	.	,00	,00	,1	,00	.	,00	,00	1,00
	Edat 16 anys Sexe Femení L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
PC + tenir que / de + INF	1,0	,00	.	.	,00	6,00						
PC + haver de + INF	6,0	4,00	2,00	12,00	,00	12,00						
PC + poder + INF	11,0	11,00	5,00	19,00	2,00	19,00						
PC+ anar a + INF	,0	,00	.	.	,00	,00						

Annex 4.45. Estudi estadístic exploratori de V14b (sexe=masculí, edat=12 anys).

	Edat 12 anys Sexe Masculí L1 Català						Edat 12 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>tenir que / de + INF + PC</i>	,1	,00	.	.	,00	1,00	,3	,00	.	.	,00	1,00
<i>haver de + INF + PC</i>	,1	,00	.	.	,00	1,00	,0	,00	.	.	,00	,00
<i>poder + INF + PC</i>	,6	,00	.	.	,00	3,00	1,7	,50	,00	2,00	,00	11,00
	Edat 12 anys Sexe Masculí L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>tenir que / de + INF + PC</i>	,5	,50	,00	,00	,00	1,00						
<i>haver de + INF + PC</i>	,0	,00	.	.	,00	,00						
<i>poder + INF + PC</i>	1,5	1,50	1,00	2,00	1,00	2,00						

Annexos

Annex 4.46. Estudi estadístic exploratori de V14b (sexe=femení, edat=12 anys).

	Edat 12 anys Sexe Femení L1 Català						Edat 12 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>tenir que / de + INF + PC</i>	,4	,00	.	.	,00	2,00	,9	,00	.	.	,00	7,00
<i>haver de + INF + PC</i>	,0	,00	.	.	,00	,00	,1	,00	.	.	,00	1,00
<i>poder + INF + PC</i>	,5	,00	.	.	,00	2,00	,5	,00	.	.	,00	2,00
	Edat 12 anys Sexe Femení L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>tenir que / de + INF + PC</i>	,8	,00	.	.	,00	4,00						
<i>haver de + INF + PC</i>	,0	,00	.	.	,00	,00						
<i>poder + INF + PC</i>	,7	,50	,00	2,00	,00	2,00						

Annex 4.47. Estudi estadístic exploratori de V14b (sexe=masculí, edat=16 anys).

	Edat 16 anys Sexe Masculí L1 Català						Edat 16 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>tenir que / de + INF + PC</i>	,1	,00	.	.	,00	1,00	,3	,00	.	.	,00	2,00
<i>haver de + INF + PC</i>	,1	,00	.	.	,00	1,00	,1	,00	.	.	,00	1,00
<i>poder + INF + PC</i>	1,8	1,00	1,00	3,00	,00	11,00	1,1	,50	,00	1,00	,00	6,00
	Edat 16 anys Sexe Masculí L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>tenir que / de + INF + PC</i>	,5	,50	,00	1,00	,00	1,00						
<i>haver de + INF + PC</i>	,0	,00	.	.	,00	,00						
<i>poder + INF + PC</i>	1,5	1,50	1,00	2,00	1,00	2,00						

Annexos

Annex 4.48. Estudi estadístic exploratori de V14b (sexe=femení, edat=16 anys).

	Edat 16 anys Sexe Femení L1 Català						Edat 16 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>tenir que / de + INF + PC</i>	,0	,00	.	.	,00	,00	,4	,00	.	.	,00	3,00
<i>haver de + INF + PC</i>	,0	,00	.	.	,00	,00	,1	,00	.	.	,00	1,00
<i>poder + INF + PC</i>	,8	1,00	1,00	2,00	,00	2,00	1,2	1,00	1,00	4,00	,00	8,00
	Edat 16 anys Sexe Femení L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>tenir que / de + INF + PC</i>	,3	,00	.	.	,00	1,00						
<i>haver de + INF + PC</i>	,4	,00	.	.	,00	1,00						
<i>poder + INF + PC</i>	1,8	1,50	,00	3,00	,00	5,00						

Annexos

Annex 4.50. Estudi estadístic exploratori de V14c (sexe=femení, edat=12 anys).

	Edat 12 anys Sexe Femení L1 Català						Edat 12 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
PC + PV + PC	,0	,00	.	.	,00	,00	,1	,00	.	.	,00	1,00
	Edat 12 anys Sexe Femení L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
PC + PV + PC	,0	,00	.	.	,00	,00						

Annex 4.51. Estudi estadístic exploratori de V14c (sexe=masculí, edat=16 anys).

Edat 16 anys Sexe Masculí L1 Català							Edat 16 anys Sexe Masculí L1 Castellà						
Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim		
PC + PV + PC	,2	,00	.	.	,00	2,00	,1	,00	.	.	,00	1,00	
							Edat 16 anys Sexe Masculí L1 Català i castellà						
							Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	
							PC + PV + PC	,0	,00	.	.	,00	,00

Annexos

Annex 4.52. Estudi estadístic exploratori de V14c (sexe=femení, edat=16 anys).

	Edat 16 anys Sexe Femení L1 Català						Edat 16 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
PC + PV + PC	,2	,00	.	.	,00	1,00	.0	,00	.	.	,00	,00
	Edat 16 anys Sexe Femení L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
PC + PV + PC	,3	,00	.	.	,00	1,00						

Annex 4.53. Estudi estadístic exploratori de V16 (sexe=masculí, edat=12 anys).

	Edat 12 anys Sexe Masculí L1 Català						Edat 12 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
Nombre de preguntes no retòriques	3,6	2,50	1,00	4,00	,00	21,00	1,9	1,00	1,00	4,00	,00	5,00
	Edat 12 anys Sexe Masculí L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
Nombre de preguntes no retòriques	3,5	3,50	2,00	5,00	2,00	5,00						

Annexos

Annex 4.54. Estudi estadístic exploratori de V16 (sexe=femení, edat=12 anys).

	Edat 12 anys Sexe Femení L1 Català						Edat 12 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
Nombre de preguntes no retòriques	2,5	2,00	1,00	5,00	1,00	7,00	1,8	2,50	,00	3,00	,00	4,00
	Edat 12 anys Sexe Femení L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
Nombre de preguntes no retòriques	2,8	2,00	2,00	10,00	,00	10,00						

Annex 4.55. Estudi estadístic exploratori de V16 (sexe=masculí, edat=16 anys).

	Edat 16 anys Sexe Masculí L1 Català						Edat 16 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
Nombre de preguntes no retòriques	4,0	3,00	3,00	5,00	,00	19,00	2,0	1,00	1,00	4,00	,00	6,00
	Edat 16 anys Sexe Masculí L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
Nombre de preguntes no retòriques	3,5	3,50	2,00	5,00	2,00	5,00						

Annexos

Annex 4.56. Estudi estadístic exploratori de V16 (sexe=femení, edat=16 anys).

	Edat 16 anys Sexe Femení L1 Català						Edat 16 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
Nombre de preguntes no retòriques	2,9	2,00	1,00	4,00	,00	10,00	2,0	1,50	1,00	3,00	,00	6,00
	Edat 16 anys Sexe Femení L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
Nombre de preguntes no retòriques	3,4	2,50	1,00	7,00	,00	9,00						

Annex 4.57. Estudi estadístic exploratori de V17 (sexe=masculí, edat=12 anys).

	Edat 12 anys Sexe Masculí L1 Català						Edat 12 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>oberta</i>	,9	,00	.	.	,00	8,00	,8	,50	,00	1,00	,00	4,00
<i>tancada</i>	2,7	2,00	2,00	4,00	,00	13,00	1,1	1,00	1,00	5,00	,00	5,00
	Edat 12 anys Sexe Masculí L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>oberta</i>	1,5	1,50	,00	3,00	,00	3,00						
<i>tancada</i>	2,0	2,00	.	.	2,00	2,00						

Annexos

Annex 4.58. Estudi estadístic exploratori de V17 (sexe=femení, edat=12 anys).

	Edat 12 anys Sexe Femení L1 Català						Edat 12 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>oberta</i>	,5	,00	.	.	,00	2,00	,6	,00	.	.	,00	2,00
<i>tancada</i>	2,1	2,00	1,00	4,00	,00	6,00	1,3	1,00	,00	2,00	,00	4,00
	Edat 12 anys Sexe Femení L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>oberta</i>	,5	,00	.	.	,00	3,00						
<i>tancada</i>	2,3	2,00	2,00	7,00	,00	7,00						

Annex 4.59. Estudi estadístic exploratori de V17 (sexe=masculí, edat=16 anys).

	Edat 16 anys Sexe Masculí L1 Català						Edat 16 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>oberta</i>	,8	1,00	1,00	3,00	,00	3,00	,4	,00	.	.	,00	2,00
<i>tancada</i>	3,2	2,00	2,00	4,00	,00	16,00	1,6	1,00	,00	3,00	,00	6,00

	Edat 16 anys Sexe Masculí L1 Català i castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>oberta</i>	1,0	1,00	.	.	1,00	1,00
<i>tancada</i>	2,5	2,50	1,00	4,00	1,00	4,00

Annexos

Annex 4.60. Estudi estadístic exploratori de V17 (sexe=femení, edat=16 anys).

	Edat 16 anys Sexe Femení L1 Català						Edat 16 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>oberta</i>	,3	,00	.	.	,00	1,00	,2	,00	.	.	,00	1,00
<i>tancada</i>	2,6	2,00	1,00	4,00	,00	9,00	1,8	1,50	,00	3,00	,00	6,00
	Edat 16 anys Sexe Femení L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>oberta</i>	,4	,00	.	.	,00	2,00						
<i>tancada</i>	3,0	2,00	1,00	7,00	,00	7,00						

Annex 4.61. Estudi estadístic exploratori de V18 (sexe=masculí, edat=12 anys).

	Edat 12 anys Sexe Masculí L1 Català						Edat 12 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>si</i>	19,3	16,00	13,00	23,00	1,00	42,00	10,0	10,00	8,00	16,00	,00	23,00
<i>mentre(s)</i>	,1	,00	.	.	,00	1,00	,3	,00	.	.	,00	2,00
<i>a no ser que</i>	,3	,00	.	.	,00	1,00	,1	,00	.	.	,00	2,00
<i>en tot cas</i>	,1	,00	.	.	,00	1,00	,0	,00	.	.	,00	,00

	Edat 12 anys Sexe Masculí L1 Català i castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>si</i>	21,0	21,00	6,00	36,00	6,00	36,00
<i>mentre(s)</i>	,0	,00	.	.	,00	,00
<i>a no ser que</i>	,0	,00	.	.	,00	,00
<i>en tot cas</i>	,0	,00	.	.	,00	,00

Annexos

Annex 4.62. Estudi estadístic exploratori de V18 (sexe=femení, edat=12 anys).

	Edat 12 anys Sexe Femení L1 Català						Edat 12 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>si</i>	12,9	12,00	8,00	20,00	4,00	34,00	10,2	10,50	6,00	14,00	2,00	20,00
<i>mentre(s)</i>	,0	,00	.	.	,00	,00	,1	,00	.	.	,00	1,00
<i>a no ser que</i>	,0	,00	.	.	,00	,00	,1	,00	.	.	,00	1,00
<i>en tot cas</i>	,1	,00	.	.	,00	1,00	,1	,00	.	.	,00	1,00

	Edat 12 anys Sexe Femení L1 Català i castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>si</i>	13,8	15,00	6,00	21,00	4,00	21,00
<i>mentre(s)</i>	,0	,00	.	.	,00	,00
<i>a no ser que</i>	,1	,00	.	.	,00	1,00
<i>en tot cas</i>	,0	,00	.	.	,00	,00

Annex 4.63. Estudi estadístic exploratori de V18 (sexe=masculí, edat=16 anys).

	Edat 16 anys Sexe Masculí L1 Català						Edat 16 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>si</i>	19,5	18,00	14,00	31,00	3,00	42,00	11,7	9,00	8,00	21,00	2,00	31,00
<i>mentre(s)</i>	,1	,00	.	.	,00	1,00	1,2	,00	.	.	,00	13,00
<i>a no ser que</i>	,3	,00	.	.	,00	3,00	,0	,00	.	.	,00	,00
<i>en tot cas</i>	,0	,00	.	.	,00	,00	,0	,00	.	.	,00	,00

	Edat 16 anys Sexe Masculí L1 Català i castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>si</i>	21,0	21,00	1,00	41,00	1,00	41,00
<i>mentre(s)</i>	,0	,00	.	.	,00	,00
<i>a no ser que</i>	,0	,00	.	.	,00	,00
<i>en tot cas</i>	3,0	3,00	,00	6,00	,00	6,00

Annexos

Annex 4.64. Estudi estadístic exploratori de V18 (sexe=femení, edat=16 anys).

	Edat 16 anys Sexe Femení L1 Català						Edat 16 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>si</i>	15,4	15,00	9,00	17,00	8,00	30,00	13,3	11,00	10,00	14,00	3,00	33,00
<i>mentre(s)</i>	,0	,00	.	.	,00	,00	,1	,00	.	.	,00	1,00
<i>a no ser que</i>	,0	,00	.	.	,00	,00	,0	,00	.	.	,00	,00
<i>en tot cas</i>	,0	,00	.	.	,00	,00	,1	,00	.	.	,00	1,00

	Edat 16 anys Sexe Femení L1 Català i castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>si</i>	19,4	14,50	11,00	34,00	10,00	41,00
<i>mentre(s)</i>	,0	,00	.	.	,00	,00
<i>a no ser que</i>	,4	,00	.	.	,00	3,00
<i>en tot cas</i>	,0	,00	.	.	,00	,00

Annex 4.65. Estudi estadístic exploratori de V22 (sexe=masculí, edat=12 anys).

	Edat 12 anys Sexe Masculí L1 Català						Edat 12 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>tenir que</i> + INF	1,6	,00	.	.	,00	13,00	2,6	2,50	,00	4,00	,00	10,00
<i>haver de</i> + INF	7,8	6,50	5,00	9,00	1,00	25,00	2,4	2,00	2,00	7,00	,00	10,00
<i>tenir de</i> + INF	,2	,00	.	.	,00	2,00	,0	,00	.	.	,00	,00
<i>caldre</i> + INF	,5	,00	.	.	,00	3,00	,1	,00	.	.	,00	1,00
	Edat 12 anys Sexe Masculí L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>tenir que</i> + INF	12,0	12,00	,00	24,00	,00	24,00						
<i>haver de</i> + INF	1,0	1,00	.	.	1,00	1,00						
<i>tenir de</i> + INF	2,5	2,50	,00	5,00	,00	5,00						
<i>caldre</i> + INF	2,5	2,50	,00	5,00	,00	5,00						

Annexos

Annex 4.66. Estudi estadístic exploratori de V22 (sexe=femení, edat=12 anys).

	Edat 12 anys Sexe Femení L1 Català						Edat 12 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>tenir que + INF</i>	3,6	,00	.	.	,00	14,00	4,9	1,00	,00	9,00	,00	21,00
<i>haver de + INF</i>	5,3	6,00	1,00	8,00	,00	15,00	6,4	6,50	,00	11,00	,00	18,00
<i>tenir de + INF</i>	1,3	,00	.	.	,00	14,00	,8	,00	.	.	,00	13,00
<i>caldre + INF</i>	,7	,00	.	.	,00	4,00	,1	,00	.	.	,00	2,00
	Edat 12 anys Sexe Femení L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>tenir que + INF</i>	4,5	3,00	,00	13,00	,00	13,00						
<i>haver de + INF</i>	3,1	1,50	,00	8,00	,00	8,00						
<i>tenir de + INF</i>	,0	,00	.	.	,00	,00						
<i>caldre + INF</i>	,5	,00	.	.	,00	3,00						

Annex 4.67. Estudi estadístic exploratori de V22 (sexe=masculí, edat=16 anys).

	Edat 16 anys Sexe Masculí L1 Català						Edat 16 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>tenir que</i> + INF	,8	,00	.	.	,00	4,00	1,7	,50	,00	2,00	,00	11,00
<i>haver de</i> + INF	9,6	6,00	5,00	15,00	1,00	38,00	4,0	3,00	2,00	8,00	,00	10,00
<i>tenir de</i> + INF	,4	,00	.	.	,00	6,00	,1	,00	.	.	,00	1,00
<i>caldre</i> + INF	,4	,00	.	.	,00	3,00	,6	,00	.	.	,00	5,00
	Edat 16 anys Sexe Masculí L1 Català i castellà											
							Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>tenir que</i> + INF							2,5	2,50	,00	5,00	,00	5,00
<i>haver de</i> + INF							2,5	2,50	1,00	4,00	1,00	4,00
<i>tenir de</i> + INF							,5	,50	,00	1,00	,00	1,00
<i>caldre</i> + INF							2,0	2,00	1,00	3,00	1,00	3,00

Annexos

Annex 4.68. Estudi estadístic exploratori de V22 (sexe=femení, edat=16 anys).

	Edat 12 anys Sexe Femení L1 Català						Edat 12 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>tenir que + INF</i>	2,5	,00	.	.	,00	18,00	1,2	,00	.	.	,00	6,00
<i>haver de + INF</i>	6,0	5,00	3,00	10,00	,00	18,00	8,1	4,50	1,00	11,00	,00	24,00
<i>tenir de + INF</i>	,4	,00	.	.	,00	4,00	,4	,00	.	.	,00	7,00
<i>caldre + INF</i>	,7	,00	.	.	,00	3,00	,2	,00	.	.	,00	1,00
	Edat 12 anys Sexe Femení L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>tenir que + INF</i>	2,1	,00	.	.	,00	9,00						
<i>haver de + INF</i>	9,5	10,00	4,00	16,00	,00	17,00						
<i>tenir de + INF</i>	,0	,00	.	.	,00	,00						
<i>caldre + INF</i>	,4	,00	.	.	,00	3,00						

Annex 4.69. Estudi estadístic exploratori de V23 (sexe=masculí, edat=12 anys).

	Edat 12 anys Sexe Masculí L1 Català						Edat 12 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>aquest + N</i>	2,8	1,00	1,00	3,00	,00	16,00	,9	,50	,00	1,00	,00	4,00
<i>aquell + N</i>	,4	,00	.	.	,00	3,00	,0	,00	.	.	,00	,00
<i>el + N + aquest</i>	,5	,00	.	.	,00	3,00	,4	,00	.	.	,00	2,00
<i>el + N + aquell</i>	,3	,00	.	.	,00	3,00	,0	,00	.	.	,00	,00
<i>aquest</i>	3,4	2,00	1,00	4,00	,00	21,00	2,9	2,00	1,00	6,00	,00	7,00
<i>aquell</i>	,3	,00	.	.	,00	1,00	,0	,00	.	.	,00	,00
<i>(un +) N + d'aquests</i>	,5	,00	.	.	,00	2,00	1,8	,50	,00	4,00	,00	6,00
<i>(un +) N + d'aquells</i>	,3	,00	.	.	,00	4,00	,1	,00	.	.	,00	1,00

Annexos

Edat 12 anys Sexe Masculí L1 Català i castellà						
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>aquest + N</i>	2,0	2,00	,00	4,00	,00	4,00
<i>aquell + N</i>	,0	,00	.	.	,00	,00
<i>el + N + aquest</i>	1,0	1,00	.	.	1,00	1,00
<i>el + N + aquell</i>	,0	,00	.	.	,00	,00
<i>aquest</i>	5,0	5,00	,00	10,00	,00	10,00
<i>aquell</i>	,0	,00	.	.	,00	,00
<i>(un +) N + d'aquests</i>	,0	,00	.	.	,00	,00
<i>(un +) N + d'aquells</i>	,0	,00	.	.	,00	,00

Annex 4.70. Estudi estadístic exploratori de V23 (sexe=femení, edat=12 anys).

	Edat 12 anys Sexe Masculí L1 Català						Edat 12 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>aquest + N</i>	1,0	1,00	,00	3,00	,00	3,00	,8	,00	.	.	,00	2,00
<i>aquell + N</i>	,1	,00	.	.	,00	1,00	,1	,00	.	.	,00	1,00
<i>el + N + aquest</i>	,1	,00	.	.	,00	1,00	,5	,00	.	.	,00	2,00
<i>el + N + aquell</i>	,3	,00	.	.	,00	1,00	,0	,00	.	.	,00	,00
<i>aquest aquell</i>	2,2	2,00	1,00	7,00	,00	7,00	4,0	2,00	1,00	8,00	,00	20,00
<i>(un +) N + d'aquests</i>	,3	,00	.	.	,00	1,00	,9	,00	.	.	,00	6,00
<i>(un +) N + d'aquells</i>	,1	,00	.	.	,00	1,00	,1	,00	.	.	,00	1,00

Annexos

Edat 12 anys Sexe Masculí L1 Català i castellà						
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>aquest + N</i>	,5	,00	.	.	,00	3,00
<i>aquell + N</i>	,1	,00	.	.	,00	1,00
<i>el + N + aquest</i>	,1	,00	.	.	,00	1,00
<i>el + N + aquell</i>	,0	,00	.	.	,00	,00
<i>aquest</i>	3,7	3,00	2,00	9,00	,00	9,00
<i>aquell</i>	,1	,00	.	.	,00	1,00
<i>(un +) N + d'aquests</i>	,8	,00	.	.	,00	3,00
<i>(un +) N + d'aquells</i>	,0	,00	.	.	,00	,00

Annex 4.71. Estudi estadístic exploratori de V23 (sexe=masculí, edat=16 anys).

	Edat 16 anys Sexe Masculí L1 Català						Edat 16 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>aquest + N</i>	2,8	2,00	2,00	4,00	,00	8,00	1,5	1,00	1,00	6,00	,00	6,00
<i>aquell + N</i>	,2	,00	.	.	,00	2,00	,0	,00	.	.	,00	,00
<i>el + N + aquest</i>	,6	,50	,00	1,00	,00	3,00	,4	,00	.	.	,00	2,00
<i>el + N + aquell</i>	,2	,00	.	.	,00	1,00	,1	,00	.	.	,00	1,00
<i>aquest aquell</i>	2,2	2,00	1,00	5,00	,00	5,00	3,0	1,00	1,00	4,00	,00	13,00
<i>(un +) N + d'aquests</i>	,8	,50	,00	1,00	,00	5,00	,6	1,00	1,00	1,00	,00	1,00
<i>(un +) N + d'aquells</i>	,1	,00	.	.	,00	2,00	,1	,00	.	.	,00	1,00

Annexos

Edat 16 anys Sexe Masculí L1 Català i castellà						
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>aquest + N</i>	2,5	2,50	2,00	3,00	2,00	3,00
<i>aquell + N</i>	,0	,00	.	.	,00	,00
<i>el + N + aquest</i>	1,0	1,00	.	.	1,00	1,00
<i>el + N + aquell</i>	,0	,00	.	.	,00	,00
<i>aquest</i>	5,5	5,50	4,00	7,00	4,00	7,00
<i>aquell</i>	,0	,00	.	.	,00	,00
<i>(un +) N + d'aquests</i>	,0	,00	.	.	,00	,00
<i>(un +) N + d'aquells</i>	,5	,50	,00	1,00	,00	1,00

Annex 4.72. Estudi estadístic exploratori de V23 (sexe=femení, edat=16 anys).

	Edat 16 anys Sexe Femení L1 Català						Edat 16 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>aquest + N</i>	2,2	1,00	,00	6,00	,00	10,00	1,9	1,00	1,00	4,00	,00	7,00
<i>aquell + N</i>	,5	,00	.	.	,00	4,00	,2	,00	.	.	,00	2,00
<i>el + N + aquest</i>	,4	,00	.	.	,00	2,00	,3	,00	.	.	,00	1,00
<i>el + N + aquell</i>	,2	,00	.	.	,00	1,00	,1	,00	.	.	,00	2,00
<i>aquest</i>	3,5	2,00	1,00	6,00	,00	13,00	2,6	1,00	1,00	3,00	,00	13,00
<i>aquell</i>	,1	,00	.	.	,00	1,00	,1	,00	.	.	,00	1,00
<i>(un +) N + d'aquests</i>	,1	,00	.	.	,00	1,00	,8	,00	.	.	,00	5,00
<i>(un +) N + d'aquells</i>	,1	,00	.	.	,00	1,00	,0	,00	.	.	,00	,00

Annexos

Edat 16 anys Sexe Femení L1 Català i castellà						
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>aquest + N</i>	1,7	2,00	,00	3,00	,00	4,00
<i>aquell + N</i>	,1	,00	.	.	,00	1,00
<i>el + N + aquest</i>	,5	,00	.	.	,00	2,00
<i>el + N + aquell</i>	,1	,00	.	.	,00	1,00
<i>aquest</i>	2,9	2,00	2,00	7,00	1,00	7,00
<i>aquell</i>	,3	,00	.	.	,00	1,00
<i>(un +) N + d'aquests</i>	2,8	1,50	,00	3,00	,00	13,00
<i>(un +) N + d'aquells</i>	,1	,00	.	.	,00	1,00

Annex 4.73. Estudi estadístic exploratori de V25 (sexe=masculí, edat=12 anys).

	Edat 12 anys Sexe Masculí L1 Català						Edat 12 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>els hi</i>	1,4	1,00	1,00	3,00	,00	5,00	,6	,00	.	.	,00	6,00
<i>els</i>	,0	,00	.	.	,00	,00	,3	,00	.	.	,00	2,00
<i>li</i>	,1	,00	.	.	,00	1,00	,5	,00	.	.	,00	2,00
<i>lis</i>	,0	,00	.	.	,00	,00	,1	,00	.	.	,00	1,00

	Edat 12 anys Sexe Masculí L1 Català i castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>els hi</i>	2,0	2,00	1,00	3,00	1,00	3,00
<i>els</i>	,5	,50	,00	1,00	,00	1,00
<i>li</i>	,5	,50	,00	1,00	,00	1,00
<i>lis</i>	,5	,50	,00	1,00	,00	1,00

Annexos

Annex 4.74. Estudi estadístic exploratori de V25 (sexe=femení, edat=12 anys).

	Edat 12 anys Sexe Femení L1 Català						Edat 12 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>els hi</i>	,5	,00	.	.	,00	3,00	,4	,00	.	.	,00	3,00
<i>els</i>	,0	,00	.	.	,00	,00	,6	,00	.	.	,00	6,00
<i>li</i>	,4	,00	.	.	,00	4,00	1,4	,00	.	.	,00	10,00
<i>lis</i>	,0	,00	.	.	,00	,00	,3	,00	.	.	,00	2,00
	Edat 12 anys Sexe Femení L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>els hi</i>	1,5	,50	,00	1,00	,00	9,00						
<i>els</i>	,4	,00	.	.	,00	2,00						
<i>li</i>	,5	,00	.	.	,00	4,00						
<i>lis</i>	,3	,00	.	.	,00	2,00						

Annex 4.75. Estudi estadístic exploratori de V25 (sexe=masculí, edat=16 anys).

	Edat 16 anys Sexe Masculí L1 Català						Edat 16 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>els hi</i>	1,6	,50	,00	1,00	,00	16,00	,6	,00	.	.	,00	5,00
<i>els</i>	,1	,00	.	.	,00	1,00	,3	,00	.	.	,00	2,00
<i>li</i>	,7	,00	.	.	,00	4,00	,7	,00	.	.	,00	5,00
<i>lis</i>	,0	,00	.	.	,00	,00	,1	,00	.	.	,00	1,00

	Edat 16 anys Sexe Masculí L1 Català i castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>els hi</i>	,0	,00	.	.	,00	,00
<i>els</i>	,0	,00	.	.	,00	,00
<i>li</i>	,5	,50	,00	1,00	,00	1,00
<i>lis</i>	,0	,00	.	.	,00	,00

Annexos

Annex 4.76. Estudi estadístic exploratori de V25 (sexe=femení, edat=16 anys).

	Edat 16 anys Sexe Femení L1 Català						Edat 16 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>els hi</i>	1,7	1,00	,00	3,00	,00	6,00	1,6	,00	.	.	,00	14,00
<i>els</i>	,1	,00	.	.	,00	1,00	,5	,00	.	.	,00	3,00
<i>li</i>	,5	,00	.	.	,00	2,00	1,0	,00	.	.	,00	5,00
<i>lis</i>	,0	,00	.	.	,00	,00	,2	,00	.	.	,00	1,00

	Edat 16 anys Sexe Femení L1 Català i castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>els hi</i>	1,0	,00	.	.	,00	4,00
<i>els</i>	,4	,00	.	.	,00	3,00
<i>li</i>	1,3	,50	,00	3,00	,00	4,00
<i>lis</i>	,0	,00	.	.	,00	,00

Annex 4.77. Estudi estadístic exploratori de V38 (sexe=masculí, edat=12 anys).

	Edat 12 anys Sexe Masculí L1 Català						Edat 12 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
Importacions	43,0	28,50	21,00	51,00	9,00	162,00	36,3	31,00	25,00	52,00	2,00	70,00
Híbrids	,0	,00	.	.	,00	,00	,3	,00	.	.	,00	3,00
Calcs	,4	,00	.	.	,00	2,00	,6	,50	,00	1,00	,00	2,00
Altres	,0	,00	.	.	,00	,00	,6	,00	.	.	,00	7,00

	Edat 12 anys Sexe Masculí L1 Català i castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
Importacions	42,5	42,50	28,00	57,00	28,00	57,00
Híbrids	,5	,50	,00	1,00	,00	1,00
Calcs	,5	,50	,00	1,00	,00	1,00
Altres	,0	,00	.	.	,00	,00

Annexos

Annex 4.78. Estudi estadístic exploratori de V38 (sexe=femení, edat=12 anys).

	Edat 12 anys Sexe Femení L1 Català						Edat 12 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
Importacions	39,3	41,00	22,00	51,00	8,00	89,00	39,3	35,50	31,00	45,00	7,00	72,00
Híbrids	,0	,00	.	.	,00	,00	,2	,00	.	.	,00	1,00
Calcs	,2	,00	.	.	,00	1,00	,5	,00	.	.	,00	3,00
Altres	,0	,00	.	.	,00	,00	,1	,00	.	.	,00	1,00

	Edat 12 anys Sexe Femení L1 Català i castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
Importacions	32,0	27,50	11,00	55,00	11,00	61,00
Híbrids	,0	,00	.	.	,00	,00
Calcs	,8	,50	,00	1,00	,00	3,00
Altres	,0	,00	.	.	,00	,00

Annex 4.79. Estudi estadístic exploratori de V38 (sexe=masculí, edat=16 anys).

	Edat 16 anys Sexe Masculí L1 Català						Edat 16 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
Importacions	40,7	29,00	26,00	36,00	16,00	117,00	35,3	32,50	14,00	53,00	9,00	75,00
Híbrids	,1	,00	.	.	,00	1,00	,0	,00	.	.	,00	,00
Calcs	,7	,00	.	.	,00	4,00	,6	,00	.	.	,00	2,00
Altres	,2	,00	.	.	,00	3,00	,1	,00	.	.	,00	1,00

	Edat 16 anys Sexe Masculí L1 Català i castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
Importacions	38,0	38,00	21,00	55,00	21,00	55,00
Híbrids	,0	,00	.	.	,00	,00
Calcs	,0	,00	.	.	,00	,00
Altres	,0	,00	.	.	,00	,00

Annexos

Annex 4.80. Estudi estadístic exploratori de V38 (sexe=femení, edat=16 anys).

	Edat 16 anys Sexe Femení L1 Català						Edat 16 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
Importacions	43,0	42,00	32,00	57,00	17,00	68,00	45,4	47,50	23,00	56,00	12,00	117,00
Híbrids	,0	,00	.	.	,00	,00	,1	,00	.	.	,00	1,00
Calcs	,5	,00	.	.	,00	2,00	,5	,00	.	.	,00	3,00
Altres	,1	,00	.	.	,00	1,00	,1	,00	.	.	,00	1,00

	Edat 16 anys Sexe Femení L1 Català i castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
Importacions	38,9	35,50	30,00	55,00	12,00	57,00
Híbrids	,0	,00	.	.	,00	,00
Calcs	,3	,00	.	.	,00	1,00
Altres	,0	,00	.	.	,00	,00

Annex 4.81. Estudi estadístic exploratori de V41 (sexe=masculí, edat=12 anys).

	Edat 12 anys Sexe Masculí L1 Català						Edat 12 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>color [blau]</i>	,4	,00	.	.	,00	3,00	,4	,00	.	.	,00	3,00
<i>[blau]</i>	,4	,00	.	.	,00	4,00	,5	,00	.	.	,00	6,00
	Edat 12 anys Sexe Masculí L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>color [blau]</i>	,0	,00	.	.	,00	,00						
<i>[blau]</i>	,5	,50	,00	1,00	,00	1,00						

Annexos

Annex 4.82. Estudi estadístic exploratori de V41 (sexe=femení, edat=12 anys).

	Edat 12 anys Sexe Femení L1 Català						Edat 12 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>color [blau]</i>	,5	,00	.	.	,00	3,00	,3	,00	.	.	,00	3,00
<i>[blau]</i>	,6	,00	.	.	,00	3,00	,6	,00	.	.	,00	5,00
	Edat 12 anys Sexe Femení L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>color [blau]</i>	,4	,00	.	.	,00	2,00						
<i>[blau]</i>	,0	,00	.	.	,00	,00						

Annex 4.83. Estudi estadístic exploratori de V41 (sexe=masculí, edat=16 anys).

	Edat 16 anys Sexe Masculí L1 Català						Edat 16 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>color [blau]</i>	,2	,00	.	.	,00	2,00	,2	,00	.	.	,00	2,00
<i>[blau]</i>	,8	,00	.	.	,00	3,00	,1	,00	.	.	,00	1,00
	Edat 16 anys Sexe Masculí L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>color [blau]</i>	,0	,00	.	.	,00	,00						
<i>[blau]</i>	,0	,00	.	.	,00	,00						

Annexos

Annex 4.84. Estudi estadístic exploratori de V41 (sexe=femení, edat=16 anys).

	Edat 16 anys Sexe Femení L1 Català						Edat 16 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
<i>color [blau]</i>	,1	,00	.	.	,00	1,00	,2	,00	.	.	,00	2,00
<i>[blau]</i>	2,0	,00	.	.	,00	7,00	,2	,00	.	.	,00	1,00
	Edat 16 anys Sexe Femení L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
<i>color [blau]</i>	,3	,00	.	.	,00	1,00						
<i>[blau]</i>	,5	,00	.	.	,00	3,00						

Annex 4.85. Estudi estadístic exploratori de V45 (sexe=masculí, edat=12 anys).

	Edat 12 anys Sexe Masculí L1 Català						Edat 12 anys Sexe Masculí L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
N. adv. -ment	2,3	1,00	1,00	5,00	,00	8,00	1,7	2,00	2,00	4,00	,00	6,00
	Edat 12 anys Sexe Masculí L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
N. adv. -ment	.0	,00	.	.	,00	,00						

Annexos

Annex 4.86. Estudi estadístic exploratori de V45 (sexe=femení, edat=12 anys).

	Edat 12 anys Sexe Femení L1 Català						Edat 12 anys Sexe Femení L1 Castellà					
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
N. adv. -ment	1,2	,00	.	.	,00	5,00	1,1	,00	.	.	,00	12,00
	Edat 12 anys Sexe Femení L1 Català i castellà											
	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim						
N. adv. -ment	3,8	,00	.	.	,00	16,00						

Annex 4.87. Estudi estadístic exploratori de V45 (sexe=masculí, edat=16 anys).

Edat 16 anys Sexe Masculí L1 Català							Edat 16 anys Sexe Masculí L1 Castellà					
Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim		Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
N. adv. -ment	4,9	3,50	1,00	6,00	,00	15,00	2,5	1,50	1,00	4,00	,00	9,00
							Edat 16 anys Sexe Masculí L1 Català i castellà					
							Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim
N. adv. -ment							3,5	3,50	,00	7,00	,00	7,00

Annexos

Annex 4.88. Estudi estadístic exploratori de V45 (sexe=femení, edat=16 anys).

Edat 16 anys Sexe Femení L1 Català						Edat 16 anys Sexe Femení L1 Castellà								
Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim	Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim			
N. adv. -ment	3,3	2,00	,00	4,00	,00	17,00	2,6	2,00	2,00	5,00	,00	7,00		
						Edat 16 anys Sexe Femení L1 Català i castellà								
						Mitjana	Mediana	95,0% CL inferior de mediana	95,0% superior de mediana	Mínim	Màxim			
						N. adv. -ment	8,0	5,50	3,00	16,00	1,00	19,00		

Annex 5. Matriu de correlacions entre les variables de l'estudi univariànt.

	L1	Edat	Sexe	TV3B	TV3C	TV3F	TV3H	TV3I	TV3J	TV3K	TV3L	TV5	TV6	TV16	TV18	TV22	TV23	TV25	TV38	TV41	TV45
L1	1	0	0,1	0	-0,1	-0,1	-0,2	-0,1	-0,2	0,1	0	0,1	-0,1	-0,2	-0,3	-0,1	0	0,1	0	-0,2	-0,1
Edat	0	1	0	0,2	0,2	0	0,2	0,3	0,2	-0,1	0	0,1	0,3	0,1	0,1	0	0,1	0,1	0	0	0,3
Sexe	0,1	0	1	0,2	0,1	-0,1	-0,1	0,1	0,1	0,2	0,1	0,1	-0,2	-0,1	-0,1	0,2	0	0,1	0	0	0
TV3B	0	0,2	0,2	1	0,1	-0,2	0,1	0,2	0,2	0,2	0,1	-0,1	-0,1	0	0,2	0,2	0,2	0,2	0	0	0,3
TV3C	-0,1	0,2	0,1	0,1	1	0,2	0	0,3	0,4	0,2	0,4	0,1	0,2	0,1	0,3	0,1	0,3	0,2	0,7	0,1	0,1
TV3F	-0,1	0	-0,1	-0,2	0,2	1	0,1	0,1	0,3	0	0,3	-0,1	0,1	0,2	0,2	0,2	0,2	0,1	0,3	0	0
TV3H	-0,2	0,2	-0,1	0,1	0	0,1	1	0,3	0,2	0	-0,1	-0,1	0,1	0,2	0,3	0	0,2	0	0	-0,1	0,2
TV3I	-0,1	0,3	0,1	0,2	0,3	0,1	0,3	1	0,4	0,2	0,3	0	0,2	0,3	0,3	0,3	0,2	0,2	0,3	0	0,2
TV3J	-0,2	0,2	0,1	0,2	0,4	0,3	0,2	0,4	1	0,2	0,4	-0,1	0,2	0,3	1	1	0,4	0,3	0,4	0	0,3
TV3K	0,1	-0,1	0,2	0,2	0,2	0	0	0,2	0,2	1	0,2	-0,1	-0,1	0,2	0,3	0,3	0,2	0,3	0,4	0,2	0,2
TV3L	0	0	0,1	0,1	0,4	0,3	-0,1	0,3	0,4	0,2	1	0	0,2	0,3	0,2	0,3	0,3	0,2	0,5	0	0,2
TV5	0,1	0,1	0,1	-0,1	0,1	-0,1	-0,1	0	-0,1	-0,1	0	1	0	0	-0,1	-0,1	-0,1	0	0	-0,1	0
TV6	-0,1	0,3	-0,2	-0,1	0,2	0,1	0,1	0,2	0,2	-0,1	0,2	0	1	0,2	0,1	0,1	0,1	0,1	0,1	0	0,1
TV16	-0,2	0,1	-0,1	0	0,1	0,2	0,2	0,3	0,3	0,2	0,3	0	0,2	1	0,2	0,2	0,1	-0,1	0,1	0	0,2
TV18	-0,3	0,1	-0,1	0,2	0,3	0,2	0,3	0,3	1	0,3	0,2	-0,1	0,1	0,2	1	0,3	0,3	0,3	0,4	0,1	0,4
TV22	-0,1	0	0,2	0,2	0,1	0,2	0	0,3	1	0,3	0,3	-0,1	0,1	0,2	0,3	1	0,3	0,2	0,3	0	0,1
TV23	0	0,1	0	0,2	0,3	0,2	0,2	0,2	0,4	0,2	0,3	-0,1	0,1	0,1	0,3	0,3	1	0,3	0,3	0	0,3
TV25	0,1	0,1	0,1	0,2	0,2	0,1	0	0,2	0,3	0,3	0,2	0	0,1	-0,1	0,3	0,2	0,3	1	0,3	0,1	0,2
TV38	0	0	0	0	1	0,3	0	0,3	0,4	0,4	0,5	0	0,1	0,1	0,4	0,3	0,3	0,3	1	0,2	0,1
TV41	-0,2	0	0	0	0,1	0	-0,1	0	0	0,2	0	-0,1	0	0	0,1	0	0	0,1	0,2	1	0,2
TV45	-0,1	0,3	0	0,3	0,1	0	0,2	0,2	0,3	0,2	0,2	0	0,1	0,2	0,4	0,1	0,3	0,2	0,1	0,2	1
D	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
V. P.	4,5	2	1,7	1,5	1,3	1,2	1	1	0,9	0,8	0,8	0,7	0,6	0,6	0,6	0,5	0,4	0,4	0,4	0,3	0,2

D: Dimensió; V. P.: Valors propis
 Valors arrodonits per qüestions d'espai.

Annex 6. Matriu de correlacions entre variables per a V23.

	L1	Edat	Sexe	V23a	V23b	V23c	V23d	V23e	V23f	V23g	V23h
L1	1	0	0,2	-0,2	-0,2	0	-0,2	0,1	-0,1	0,2	-0,1
Edat	0	1	0	0,2	0,1	0	0	-0,1	0	0	0
Sexe	0,2	0	1	-0,1	0	-0,1	0	0	-0,1	0	-0,1
V23a	-0,2	0,2	-0,1	1	0,1	0,1	0	0,2	0	0	-0,1
V23b	-0,2	0,1	0	0,1	1	-0,1	0,2	0,1	0,2	-0,1	0,1
V23c	0	0	-0,1	0,1	-0,1	1	0,1	0,1	0	0,2	-0,1
V23d	-0,2	0	0	0	0,2	0,1	1	0,1	0,4	0,1	0,4
V23e	0,1	-0,1	0	0,2	0,1	0,1	0,1	1	0,1	0,1	-0,1
V23f	-0,1	0	0	0	0,2	0	0,4	0,1	1	0	0,2
V23g	0,2	0	0	0	-0,1	0,2	0,1	0,1	0	1	0
V23h	-0,1	0	-0,1	-0,1	0,1	-0,1	0,4	-0,1	0,2	0	1
D	1	2	3	4	5	6	7	8	9	10	11
V. P.	1,9	1,4	1,4	1,2	1	0,9	0,8	0,7	0,7	0,6	0,5

V23a: *aquest* + N; V23b: *aquell* + N; V23c: *el* + N + *aquest*; V23d: *el* + N + *aquell*; V23e: *aquest*; V23f: *aquell*; V23g: *(un +) N* + *d'aquests*; V23h: *(un +) N* + *d'aquells*
D: Dimensió; V. P.: Valors propis
Valors arrodonits per qüestions d'espai.

© Roser Giménez García, 2022.

