

Amb aquest escrit declaro que sóc l'autor/autora original d'aquest treball i que no he emprat per a la seva elaboració cap altra font, incloses fonts d'Internet i altres mitjans electrònics, a part de les indicades. En el treball he assenyalat com a tals totes les citacions, literals o de contingut, que procedeixen d'altres obres. Tinc coneixement que d'altra manera, i segons el que s'indica a l'article 18, del capítol 5 de les Normes reguladores de l'avaluació i de la qualificació dels aprenentatges de la UB, l'avaluació comporta la qualificació de "Suspens".

Barcelona, a 18 de juny de 2023

Signatura:



Sofía Rodríguez García

Agraïments

Anem per parts.

-A la meua família; m'heu suportat econòmicament durant tot el temps a fora i vau respectar i celebrar la meua decisió d'anar a estudiar una carrera que potser no és la que tot pare desitjaria pel seu fill. Gràcies per la vostra confiança en mi. A Abel, gràcies per portar-me a l'aeroport per anar a Donostia, i a Martín, per fer la portada del treball.

-A les meues tutores del TFG, a Itziar per l'entusiasme amb què vas rebre que volguera fer un TFG "sobre euskara i lingüística computacional, però no sé per on encaminar-lo" i a Begoña per donar la idea del que ha sigut finalment el treball; a mi no se m'haguera acudit.

-A l'Etxepare Euskal Institutua per becar el meu treball i sobretot al grup Ixa de l'EHU pel caliu amb què em vau rebre durant la meua estada allà. Van ser poquets dies, però per mi l'experiència va ser un regal.

-Als meus amics de Barcelona per tornar aquest poti-poti irrespirable en una cosa prou pareguda a una llar. Vull donar especialment les gràcies a la meua companya de pis per tindre cura de mi durant mesos i evitar que caiguera sovint en bucles de planys i plors.

-A la meua parella i als meus amics de Castelló. Trobar-vos a faltar em dona una força que no imaginava tindre abans de deixar-vos.

Resum

Aquest treball presenta una anàlisi quantitativa del lèxic i de les mètriques de llegibilitat dels comunicats d'ETA al període 2006-2018 en castellà i èuscar. L'objectiu és establir si existeixen diferències de llegibilitat entre els comunicats de manera diacrònica, així com comparar els comunicats en castellà i en èuscar entre si, i també contrastar els comunicats en castellà, diferenciant els que són originals d'ETA en castellà i els que són traduccions d'altres mitjans de comunicació. A més a més, hem analitzat el lèxic quantitativament per relacionar-lo amb el context temporal dels comunicats i amb el llenguatge de la política o el nacionalisme. Per realitzar l'estudi, hem desenvolupat un corpus bilingüe dels comunicats disponibles a les dues llengües al període temporal esmentat. El resultat ha estat la creació del corpus bilingüe en èuscar i castellà d'ETA més extens que existeix actualment, amb un total de 36 textos; 18 a cada llengua.

Paraules clau: Llegibilitat, *translationese*, anàlisi quantitatiu, freqüència, ETA.

Laburpena

Lan honek ETAREN 2006-2018 aldiko komunikatuen lexikoaren eta irakurgarritasun-metriken azterketa kuantitatiboa aurkezten du, gaztelaniaz eta euskaraz. Helburua da zehaztea ea irakurgarritasun-desberdintasunik dagoen komunikatuen artean diakronikoki, bai eta gaztelaniazko komunikatuak eta euskarazkoak elkarren artean konparatzea ere, eta gaztelaniazko komunikatuak kontrastatzea, ETAREN jatorrizkoak eta beste komunikabide batzuen itzulpenak bereiziz. Gainera, lexikoa kuantitatiboki aztertu dugu komunikatuen denbora-testuinguruarekin eta politikaren edo nazionalismoaren hizkuntzarekin lotzeko. Azterketa egiteko, aipatutako aldian bi hizkuntzetan eskuragarri dauden komunikatuen corpus elebiduna garatu dugu. Horren ondorioz, ETAREN euskarazko eta gaztelaniazko corpus elebiduna sortu da, egungoa baino zabalagoa. Guztira 36 testu daude, 18 hizkuntza bakoitzean.

Hitz gakoak: Irakurgarritasuna, *translationese*, azterketa kuantitatiboa, maiztasuna, ETA.

Índex

Introducció:.....	7
1. Marc teòric:	9
1.1. Context històric	9
1.2. La llengua d'ETA a la seua producció escrita	11
1.3. Els comunicats, importància i difusió	12
1.4. Rellevància de l'estudi: eix temporal i criteris de disseny del corpus	13
1.5. Anàlisi quantitativa, estratègies de legitimitació.....	13
1.6. Llegibilitat	14
1.6.1. El cas de les traduccions.....	15
1.7. Context i explicació dels comunicats triats.....	16
2. Metodologia	18
2.1. Elaboració del corpus	18
2.2. Processament dels textos	20
2.2.1. MultiAzterTest.....	21
2.2.2. RStudio	22
2.2.3. UdPipe	22
2.2.4. Voyant Tools	23
2.2.5. AntConc.....	23
3. Resultats	24
3.1. Comparació de mètriques de llegibilitat	24
3.1.1. Comparació interlingüística.....	24
3.1.2. Comunicats en castellà, traduccions i originals.....	26
3.1.3. Canvis diacrònics.....	31
3.1.3.1. Resultats en castellà	32
3.1.3.2. Resultats en èuscar.....	33
3.2. Comparació del lèxic.....	35
3.2.1. El cas d'eta i ETA en èuscar.....	36
3.2.2. Evolució de termes en èuscar	38
3.2.2.1. Referències a França i a Espanya	39
3.2.2.2. Solució i diàleg	40
3.2.2.3. Polític i democràtic	41
3.2.2.4. Lluita i conflicte.....	42
3.2.2.5. Camí i pas	43

3.2.2.6. Construir i pau	43
4. Discussió	45
4.1. Comparació interlingüística.....	45
4.2. Comunicats en castellà, traduccions i originals	46
4.3. Canvis diacrònics	46
4.4. Lèxic més freqüent	47
4.5. L'ús d'ETA	48
4.6. Lèxic, estratègies i metàfores	49
5. Conclusions	51
6. Bibliografia.....	52

Introducció:

Tot i que l'any 2011 queda a la memòria col·lectiva com l'any de la fi d'Euskadi Ta Askatasuna, aquesta no es va dissoldre formalment fins a l'any 2018. La història dels seus últims anys és recent i també ho són la majoria dels treballs que se n'han fet. Encara que el tema ha atret l'atenció d'investigadors provinents del món de la comunicació o el periodisme, hi ha una manca d'anàlisis purament lingüístiques. Aquest treball de fi de grau pretén cobrir, dins de les seues xicotetes possibilitats, part d'aquest buit.

Al següent treball, analitzarem els comunicats disponibles en castellà i èuscar des de 2006 fins a la seua desaparició a 2018.

A la primera part, farem un xicotet recorregut pel marc teòric del treball. Primer, mostrarem un breu repàs pel context històric de l'organització des dels seus orígens, i després analitzarem la seua llengua d'ús escrit. A continuació, exposarem la rellevància dels comunicats com a forma de comunicació externa i també explicarem la importància del marc temporal triat i dels criteris de selecció del corpus. Finalment, explicarem els conceptes amb els quals treballarem el corpus: l'anàlisi quantitativa de lèxic i l'avaluació de la llegibilitat; aquesta última, en concret, contrastant també les traduccions i els originals. A més, hem inclòs un apartat on expliquem breument el contingut dels comunicats que hem analitzat amb un petit context històric.

A la segona part, explicarem la metodologia utilitzada: primer, comentarem l'elaboració del corpus, i a continuació, les eines que hem utilitzat per processar-lo, explicant els resultats que esperàvem de cadascuna d'elles.

A la tercera part, exposarem els resultats. Començarem per la part de l'avaluació de la llegibilitat: primer, comparant els nivells extrets en castellà i en èuscar; després, comparant les mètriques entre els comunicats originals en castellà d'ETA i les traduccions de comunicats en èuscar fetes per mitjans de comunicació; finalment, contrastarem les diferències entre els comunicats anteriors i posteriors a l'any 2011, primer en èuscar i posteriorment en castellà. Més tard, mostrarem els resultats relacionats amb el lèxic: primer mostrarem una llista de les paraules més freqüents a cada llengua; a continuació, dedicarem un apartat a explicar com s'han diferenciat les referències a ETA de la conjunció "eta" en èuscar, a banda de la posició als textos de les referències al seu nom; en últim lloc, mostrarem per parelles els mots que hem trobat que mostraven una evolució més relacionada amb el context històric.

A la quarta part, discutirem els resultats obtinguts. Primer, relacionarem els nivells de llegibilitat amb la baixada històrica percebuda de la complexitat del discurs polític. Després, contrastarem els resultats aconseguits pels originals i les traduccions amb el concepte de *translationese*, que desenvoluparem més endavant. A continuació, examinarem la rellevància de les dades a la comparació diacrònica dels comunicats. Finalment, repassarem el lèxic mostrat als resultats i comentarem la seua relació amb l'estil de l'IRA, de Rodríguez Zapatero al seu discurs antiterrorista o de la política occidental en general.

A la cinquena part mostrarem un resum de les conclusions i diverses propostes per estudis posteriors.

1. Marc teòric:

1.1. Context històric

ETA (Euskadi Ta Askatasuna, “Euskadi i Llibertat”) va ser una organització armada nacionalista basca operativa des de l’any 1958 i dissolta oficialment el 3 de maig de 2018. Va ser focus d’una gran cobertura mediàtica durant la major part de la seua activitat i és, juntament amb l’IRA, un dels grups armats nacionalistes més notoris de la història europea recent.

Tot i tenir diverses escissions al llarg del seu recorregut, es va mantenir com a objectiu principal la construcció d’un estat basc socialista independent de França i Espanya, que cobrira l’àrea històricament coneguda com a Euskal Herria i formada per quatre regions pertanyents a l’estat espanyol (Biscaia, Guipúscoa, Àlaba i Navarra) i tres al francès (Lapurdi, Baixa Navarra i Zuberoa).

ETA va nàixer a 1958 com una escissió d’Ekin, un grup d’estudi juvenil de cultura i llengua basques sorgit durant el franquisme, en un principi associat al Partit Nacionalista Basc. ETA va distanciar-se del PNV i va apostar per l’ús de la violència per aconseguir la independència d’una nació basca aconfessional i laica, contrària a la tendència catòlica del PNV. No tenia una proposta com a partit polític, sinó que es definia com “un moviment basc revolucionari d’alliberament nacional, creat a la resistència patriòtica” (Bilbao Ariztimuño, 2008). Tot i això, van passar deu anys entre la creació d’ETA i el seu primer atemptat, a 1968 (González Aparicio, 2017), ja que el seu començament com a grup es caracteritza pel debat entre visions ideològiques que diferien en alguns aspectes, com el de la via armada. Amb l’escissió de la secció més propera al moviment obrer i l’assassinat de Javier Etxebarrieta, autor material de la primera mort d’ETA en disparar contra un agent de la Guàrdia Civil a un control, es va accelerar l’activitat armada posterior, materialitzada el mateix any en l’atemptat contra Melitón Manzananas, cap de la Brigada Político-social a Guipúscoa i responsable de tortures contra opositors al règim a Euskadi.

L’any 1970, es va produir el Procés de Burgos, on es va jutjar a setze membres de la banda per tres assassinats, però la denúncia dels acusats de l’opressió a Euskadi va tenir un gran ressò fins i tot fora d’Espanya, i van acabar sent indultats per Francisco Franco. Posteriorment, el grup més obrerista d’ETA a aquell moment, ETA-VI, se’n va escindir, i el sector militarista restant, ETA-V, va reprendre les sigles originals. Després de

l'atemptat de Carrero Blanco a 1973, ETA torna a sofrir una escissió a 1974 després de l'atemptat a la cafeteria Rolando, que es va saldar amb 13 morts. Van distanciar-se dos grups: ETA politicomilitar, que es va negar a reconèixer l'atemptat, i ETA militar, més partidària a la vida armada, que volia reivindicar-los. Aquesta va ser l'última gran escissió d'ETA: mentre que ETA politicomilitar es va vincular amb el partit socialista Euskadiko Ezkerra, ETA militar va reprendre l'activitat armada amb força després de l'Amnistia General de 1977: 1978, 1979 i 1980 són els anys on es produeixen més atemptats mortals per part d'ETA.

L'any 1978 es va fundar el partit Herri Batasuna, partit abertzale afí a les idees d'ETA militar (ara ja, ETA) i que va adquirir rellevància durant les dècades dels 80 i els 90, fins a la seua il·legalització a 2003 per part del Tribunal Suprem arran de la Llei de Partits de l'any anterior en provar-la com a instrument polític d'ETA. Per una altra banda, diverses accions preses posteriorment per ETA van minvar el suport social que tenia a Euskadi, com va ser l'assassinat de Yoyes (exmembre del grup) a 1986 o, finalment, el de Miguel Ángel Blanco, regidor del Partit Popular d'Ermua, a 1997, arran del qual es van organitzar diverses mostres de repulsa que es van repetir a partir de llavors a atemptats posteriors.

Amb l'arribada al govern de José Luís Rodríguez Zapatero a 2004, es va produir un punt d'inflexió a la història d'ETA (Hernández Valls, 2017). L'executiu socialista va decantar-se per una política de negociació des del principi de la legislatura, i tot i que no va tenir conseqüències a curt termini, va ser determinant en els esdeveniments posteriors. ETA va anunciar un alto el foc permanent en març de 2006 per impulsar el procés democràtic a Euskal Herria, després de gairebé tres anys sense víctimes mortals. Així i tot, el 30 de desembre del mateix any, es va produir un atemptat a la terminal T4 de l'aeroport de Barajas a Madrid al qual van morir dues persones i que ETA va reivindicar poc després com a mostra de desacord amb la política duta a terme pel govern. Alguns experts en la lluita antiterrorista han considerat aquest esdeveniment com el principi del fi de la Banda (Hernández Valls, 2017); les negociacions amb el govern espanyol es van tallar des d'aleshores, i als anys posteriors, es va observar un canvi de tendència a l'esquerra abertzale cap a propostes polítiques pacífiques que es va materialitzar a la declaració d'Altsasu de novembre de 2009. Finalment, el setembre de 2010, ETA anuncia un alto el foc i, un any després, l'octubre de 2011, el cessament definitiu de la seua activitat armada.

Els anys posteriors es van caracteritzar per ser un període d'estancament de la seua activitat; el govern del partit popular de Mariano Rajoy (arribat al poder el 2011) va

declinar negociar-hi com ho haguera fet l'executiu anterior, i la banda va frenar la incorporació de nous membres o les estratègies d'extorsions. El procés de desarmament posterior es va allargar per la falta d'acord i negociació, mentre es continuaven produint desmantellaments d'arsenals, detencions d'alts càrrecs de l'organització i entregues simbòliques d'armaments. Finalment, a abril del 2017, es van acollir a l'estratègia proposada pel govern basc anteriorment de fer l'entrega d'armes a la justícia mitjançant mediadors, i el 7 d'abril van enviar un comunicat a la BBC declarant-se una organització desarmada. Posteriorment, van anunciar el començament d'un debat intern on van plantejar als membres la fi de la banda perquè fora vota i acceptada. A finals d'abril del 2018, ETA va publicar un comunicat lamentant el mal causat durant la seua història, i poc després, a principis de maig, va anunciar la seua dissolució i el desmantellament de totes les seues estructures internes.

1.2. La llengua d'ETA a la seua producció escrita

Durant la seua activitat, va produir una gran quantitat de documentació, ja fora en forma de pamflets, missatges, bolletins interns (com Zutik o Zutabe) o comunicats a premsa. Aquesta producció escrita ha estat fonamentalment en èuscar, corresponent amb l'ús de la llengua pròpia com un dels elements centrals del nacionalisme basc (Garmendia i Urteaga, 2017). Tot i això, a causa del ressò social de la seua activitat a la resta de l'estat espanyol, ha sigut freqüent la presència del castellà, ja fora utilitzat des de dins del grup o a traduccions fetes principalment per mitjans de comunicació per fer arribar aquests a un públic no basc (o no parlant d'èuscar). També ha sigut notable l'ús del francès en dirigir-se a l'estat francès, i en casos més concrets, l'anglès per a comunicats d'abast més internacional.

L'èuscar és, a més, una llengua minoritzada, i la pràctica totalitat dels seus parlants és bilingüe d'aquesta llengua amb el castellà o el francès. A diverses parts del qual es considera actualment l'àrea lingüística històrica d'Euskal Herria hi ha hagut un retrocés de la llengua basca, com ha sigut el cas de la substitució cap al castellà a la major part d'Àlaba des del segle XVI. L'aparició d'ETA s'enquadra, a més, a l'època del franquisme, que es va caracteritzar per una política lingüística homogeneïtzadora que va precaritzar encara més la situació del basc a favor de la supremacia del castellà com llengua de l'estat. Així, ja en la primera assemblea d'ETA, a 1962, l'èuscar es presenta com la llengua pròpia i nacional d'Euskal Herria, però per la situació sociolingüística del moment, s'hi reconeix la presència del castellà i el francès. Tot i això, l'atenció principal

era cap a l'ús de l'èuscar. Arran d'un debat intern durant 2009, es va reconèixer explícitament l'èuscar com llengua d'ús intern de l'organització, mentre que la comunicació externa podia donar-se en altres llengües; així, el coneixement d'èuscar era imprescindible per ocupar càrrecs dins ETA i era precis un compromís d'aprendre'l a curt termini per part dels que no el tingueren com a llengua (Vasco Press, 2009).

Així, és d'esperar que els membres d'ETA i, per tant, els encarregats de la seua producció escrita, tingueren competència en aquestes llengües, a banda de l'èuscar, i que tot i que la major part de la producció fora en aquesta última, podia haver-hi diferents tries lingüístiques per a comunicats externs.

1.3. Els comunicats, importància i difusió

ETA comptava amb una secció o aparell polític encarregat d'elaborar els documents de la banda, tant els bolletins interns com els comunicats externs (Hernández Valls, 2017). Amb el pas dels anys i la caiguda de l'activitat armada, la seua rellevància va anar creixent fins a marcar també l'estratègia i decisions. Els comunicats complien la funció de pressionar als governs espanyol i francès, llançar missatges, incidir a l'opinió pública o comunicar les seues conclusions. El seu propòsit principal era el d'informar sobre les accions i decisions preses, però com passa amb els comunicats d'altres grups com l'IRA, aquesta narració serveix també per a donar una visió positiva d'ells mateixos i una de negativa del contrari, i per tal, de promocionar la seua actuació (Cardillo, 2017).

A banda, els comunicats tenen diferents funcions segons el context al qual es van publicar. González Aparicio (2017) distingeix tres tipus de comunicats segons la seua temàtica a l'última etapa d'ETA: relacionats amb l'estat espanyol, amb la violència (reivindicacions d'atemptats, principis o finals de treves) i identitaris (publicats en festivitats o dies assenyalats, textos sobre la societat bascs). També destaca que, segons la visió de Foucault (1970) del concepte de discurs, els comunicats d'ETA mostren els diferents objectes de desig i poder propis i dels seus receptors: l'estat espanyol i francès i la ciutadania basca. Els autors parlen de temes tabús davant d'actors que tenen el dret d'expressar-se en qualsevol circumstància que els falta a ells.

Després del tancament dels periòdics Egin (1998, ordre del jutge Baltasar Garzón) i Egunkaria (2003, per ordre del jutge Juan del Olmo), el diari Gara ha estat el mitjà principal per la publicació dels comunicats durant els últims anys d'ETA.

1.4. Rellevància de l'estudi: eix temporal i criteris de disseny del corpus

És important destacar diferents aspectes dels textos triats per l'estudi.

En primer lloc, era necessari que foren comunicats amb una versió en èuscar i una altra en castellà. Aquest primer criteri ha reduït de manera considerable la quantitat de textos que s'han pogut seleccionar: la majoria dels textos no compten amb una versió bilingüe sencera, ja que se solien publicar en èuscar i els periòdics en castellà que en feien eco no publicaven una traducció literal d'aquest, sinó una explicació resumida del seu contingut barrejada amb explicacions i nocions pròpies dels autors. No són, per tant, textos amb una correspondència.

En segon lloc, era crucial que tingueren una estructura de comunicat. Les publicacions externes d'ETA no sempre han tingut aquest format, de fet, segons González Aparicio (2017), abans de 1998, ETA utilitzava altres canals de comunicació com els pamflets per enviar els seus missatges, i els comunicats a premsa van ser una estratègia posterior.

En tercer lloc, com hem explicat al primer apartat d'història, el trencament de la treva començada en març de 2006 a finals del mateix any va suposar, segons alguns, el començament de l'últim tram històric d'ETA. Per això, el període històric triat per fer-ne el corpus i anàlisis posteriors és des de 2006 fins a 2018, any de la dissolució de la banda.

Així, la realització d'aquest treball ha tingut com a resultat també la creació del que és, ara com ara i segons el nostre coneixement, el corpus bilingüe èuscar-castellà de comunicats d'ETA més extens desenvolupat, format per 18 comunicats en les dues llengües; per tant, un total de 36 textos.

Tot i això, remarcuem de nou que açò només és una petita mostra dels comunicats totals (normalment en èuscar i sense una traducció literal al castellà o altres llengües) publicats en la franja 2006-2018. Segons Hernández Valls (2017), només des de 2006 al gener de 2014, ETA va publicar 51 comunicats.

1.5. Anàlisi quantitativa, estratègies de legitimació

Aquest treball s'emmarca dins la lingüística de corpus, el qual el diferencia d'altres estudis fets anteriorment sobre els comunicats d'ETA (Hernández Valls, 2017; González Aparicio, 2017).

La lingüística de corpus està relacionada necessàriament amb els mètodes quantitius (García-Miguel, 2022) i una de les mesures importants que proporcionen aquests corpus

són les freqüències d'ocurrència, siga de certes paraules o construccions gramaticals. A més, pel que fa als estudis diacrònics, els estudis de freqüències més interessants estan relacionats amb l'augment o disminució d'ús al llarg del temps d'algun terme. Si ens trobem, com és el cas, davant de diferents textos individuals relativament no molt llargs, allò que pot variar més entre cadascú seran els elements lèxics, condicionats per la temàtica del text (García-Miguel, 2022). Com passa a Bustos Gavilán (2022), un acostament quantitatiu a un corpus de textos mirat de manera diacrònica pot reflectir l'evolució de temàtiques i d'estratègies lingüístiques usades pels autors.

Diversa literatura anterior sobre el discurs al terrorisme o a la política explica com les diferents estratègies discursives utilitzades tenen com a objectiu la legitimació de les accions i decisions preses (El-Nashar i Nayef, 2019; Abdelrahim, 2019; Hellín-García, 2013; Koteyko i Ryazanova-Clarke, 2009).

L'estudi quantitatiu servirà per a relacionar algunes d'aquestes estratègies amb el lèxic emprat: en concret, estudiarem també la presència de metàfores de construcció i viatge, que s'han vist molt emprades al discurs polític europeu (Koteyko i Ryazanova-Clarke, 2009), i, en concret, al discurs antiterrorista de J.R. Zapatero durant el seu mandat (Hellín-García, 2013).

1.6. Llegibilitat

La complexitat textual és un constructe multidimensional dels textos i una qualitat estructural independent del contingut d'aquests (Tolochko, Song i Boomgaarden, 2019). És esperable que una informació complexa requereixca més recursos cognitius per ser processada i retinuda: per tant, un missatge més simple és més fàcil de processar i ser comprés pel lector. A estudis anteriors, s'hi ha vist que la complexitat textual afecta al coneixement factual extret (Eveland i Cortese, 2004; Kleinnijenhuis, 1991).

En textos de temàtica política, això pot tenir conseqüències com una major possibilitat de compromís i aprenentatge per part del lector (Tolochko, Song i Boomgaarden, 2019). De fet, s'ha observat que la complexitat del discurs polític pareix haver caigut als últims anys, de manera que caldria un menor nivell poder comprendre'ls (Benoit, Munger i Spirling, 2019). Així, una baixada de la sofisticació lingüística tindria com a benefici una major claredat pel lector.

A Benoit, Munger i Spirling (2019), es diferencien diversos conceptes a l'hora d'explicar aquest canvi de tendència dels textos polítics. Per una part, la sofisticació està relacionada

amb el grau de complexitat del sistema de creences de l'individu, i no té per què tenir una representació directa en l'àmbit textual. Per una altra, existeix la complexitat textual, definida com una característica pròpia del text o "la quantitat de recursos computacionals necessaris per a processar un document" (Gibson, 1998, p.2). No obstant això, Benoit, Munger i Spirling (2019) discuteixen que la mesura de la llegibilitat o dificultat textual (*readability*) podria combinar d'alguna manera tots dos conceptes (com és de complex allò del qual parla, o sofisticació, i com és de complexa la manera en la qual ho fa, o complexitat), de manera que mesuràriem directament la dificultat a l'hora de llegir i comprendre un text. Aquesta llegibilitat es compon de factors que influeixen com entén el lector el text (Souza i Borba, 2021), i no ha d'entendre's, en un principi, com una marca d'adequació o de qualitat per part del text, sinó que revela els elements que podrien dificultar-ne la interpretació (Souza i Borba, 2021).

Algunes de les mètriques tradicionals per la llegibilitat, aplicada en un primer moment a la compressió de textos educatius a l'escola, inclouen la llargària de la frase, el nombre de síl·labes per paraula, les classes gramaticals presents o la (mitjana) familiaritat dels mots (Benoit, Munger i Spirling, 2019), però s'ha qüestionat fins a quin punt és possible aplicar les mateixes mètriques de llegibilitat per diferents tipus de text, sense tindre en compte el context amb el qual s'utilitzen (Souza i Borba, 2021). Altres mètriques proposades a Benoit, Munger i Spirling (2019) per a textos polítics en específic serien l'ús de paraules més llargues (relacionat amb la mètrica tradicional del nombre de síl·labes per paraula), l'ús de paraules relativament desconegudes (familiaritat dels mots), llargària de les frases i estructures sintàctiques i gramaticals més complexes (per exemple, l'ús de subordinades).

1.6.1. El cas de les traduccions

Diversos estudis (Baroni i Bernardini, 2006; Kurokawa, Goutte i Isabelle, 2009) exploren com la llengua original té un impacte en la qualitat dels textos traduïts i parlen d'una propietat de les traduccions anomenada *translationese*. Aquesta provoca que l'experiència de lectura d'un text traduït varie respecte a la d'un text en la llengua original (Kurokawa, Goutte i Isabelle, 2009). En aquest sentit, la *translationese* també és l'empremta que deixa la llengua original en la traducció posterior (Baroni i Bernardini, 2006), i diverses hipòtesis apunten que les traduccions comparteixen algunes característiques estructurals que les fan detectables mitjançant la distribució de paraules funcionals, categories morfosintàctiques, pronoms personals i adverbis. Alguns d'aquests

trets serien una menor densitat lèxica, la freqüència de bigrames formats per paraules funcionals o la presència de textos més conservadors i explícits.

Pel que fa a la complexitat textual o llegibilitat, s'ha observat que les traduccions, comparades amb els textos no traduïts, tenen menys complexitat, la qual cosa té un reflex, especialment, a la subordinació i la complexitat de les frases (Liu i Afzaal, 2021). Tot i això, la variació de la complexitat sintàctica també està influenciada pel gènere textual de la traducció, de manera que la prosa genèrica o l'escriptura acadèmica traduïdes solen perdre complexitat, mentre que la ficció traduïda en guanya.

1.7. Context i explicació dels comunicats triats

A continuació, mostrarem una petita explicació de les dates on es van publicar els comunicats que apareixen a l'estudi i dels seus continguts:

- 22-03-2006: ETA anuncia un alto el foc permanent arran del procés de diàleg amb el govern espanyol, autoritzat des de maig de 2005.
- 23-03-2006: Es publica una extensió del comunicat, amb la mateixa temàtica.
- 24-09-2006: ETA llig un comunicat per commemorar, com a homenatge, el Gudari Eguna (el dia del soldat).
- 04-01-2007: ETA reivindica l'atemptat de Barajas produït l'anterior 30 de desembre de 2006.
- 05-06-2007: ETA posa fi oficialment a la treva i torna a l'activitat armada.
- 31-12-2009: ETA publica un comunicat on reflexiona sobre el paper de l'esquerra abertzale, arran de la declaració d'Altsasu publicada l'anterior 14 de novembre, on aquesta parlava de la necessitat de la fi de la violència armada.
- 04-04-2010: ETA publica un comunicat explicant els fets que van envoltar la mort de la seua última víctima, un agent de la policia francesa, durant un tiroteig l'anterior 16 de març.
- 05-09-2010: ETA anuncia la seua decisió de no portar a terme més accions armades ofensives i de participar en un canvi polític cap a un procés democràtic.
- 19-09-2010: ETA es reafirma en el fi de les accions armades i mostra el seu acord amb la Declaració de Brussel·les, presentada el 29 de març del mateix any al Parlament Europeu, on diversos membres expressaven el seu suport cap a la nova via presa per l'esquerra abertzale.
- 08-01-2011: ETA anuncia un alto el foc permanent, general i verificable.

- 20-10-2011: ETA anuncia el cessament definitiu de la seua activitat armada.
- 04-2012: ETA es dirigeix a la Unió Europea per demanar-los que pressionen perquè s'obre un diàleg directe amb els governs espanyol i francès.
- 17-03-2013: ETA publica un comunicat on recrimina al govern espanyol la seua negativa a negociar amb la cúpula d'ETA a Oslo.
- 27-09-2015: ETA publica un comunicat al Gudari Eguna on declara mantenir els seus compromisos, tot i que avisen que la detenció de la seua cúpula l'anterior 22 de setembre pot complicar el procés de desarmament.
- 27-03-2016: En un comunicat publicat a l'Aberri Eguna (el dia de la pàtria basca), ETA reconeix no ser al lloc on desitjaria per no haver-se iniciat el procés de diàleg i pau necessari amb el govern.
- 07-04-2017: ETA envia un comunicat a la BBC on es declara com una organització desarmada davant la comunitat internacional, entregant la totalitat del seu armament l'endemà.
- 08-04-2018: ETA publica un comunicat on demana perdó a les víctimes i reconeix la seua responsabilitat al mal patit durant el recorregut de la seua activitat armada.
- 03-05-2018: ETA anuncia la seua dissolució com a organització.

2. Metodologia

2.1. Elaboració del corpus

Per la creació del nostre corpus, ens hem guiat amb diferents criteris, tant d'historiogràfics com de més tècnics.

La primera decisió va ser el rang de temps per cobrir abans de començar la recopilació de textos. Es va decidir el període de 2006 a 2018 basant-nos en criteris d'altres treballs (Hernández Valls, 2017), que la consideren com l'última etapa d'ETA. A més, és sobretot a partir del començament del segle XXI que els seus comunicats a premsa comencen a ser una manera més regular de comunicació estàndard (González Aparicio, 2017).

Amb una temporalitat ja acotada, els textos havien de complir altres requisits; en concret, que foren textos digitalitzats, pel fet que per la magnitud del treball no podíem assumir la càrrega de buscar i digitalitzar textos només disponibles en paper; que foren textos disponibles en línia, ja que aquests eren els més accessibles a l'hora de fer recerca; i que foren textos disponibles en èuscar i en castellà (fora aquesta una versió produïda per ETA mateixa o bé una traducció d'algun periòdic o col·lectiu a partir de l'original en èuscar).

Ens vam centrar, a més, en les produccions d'ETA en forma de "comunicat"; és a dir, aquells documents que feien arribar a la premsa per tal de ser publicats i llegits per la població (ja fora aquesta els ciutadans d'Euskal Herria o de l'estat espanyol, o bé d'altres com figures públiques de la política o la Comunitat Europea). Per tant, vam descartar seleccionar altres produccions escrites d'ETA durant aquesta època que pertanyeren, per exemple, a Zutabe o altres documents de distribució interna.

La recerca de comunicats en línia amb aquests criteris va tenir com a resultat un corpus bilingüe format per 36 textos en èuscar i castellà (18 comunicats amb la versió en cada llengua), de l'any 2006 fins a l'any 2018; alguns anys no han quedat coberts, ja que no s'hi ha trobat comunicats en la versió en castellà, generalment, mentre que a altres anys se'n trobava més d'un. Com hem explicat anteriorment, és el corpus bilingüe en èuscar i castellà de comunicats d'ETA més extens fins ara.

A continuació, mostrem la llista de comunicats del corpus amb la seua data (en format dia, mes i any) i en ordre cronològic, primer amb la versió en èuscar i després en castellà. Els títols corresponen als posats per ETA en èuscar i en castellà si són originals seus, i en cas de ser comunicats en castellà d'altres fonts, hem posat els títols triats per aquestes.

Les dates corresponen a les declarades per ETA per cada comunicat, no a les dates de publicació en els diferents mitjans.

1. ETAREN Mezua Euskal Herriari, 22/03/2006.
2. Mensaje de Euskadi Ta Askatasuna al pueblo vasco, 22/03/2006.
3. ETAREN Adierazpena Euskal Herriari, 23/03/2006.
4. Segundo mensaje de Euskadi Ta Askatasuna al pueblo vasco, 23/03/2006.
5. ETAREN Mezua Gudari Egunean, 24/09/2006.
6. Mensaje de ETA en el día del guerrero vasco, 24/09/2006.
7. Euskadi Ta Askatasunaren Argiria Euskal Herriari, 04/01/2007.
8. Texto Íntegro del comunicado, 04/01/2007.
9. Euskadi Ta Askatasunaren Agiria Euskal Herriari, 05/06/2007.
10. Comunicado de ETA, 05/06/2007.
11. ETAREN agiria Euskal Herriari, 31/12/2009.
12. Comunicado de ETA a Euskal Herria, 31/12/2009.
13. Euskadi Ta Askatasunaren Agiria Euskal Herriari, 04/04/2010.
14. Traducción al castellano del comunicado de ETA, 04/04/2010.
15. Euskadi Ta Askatasunaren adierazpena, 05/09/2010.
16. Declaración de ETA, 05/09/2010.
17. Euskadi Ta Askatasunaren Agiria Nazioarteari, 19/09/2010.
18. Comunicado de ETA a la Comunidad Internacional, 19/09/2010.
19. ETAREN Adierazpena, 08/01/2011.
20. Declaración de ETA, 08/01/2011.
21. ETAREN Adierazpena, 20/10/2011.
22. Declaración de ETA, 20/10/2011.
23. ETAREN Agiria Europako Komunitateari, 04/2012.
24. Comunicado de ETA a la Comunidad Europea, 04/2012.
25. ETAREN Agiria Euskal Herriari, 17/03/2013.
26. Comunicado de ETA, 17/03/2013.
27. Euskadi Ta Askatasunaren Agiria Euskal Herriari, 27/09/2015.
28. Comunicado de Euskadi Ta Askatasuna al Pueblo Vasco, 27/09/2015.
29. Euskadi Ta Askatasunaren Agiria Aberri Egunean, 27/03/2016.
30. Comunicado de Euskadi Ta Askatasuna en el día del Aberri Eguna, 27/03/2016.
31. ETAREN agiria armagabetzeaz, 07/04/2017.

32. Comunicado íntegro de ETA en el que se declara ‘desarmada’, 07/04/2017.
33. ETAk Euskal Herriari eragindako kalteari buruzko adierazpena, 08/04/2018.
34. ETA al pueblo vasco, declaración sobre el daño causado, 08/04/2018.
35. ETAk Euskal Herriari, azken adierazpena, 03/05/2018.
36. Declaración final de ETA al pueblo vasco, 03/05/2018.

2.2. Processament dels textos

Per preparar els textos per una anàlisi posterior, vam emmagatzemar cadascú en un arxiu de text pla i vam identificar-los al nom per la data i la llengua. Després, vam separar els textos per frases; és a dir, introduint un salt de línia després de cada punt per tenir un text final constituït per frases diferenciades.

Vam introduir les nostres dades en diferents eines per portar a terme anàlisis de diversos tipus:

- MultiAzterTest, per extraure mètriques quantitatives i qualitatives del nivell de llengua o llegibilitat dels comunicats en èuscar i en castellà.
- RStudio, per portar a terme comparacions estadístiques entre els dos grups històrics o lingüístics i comprovar si es tractaven de diferències significatives, a més de generar-ne els gràfics corresponents.
- UdPype, per lematitzar els textos i poder analitzar-ne el lèxic. La importància de la lematització en èuscar a l’hora de fer estudis de vocabulari es discutirà més endavant.
- Voyant Tools, per extraure analítiques de lèxic a partir dels textos lematitzats; en concret, freqüències de vocabulari en total, evolució històrica de la freqüència d’alguns termes concrets, paraules pròpies de cada comunicat segons la seua temàtica i mètriques de densitat lèxica.
- AntConc, també com a eina de suport a l’hora de fer quantificació de lèxic, a més de per analitzar algunes col·locacions i n-grames, tenint en compte que es tractava de textos lematitzats.
- Weka, per establir quines mètriques tretes de MultiAzterTest eren més significatives a l’hora de comparar dos grups temporals (abans i després de 2011) o lingüístics (comunicats en castellà originals o traduïts).

2.2.1. MultiAzterTest

És una eina de codi obert disponible en versió web que pot extraure més de 125 mètriques de cohesió, llegibilitat i estilometria en èuscar, castellà i anglés (Bengoetxea i Gonzalez-Dios, 2021). L’hem utilitzat per extraure 141 mètriques en castellà i 125 en èuscar, a banda de per obtenir la mètrica de “nivell del text”, que avalua cada text com “elemental” o “avançat” a partir de les altres mètriques extretes.

Les mètriques extretes es poden categoritzar en diferents grups, en concret:

- Level of Text: A partir del càlcul de les altres mètriques, dona una etiqueta de “Elementary” o “Advanced” segons la llegibilitat estimada de cada comunitat.
- Descriptive: Dades de quantificació de paraules, frases o lletres en diferents contextos.
- Lexical Diversity: Mètriques de diversitat lèxica per diverses categories gramaticals.
- Word Frequency: Freqüència de paraules per frase i altres mètriques relacionades amb la presència de paraules inusuals.
- Word Information: Quantificació de diferents paraules per categoria lèxica i altres paràmetres (distingint *content words* (paraules lèxiques, amb contingut semàntic), verbs en diferents modes...).
- Syntactic Complexity: Complexitat de les oracions i sintagmes, presència de negació, verbs en diferents formes...
- Word Semantic Information: Presència de polisèmia i hiperonímia en verbs i noms.
- Referential Cohesion: Nivell de cohesió o coincidència (*overlap*) per diferents categories.
- Semantic Overlap: Mètriques relacionades amb la semblança semàntica entre frases.
- Discourse Connectives: Quantificació total i relativa dels diferents connectors discursius.

Les dades extretes per cada comunicat es van ajuntar a un full de càlcul, etiquetades per data i llengua, i es va procedir a una primera observació i generació de gràfiques per veure alguna tendència en el temps d’alguna de les mètriques per una de les dues llengües.

La comparació directament interlingüística entre els mateixos comunicats no resultava fructuosa, pel fet que en les diferències entre les mètriques en castellà i en èuscar incideixen les diferències tipològiques de cada llengua: l'èuscar, sent una llengua aglutinant, presentarà, per exemple, densitats lèxiques superiors a les del castellà, una llengua flexiva, perquè té una densitat d'informació per paraula major (Mi Oh, Pellegrino, Marsico i Coupé, 2013).

Tot i això, sí que vam parar atenció a les diferències de mètriques entre els comunicats que considerem traduccions de periòdics i els comunicats que són originals d'ETA en castellà.

2.2.2. RStudio

Utilitzant Rstudio (RStudio Team, 2020), hem realitzat tests de normalitat i significança de les dades per intentar establir si hi havia diferències significatives entre dos grups temporals: els comunicats des de 2006 fins a 2010 i els comunicats de 2011 fins a 2018 (després de l'alto el foc definitiu i revisable). Hem fet el mateix entre els comunicats en castellà originals d'ETA o traduïts per periòdics o altres. També hem creat gràfics (*boxplots*) a partir dels resultats per visualitzar les mètriques entre els dos grups i comprovar si aquesta significança és visible.

2.2.3. UdPipe

UdPipe és un processador del llenguatge natural que permet la tokenització, etiquetatge de *PoS* (Parts of Speech), lematització i anàlisi de dependències en diferents llengües a partir de corpus lingüístics etiquetats (Straka, Hajič i Strakov, 2016). Al nostre cas, hem utilitzat l'eina UdPipe entrenada amb el corpus Ancora pel castellà i el corpus BDT en èuscar.

L'objectiu era extraure els lemes per poder-ne fer una anàlisi de freqüències de lèxic que poguera, entre altres, ser comparable entre les dues llengües. Com l'èuscar és una llengua aglutinant que permet una gran flexió, els arrels poden aparèixer en moltes formes diferents segons el sufix de cada paraula en concret, que pot variar en funció del nombre, el cas, el temps i aspecte verbal, entre altres (Arregi i Fernández, 2002). Per tant, el lema d'aquestes paraules és allò que porta un pes semàntic, i perquè les diferents formes flexionades d'un mateix lema no interferisquen a l'hora de fer una quantificació de la presència d'un mot, la lematització és el procés que ens permetrà obtenir les paraules a la seua forma no flexionada. El castellà, tipològicament diferent, no tindria tanta necessitat

d'una lematització per poder fer-ne una anàlisi de freqüències fiable en no presentar tantes variacions a partir del mateix lema, però s'ha optat per passar els textos en castellà pel mateix procés per tal d'aconseguir resultats comparables, lematitzats, en totes dues llengües.

Així, es van extraure els lemes de tots els comunicats i es van col·locar a nous fitxers de text etiquetats per llengua i data per fer-ne l'anàlisi posterior.

2.2.4. Voyant Tools

Voyant Tools és una aplicació per fer anàlisis de corpus lingüístics; és de codi obert i es troba disponible en versió web (Sinclair i Rockwell, 2016).

Permet la visualització del lèxic més freqüent tant en total com per cada text, així com fer anàlisis de freqüències relatives per a mots concrets comparant-los entre els diferents documents, amb el qual podem observar l'evolució al llarg del temps de paraules que puguen tenir interès. A més, també mostra en quina posició dins dels textos és més freqüent que aparega un mot (si, per exemple, tendeix més a ser al principi o a la fi). L'aplicació també mostra el lèxic diferenciat més important per cada text (les paraules que els distingeixen dels altres), i quantifica les frases més freqüents. També és capaç de mostrar els contextos d'un mot i les seues correlacions (paraules amb les quals potser no formen n-grames, però amb les quals tendeixen a ser a prop).

Finalment, genera gràfics dels diferents resultats que proporciona per poder visualitzar-los de manera més clara, com diagrames de línies (per freqüències relatives comparades entre documents) o núvols de paraules (per mostrar freqüències totals).

2.2.5. AntConc

Antconc (Anthony, 2021) és també una aplicació de codi obert que permet fer anàlisis pròpies de la lingüística de corpus.

Presenta les concordances totals dels diferents textos, la seua posició al text i els contextos on apareix. A més, pot extraure n-grames de diferents llargàries (nosaltres ens hem centrat als trigrames i sobretot als bigrames) en posicions anteriors o posteriors a la paraula buscada. En aquest sentit, ens ha servit com a complement per l'anàlisi de contingut i freqüències que hem fet també amb Voyant Tools.

3. Resultats

3.1. Comparació de mètriques de llegibilitat

3.1.1. Comparació interlingüística

El resultat més significatiu trobat en comparar els comunicats en èuscar i en castellà amb MultiAzterTest ha estat l'etiqueta de “Level of Text”, la qual s'estima a partir d'entre 125 mètriques per l'èuscar i 141 pel castellà. Aquesta etiqueta funciona com un classificador de llegibilitat i nivell de dificultat d'un text (citar article MultiazterTest), i diferencia un nivell elemental (“Elementary”) o avançat (“Advanced”) dels textos.

Al nostre cas, dels 18 textos en èuscar, 17 han rebut l'etiqueta d'elementals, i 1 d'avançat. Per l'altra banda, dels 18 textos en castellà, 16 han rebut l'etiqueta d'avançats i 2 d'elementals.

Data	Level of text	
	Èuscar	Castellà
22/03/2006	Elementary	Advanced
23/03/2006	Elementary	Advanced
24/09/2006	Elementary	Advanced
04/01/2007	Elementary	Advanced
05/06/2007	Elementary	Advanced
31/12/2009	Elementary	Advanced
04/04/2010	Elementary	Elementary
05/09/2010	Elementary	Advanced
19/09/2010	Advanced	Advanced
08/01/2011	Elementary	Advanced
20/10/2011	Elementary	Elementary
abr-12	Elementary	Advanced
17/03/2013	Elementary	Advanced
27/03/2015	Elementary	Advanced
27/03/2016	Elementary	Elementary
07/04/2017	Elementary	Advanced
08/04/2018	Elementary	Advanced
03/05/2018	Elementary	Advanced

Taula 1: Comunicats per data, mètrica “Level of Text” pel comunicat en cada llengua

Com podem observar a la taula, la majoria dels textos en èuscar han sigut classificats com “Elementary”, mentre que la majoria dels textos en castellà (que estan constituïts, recordem, tant per originals escrits per ETA com per traduccions fetes per altres organismes) tenen la classificació de “Advanced”.

Així, podríem establir en un primer moment que els textos escrits en castellà tenen un nivell més alt de complexitat i llegibilitat que els escrits en èuscar.

Els dos textos en castellà que reben l'etiqueta d'elementals són justament, a més, traduccions de periòdics, no originals d'ETA. Com discutirem a l'apartat següent, en mostrar la llista de textos en castellà diferenciant-ne les traduccions dels originals, les traduccions tendeixen a presentar mètriques de complexitat menors a causa de diferents efectes de la *translationese* o traduccionisme, com un estil més explícit i conservador o una densitat lèxica menor que els originals (Baroni i Bernardini, 2006).

Per una altra banda, en observar les freqüències de paraules inusuals entre les diferents llengües, esperàvem obtenir un valor major de l'èuscar per qüestions tipològiques: com es pot veure a les taules anteriors, els valors de riquesa lèxica són majors per l'èuscar per ser una llengua aglutinant que presenta menys paraules funcionals que el castellà. No obstant això, a les mètriques de MultiAzterTest referides a la freqüència de paraules inusuals, només es comptabilitzaven paraules lèxiques o de contingut. D'aquesta manera, hem considerat que els comunicats en castellà tendeixen a presentar més paraules inusuals diferents entre si que en èuscar. Els casos on el castellà ha tingut resultats més baixos que l'èuscar han estat, precisament, originals d'ETA en castellà (excepte el cas del comunicat del 24 de setembre de 2006, traduït per EITB).

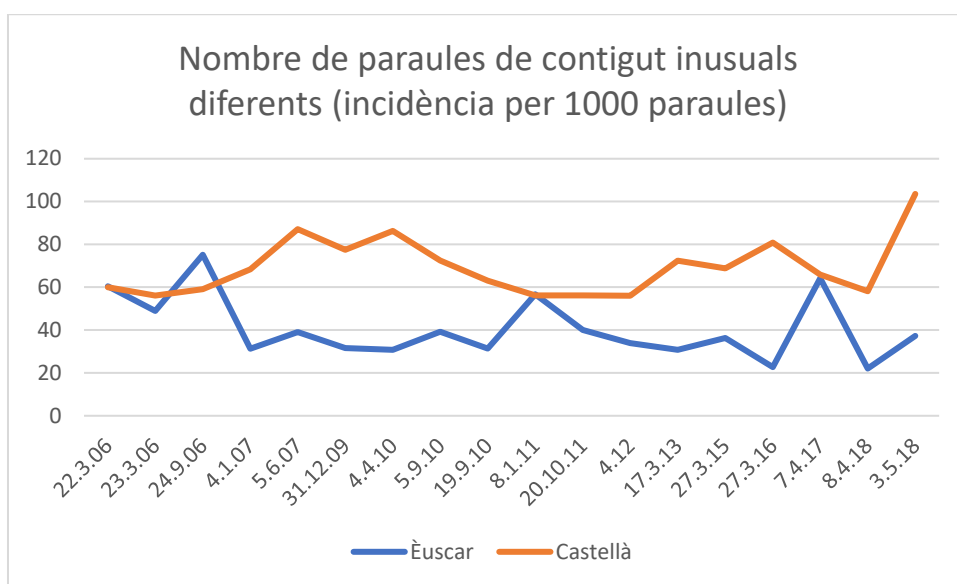


Figura 1: Incidència per 1000 paraules de les paraules de contingut inusuals diferents en èuscar i castellà

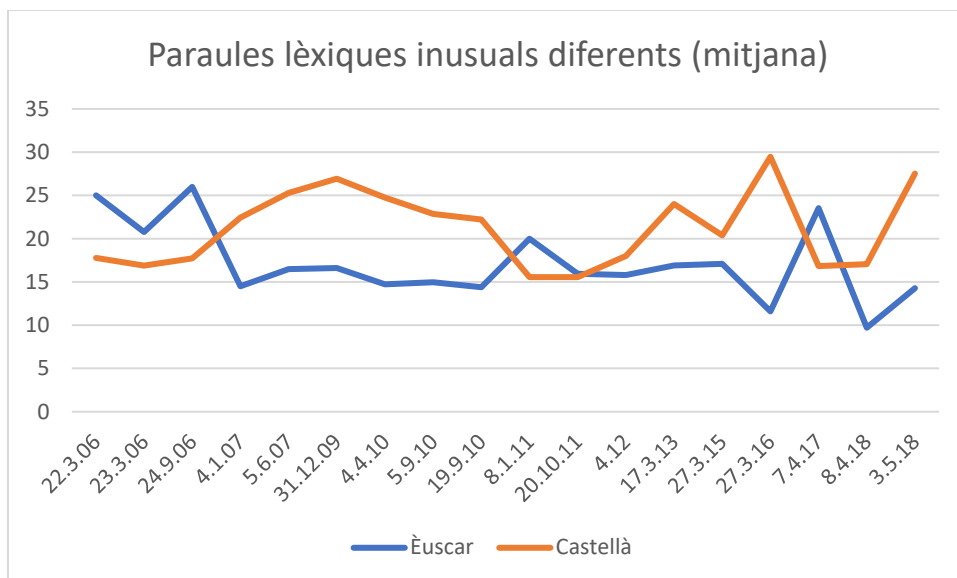


Figura 2: Mitjana de les paraules lèxiques inusuals diferents en èuscar i castellà

A la discussió, analitzarem la importància que poden tenir aquests resultats a l'hora de fer hipòtesis sobre les llengües d'ús dins d'ETA, a més de contrastar la fiabilitat dels paràmetres de MultiAzterTest a l'hora de classificar els nivells de diferents llengües.

3.1.2. Comunicats en castellà, traduccions i originals

Mensaje de Euskadi Ta Askatasuna al pueblo vasco	22/03/2006	Original
Segundo mensaje de Euskadi Ta Askatasuna al pueblo vasco	23/03/2006	Original
Mensaje de ETA en el día del guerrero vasco	24/09/2006	Traducció
Texto íntegro del comunicado	04/01/2007	Traducció
Comunicado de ETA	05/06/2007	Traducció
Comunicado de ETA a Euskal Herria	31/12/2009	Traducció
Traducción al castellano del comunicado de ETA	04/04/2010	Traducció
Declaración de ETA	05/09/2010	Traducció
Comunicado de ETA a la Comunidad Internacional	19/09/2010	Traducció
Declaración de ETA	08/01/2011	Original
Declaración de ETA	20/10/2011	Original
Comunicado de ETA a la Comunidad Europea	04/2012	Original
Comunicado de ETA	17/03/2013	Traducció
Comunicado de Euskadi Ta Askatasuna al Pueblo Vasco	27/03/2015	Original
Comunicado de Euskadi Ta Askatasuna en el día del Aberri Eguna	27/03/2016	Traducció

Comunicado íntegro de ETA en el que se declara 'desarmada'	07/04/2017	Original
ETA al pueblo vasco, declaración sobre el daño causado	08/04/2018	Traducció
Declaración final de ETA al pueblo vasco	03/05/2018	Original

Taula 2: Comunicats en castellà, diferenciant els originals dels traduccions

Com ja hem avançat, és esperable que una traducció tinga unes mètriques menors en comparació als originals pel fet d'haver-se escrit a partir d'una altra llengua i adaptar-se amb major o menor èxit.

Amb l'eina Voyant Tools, hem pogut comparar la riquesa lèxica dels diferents comunicats en castellà i fer-ne un rànquing de més a menys. A continuació, una taula amb els resultats ordenats:

07-04-2017	0,535
08-01-2011	0,488
03-05-2018	0,48
22-03-2006	0,47
20-10-2011	0,463
24-09-2006	0,462
05-09-2010	0,459
27-03-2015	0,425
05-06-2007	0,414
23-03-2006	0,397
04-2012	0,391
08-04-2018	0,386
04-04-2010	0,381
17-03-2013	0,348
04-01-2007	0,327
19-09-2010	0,285
27-03-2016	0,271
31-12-2009	0,26

Taula 3: Taula de ràtio type/token per data de comunicat, en castellà. En verd, els originals d'ETA en castellà; en roig, traduccions fetes a partir d'originals d'ETA en èuscar

El valor extret a la taula és el resultat de la divisió entre els types (el número les paraules diferents) i els tokens (el número de paraules en total) dels textos lematitzats en castellà (per poder comparar-los posteriorment amb els textos lematitzats en èuscar). Aquest valor és la densitat lèxica del text, i està relacionat amb la riquesa de vocabulari que hi trobem, tot i que el resultat també està influït per factors com la llargària d'un text.

Com podem observar, tot i que hi ha cert grau de barreja entre originals i traduccions per la meitat de la taula, pareix que els cinc primers textos amb major riquesa lèxica són,

efectivament, originals d'ETA, mentre que els últims set textos amb menor riquesa són traduccions; tot i això, les altres traduccions tenen posicions de sisé, seté o nové grau, i altres originals, de huité, desé o desé primer (de díhuit).

Tot i això, en examinar els mateixos resultats en èuscar, hem obtingut la següent taula:

07-04-2017	0,598
24-09-2006	0,586
08-01-2011	0,563
22-03-2006	0,552
20-10-2011	0,547
05-06-2007	0,544
03-05-2018	0,539
27-03-2015	0,516
04-2012?	0,489
08-04-2018	0,47
05-09-2010	0,464
23-03-2006	0,463
04-04-2010 (boltxe kolektiboa)	0,442
04-01-2007	0,435
19-09-2010	0,403
17-03-2013	0,400
31-12-2009	0,375
27-03-2016	0,372

Taula 4: Taula de ràtio type/token per data de comunicat, en èuscar. En verd, els textos que tenen una versió original d'ETA en castellà; en roig, els textos que ETA només va escriure en èuscar i que per tant tenen una versió traduïda al castellà

Tot i que trobem alguns textos amb traducció al castellà més dalt en èuscar que en castellà, els comunicats amb menys densitat lèxica pareixen ser els mateixos tant en èuscar com en castellà, de manera que la menor riquesa dels textos traduïts en castellà pot deure's no del tot al fet que siguin traduccions (i, per tant, tinguen una menor riquesa lèxica per efectes del traduccionisme), sinó perquè són traduccions, justament, dels textos que ja tenien la menor riquesa lèxica en èuscar; per tant, l'únic que han fet ha sigut traduir eixa riquesa menor al castellà (i no baixar-la, com podríem esperar que passara, en ocórrer la traducció). Tot i així, sí que veiem que en èuscar puja en el rànquing la riquesa del comunicat del 24 de setembre de 2006 o la del 5 de juny de 2007, o siga que és possible que aquests comunicats s'hagen vist afectats més per una baixada de la riquesa en ser traduïts.

També podem observar que els resultats de la riquesa són superiors en general a les seues versions en castellà. Això s'explica per les diferències tipològiques entre ambdues

llengües. L'èuscar, en ser una llengua aglutinant, té una presència major de paraules de categoria semàntica, ja que sovint porten incorporades en la seua declinació la informació gramatical que el castellà, com a llengua flexiva, expressa amb paraules de categoria funcional o *stopwords*, les quals apareixen sovint, repetint-se (encapçalen, en general, la llista de paraules més freqüents d'un text) i baixant aixina el valor de la riquesa lèxica.

A banda d'aquests resultats en taules, també hem fet comparació per grups amb l'ajuda de Rstudio, per visualitzar si les diferències del valor de la riquesa lèxica entre grups eren significatives. Després de fer un test de normalitat i comprovar que les dades tenien una distribució normal amb el test de Shapiro-Wilk ($W = 0.95144$, $p = 0.4481$), vam realitzar un T-test i vam determinar que la diferència entre tots dos grups era significativa ($W = 21.472$, $p = 9.344e-14$), tot i que el gràfic resultant no va ser gaire aclaridor. La mitjana del grup de textos originals en castellà és major a la de les traduccions. Es percep una gran variació intergrup, però hem pogut establir la igualtat de variàncies amb el test de Bartlett ($F(18, 1) = 0.017036$, $p = 0.8962$).

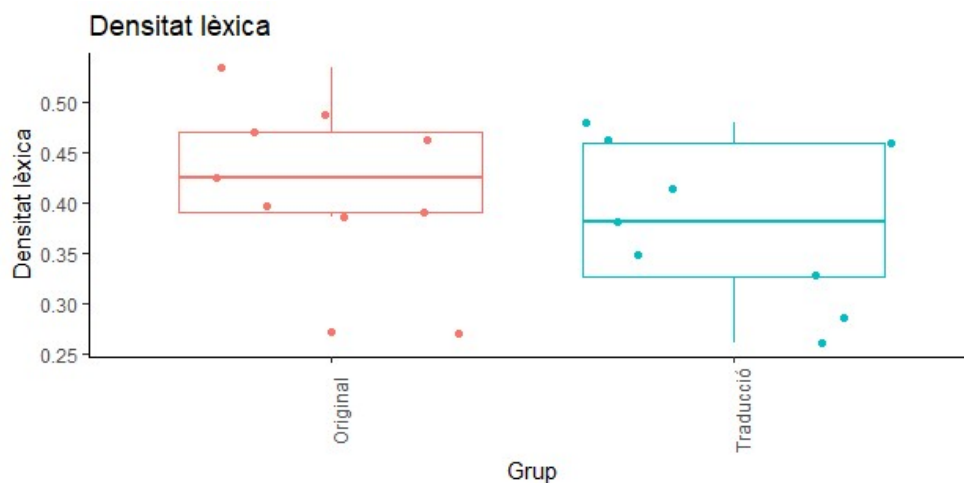


Figura 3: Gràfic del valor de riquesa lèxica de dos grups: el dels comunicats originals en castellà, i el dels traduïts provinents d'un original en èuscar

Hem observat altres mètriques amb les dades que ha donat MultiAzterTest. Les densitats de les diferents categories gramaticals també són més altes pels originals (nominal i adjectival), excepte per la densitat adverbial (major a les traduccions) i la verbal (a la qual no hi ha diferències significatives entre tots dos grups). No obstant això, no hem pogut demostrar la igualtat de variàncies a la densitat adverbial ($F(18, 1) = 3.4448$, $p = 0.06345$); la desviació estàndard al grup dels originals és prou alta, i la diferència entre les mitjanes dels grups és molt petita.

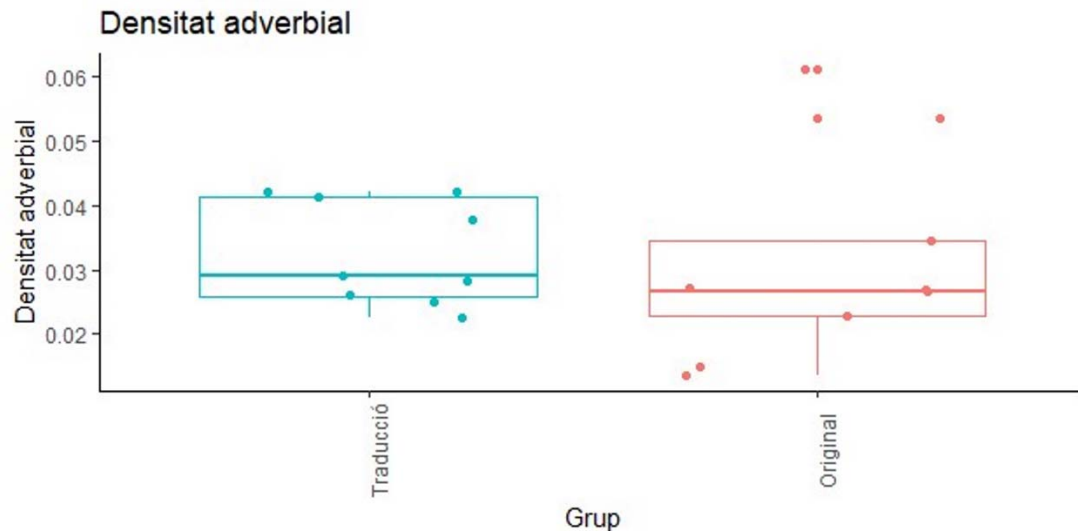


Figura 4: Gràfic del valor de densitat adverbial de dos grups: el dels comunicats originals en castellà, i el dels traduïts provinents d'un original en èuscar

Altres mètriques que vam comparar van ser les relacionades amb la complexitat sintàctica. De les 13 mètriques que vam comparar, 10 van oferir resultats amb diferències significatives per cada grup: el grup d'originals era superior a 5 mètriques i el de les traduccions, a les altres 5; en concret:

- Originals: Mitjana de nivells de profunditat a l'arbre de dependències, mitjana de proposicions per frase, mitjana de sintagmes verbals per frase, mitjana de sintagmes nominals per frase, densitat de sintagmes nominals.
- Traduccions: Mitjana de paraules abans del verb principal, nombre de subordinades (incidència per 1000 paraules), signes de puntuació per frase (mitjana), densitat de sintagmes verbals i nombre de paraules negatives (incidència per 1000 paraules).

Finalment, vam comparar també els resultats amb les mètriques més significatives pel castellà segons (Bengoetxea i Gonzalez-Dios, 2021), com apareix després a la taula 5. La diferència entre els grups va resultar significativa per les 10 mètriques, i també ací, cada grup té 5 mètriques on destaca sobre l'altre:

- Originals: Mitjana de sintagmes verbals per frase, semblança entre paràgrafs adjacents (desviació estàndard), llargària de les frases sense stopwords (mitjana), Lemma TTR (ràtio token/type amb els lemes) i mitjana de la llargària de les frases.
- Traduccions: Mitjana de paraules lèxiques inusuals diferents, paraules de contingut inusuals diferents (incidència per 1000 paraules), semblança entre

paràgrafs adjacents (mitjana), semblança entre paràgrafs (mitjana) i nombre de frases (incidència per 1000 paraules).

3.1.3. Canvis diacrònics

Hem observat també si hi havia canvis diacrònics per les mètriques concretes extretes. MultiAzterTest és capaç de traure més mètriques pel castellà que per l'èuscar, però per aquest estudi hem seleccionat només les mètriques que comparteixen tots dos. D'aquestes 125 mètriques, per èuscar i castellà, ens hem centrat en un principi en les 10 mètriques més predictives a l'hora de classificar el nivell de complexitat textual a MultiAzterTest. Hem seleccionat les 10 mètriques més predictives basades en ràtios, i no en números quantitativs (ja que aquestes xifres varien en funció de la llargària del text, la qual pot diferir entre diferents comunicats).

A continuació, la llista de les 10 mètriques per cada llengua (Bengoetxea i Gonzalez-Dios, 2021):

Castellà	Èuscar
Sintagmes verbals per frase (mitjana)	Sintagmes verbals per frase (mitjana)
Paraules rares distintes (mitjana)	Llargària del lema (mitjana)
Nombre de paraules rares diferents (incidència per 1000 paraules)	Llargària de les frases sense stopwords (mitjana)
Semblança entre paràgrafs adjacents (mitjana)	Llargària de les paraules (mitjana)
Semblança entre paràgrafs adjacents (desviació estàndard)	Nombre de frases (incidència per 1000 paraules)
Llargària de les frases sense stopwords (mitjana)	Llargària de les paraules sense stopwords (mitjana)
Llargària de la frase (mitjana)	Llargària de la frase (mitjana)
Lemma TTR (ràtio entre token/type)	Nombre de paràgrafs (incidència per 1000 paraules)
Semblança entre paràgrafs (mitjana)	Nombre de síl·labes per paraula (desviació estàndard)
Nombre de frases (incidència per 1000 paraules)	Profunditat de les subordinades per frase (mitjana)

Taula 5: Taula de mètriques (ràtios) més significatives per determinar la complexitat textual en èuscar i castellà amb MultiAzterTest

A continuació, hem dividit els comunicats en dos grups, segons si els comunicats provenien d'abans de 2011 (grup 1) o a partir de 2011, amb l'alto el foc definitiu (grup 2). S'han fet tests de normalitat (Test de Shapiro-Wilk), significança (T-test o Wilcoxon, depenent de la normalitat) i igualtat de variàncies (Test de Levene o de Bartlett segons la normalitat) amb les dades per determinar si existien diferències significatives entre els

- Grup 1: CRFCWOa (mesura de cohesió entre les frases d'un mateix paràgraf o tros de text semblant), densitat verbal, nombre de paraules inusuals (incidència per 1000 paraules), nombre d'adverbis (incidència per 1000 paraules), mitjana de paraules inusuals.
- Grup 2: Desviació estàndard de la mitjana de síl·labes per paraula, mitjana dels valors d'hiperònims dels verbs a WordNet, LAdjTTR (ràtio type/token) dels lemes dels adjectius), densitat adjectival i mitjana de síl·labes per paraula.

3.1.3.2. Resultats en èuscar

Per les mètriques en èuscar, també hem comparat primer els resultats segons les 10 mètriques més significatives. En aquest cas, 9 mètriques han donat diferències significatives entre grups: en 6 d'elles, ha destacat el grup 2, i en 3, el grup 1:

- Grup 1: Llargària dels lemes (mitjana), llargària de les paraules (mitjana) i nombre de paràgrafs (incidència per 1000 paraules).
- Grup 2: Sintagmes verbals per frase (mitjana), llargària de les frases sense stopwords (mitjana), llargària de les frases (mitjana), llargària de les paraules sense stopwords (mitjana), nombre de síl·labes per paraula (desviació estàndard) i profunditat de les subordinades per frase (mitjana).

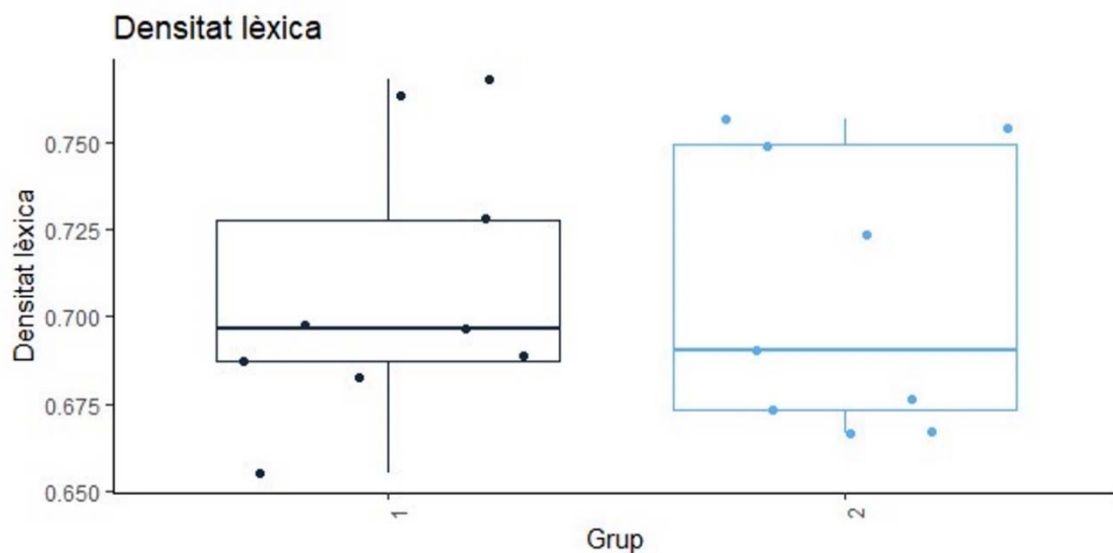


Figura 6: Densitat lèxica entre el grup 1 (abans de 2011) i el grup 2 (després de 2011) en èuscar. A diferència dels comunicats en castellà, a l'èuscar és el primer grup el que té una densitat lèxica major. Tot i que segons el t-test, hi ha una diferència significativa entre tots dos ($W= 80.179$, $p < 2.2e-16$), aquesta és molt petita, sobretot si mirem a la gran variabilitat intragrupal. El grup 1 també té més densitat nominal i adjectival, mentre que el grup 2 en

té més d'adverbial. La densitat verbal, com al cas anterior, no ha donat diferències intragrups importants.

Hem tornat a mirar amb Weka quines eren les 25 mètriques més útils a l'hora d'establir diferències entre els grups, i a diferència de l'espanyol, la major part ha donat un resultat significatiu (19 mètriques de 25). D'eixes 19, el grup 2 ha tingut una mitjana superior a 14, i el grup 1 a 5.

- Grup 1: Verbs en infinitiu (incidència per 1000 paraules), pronoms (incidència per 1000 paraules), paraules lèxiques inusuals (mitjana), adjectius inusuals (incidència per 1000 paraules) i llargària de les paraules sense stopwords (desviació estàndard).
- Grup 2: Nombre de connectius (incidència per 1000 paraules), connectius lògics (incidència per 1000 paraules), connectius temporals (incidència per 1000 paraules, semblança semàntica entre frases adjacents (desviació estàndard), semblança semàntica entre paràgrafs adjacents (desviació estàndard), llargària de les frases (desviació estàndard), mitjana de síl·labes per paraula (desviació estàndard), densitat adverbial, llargària de les paraules (desviació estàndard), semblança semàntica entre qualsevol parell de frases a un paràgraf (desviació estàndard), nombre de verbs en indicatiu (incidència per 1000 paraules), NTTR (ràtio type/token nominal), mitjana de llargària de paraules sense stopwords (desviació estàndard).

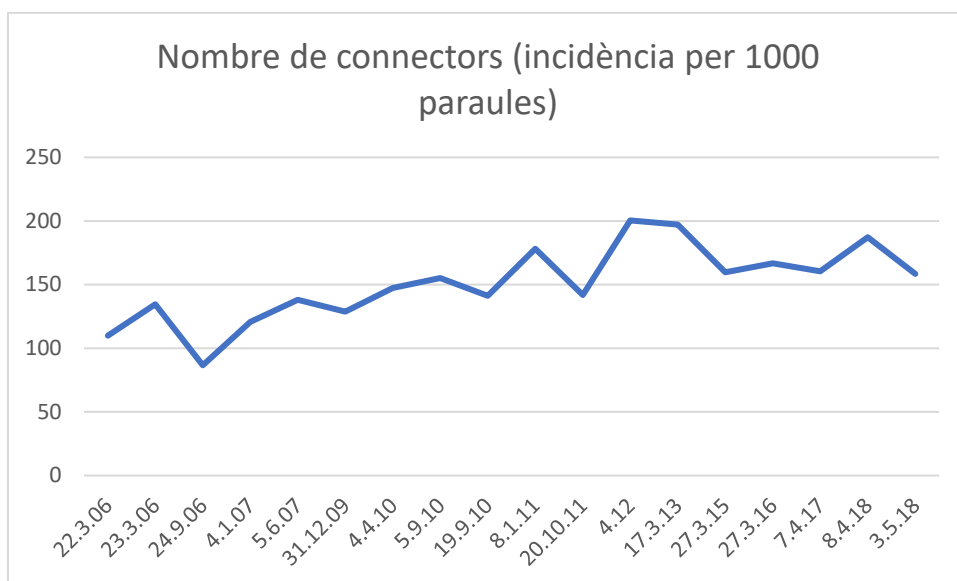


Figura 7: incidència per 1000 paraules del nombre de connectors als comunicats en èuscar

El cas de la freqüència de connectors als comunicats és la mètrica a la qual podem observar una evolució més palpable al llarg del temps, amb una crescuda que arriba al seu màxim al comunicat de 2012.

A altres casos, la mitjana del segon grup és major al del primer, però aquesta pujada no és tan constant en el temps, com és el cas de la desviació estàndard de la semblança semàntica entre frases adjacents.

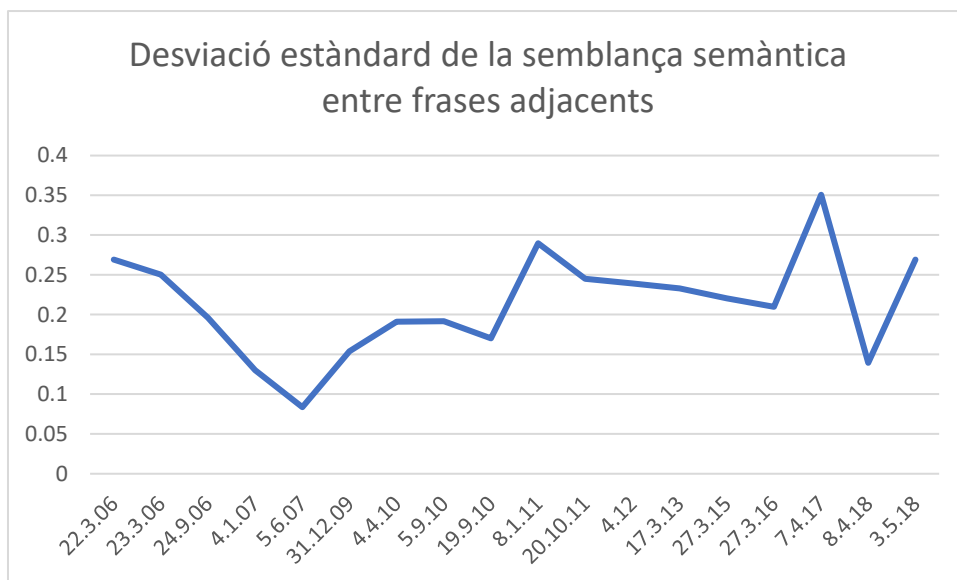


Figura 8: Desviació estàndard de la semblança semàntica entre frases adjacents als comunicats en èuscar

3.2. Comparació del lèxic

Després de la lematització dels textos, s'ha procedit a extraure'n les freqüències lèxiques més importants. És rellevant, doncs, remarcar que les freqüències obtingudes són a partir dels lemes dels textos, i no dels comunicats en la seua forma original.

S'han extret les 60 paraules més freqüents de tots els comunicats en castellà i en èuscar, destacant les paraules amb pes semàntic per diferenciar-les de les paraules de categoria funcional (que solen aparèixer entre les més freqüents pel seu ús).

A continuació, la llista de les paraules més freqüents per cada llengua, amb el número de freqüència total entre parèntesis:

En castellà:

el (1545); de (855); y (510); que (440); en (387); a (317); él (300); haber (259); ser (217); uno (208); para (178); euskal (157); herria (155); eta (153); no (147); del (140); este (115); con (114); por (103); tener (101); su (100); vasco (95); proceso (94); pueblo (84);

al (84); todo (78); político (78); hacer (75); conflicto (74); estar (71); ese (71); democrático (64); dar (62); nuestro (56); como (56); querer (54); yo (53); más (49); otro (47); futuro (46); lucha (45); poder (44); estado (43); solución (42); sin (42); mucho (41); gobierno (41); españa (38); español (37); también (36); deber (36); compromiso (35); paz (34); decisión (34); ta (33); camino (33); nuevo (32); armado (32); situación (31).

En èuscar:

eta (733); izan (528); edun (427); herri (290); euskal (283); ukan (210); ez (152); egin (142); politiko (107); nahi (101); prozesu (97); bera (88); hori (83); behar (81); bat (79); ere (78); egon (77); gatazka (76); espainia (76); bide (75); demokratiko (74); herritar (73); eman (68); erabaki (59); hau (58); gu (58); askatasun (58); borroka (57); berri (57); konponbide (52); hartu (52); guzti (51); gobernu (50); arte (49); beste (48); konpromiso (45); frantzia (44); ta (40); aukera (40); gabe (39); aurre (38); egoera (37); etorkizun (35); bake (35); lortu (34); gora (34); eskubide (34); armatu (34); egun (33); utzi (31); eraiki (31); sozialista (30); ondorio (30); estatu (30); ari (30); ezker (29); urrats (28); sortu (28); euskadi (28).

Amb case sensitive:

eta (558); izan (528); edun (427); herri (290); euskal (283); ukan (210); ETA (165); ez (152); egin (142); politiko (107); nahi (101); prozesu (97); bera (88); hori (83); behar (81); bat (79); ere (78); egon (77); gatazka (76); espainia (76); bide (75); demokratiko (74); herritar (73); eman (68); erabaki (59); hau (58); gu (58); askatasun (58); borroka (57); berri (57); konponbide (52); hartu (52); guzti (51); gobernu (50); arte (49); beste (48); konpromiso (45); frantzia (44); ta (40); aukera (40); gabe (39); aurre (38); egoera (37); etorkizun (35); bake (35); lortu (34); gora (34); eskubide (34); armatu (34); egun (33); utzi (31); eraiki (31); sozialista (30); ondorio (30); estatu (30); ari (30); ezker (29); urrats (28); sortu (28); euskadi (28).

Com era esperable, el vocabulari més freqüent té una temàtica fortament política i tècnica, ajuntant termes polítics, identitaris o metafòrics, com discutirem més endavant.

3.2.1. El cas d'eta i ETA en èuscar

En un primer moment, en fer la llista de freqüències en èuscar amb Voyant Tools, vam observar que categoritzava una sola categoria d'“eta”. Com havíem fet aquesta anàlisi a partir dels textos lematitzats, tots en minúscules, no podíem diferenciar “eta” com a

conjunció del base (“i”, molt freqüent) i “ETA” com a nom propi de la banda. Per poder diferenciar-les, vam analitzar els textos originals amb l’eina AntConc, que sí que permet una opció de *case-sensitive* (diferenciar una paraula en funció de si va en majúscula o no). Així, vam poder comptabilitzar “eta” i “ETA” de manera diferenciada, amb les freqüències que els hem donat a la llista.

Així, per a ETA (la qual té el pes semàntic important), les dades resultants es van quantificar tant de manera total (el nombre de vegades que apareix ETA per comunicat) i de manera relativa (dividint el nombre total anterior entre el nombre de paraules totals de cada comunicat, per reduir l’efecte de les diferents llargàries dels textos als resultats):

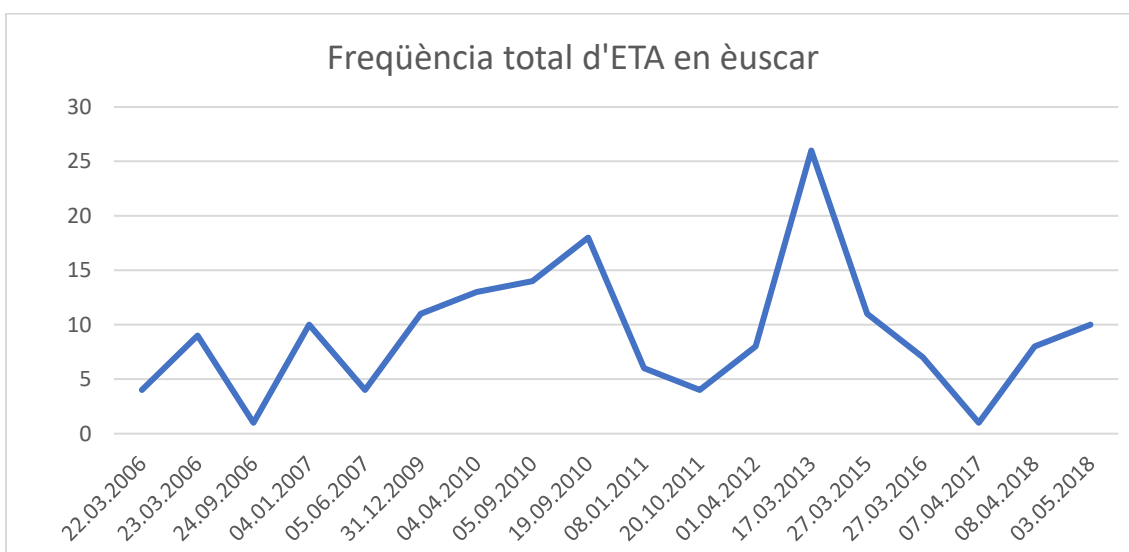


Figura 9: Nombre total d’aparicions d’ETA per cada comunicat en èuscar

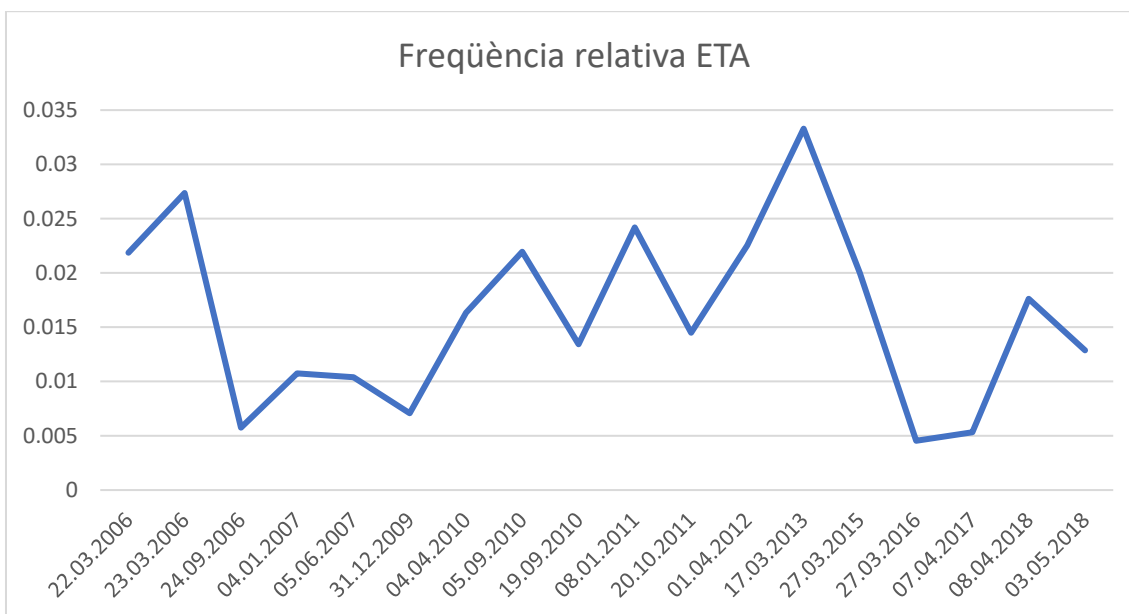


Figura 10: Nombre total d’aparicions d’ETA entre el nombre total de paraules de cada comunicat en èuscar

Es pot observar un pic d'aparicions de mencions a ETA al comunicat del 17 de març de 2013, coincidint amb el trencament del diàleg durant la negociació pel desarmament amb el govern de M. Rajoy. Els comunicats amb unes freqüències més baixes són els de l'Aberri Eguna de 2016 i el del 7 d'abril de 2017, on la banda es va declarar desarmada.

Altres mencions que es fan al nom propi d'ETA, en els comunicats en èuscar, inclouen les expressions E.T.A. (15) i Euskadi Ta Askatasuna (28), nom complet. La versió E.T.A. apareix exclusivament com a signatura final en la major part dels comunicats, mentre que la versió completa Euskadi Ta Askatasuna sol aparèixer bé com a firma però també al principi dels comunicats.

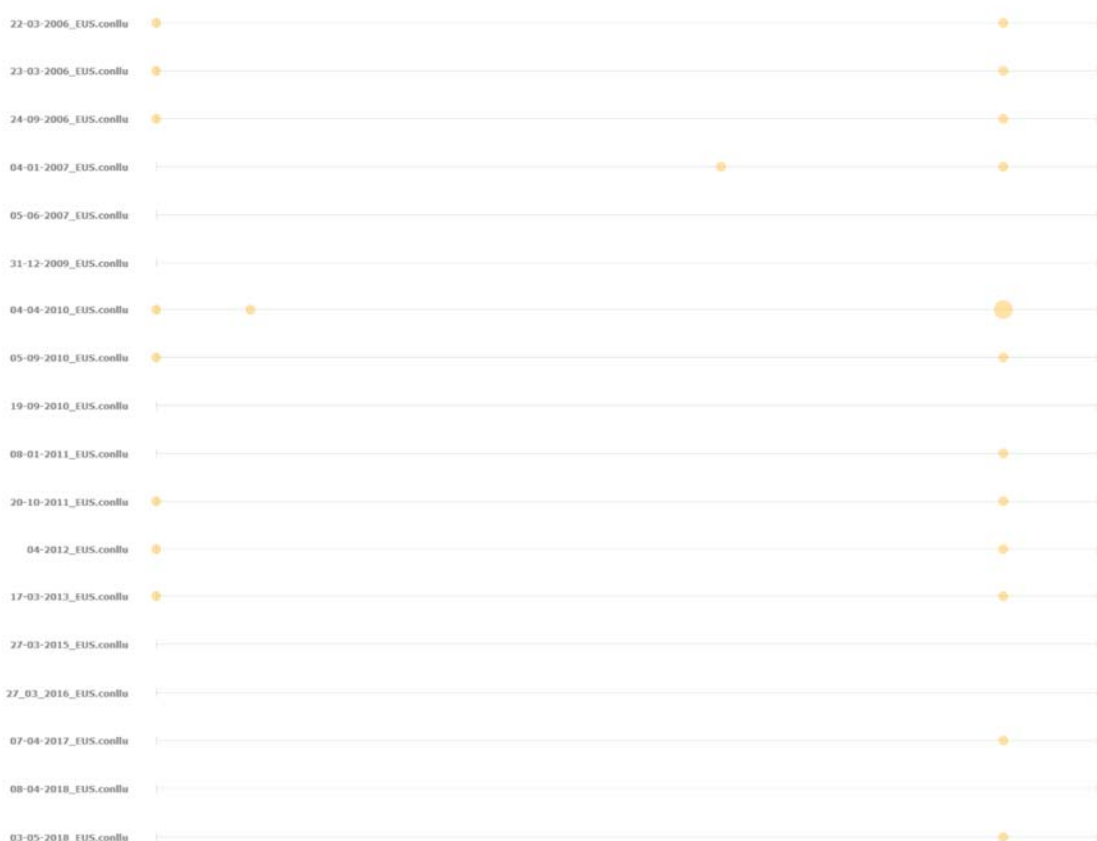


Figura 11: Posició del terme “Euskadi Ta Askatasuna” per comunicat en èuscar

3.2.2. Evolució de termes en èuscar

A partir dels termes més freqüents, hem analitzat si hi havia canvis de tendència per algunes paraules al llarg del període estudiat. També hem parat atenció a altres termes que no es trobaven a la llista dels 60 termes més utilitzats, però que hem considerat importants per la seua relació temàtica amb els comunicats. Finalment, hem analitzat l'evolució d'alguns bigrames per observar col·locacions usades amb més freqüència.

És rellevant recordar que els comunicats en èuscars són només una mostra dels comunicats totals que es van fer durant aquesta època; en concret, solament constitueixen els comunicats amb una versió al castellà. Tot i això, existeixen més comunicats fets des de 2006 fins a 2018 amb autoria d'ETA, i aquests resultats no han de considerar-se una mostra total del lèxic d'ETA durant aquest temps, sinó només dels comunicats que van tenir una versió en castellà i, per tant, major repercussió.

3.2.2.1. Referències a França i a Espanya

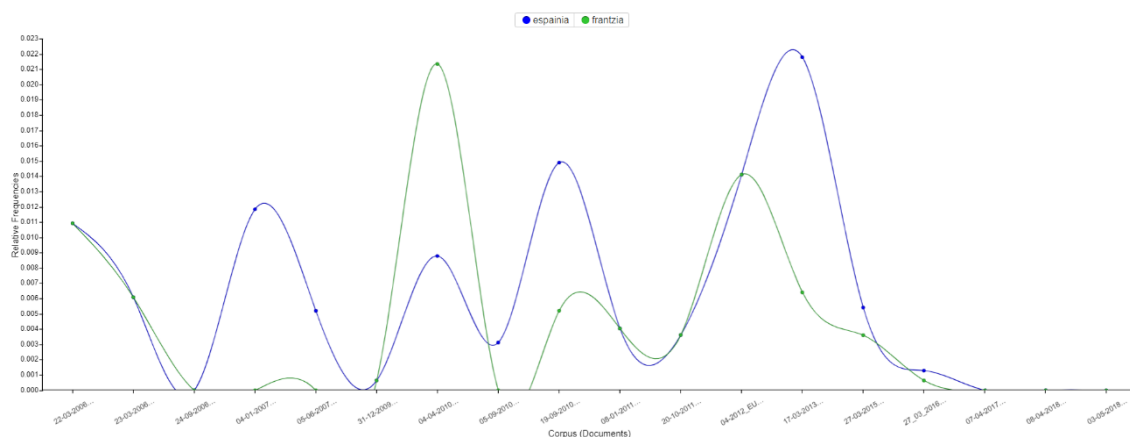


Figura 12: Freqüència relativa de “espania” (blau) i “frantzia” (verd) dels comunicats d'ETA lematitzats en èuscar

La freqüència de cada terme és variable per comunicat, encara que sí que es percep que les referències a Espanya són en general superiors a França, exceptuant el comunicat del 4 d'abril de 2010, coincidint amb la mort d'un agent de la policia francesa en la localitat de Dammarie-lès-Lys durant un tiroteig amb ETA, on la banda va explicar la seua versió dels fets. El pic de referències a Espanya el trobem al comunicat del 17 de març de 2013, arran del fracàs del diàleg entre el govern espanyol i Euskadi Ta Askatasuna. Tot i això, a partir d'aquest últim, les referències a Espanya baixen dràsticament als següents comunicats; als tres últims, datats entre 2017 i 2018, no n'hi ha una sola referència.

3.2.2.2. Solució i diàleg

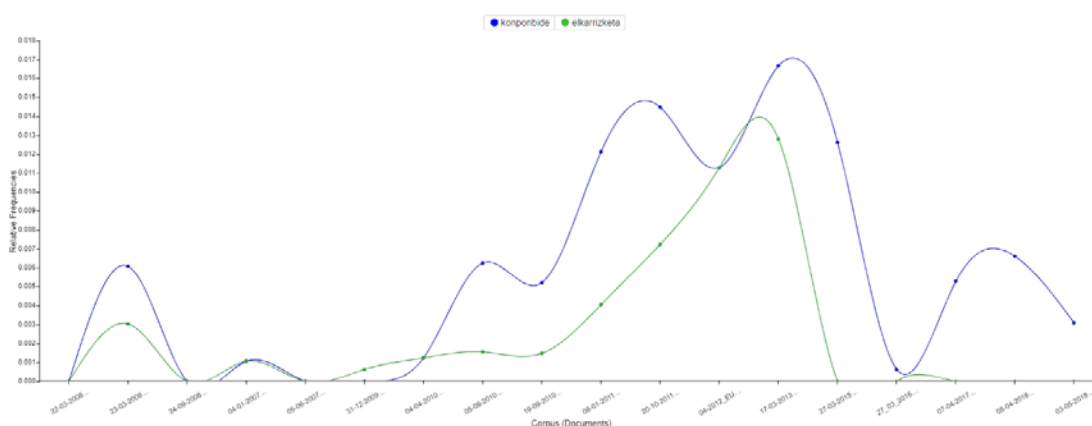


Figura 13: Freqüència relativa de “konponbide” (blau) i “elkarrizketa” (verd) dels comunicats d’ETA lematitzats en èuscar

Els mots “konponbide” (‘solució’) i “elkarrizketa” (‘diàleg’) apareixen un total de 52 i 24 vegades respectivament; la primera és, a més, la trigèsima paraula més utilitzada a tot el corpus. Totes dues presenten un creixement prou sostingut (sobretot al cas de “elkarrizketa”) al període 2010-2013, coincidint amb el plantejament de l’alto el foc permanent. El pic de freqüència relativa de totes dues paraules coincideix amb el comunicat del 17 de març de 2013, on es va trencar temporalment el diàleg entre el govern espanyol i ETA. A partir de 2015 hi ha una caiguda de l’ús de les dues, especialment de “elkarrizketa” (‘diàleg’); “konponbide” (‘solució’) sí que té alguna presència als comunicats de 2017 i 2018. Alguns bígrams repetits són “bake elkarrizketa”(2) (‘diàleg de pau’), “elkarrizketa gune”(3) (‘espai de diàleg’), “konponbide prozesu”(11) (‘procés de solució’) o “konponbide demokratiko”(6) (‘solució democràtica’), “konponbide eman”(4) (‘donar solució’) i “konponbide justu”(4) (‘solució justa’); a banda del trígram “elkarrizketa eta negoziazio”(7) (‘diàleg i negociació’).

3.2.2.3. Polític i democràtic

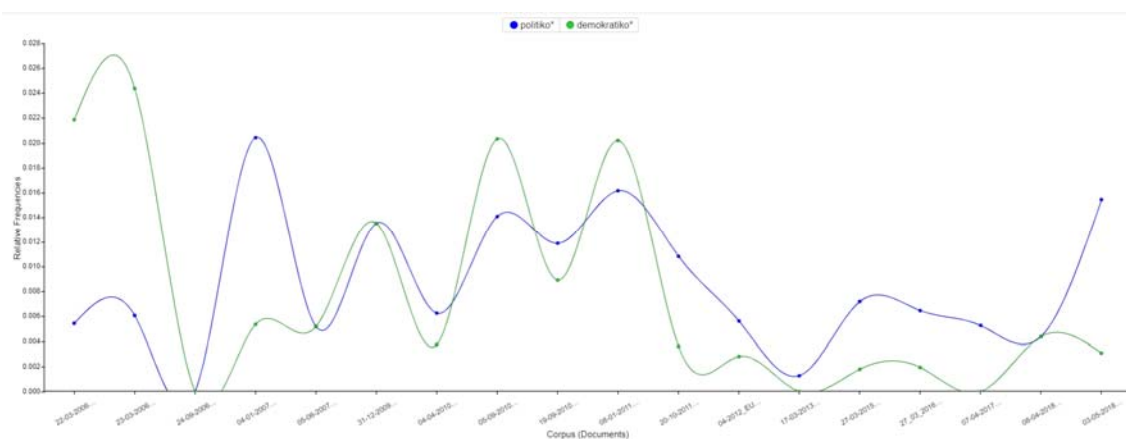


Figura 14: Freqüència relativa de “politiko” (blau) i “demokratiko” (verd) dels comunicats d’ETA lematitzats en èuscar

“Politiko” i “demokratiko” són els dos adjectius més utilitzats als comunicats; “politiko”, amb 107 aparicions, és la desena paraula més freqüent (la quarta sense tenir en compte les paraules de categoria funcional), i “demokratiko”, la vigèsima segona, té 74 ocurrències. Com podem observar, totes dos tenen una presència variable abans de 2011, després del qual experimenten una baixada fins a 2013; “politiko”, però, torna a usar-se amb alguna freqüència més, notablement a l’últim comunicat de 2018 que anuncia la dissolució d’ETA.

“Demokratiko” presenta el seu pic més alt al comunicat del 23 de juny de 2006, seguit dels comunicats del 22 de juny de 2006, 5 de setembre de 2010 i 8 de gener de 2011. Curiosament, aquests quatre fan referència a alts al foc: el primer va donar principi a un “alto el foc permanent” que va durar oficialment del 22 de juny de 2006 (dia de publicació d’una primera versió d’aquest comunicat i que té una presència prou similar de “demokratiko”, el segon de la llista) fins al 5 de juny de 2007 (data on publiquen un comunicat que és al nostre corpus i que té una presència prou menor de “demokratiko”); el tercer, datat el 5 de setembre de 2010, és un preludi d’allò que seria l’alto el foc permanent, general i verificable definitiu de 2011, el quart de la llista.

“Politiko”, per una altra banda, presenta un pic important al comunicat del 4 de gener de 2007 que reivindica l’atemptat a l’aeroport de Barajas, comés el 30 de desembre de 2006.

Com a bigrams, destaquen “gatazka politiko”(10) (‘conflicte polític’), “eztabaida politiko”(7) (‘debat polític’), “aldaketa politiko”(6) (‘canvi polític’), “eragile politiko”(5) (‘agent polític’), “preso politiko” (5) (‘pres polític’) i “proiektu politiko”(5) (‘projecte polític’). Amb “demokratiko”, tenim “prozesu demokratiko”(31) (‘procés democràtic’),

“konponbide demokratiko”(6) (‘solució democràtica’) i “egoera demokratiko”(5) (‘situació democràtica’). Com a trigràma, tenim “marko demokratiko eraiki”(3) (‘construir un marc democràtic’), que apareix als comunicats d’alto el foc de març de 2006 i de setembre de 2010.

3.2.2.4. Lluita i conflicte

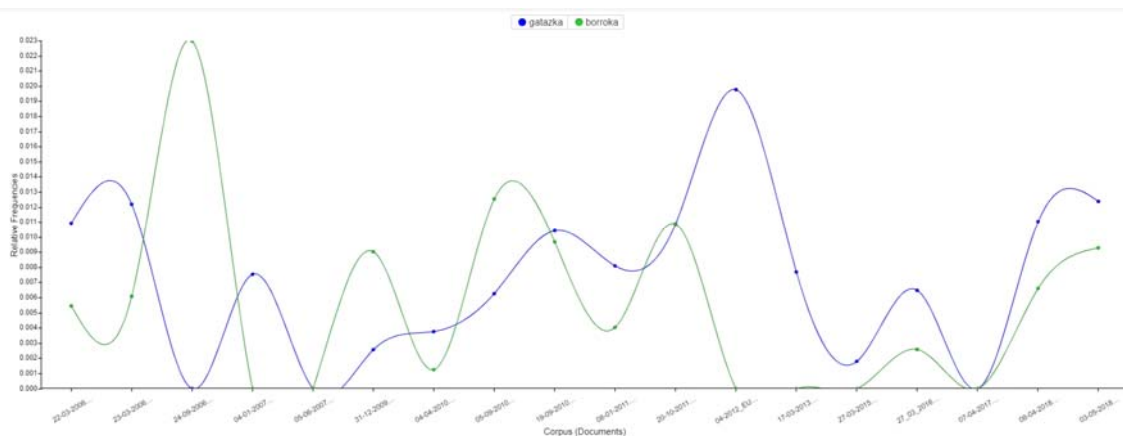


Figura 15: Freqüència relativa de “gatazka” (blau) i “borroka” (verd) dels comunicats d’ETA lematitzats en èuscar

“Gatazka” (‘conflicte’), que apareix 76 vegades i “borroka”(‘lluita’), que n’apareix 57, es troben entre les 30 paraules més freqüents del corpus. Totes dues experimenten una baixada de freqüència a partir de 2011-2012 i després tornen a aparèixer als últims comunicats de 2017 i 2018.

El pic de la freqüència relativa de “borroka” és al comunicat del 24 de setembre de 2006, durant el Gudari Eguna. Abans de l’any 2011, “gatazka” i “borroka” apareixen de manera variable, sent cadascuna més utilitzada que l’altra segons el cas, però a partir de l’any 2012, “gatazka” és l’opció preferida. “Gatazka”, per la seua banda, té el seu pic al comunicat d’abril de 2012, on ETA es dirigeix a la comunitat europea perquè impulsara el diàleg entre França i Espanya per trobar una solució al conflicte o “gatazka”. L’ús de “gatazka” creix de manera continuada des de 2009 fins aquest comunicat de 2012.

Els bigrames més freqüents amb “borroka” són “borroka armatu”(14) (‘lluita armada’), “borroka egin” (4) (‘lluitar’), “gatazkaren ondorio”(14) (‘conseqüència del conflicte’), “gatazka politiko”(10) (‘conflicte polític’), “gatazkaren konponbide”(8) (‘solució al conflicte’), “gatazka gainditu” (5) (‘superar el conflicte’). Com a trigràma tenim “ahalegin eta borroka”(3) (‘esforç i lluita’).

3.2.2.5. Camí i pas

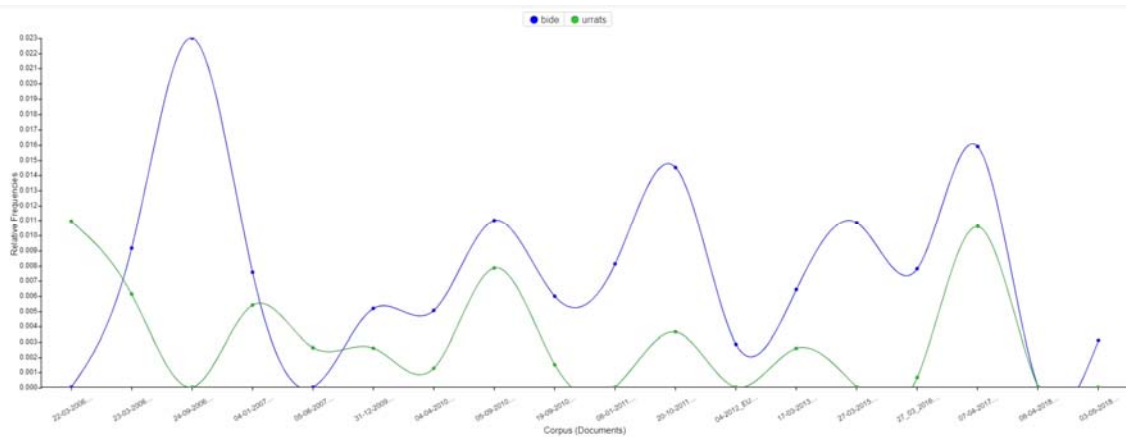


Figura 16: Freqüència relativa de “bide” (blau) i “urrats” (verd) dels comunicats d’ETA lematitzats en èuscar

“Bide” (‘camí’) i “urrats” (‘pas’) estan entre les 60 paraules més freqüents del corpus (“bide” ocupa el lloc 21) i tenen un nombre d’ocurrències de 75 i 28 en total respectivament. La presència de cadascuna és prou variant durant el temps, que “bide” sol ser-hi més present, però totes dues apareixen en un grau o un altre en gairebé tots els comunicats. “Bide” té el seu pic al 24 de setembre de 2006, al Gudari Eguna, i també té una freqüència relativa destacada al comunicat del 20 de novembre de 2011, on es declarava un fi definitiu a la lluita armada i el del 7 d’abril de 2017, on s’anuncia el desarmament total a la comunitat internacional. “Urrats”, tot i tindre una presència menor en comparació, destaca també a aquest últim comunicat de desarmament del 2017.

Els bigrames més freqüents amb lèxic de pes semàntic per “negozioaren bide”(5) (‘el camí de la negociació’), “urrats egin”(5) (‘fer passes’) i “urrats sendo”(4) (‘pas ferm’).

3.2.2.6. Construir i pau

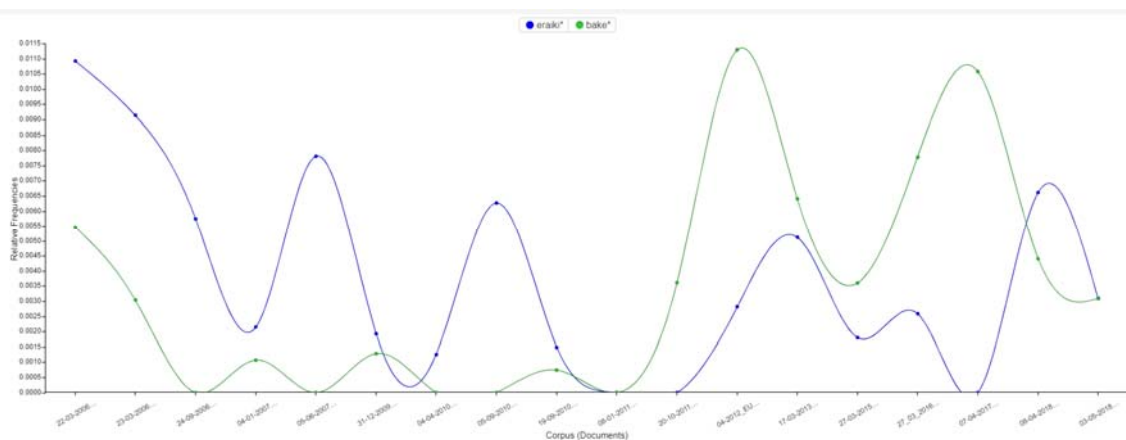


Figura 17: Freqüència relativa de “eraiki” (blau) i “bake” (verd) dels comunicats d’ETA lematitzats en èuscar

“Eraiki” (‘construir’) i “bake” (‘pau’) apareixen entre les paraules més freqüents del corpus: “bake” apareix un total de 35 vegades i “eraiki”, 31. La presència de “eraiki” és irregular durant el temps: es percep una freqüència relativa un poc major als comunicats que anuncien la treva de 2006, però en general les aparicions varien per comunicat i tenen una freqüència similar quan apareixen. “Bake” té una presència molt baixa al primer interval de temps (tot i que és un poc major als comunicats de treva de 2006) però una gran pujada a partir del comunicat de fi de la lluita arma de novembre de 2011; la freqüència es manté més alta que abans fins a la fi dels comunicats. Destaca la seua presència als comunicats d’abril de 2012 (on busquen establir diàlegs amb els governs espanyol i francès) i el d’abril de 2017 (on es declaren com una organització desarmada). Destaquen els bigrames “bake eraiki” (4) (‘construir la pau’), “bake elkarrizketa” (2) (‘diàleg de pau’), “bake lortu” (2) (‘aconseguir la pau’), “bake prozesu” (2) (‘procés de pau’), i els trigramas “marko demokratiko eraiki” (3) (‘construir un marc democràtic’), “bakea eta askatasuna” (6) (‘pau i llibertat’) i “askatasuna eta bakea” (3) (‘llibertat i pau’); en aquestes dues últimes veiem com la pau i la llibertat es mencionen sovint juntes.

Només hem mostrat els resultats de la comparació de lèxic en èuscar, ja que els comunicats en castellà oferien unes conclusions prou similars, llevat d’alguna decisió de traducció que fa diferir les freqüències, però en tot cas, gens rellevant.

4. Discussió

4.1. Comparació interlingüística

Com hem vist anteriorment, pràcticament la totalitat dels comunicats en castellà són descrits com “avançats” segons la seua llegibilitat (més complexos), mentre que gairebé tots els comunicats en èuscar analitzats reben l’etiqueta d’“elementals”. Aquest és l’únic valor que hem trobat comparable entre totes dues llengües; a causa de les diferències tipològiques que afecten les diferents mètriques de cadascú (Arregi i Fernández, 2002), no hem comparat altres mètriques entre si, ja que els seus valors estan condicionats a les estructures lingüístiques respectives, més enllà de les diferències reals de com estan escrits.

Per una banda, l’èuscar, en ser una llengua minoritzada, té una presència escrita menor a la d’altres llengües, i això pot dificultar el desenvolupament d’eines per avaluar la llegibilitat textual, sumant, a més, el fet que és una llengua aïllada amb poca semblança respecte a altres llengües (Madrazo Azpiazu i Pera, 2020). Tot i això, els resultats a Bengoetxea i Gonzalez-Dios (2021) destaquen que la fiabilitat de les mètriques extretes per l’èuscar amb MultiAzterTest és molt alta, aixina com el grau d’encert a l’hora d’assignar entre un nivell “elemental” o “avançat” per a cada text.

En general, les mètriques més predictives a l’hora de mesurar la llegibilitat a MultiAzterTest són les de llargària i vocabulari.

Per una banda, s’ha observat que hi ha una tendència cap a la simplificació al discurs polític (Benoit, Munger i Spirling, 2019), el qual tindria com a conseqüència una major claredat pel lector i una necessitat d’un nivell de lectura menor per resultat comprensible. Com s’ha vist a González Aparicio (2017), el lèxic utilitzat per ETA és de tipus tècnic i té una intenció de projectar objectivitat com a estratègia discursiva de legitimitat. Aquest estil lèxic triat pot haver influït a les mètriques de llegibilitat obtingudes amb el còmput global. Com s’ha esmentat a l’apartat de resultats, els comunicats en castellà tendeixen a utilitzar més paraules de contingut o lèxiques diferents entre sí que els comunicats en èuscar. Una possible explicació per això és una opció estilística concreta per escriure en èuscar en comparació al castellà, a banda que una part important dels comunicats en castellà són traduccions no fetes per ETA, de manera que aquest estil pot haver variat. Destaca el fet que els comunicats amb valors menors de paraules inusuals en castellà, que han sigut superats pels valors en èuscar, eren en general originals d’ETA en castellà, el

qual reforça més la idea que són sobretot les traduccions fetes per altres mitjans les que adopten un lèxic més variat en castellà.

4.2. Comunicats en castellà, traduccions i originals

Les diferències de mètriques entre traduccions i originals en castellà no pareixen ser molt significatives. Mentre que els originals han destacat a mètriques de densitat lèxica, al nombre de sintagmes verbals o nominals, als nivells de profunditat als arbres de dependència i a la llargària de les frases, les traduccions tenien més paraules inusuals, més semblança entre els paràgrafs, més paraules negatives o més signes de puntuació i nombre de frases (el qual té sentit en comparació; si l'altre grup té frases més llargues, aquest altre tindrà més frases). Mostren característiques un poc diferents, però no podem establir que uns siguin clarament més complexos que els altres. Així i tot, de tots els comunicats, només dos han sigut categoritzats com “elementals” segons MultiAzterTest, i aquests han sigut les dues traduccions, de manera que, de 10 traduccions, 8 han sigut més complexes i 2 més simples, i dels 8 originals, tots han sigut més complexos. L'efecte de la *translationese* més evident concordant amb la concepció de Baroni i Bernardini (2006) és que les traduccions han presentat una diversitat lèxica menor a la dels originals, aixina com una menor diversitat d'adjectius i noms. Tot i això, hem pogut comprovar que aquestes traduccions, tot i presentar una menor densitat lèxica, sí que tenen més freqüència de paraules lèxiques inusuals. Les traduccions també presenten més subordinades, cosa no esperable en principi, però els originals han tingut una major llargària de frases i més nivells de profunditat a l'arbre de dependències. Els efectes de la traducció en aquests casos, doncs, es percep de manera molt lleu.

4.3. Canvis diacrònics

L'evolució històrica de les mètriques en castellà no ha donat resultats molt significatius, i aquests són també qüestionables per qüestions de *translationese*: la majoria dels comunicats de la primera etapa (2006-2010) són traduccions, mentre que els de la segona (2011-2018) són originals. Per tant, el primer grup històric ha destacat en mètriques paregudes a les del grup de traduccions, com són la freqüència de paraules inusuals o la semblança semàntica entre diferents paràgrafs. El grup 2 ha tingut, com al grup dels originals, valors més alts relacionats amb la densitat adjectival, a banda de més síl·labes per paraula (és a dir, paraules més llargues).

En èuscar no hi entra en joc la traducció, ja que tots són originals, i les mètriques més significatives segons (Bengoetxea i Gonzalez-Dios, 2021), aixina com les mètriques que

més diferències donaven entre els dos grups segons Weka, han destacat més al segon grup temporal. Pel que fa a les mètriques més importants a l'hora de determinar la llegibilitat, el grup 2 ha destacat en la llargària de les frases, els sintagmes verbals, la llargària de les paraules (excloent les paraules funcionals) i la profunditat de les subordinades. El grup 1, en canvi, ha destacat a la llargària dels lemes i de les paraules (incloent-hi les de categoria funcional), així com al nombre de paràgrafs. A més a més, amb Weka s'hi ha trobat més mètriques amb diferències significatives que en castellà: de les 25 mètriques recomanades, 19 han mostrat resultats significatius, i en 14 d'aquests hi ha destacat el segon grup. El primer grup mostra més verbs en infinitiu, pronoms i paraules inusuals, mentre que el segon grup ha destacat per mètriques relacionades amb la densitat lèxica, la desviació estàndard de diferents valors (la semblança semàntica entre frases i paràgrafs o la llargària de frases i paraules), els verbs en indicatiu i especialment per l'ús de connectors total. És aquest últim cas al qual es veu més palpablement un augment sostingut al temps del valor, i passa també amb els connectors de tipus temporal i lògics. És aquesta la diferència més important entre tots dos grups, tot i que a (Bengoetxea i Gonzalez-Dios, 2021), els connectors discursius no es consideren un dels factors més importants a l'hora de determinar la llegibilitat, encara que es situen més bé en la meitat de la classificació en comparació als altres tipus de mètriques. Un major grau de desviació estàndard tampoc és directament signe de major llegibilitat, ja que entra en joc amb altres mesures. Així, amb les mètriques restants, el grup 2 pareix tenir algunes mètriques més grans que el primer, però és un resultat, com els altres, relatiu a la petita mostra amb la qual s'ha fet l'estudi. A més, hem de recordar que l'únic comunicat que ha sigut classificat com a "avançat" pertany al grup 1 (és, de fet, l'últim comunicat que tenim de 2010 i, per tant, del primer grup temporal).

4.4. Lèxic més freqüent

Com s'ha pogut observar a les paraules més utilitzades en totes dues llengües, hi ha una presència alta de paraules de categoria funcional, és a dir, les paraules que no tenen informació semàntica sinó que compleixen una funció merament gramatical, com poden ser determinants o conjuncions, però també verbs auxiliars, com ara "izan" o "ukan" en èuscar. És esperable trobar-les a les posicions més altes dels estudis de freqüències pel seu paper gramatical. De les 60 paraules més freqüents extretes, eren paraules funcionals o stopwords 35 d'elles en castellà i 30 en èuscar, el qual es pot explicar per les diferències tipològiques de cada llengua: per exemple, en èuscar, una llengua aglutinant, el genitiu

s'expressa amb un sufix de cas al posseïdor, mentre que al castellà s'expressa principalment amb la paraula funcional “de”, preposició.

Pel que fa a les paraules lèxiques, podem veure, com s'ha comentat a altres estudis sobre comunicats d'ETA (González Aparicio, 2017; Hernández Valls, 2017), que el vocabulari emprat és prou tècnic, i que destaca per la presència de conceptes de la ciència política, la governança o el dret (González Aparicio, 2017), com podem veure a l'ús de “politiko”, “demokratiko” i els altres mots destacats als resultats. Això es relaciona directament amb una intenció de presentar les idees pròpies com objectives i legitimar-les davant els altres. Paraules com “compromís”, “pau” i “procés” són també les més utilitzades als comunicats de l'IRA des de l'anunci del seu alto el foc l'any 1994 fins a la fi de la lluita armada el 2005 (Cardillo, 2017). En aquest sentit, els comunicats tenen la funció de presentar positivament les accions i decisions preses pel grup; a banda, en tots dos casos es tracta de vocabulari emprat durant els seus respectius processos finals com a grup armat de desmilitarització i de diàleg amb els governs, i part de la seua legitimació es basa a mostrar-se col·laboratiu durant la fi de la seua activitat.

4.5. L'ús d'ETA

De la mateixa manera que passa amb els comunicats de l'IRA a Cardillo (2017), el nom de la banda és una de les paraules més freqüents: ETA apareix mencionada un total de 165 vegades a tot el corpus; 208 si hi sumem les altres versions amb les quals es esmentada, E.T.A. i Euskadi Ta Askatasuna, el nom complet. És la tercera paraula més freqüent, per darrere de “euskal” i “herria”, que apareixen la major part de les voltes juntes al bigrama “euskal herria” (194 vegades); als comunicats de l'IRA, el nom de la banda és la paraula més freqüent, mentre que “irlandés” es la novena. E.T.A. i Euskadi Ta Askatasuna apareixen al principi o a la fi dels comunicats, bé a la frase utilitzada normalment al principi dels comunicats, “Euskadi Ta Askatasunak, nazio askapenerako euskal erakunde sozialista iraultzaileak” (“Euskadi Ta Askatasuna, organización socialista revolucionaria vasca de liberación nacional”) o bé com a firma final, després d'una consigna sovint utilitzada baix el comunicat, “Gora Euskal Herria askatuta! Gora Euskal Herria sozialista! Jo ta ke independentzia eta sozialismoa lortu arte!” (“Visca Euskal Herria lliure! Visca Euskal Herria socialista! Amb força fins a aconseguir la independència i el socialisme!”). És sobretot per aquesta raó que “sozialista” és una paraula tan freqüent, per la seua aparició regular al principi i al final dels comunicats, dins de les fórmules indicades (de les 30 vegades que apareix, 15 apareixen a la construcció

“nazio askapenerako euskal erakunde sozialista iraultzaileak” del principi i 14 a la frase final “Gora euskal herria sozialista!”. El lligam del socialisme amb la banda o amb Euskal Herria són propis de la ideologia d’ETA des dels seus inicis, per tant, era esperable trobar col·locacions on aparegueren junts.

4.6. Lèxic, estratègies i metàfores

Les referències als estats espanyol i francès són habituals per ser els escenaris principals de l’actuació d’ETA, als quals es troben els diferents territoris que conformen Euskal Herria. Les mencions a Espanya (76) són més freqüents que a França (44), el qual té sentit si tenim en compte que l’últim tram d’Euskadi Ta Askatasuna està marcat pel diàleg amb el govern espanyol de Zapatero: de les 76 referències a Espanya, 32 són a la construcció “espainiako governu” (govern d’Espanya). També de les 44 mencions a França, el bigrama més important és “frantziako governu” (12 vegades), seguit de “frantziako polizia”, que apareix 8 vegades, totes elles al comunicat del 4 d’abril de 2010, on s’explicava la mort d’un gendarme francès a un tiroteig amb ETA (i qui va resultar ser la seua última víctima). Com ja hem vist als resultats, és aquest comunicat on la freqüència relativa de les mencions a França és més alta, comparat inclús amb altres comunicats on també es feia referència al govern francès. A més a més, Espanya i França apareixen esmentades juntes a 22 ocasions (com a “frantzia eta espainia” o “espainia eta frantzia”), i es fa sempre, de nou, per parlar d’ells com a govern, estat, o per mencionar als seus mandataris. En aquest sentit, la pràctica totalitat de les mencions a aquests estats té a veure amb els seus governants, i no amb els seus ciutadans.

Una de les troballes més curioses és el creixement de l’ús de “konponbide” i “elkarrizketa” a partir de 2010 fins a 2013 i que podem relacionar directament amb les conseqüències que va tenir la declaració d’Altsasu a 2009 al context posterior a l’esquerra abertzale i al fi de la lluita armada. El període està marcat per la mediació internacional, el plantejament de l’alto el foc definitiu i la negociació entre ETA i l’estat espanyol sobre les conseqüències del conflicte (Ripoll García, 2017). També el discurs del president del govern espanyol tenia una marcada tendència a parlar de diàleg (Hellín-García, 2013) i utilitzava diferents estratègies discursives (especialment, les metàfores de construcció) per legitimar-lo com una de les claus per la fi del terrorisme.

L’ús del terme “gatazka” (conflicte) és habitual als comunicats d’ETA per parlar de la situació patida al país basc (Hernández Valls, 2018), tot i que veiem una pujada a partir també del període esmentat, relacionada amb la fi de la lluita armada. Com s’ha vist als

resultats, es parla d'un conflicte polític per al qual es busquen solucions (com hem vist, també augmenta l'ús de "solució" especialment durant aquesta època). Apareix doncs un llenguatge més conciliador i suavitzat en sincronia amb el context temporal. "Borroka", en canvi, té una presència més aviat variable, però aquesta és menor que "gatazka" a partir de 2011. Tot i que podríem traduir-la com "lluïta", és una paraula amb una forta significació a la història de l'esquerra abertzale, per tant, no és estrany que aparega amb una freqüència molt alta al comunicat del Gudari Eguna de 2006, que segons González Aparicio (2017), es podria considerar un comunicat de caràcter més aviat identitari pels temes que tracta. Tot i això, en altres dies commemoratius importants d'anys posteriors a l'alto el foc definitiu, com al Gudari Eguna de 2015 i o l'Aberri Eguna 2016, els comunicats corresponents no presenten cap aparició de "borroka"; no ho fa cap dels comunicats des de 2012 fins a 2017, quan ETA procedeix a l'entrega de tot el seu armament. Tot i que reapareixen als últims comunicats, el canvi de tendència respecte a l'ús (o més prompte, el no ús) d'aquest mot és notable.

Els usos de "bide" i "urrats" són metafòrics: no fan referència a camins i passos reals, sinó que serveixen per a parlar de processos i etapes d'aquests. Segons Johnson (1987), la metàfora de viatge és "PURPOSEFUL ACTIVITY IS TRAVELLING ALONG A PATH TOWARDS A DESTINATION" ("una activitat amb propòsit és viatjar al llarg d'un camí cap a una destinació"), a partir de la recerca de lingüistes cognitivistes a l'hora de definir les diferents metàfores. Les metàfores viatge/camí/carretera són i han sigut prou populars al discurs polític occidental (Koteyko i Ryazanova-Clarke, 2009), i com podem observar, també són prou presents al lèxic d'ETA als comunicats del període estudiat. "Bide" com a camí és l'exemple més representatiu d'aquest tipus de metàfora, però "urrats", com pas, també hi està relacionat amb la noció de viatge i de camí com escenari d'aquest viatge. En aquest sentit, l'alta freqüència de "prozesu" (procés), que apareix 97 vegades al corpus (és la cinquena paraula lèxica més freqüent), també acompanya la noció del viatge/camí a recórrer.

El cas de "eraiki" també fa referència a una metàfora: no parlen d'una construcció física de cap cosa, sinó de la construcció de termes abstractes. La noció metafòrica de la construcció, com la del viatge, és productiva al discurs polític occidental (Koteyko i Ryazanova-Clarke, 2009); de fet, s'ha trobat prou utilitzada pel discurs de Zapatero durant els anys del seu mandat en relació amb el terrorisme (Hellín-García, 2013). Aquestes i altres metàfores tenen poder polític per la seua capacitat de persuasió (Mussolf

2004; Goatly 2007), i participen en l'estratègia de legitimació de donar una visió pròpia positiva (o també de donar-ne una de negativa per l'adversari). A la seua formulació de les diferents metàfores, Lakoff i Johnson (1980) van proposar ARGUMENT IS A BUILDING (l'argument és una construcció) i THEORIES ARE BUILDINGS (les teories són construccions), de manera que aquests estiguen basats en idees i fets sòlids. És aquesta noció de solidesa la que acompanya l'ús d'aquestes metàfores a l'hora de descriure processos i accions. Per exemple, trobem el bigrama “bake eraiki” (4) (construir la pau) o el trigràma “marko demokratiko eraiki” (3) (construir un marc democràtic); aquests últims apareixen especialment als comunicats de l'alto el foc de 2006.

Finalment, l'evolució del terme “bake” (pau) és una mostra clara del desenvolupament d'aquests últims anys. Abans del 2011, els únics comunicats on té una presència més rellevant són els de març de 2006, quan van declarar l'alto el foc d'aquell any. No és fins 2011, i especialment 2012, que hi retorna amb força i hi queda fins als últims comunicats. També va ser un dels mots més utilitzats després de l'alto el foc de l'IRA (Cardillo, 2017), i està relacionat amb el llenguatge posterior a la fi de la lluita armada, com ho estan també altres mots com “konponbide”, “elkarrizketa” o “gatazka”, en línia amb el lèxic utilitzat pel govern espanyol de l'època (Hellín-García, 2013).

5. Conclusions

No s'ha trobat diferències especialment notables pel que fa a la llegibilitat dels comunicats i les mètriques que la conformen al llarg del temps, exceptuant, pot ser, l'augment visible de connectors en èuscar.

Els resultats apunten que el grup dels originals en castellà tindria algunes mètriques que els farien més complexos respecte a les traduccions, el qual reforçaria la idea de la *translationese* com un element que afecta la complexitat dels textos traduïts. No obstant això, la petita mida de la mostra ens obliga a relativitzar aquests resultats.

Pel que fa a la comparació interlingüística, els textos en castellà pareixen ser més complexos que els textos en èuscar. De nou, la quantitat de textos processats és prou baixa per a establir patrons grans, però pareix veure's una tendència cap a textos més simples en èuscar.

El lèxic, com hem vist, sí que pareix haver mostrat alguns resultats que podem relacionar amb el context de publicació dels comunicats: l'ús d'un lèxic més conciliador a partir de

la fi de la lluita armada; l'ús les metàfores de construcció i viatge, pròpies del discurs polític occidental actual del govern espanyol dels últims anys abans l'alto el foc; l'ús de lèxic tècnic per mostrar objectivitat; tot són estratègies compatibles amb la cerca de legitimitat que apareixen al discurs públic de la política i d'altres bandes armades i moviments socials.

En estudis posteriors, seria interessant desenvolupar un corpus amb més textos i més recorregut històric per comprovar si podem observar més diferències diacròniques, tant en l'àmbit lèxic com de llegibilitat. Un corpus més gran de comunicats en castellà també podria oferir resultats més concloents respecte a la influència de la *translationese* als comunicats en ser traduïts per altres mitjans. Fins i tot, una cerca detallada de diferents mètriques relacionades amb els efectes de la traducció ens podria acostar a saber en quina llengua escrivia ETA els originals; en un principi aquests són en èuscar, però la història de l'organització és llarga i no podem descartar que hi hagen existit diferents estratègies a l'hora de redactar-los. No hem d'oblidar que l'èuscar ha tingut una forta recessió a Euskal Herria i que no tothom hi ha tingut accés a la llengua.

Un altre aspecte que no s'ha mencionat en aquest estudi ha sigut els diferents autors de comunicats que ha tingut al llarg del temps. Amb un estudi de més recorregut, es podrien establir perfils lingüístics relacionant-los amb l'autor o autors de cada etapa, diferenciant estils personals.

Finalment, una tasca que queda per fer amb el corpus creat, bilingüe, és convertir-lo en un corpus paral·lel on establir les correspondències del comunicat en cada llengua.

6. Bibliografia

- Abdelrahim, Y. A. A. (2019). Visual Analysis of ISIS Discourse Strategies and Types in Dabiq and Rumiyah Online Magazines. *Visual Communication Quarterly*, 26(2), 63-78. <https://doi.org/10.1080/15551393.2019.1586546>
- Anthony, L. (2021). *AntConc* (Versió 3.5.8.) [Computer Software]. Tokio, Japó: Universitat de Waseda. <https://www.laurenceanthony.net/software>.
- Arregi, O. i Fernández, I. (2002). Clasificación de documentos escritos en euskara: impacto de la lematización. A E. Sanchis, L. Moreno i I. Gil (Eds.), *I Jornadas de Tratamiento y Recuperación de Información (JOTRI)* (p. 29-36). València: Universitat Politècnica de València.

- Baroni, M. i Bernardini, S. (2006). A new approach to the study of translationese: Machine-learning the difference between original and translated text. *Literary and Linguistic Computing*, 21(3), 259-274. <https://doi.org/10.1093/lc/fqi039>
- Bengoetxea, K. i Gonzalez-Dios, I. (2021). Multiaztertest: A multilingual analyzer on multiple levels of language for readability assessment. *arXiv preprint arXiv:2109.04870*. <https://doi.org/10.48550/arXiv.2109.04870>
- Benoit, K., Munger, K. i Spirling, A. (2019). Measuring and explaining political sophistication through textual complexity. *American Journal of Political Science*, 63(2), 491-508. <https://doi.org/10.1111/ajps.12423>
- Bilbao Ariztimuño, K. (2008). *Crónica de una izquierda singular (De ETA-Berri a EMK/MC y a Zutik-batzarre)*. Edició Electrònica.
- Bustos Gavilán, D. (2022). *Acercamiento cuantitativo al discurso político chileno en democracia* [Treball de fi de grau]. Universitat de Barcelona.
- Cardillo, N. (2017). The power of language: a critical discourse analysis of IRA's statements. *International Journal of Philosophy*, 23, 151-169. <https://doi.org/10.7413/18281567095>
- El-Nashar, M. i Nayef, H. (2019). 'Cooking the Meal of Terror' Manipulative Strategies in Terrorist Discourse: A Critical Discourse Analysis of ISIS Statements. *Terrorism and Political Violence*, 34(1), 155-175. <https://doi.org/10.1080/09546553.2019.1676238>
- Eveland Jr, W. P. i Cortese, J. (2004). How web site organization influences free recall, factual knowledge, and knowledge structure density. *Human Communication Research*, 30(2), 208-233. <https://doi.org/10.1111/j.1468-2958.2004.tb00731.x>
- Foucault, M. (1970). *El orden del discurso*. Buenos Aires: Tusquets.
- Frank, E., Hall, M. A. i Witten, I. H. (2016). *The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"* (4.^a ed.). Morgan Kauffman.
- García-Miguel, J. M. (2022). Lingüística de corpus: de los datos textuales a la teoría lingüística. *Estudios de lingüística del español*, 45, 11-42. <https://doi.org/10.36950/elies.2022.45.8848>

- Garmendia, H. i Urteaga, E. (2017). La langue basque dans le discours nationaliste basque. *Cahiers internationaux de sociolinguistique*, 11, 169-199. <https://doi.org/10.3917/cisl.1701.0169>
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68(1), 1-76. [https://doi.org/10.1016/s0010-0277\(98\)00034-1](https://doi.org/10.1016/s0010-0277(98)00034-1)
- Goatly, A. (2007). *Washing the Brain: Metaphor and Hidden Ideology*. John Benjamins Publishing.
- González Aparicio, C. J. (2015). *De la violencia de ETA al terrorismo de España: Análisis de los comunicados de ETA* [Tesi de màster]. Universidad de Chile.
- Hellín-García, M. J. (2013). Legitimization and delegitimization strategies on terrorism: A corpus-based analysis of building metaphors. *Pragmatics*, 23(2), 301-330. <https://doi.org/10.1075/prag.23.2.05hel>
- Hernández Valls, F. (2017). *Terrorismo y propaganda en los medios de comunicación. El final de ETA a través de la prensa* [Tesi doctoral]. Universidad Complutense de Madrid.
- Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason*. University of Chicago Press.
- Kleinnijenhuis, J. (1991). Newspaper complexity and the knowledge gap. *European journal of communication*, 6(4), 499-522. <https://doi.org/10.1177/0267323191006004006>
- Koteyko, N. i Ryazanova-Clarke, L. (2009). The Path and Building Metaphors in the Speeches of Vladimir Putin: Back to the Future? *Slavonica*, 15(2), 112-127. <https://doi.org/10.1179/136174209x12507596634810>
- Kurokawa, D., Goutte, C. i Isabelle, P. (2009). Automatic detection of translated text and its impact on machine translation. A *Proceedings of Machine Translation Summit XII*. Ottawa, Canada.
- Madrazo Azpiazu, I. i Pera, M. S. (2019). Is cross-lingual readability assessment possible? *Journal of the Association for Information Science and Technology*, 71(6), 644-656. <https://doi.org/10.1002/asi.24293>

- Mi Oh, Y., Pellegrino, F., Marsico, E. i Coupé C. (2013). A Quantitative and Typological Approach to Correlating Linguistic Complexity. A T. Wielfaert, K. Heylen i D. Speelman (Eds.), *Proceedings of the 5th Conference on Quantitative Investigations in Theoretical Linguistics* (p. 71-75). Universitat de Leuven.
- Musolff, A. (2002). *Metaphor and Political Discourse*. Palgrave Macmillan.
- Lakoff, G. i Johnson, M. H. (1980). *Metaphors We Live By*. University of Chicago Press.
- Liu, K. i Afzaal, M. (2021). Syntactic complexity in translated and non-translated texts: A corpus-based study of simplification. *PLoS ONE*, 16(6): e0253454. <https://doi.org/10.1371/journal.pone.0253454>
- Ripoll García, C. V. (2017). *La violencia terrorista de ETA y su influencia en la opinión pública* [Treball de fi de grau]. Universitat Miguel Hernández.
- RStudio Team. (2020). *RStudio: Integrated Development Environment for R* (Versió 4.2.2.) [Computer Software] RStudio, PBC, Boston. Disponible a <http://www.rstudio.com/>.
- Souza, J. A. S. D. i Borba, J. A. (2021). Readability as a measure of textual complexity: determinants and evidence in Brazilian companies. *Revista Contabilidade & Finanças*, 33, 112-129. <https://doi.org/10.1590/1808-057x202114180>
- Stéfan, S. i Rockwell, G. (2016). *Voyant Tools* (Versió 2.0) [Versió web]. Disponible a <http://voyant-tools.org/>.
- Straka, M., Hajic, J. i Straková, J. (2016, May). UDPipe: trainable pipeline for processing CoNLL-U files performing tokenization, morphological analysis, pos tagging and parsing. A N. Calzonari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk i S. Piperidis (Eds.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (p. 4290-4297). Portorož, Eslovènia.
- Tolochko, P., Song, H. i Boomgaarden, H. (2019). “That looks hard!”: Effects of objective and perceived textual complexity on factual and structural political knowledge. *Political Communication*, 36(4), 609-628. <https://doi.org/10.1080/10584609.2019.1631919>

Vasco Press. (2009, 19 de setembre). ETA establece que el euskera será la única lengua interna de la banda. *El Correo*.
<https://www.elcorreo.com/vizcaya/20090919/mas-actualidad/politica/establece-euskera-sera-unica-200909191929.html>