

Grau de Lingüística

Treball de Fi de Grau

Curs 2019-2020

**Marcadors lingüístics de la personalitat antisocial: una
detecció mitjançant l'aprenentatge automàtic**

Júlia Donisa Barceló

Tutora: Irene Castellón Masalles

Barcelona, 12 de juny de 2020



Declaració d'autoria

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 12 de juny de 2020

Signatura:

Resum

El llenguatge utilitzat durant el procés comunicatiu pot ser indicatiu tant de trets de la personalitat com d'estats emocionals, identitaris i cognitius. Si bé aquesta relació ha estat explorada tant en patologies mentals com en fenòmens socials, és poca la investigació que s'ha efectuat respecte als trastorns de la personalitat, en concret pel que fa a la personalitat antisocial.

En aquest estudi s'ha buscat identificar marcadors lingüístics característics de la personalitat sociopàtica mitjançant l'anàlisi estadístic i tasques computacionals pertanyents al camp de la mineria de dades i l'aprenentatge automàtic. Els resultats obtinguts apunten cap a una tendència a l'evitació dels pronoms de primera persona de plural i a mots referents a qualsevol tipus d'afiliació social; juntament amb una alta presència, en canvi, de pronoms de primera persona singular.

Paraules clau: psicopatia, trastorn antisocial, aprenentatge automàtic, mineria de dades

Abstract

Language used during the communicative process can be an indicator of personality traits and emotional, identity or cognitive states. Although this relation has been explored in terms of mental pathologies and social phenomena, there has been little investigation in respect of personality disorders, especially in which refers to antisocial personality disorder.

This study has been aimed at searching for the identification of linguistic markers that are characteristic of sociopathic personality by means of statistical analysis and computational tasks proprium to data mining and machine learning fields. The results obtained point towards an avoidance tendency of singular first person pronouns and words referring to social affiliation; together with a high presence of singular first person pronouns.

Keywords: psychopathy, antisocial disorder, machine learning, data mining

ÍNDIX

1. INTRODUCCIÓ	4
1.1. Objectius i limitacions	5
2. MARC TEÒRIC	6
2.1. Marcadors lingüístics de la personalitat antisocial	6
3. METODOLOGIA	9
3.1. Hipòtesi	9
3.2. Creació dels corpus	9
3.3. Importació de les dades sense processar a WEKA	10
3.3.1. Preprocessament de les dades	10
3.3.1.1. Refinament d'atributs	12
3.3.2. Elecció del classificador	13
3.4. Importació de les dades processades a WEKA	14
3.4.1. Tractament dels corpus amb LIWC i SPSS	14
3.4.2. Preprocessament de les dades	16
3.4.3. Classificació	17
3.4.3.1. Classificació amb J48	17
3.4.3.2. Classificació amb RandomTree	18
3.4.3.3. Classificació amb SMO	20
4. RESULTATS I DISCUSSIÓ	21
5. CONCLUSIONS	26
6. REFERÈNCIES BIBLIOGRÀFIQUES	28
7. ANNEX	31

1. INTRODUCCIÓ

El llenguatge produït durant el procés comunicatiu pot ser indicatiu tant de trets de la personalitat com d'estats emocionals, identitaris i cognitius (Pennebaker, 2011; Pennebaker & King, 1999; Pennebaker, Mehl, & Niederhoffer, 2003, Gottschalk & Betchel, 1995). Diversos estudis han examinat la relació entre l'ús lingüístic i diferents trastorns, com ara la depressió (e.g. Pennebaker & Graybeal, 2001) o la paranoia (e.g. Oxman, Rosenberg, Schnurr & Tucker, 1988); així com entre la personalitat i l'ús lingüístic (e.g. Oberlander & Gill, 2006). No obstant, la relació del trastorn antisocial de la personalitat —popularment conegut amb el terme de *sociopatia* o *psicopatia*— amb l'ús lingüístic ha estat, fins ara, poc examinada (Hare, 1993).

Els trastorns de la personalitat, segons l'APA (*American Psychiatric Association*) són conductes definides com “patrons persistents d'experiència interna i comportament que s'aparta marcadament de les expectatives de la cultura de la persona que l'exhibeix”. N'hi ha de diversos tipus, però el grup que mostra els malajustaments més inconfusibles a la vida social és el d'aquelles persones que presenten personalitats psicopàtiques (Bromberg, 1948).

Hare (2003) senyala que “la personalitat psicopàtica presenta combinacions específiques de característiques cognitives, socials i emocionals que la diferencien de la població general”. Concretament, es caracteritza per presentar una capacitat disminuïda per a la sensibilitat moral, una visió del món i de les persones merament instrumental (Porter & Woodworth, 2007) i una marcada tendència narcíssista. Estudis elaborats amb l'ajuda de tècniques de neuroimatge suggereixen que tals particularitats podrien deure's a reduccions de matèria gris en les àrees frontal i temporal (de Oliveira-Souza et al., 2008) i anomalies al còrtex prefrontal i al cos callós (Raine et al., 2004).

Tant els estudis previs dedicats a retratar amb detall la personalitat antisocial i l'ús lingüístic que la caracteritza (v. Ceckley, 1941; Rieber & Vetter, 1994) com els estudis experimentals que se n'ha fet al respecte (v. Henning, 2017; Hancock, Woodworth & Porter, 2011; Hancock, Woodworth & Boochever, 2018; Endres, 2004; Porter & Woodworth, 2007) suggereixen, en efecte, que existeixen unes característiques cognitives i socials definitòries d'aquest trastorn que poden materialitzar-se en patrons lingüístics concrets, als quals anomenarem *marcadors lingüístics*.

1.1. Objectius i limitacions

En aquest estudi es pretén recopilar els marcadors lingüístics que s'han atribuït a la personalitat sociopàtica al llarg de la literatura i aplicar tècniques de processament del llenguatge natural, com ara l'aprenentatge automàtic supervisat i la minera de dades, en textos pertanyents a individus amb aquest trastorn. L'objectiu últim és comprovar si, efectivament, tals patrons lingüístics són definitoris i tenen més presència en el discurs antisocial que no pas en el discurs d'una personalitat típica. Cal insistir en el fet que el que es vol és avançar cap a la caracterització d'un tipus de discurs en concret i no pas en la diagnosi de cap trastorn, ni molt menys en la detecció dels individus amb personalitat sociopàtica entesos com a "perill potencial per a la societat", tal com succeeix sovint en la cultura popular.

Amb tot, existeixen certes limitacions en l'estudi. Primer de tot, les dades textuais que s'ha fet servir són poc extenses a causa de la dificultat que suposa l'accés directe a persones diagnosticades amb aquest trastorn. En conseqüència, s'ha obtingut les dades via Internet, com es detalla en l'apartat referent a la metodologia.

Atès que la gran majoria d'estudis que examinen els marcadors lingüístics de la personalitat psicopàtica es basen en narratives d'individus no insertats en la societat que descriuen els seus actes criminals, es pot plantejar com a limitació d'aquest estudi el fet de no disposar exclusivament d'aquest tipus de discurs. No obstant, aquest fet no té per què suposar una barrera, ans al contrari: ens brinda la possibilitat d'examinar els trets del llenguatge sociopàtic en totes les seves vessants discursives.

De la limitació anterior en deriva una altra: tal com postula Henning (2017), "l'escriptura pot ser diferent [de la llengua parlada] en el sentit que una persona pot editar les seves preferències. El nombre de filtres pot incrementar-se, i en conseqüència amagar o suavitzar una intenció genuïna". Tanmateix, tal com fa notar el mateix autor, "els anàlisis psicològics moderns es fan o bé per una sèrie d'entrevistes o bé per inventaris de personalitat escrits [...], i una persona sempre té la oportunitat d'amagar o suavitzar les seves declaracions a un analista. En altres paraules: quan una persona sap que està sota el judici d'un professional mèdic especialitzat, sempre pot exercir restriccions en el que diu" (p. 15).

2. MARC TEÒRIC

2.1. Marcadors lingüístics de la personalitat antisocial

En aquest estudi s'ha efectuat una tasca de recopilació de les característiques més definitòries de la personalitat sociopàtica, així com dels marcadors lingüístics que s'han atribuït a aquesta al llarg de la literatura (v. Ceckley, 1998; Rieber & Vetter, 1994; Henning, 2017; Hancock, Woodworth & Porter, 2011 o Endres, 2004 i Porter & Woodworth, 2007, entre d'altres).

Si bé hi ha trets psicopàtics que no es poden relacionar amb cap marcador lingüístic (com ara la falta de sinceritat patològica) o que, atesa la naturalesa d'aquest estudi, no es poden quantificar (com la presència de disfluències en el discurs i de pauses més llargues del normal), existeixen moltes característiques que sí que estan concretades en patrons d'ús lingüístic determinats que són observables en fonts textuais.

L'individu amb trastorn antisocial de la personalitat presenta una marcada tendència a la cerca constant d'adrenalina i, sobretot, de poder. Tal com exposen Rieber & Vetter (1994),

[Els individus sociòpates] no només mostren un interès extrem per les relacions de poder i en la pròpia obtenció del màxim poder possible, sinó que semblen determinats a usar-lo amb finalitats destructives. [El poder] és entès com un atribut que només pot ser experimentat si és en el context de la victimització: si un és fort, és natural que algú altre en surti perjudicat (p. 12).

Aquesta atracció pel concepte de *poder* suggereix que les persones amb trastorn psicopàtic es refereixen més sovint a termes relacionats amb aquest (com ara *dominar*, *lideratge* o *sotmetre*).

Així mateix, presenta un coneixement superficial respecte al significat de termes que no formen part de la seva realitat, sobretot pel que fa a les emocions. De fet, un fet que el caracteritza és que “els seus crims no són perpetrats per sentiments com la venjança, la ràbia, la por o la gelosia; sinó que solen ser accions planejades i raonades que els beneficien en termes sexuals, econòmics o materials” (Rieber & Vetter, 1994). Aquesta característica es tradueix en diferents particularitats textuais: per començar, en l'ús d'un llenguatge que denota distanciament psicològic i falta de responsabilitat a través d'un ús més generalitzat del temps passat (Hancock et al., 2013). D'altra banda, en una major aparença de termes referits a necessitats més fisiològiques que espirituals (conceptes

pertanyents als camps semàntics de diners, refugi, sexe, menjar, roba i similars tindrien més pes que no pas termes socials, familiars, religiosos o espirituals) que, de rebot, podria veure's reflectida també en una presència major de termes referents als cinc sentits (Au, 2009). Per últim, també es reflecteix en en una producció de menys —i de menys intenses— paraules emocionals (Hancock et al. 2013).

Una altra característica preponderant de la personalitat psicopàtica és la tendència egocèntrica i narcissista. L'individu antisocial té poca o nul·la dependència dels altres (Huprich, Gacono, Schneider & Bridges 2004) i presenta una notable afició per parlar d'un mateix i del seu passat; d'explicar-se i de justificar-se. (Hancock et al., 2013).

Endress (2004), en el seu treball de caracterització del llenguatge psicopàtic, relacionà aquesta tendència narcissista amb la teoria del desenvolupament de l'ego de Loevinger (1976) a causa dels paral·lels sorprenents que presentava amb aquest trastorn. Loevinger (1976) postula que “l'ego madura i evoluciona a través de diferents estadis com a resultat de la interacció dinàmica entre el jo intern i l'entorn extern”. Endress (2004) trobà molta similitud entre els dos estadis més baixos descrits pel psicòleg i els aspectes emocionals i conductuals de la psicopatia:

L'estadi inicial del desenvolupament de l'ego és definit per un mode de funcionament interpersonal egocèntric i dependent, i per un enfocament en els impulsos corporals. En l'estadi impulsiu hom veu el món només en termes de com les coses l'afecten a ell. En l'estadi següent, l'auto-protector, l'individu és descrit com a manipulador i precabut, enfocat en l'evasió de les problemàtiques i en el control sobre els altres (p. 3).

S'ha suggerit que aquesta naturalesa egocèntrica és reflectida per un recompte més elevat de pronoms en el discurs psicopàtic, dins dels quals hi trobaríem més termes de primera persona i menys termes dirigits als altres (Henning, 2017). Així mateix, la conducta evasiva que s'especifica en l'estadi auto-protector de Loevinger es pot relacionar amb el tret que postula Rieber & Vetter (1994) de l'alta presència de negadors en el discurs psicopàtic com a reflex d'una tendència protectora a esdevenir conscient del conflicte intern i a gestionar la culpa.

Per últim, estudis recents suggereixen que “el recompte de paraules quantitatives pot ser utilitzat com una eina en la identificació i examen de processos psicològics anormals” (Hancock et al., 2013).

	Marcador lingüístic	Exemples
Alta presència	Termes referents al <i>poder</i>	<i>domination, administrate, lead, submission</i>
	Termes en temps passat	<i>was, went, did, said, have seen</i>
	Termes referents a necessitats fisiològiques	<i>drink, shoes, buy, duties, room</i>
	Termes referents als cinc sentits	<i>see, observe, hear, touch, feel</i>
	Pronoms de primera persona exclusius	<i>I, me, mine, my,</i>
	Negacions	<i>don't, not, no, ain't</i>
Baixa presència	Termes quantitius	<i>few, many, much</i>
	Termes emocionals	<i>rage, love, friend, empathy, anxious</i>
	Termes referents al present o al futur	<i>will be, are, now, walking</i>
	Pronoms de segona o tercera persona / primera inclusius	<i>your, yours, he, you, we, them</i>

Fig. 1. Marcadors lingüístics característics de la personalitat antisocial

3. METODOLOGIA

Per tal de procedir a la detecció dels esmentats marcadors, s'ha optat per la creació d'un corpus textual, del qual s'extreurà la informació que després s'analitzarà tant estadísticament com amb eines computacionals pertanyents al camp del *machine learning*. En aquest apartat es detalla tant la hipòtesi inicial amb què es va partir com el procés que ens ha portat fins l'obtenció dels resultats.

3.1. Hipòtesi

En aquest estudi s'assumeix la *hipòtesi nul·la*, és a dir, s'adopta la suposició contrària a allò que es vol provar i no es passa a valorar la *hipòtesi alternativa* fins que els resultats no demostren que el punt de partida era equivocat. En el cas que ens ocupa, la hipòtesi nul·la es tradueix en la suposició de que no existeixen diferències significatives entre el llenguatge utilitzat per persones amb trastorn antisocial i el llenguatge usat per les que no el presenten.

El motiu pel que s'ha triat acollir-se a aquest tipus d'hipòtesis pròpia del camp de l'estadística inferencial és que, en aquest estudi, es treballarà parcialment amb anàlisis estadístiques.

3.2. Creació dels corpus

Per tal de detectar els marcadors ha estat necessària la creació d'un *corpus psicopàtic* i d'un *corpus no psicopàtic* amb què comparar-lo. Atès que accedir a individus amb trastorn de la personalitat antisocial resultava complicat, es va optar per recopilar textos d'índole diversa disponibles a la xarxa (entrevistes, manifestos, transcripcions de declaracions, cartes, notes, entrades de blog) escrits o proferits per persones amb trastorn

sociopàtic. Algunes d'aquestes persones són conegudes pels seus actes criminals, d'altres porten una vida convencional i estan perfectament insertades en la societat¹.

A diferència d'estudis similars en què el *corpus no psicopàtic* s'ha construït principalment a base d'articles periodístics (e.g., Henning, 2017), en aquest cas s'ha procurat que les tipologies textuais fossin les mateixes que les del *corpus psicopàtic* (reculls de cartes, entrevistes a personalitats diverses, etc.) per tal d'evitar extreure conclusions que raurien en la diferència estilística. Val a dir que, no obstant i dins d'aquest marc, les seleccions han estat del tot atzaroses. S'han dut a terme mitjançant motors de cerca d'internet i diversos *corpora* d'accés lliure com *COCA (Corpus of Contemporary American English)*, *iWeb* o *GloWbE (Corpus of Global Web-Based English)*.

Després de passar el *corpus psicopàtic* i el *corpus no psicopàtic* a text pla, es passà a deliberar com havien de ser tractades aquestes dades. Es considerà tant l'extracció dels marcadors mitjançant expressions regulars com l'ús de tècniques de *machine learning* que permetessin que fos el sistema el que analitzés les dades i trobés patrons potencials. Atès que l'aprenentatge automàtic és un camp en auge “àmpliament aplicat en diverses àrees, inclosa la classificació i anàlisi de textos” (Ikonomakis, Kotsiantis & Tampakas, 2005), es va creure oportú considerar la seva aplicació; de manera que es procedí a la cerca d'un programari que permetés aplicar aquestes tècniques al present estudi.

Es va escollir el software WEKA (*Waikato Environment for Knowledge Analysis*), atesa la seva interfície senzilla i la seva enorme potencialitat. Es tracta d'una plataforma de software desenvolupada per la Universitat de Waikato i distribuïda sota la llicència GNU-GPL. WEKA permet extreure informació d'un conjunt de dades mitjançant algoritmes i tècniques diverses per arribar a un modelat predictiu, és a dir, és utilitzat per la mineria de dades a partir de tècniques diverses d'aprenentatge automàtic.

En funció del problema que se'ns presenta i del tipus d'estratègia amb què decidim abordar-lo (problema de classificació, problema de *clustering* o problema de regressió) escollirem un tipus determinat de tècnica d'aprenentatge o algoritme: supervisat, no supervisat o bé reforçat.

Atès que el cas que ens ocupa és un problema de classificació —el que cerquem és identificar a quin conjunt de subpoblacions pertanyen les observacions recopilades—,

¹ És el cas de M. E. Thomas, escriptora del llibre *Confessions of a Sociopath* i autora del blog *Sociopath World*; o del neurocientífic James H. Fallon, el qual ha dut a terme xerrades sobre el trastorn antisocial de la personalitat i ha estat entrevistat en diverses ocasions pels mitjans tradicionals.

s'ha escollit l'aprenentatge supervisat. En aquest tipus d'aprenentatge, els algoritmes s'entrenen amb dades ja classificades per tal de trobar una funció que, donat un *input* de dades noves no catalogades, les classifiqui correctament. En altres paraules, l'algoritme s'entrena amb un 'històric' de dades i així 'aprèn' a assignar l'etiqueta de sortida adequada a un nou valor, és a dir, prediu el valor de sortida.

Abans d'aprofundir més en les tècniques utilitzades per tractar les nostres dades, és necessari explicar com es va *preprocessar* la informació i quins van ser els passos previs a la construcció del classificador.

3.3. Importació de les dades sense processar a WEKA: un primer apropament

3.3.1. Preprocessament de les dades

En aquesta primera presa de contacte amb el software, les dades van ser importades sense abans haver estat processades per cap altre programari; és a dir, senzillament s'hi va introduir tot el corpus textual que s'havia recopilat. Per tal de fer-ho, es va crear un document en format *.arff* (*attribute-relation file format*), un tipus de fitxer de text ASCII especialment dissenyat per a l'ús de WEKA que descriu una llista d'instàncies que comparteixen un conjunt d'atributs.

El fitxer *.arff* conté, d'una banda, la capçalera o *header* i, d'altra banda, la informació o *data*, on es recullen les instàncies. A la capçalera es declara el nom de la relació i s'especifiquen i defineixen la naturalesa dels atributs, que són els trets potencials d'un fenomen en concret (en el cas que ens ocupa, aquests serien els diferents marcadors lingüístics) en els quals l'algoritme es basa per classificar instàncies (en el nostre cas, les frases). Com es pot observar en la Figura 1, s'ha declarat que el nostre dataset presenta *string attributes*, és a dir, atributs que contenen valors textuais i no numèrics ni nominals; per bé que encara no estiguin especificats (més endavant veurem per què). El que sí que s'especifica és l'atribut de classe, que determina les categories en què l'input es dividirà ({psico,nopsico}).

D'altra banda, en la part on es declara la informació (@data), cada instància amb què l'algoritme s'entrenarà és representada per una cadena de text envoltada per cometes simples i seguida per la classe a què pertany. El que s'ha fet per tal d'aconseguir-ho és separar els diferents textos que formen els corpus en frases de longituds diverses, i a cadascuna d'elles se li ha assignat una classe o una altra en funció de a quina categoria

pertany l'individu que l'ha proferit (*psicopàtic* o *no psicopàtic*), tot això amb ajuda de les eines que ofereix UNIX. Es pot observar part del fitxer *arff* a la Figura 2.

Un cop completat el fitxer, es va procedir a obrir-lo des del software i a examinar les dades que s'hi havien introduït. Una de les moltes aplicacions que incorpora WEKA és l'ús dels anomenats *filtres*, que són eines dissenyades per ajudar amb la preparació del *dataset* amb les quals aquest es pot alterar per a fins diversos, abans de fer-lo servir amb algun algoritme.

Tal com s'ha explicat anteriorment, en un principi els atributs amb què l'algoritme ha de treballar no s'havien declarat: abans s'havia de trobar la manera de fer que cada paraula de cada de text introduïda comptés com a atribut potencial per a la nostra classificació. Això es va aconseguir amb l'aplicació del filtre *StringToWordVector*.

```
@relation 'corpus'

@attribute text string
@attribute class {psico,nopsico}

@data
'Dear Cedric, A strange thing happened to me today',NoPsico
'I saw a big thundercloud move down over Half Dome, and it was so big and clear and brilliant that it
made me see many things that were drifting around inside of me; things that related to those who are
loved and those who are real friends',NoPsico
'For the first time I know what love is; what friends are; and what art should be',NoPsico
```

Fig. 2. Fitxer *.arff*

Per tal d'extreure informació d'un text, aquest s'ha de convertir en un conjunt de números, és a dir, en un vector. Així, el procés de convertir paraules en números s'anomena vectorització; i el conjunt de tècniques d'aprenentatge automàtic que s'ocupen d'aquesta conversió s'integren dins el terme de *word embedding* o *word vectorization*.

Simplificadament, un vector és una seqüència finita o *n-tupla*, composta per *n* nombres reals ordenats. Cada número representa una paraula del text, per bé que la no-presència de paraules també pot estar representada. Depenent de la tècnica de vectorització (*Bag of Words*, *TF-IDF Vectorizer*, *Count Vectorizer*, etc.) els valors numèrics expressaran des de l'existència fins la freqüència d'aparició de cada *token*.

A WEKA, aquest procés s'implementa de manera automàtica, però el software ens permet ajustar múltiples paràmetres del filtre, com per exemple el *LowerCaseConversion*, que

permet no fer distincions entre majúscules i minúscules. Altres possibilitats que ens ofereix el filtre *StringToWordVector* són transformar les cadenes de text en *n-grames* o eliminar les paraules buides², encara que en aquest treball no han estat utilitzades.

Un cop efectuada la vectorització, podem veure en la interfície del software que cada paraula del corpus textual, entesa com a *type*, s'ha convertit en un atribut (v. Figura 3). De la llista paraules generada, no obstant, s'ha d'eliminar les paraules que no ens poden servir com a atributs atesa la seva arbitrarietat i dependència absoluta del context: és el cas de noms de marques, noms propis, noms d'accidents o llocs geogràfics o xifres. Com que elaborar una llista de paraules buides amb totes els mots a evitar resultava extremadament laboriós, es va procedir a l'eliminació manual dels atributs directament de la llista generada. Al final d'aquest procés, el conjunt d'atributs va quedar en un total de 1340.

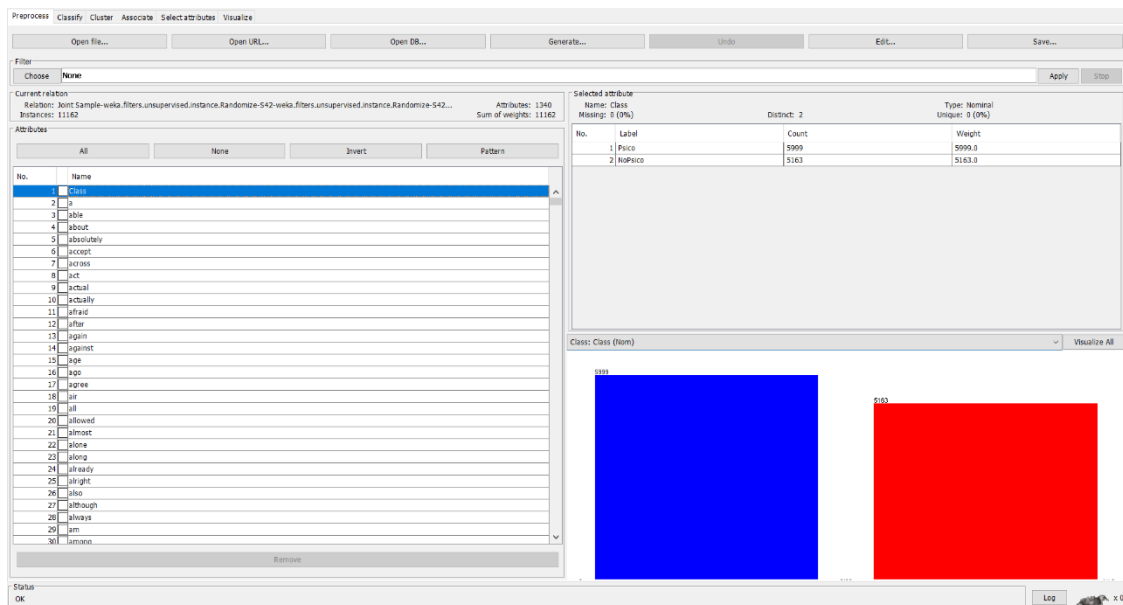


Fig. 3. A l'esquerra, llista d'atributs que resulta de la implementació del filtre *StringToWordVector*

3.3.1.1. Refinament d'atributs

En tasques de *data mining* es generen llistes interminables d'atributs, tanmateix, el més probable és que no els necessitem tots per dur a terme la classificació. És més: la presència d'atributs redundants o irrelevantes poden esbiaixar els resultats i empitjorar l'efectivitat

² Les paraules buides (en anglès *stopwords*) són llistes de paraules que es filtren durant el processament de llenguatge natural (Rajaraman & Ullman, 2012). Les paraules que s'incorporen a la llista són escollides en funció del tipus de tasca que es vol dur a terme.

de la tasca; motiu pel qual és recomenable treballar únicament amb els trets més rellevants de la nostra llista.

Hi ha diversos tipus de mètodes de refinament, els quals es basen primordialment en teories estadístiques. En aquest estudi es consideren els dos tipus de tècniques de selecció d'atributs més populars (Gnanambal, Thangraj, Meenatchi & Gayathri, 2018): el mètode *Wrapper*, que utilitza algorismes per a trobar possibles subconjunts d'atributs i evalua cada subconjunt en funció de l'èxit del classificador; i el mètode *Filter*, que agafa propietats intrínseques dels atributs (e.g., la “rellevància” dels trets) i els selecciona en funció de la puntuació que treuen en diversos tests estadístics.

Ambdós tipus de mètodes tenen diferents avantatges i inconvenients. Atès que què la tècnica *Filter* és molt més ràpida i presenta menys tendència al sobreajustament, es procedirà a utilitzar-la per a la selecció de trets. Concretament, s'escollirà un evaluador d'atributs basat en la proporció de guany d'informació³ i un *ranker* que classificarà els atributs en base a les seves avaluacions individuals. La manera a procedir serà la següent: el *dataset* es sotmetrà a un algorisme de classificació, després es refinarà la llista d'atributs i es tornarà a fer córrer l'algorisme per observar si la classificació ha millorat. Aquest procediment es repetirà indeterminadament, de manera que la quantitat d'atributs s'anirà fent cada cop més reduïda, fins que s'hagin obtingut els resultats òptims.

3.3.2. Elecció del classificador

Tal com postula el teorema *No Free Lunch* (Wolpert, 1997), resulta complicat discernir quin algorisme d'entre tots els existents pot funcionar en cada situació donada. És per aquest motiu que, en aquestes tasques, s'ha de triar diferents classificadors i després seleccionar el que aconsegueixi els resultats més exitosos.

En el cas que ens ocupa, es va escollir entrenar les dades amb tres dels algorismes més utilitzats en el camp de l'aprenentatge automàtic, en especial pel que fa als problemes de classificació. D'una banda, s'han fet servir dos arbres de decisió: el *J48* i el *RandomTree*. Un arbre de decisió és un tipus d'algorisme d'aprenentatge supervisat que divideix la població en subpoblacions basant-se en els atributs més diferenciadors que troba en l'input. Els avantatges d'aquest tipus de tècniques és que l'*output* que generen és possible

³ La proporció de guany d'informació pretén determinar quin atribut és més útil per discriminar entre les classes, i ho fa tenint en compte l'entropia. L'entropia és utilitzada -simplificant en gran mesura- per a mesurar el nivell d'“impuresa” d'un grup de dades: una mostra alta en impuresa serà aquella que presenti exemples variats, pertanyents a diverses classes. Així, podem dir que, com més alta l'entropia, més alt serà el contingut informatiu d'unes dades.

d'entendre i representen una manera molt eficient d'identificar variables significants; però un dels majors inconvenients és que presenten molta tendència al sobreajustament. D'altra banda, s'ha fet servir un altre tipus d'algoritme d'aprenentatge supervisat; l'*SMO*, el qual presenta resultats notables en la majoria de casos i generalment requereix poques dades per entrenar el model, tot i que la seva lògica operant no és transparent com la d'un arbre de decisió.

Si bé es va procedir a entrenar els algoritmes amb les dades textuais, existien moltes imprecisions en el model que perjudicaven la seva fiabilitat. Per exemple; si bé els mateixos *types* eren agrupats en un sol atribut, no era possible procedir amb la tasca de lematització; i casos com el de les contraccions eren intractables, ja que el sistema les entenia com si fossin una sola paraula (i, en conseqüència, segments com *you'll* es comprenien com un sol atribut, i no pas com a paraules i atributs separats). Per tal de solucionar aquests impediments, es passà a considerar l'opció d'introduir les dades de nou, però aquest cop fent ús d'eines complementàries que potencialment enriquirien d'allò més el nostre estudi.

3.4. Importació de les dades processades a WEKA

Si bé en l'anterior apropament havíem introduït les dades textuais a WEKA sense cap tractament previ, en aquest cas s'efectuarà una transformació de les dades textuais en dades numèriques abans de procedir a la importació.

3.4.1. Tractament dels corpus amb LIWC i SPSS

LIWC (Linguistic Inquiry and Word Count; Pennebaker, Francis, Booth, 2001) és una eina software d'anàlisi textual que calcula el percentatge d'aparició de determinades paraules dins d'un text, d'acord amb diverses categories. El diccionari del programa és el pal de paller de l'estratègia d'anàlisi: està compost per 6.400 paraules i lemes, i cada entrada refereix, de manera addicional, a una o més categories de paraules. Per exemple, la paraula *cried* és part de diverses de les categories que presenta LIWC: *negative emotion, verb and past focus*, de manera que, si tal paraula fos present en el text, el percentatge de cadascuna d'aquestes categories augmentaria. Així mateix, tal com s'ha comentat, el diccionari també inclou lemes, fet que ens permet detectar *paraules target* que tinguessin, posem per cas, les mateixes primeres cinc lletres (és a dir, l'entrada del lema tindria la forma *hungr**, i es recollirien dins d'aquesta paraules com ara *hungry, hungrier* o *hungriest* (Pennebaker, Boud, Jordan & Blackburn, 2015). LIWC no només

presenta categories estructurals que fan referència a dimensions lingüístiques (com ara ‘preposicions’, ‘articles’, ‘negacions’ o ‘pronoms personals’), sinó que en presenta moltes referents a processos psicològics (‘emoció positiva’, ‘dubte’, ‘processos perceptuals’ o ‘orientació en el passat’ en són alguns exemples) i preocupacions personals (com ara ‘feina’, ‘religió’, ‘mort’ o ‘casa’). El conjunt d’aquestes característiques fan del software una opció molt útil per l’estudi que ens ocupa, especialment pel component psicològic que aquest ofereix.

Els corpus foren introduïts a LIWC, que retornà una matriu emplenada amb el percentatge de presència de termes referents a les diverses categories en cadascun dels fitxers. Aquesta fou guardada en format *.xls* (vg. Fig. 4) i posteriorment introduïda al programa d’anàlisi estadístic SPSS, amb el qual s’efectuà una prova no paramètrica específica per a mostres independents i de distribució no normal (*Test de U de Mann-Whitney*), per tal de comprovar si existien variables en què els dos grups presentessin unes mitjanes prou disperses com per assumir significància estadística.

	A	B	C	D	E	F	G	H	I	J
1	Filename	Segment	WC	Analytic	Clout	Authentic	Tone	Dic	function	pronoun
2	np_blog1.t	1	7273	55,37	63,77	61,87	43,01	89,92	55,09	16,90
3	np_blog2.t	1	8249	71,79	70,28	68,73	60,47	86,90	51,78	14,34
4	np_blog3.t	1	9122	65,07	77,83	43,12	62,16	87,67	53,72	13,81
5	np_blog4.t	1	4846	77,98	85,84	50,32	86,52	86,19	50,45	12,03
6	np_CHAE1	1	6057	79,54	80,96	12,73	54,68	83,28	51,53	13,55
7	np_CHAE2	1	7607	66,43	75,46	24,67	49,77	85,37	54,37	15,34
8	np_CHAE3	1	11324	74,26	69,80	34,21	50,32	85,13	52,57	13,79
9	np_luther.	1	10682	83,39	73,51	31,82	29,53	84,11	52,89	11,99
10	np_mande	1	3769	39,06	70,14	54,15	72,91	90,02	60,94	18,89
11	np_muhan	1	2414	28,86	68,00	34,44	48,86	85,25	51,57	17,69
12	np_story.t	1	10809	84,35	78,78	19,32	46,02	81,01	50,39	12,46
13	np_teresa.	1	2161	37,14	86,77	17,57	54,52	96,16	61,78	20,31
14	p_blog.txt	1	21782	59,85	54,67	49,35	40,92	89,00	56,14	15,41

Fig. 4. Output parcial de l’anàlisi de LIWC

Resumen de contrastes de hipótesis

	Hipótesis nula	Prueba	Sig.	Decisión
1	La distribución de Clout es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,001 ^a	Rechace la hipótesis nula.
2	La distribución de Authentic es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,025 ^a	Rechace la hipótesis nula.
3	La distribución de Tone es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,020 ^a	Rechace la hipótesis nula.
4	La distribución de I es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,020 ^a	Rechace la hipótesis nula.
5	La distribución de we es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,016 ^a	Rechace la hipótesis nula.
6	La distribución de male es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,025 ^a	Rechace la hipótesis nula.
7	La distribución de see es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,012 ^a	Rechace la hipótesis nula.
8	La distribución de affiliation es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,025 ^a	Rechace la hipótesis nula.
9	La distribución de money es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,016 ^a	Rechace la hipótesis nula.

Fig. 5. Variables per a les quals s’assumeix significància estadística (prova *u* de Man Whitney)

Les variables que passaren el test foren les categories *Clout, Authentic, Tone, I, we, male, see, affiliation* i *money* (v. Fig. 5). En estadística, el *valor p* és el nivell de significància més petit que condueix al rebuig de la hipòtesi nul·la. És a dir: si una prova de contrast d'hipòtesi com la que s'ha dut a terme proporciona un *valor p* inferior a 0.05, s'assumeix que la hipòtesi nul·la no es vàlida i es passa a assumir la hipòtesi alternativa: que hi ha significància. Com més baix és el nivell de la significació, més forta serà l'edivència de que un fet no es deu a l'atzar. Tanmateix, tant la naturalesa d'aquests atributs com la valoració d'aquests com a significatius es passarà a valorar un cop finalitzat l'estudi en la seva totalitat.

3.4.2. Preprocessament de les dades

Tal com s'ha apuntat prèviament, si bé primer havíem importat les dades textuais a WEKA sense sotmetre-les a cap tipus de tractament, en aquest cas primer hem transformat les dades textuais en dades numèriques (percentatges de cada categoria en un text) i després les hem introduït al programari. Acte seguit, s'ha procedit a exactament el mateix procés sistemàtic de refinament d'atributs i d'entrenament amb els diversos algoritmes esmentats al punt 3.3.2.

Abans de procedir a la classificació s'ha escollit de quina manera es distribuïrien les dades per a entrenar l'algoritme i també per testar el seu aprenentatge. Un dels mètodes més utilitzats és el de dividir la totalitat del corpus en *corpus d'aprenentatge* (input textual que s'anota i a partir del qual el classificador extreu la informació) i en *corpus de verificació* (input textual que no està classificat, sobre el qual l'algoritme aplica tot el que ha après). Tanmateix, atès que en aquest estudi es treballarà amb arbres de decisió i aquests són molt propensos al sobreajustament⁴, s'ha optat per un altre tipus de distribució molt utilitzada per evitar aquesta problemàtica: la validació creuada de 10 iteracions.

En aquest mètode, les dades es divideixen en 10 subconjunts; el primer d'aquests és utilitzat com a prova i la resta com a test. Acte seguit, el segon d'aquests és usat com a prova i la resta com a test, i així successivament fins arribar a les 10 iteracions. D'aquesta manera, el corpus de verificació roman ocult en tot moment i s'evita la memorització.

⁴ El sobreajustament es dona quan un algoritme no aprèn a la tendència que busquem que aprengui, sinó que aprèn a *memoritzar* els passos que ha de seguir per aconseguir ser exitós donades determinades dades. El sobreajustament pot ser descobert si l'algoritme testeja un nou *dataset* (que no és a partir del qual ha memoritzat a procedir) i el percentatge d'èxit decau notablement.

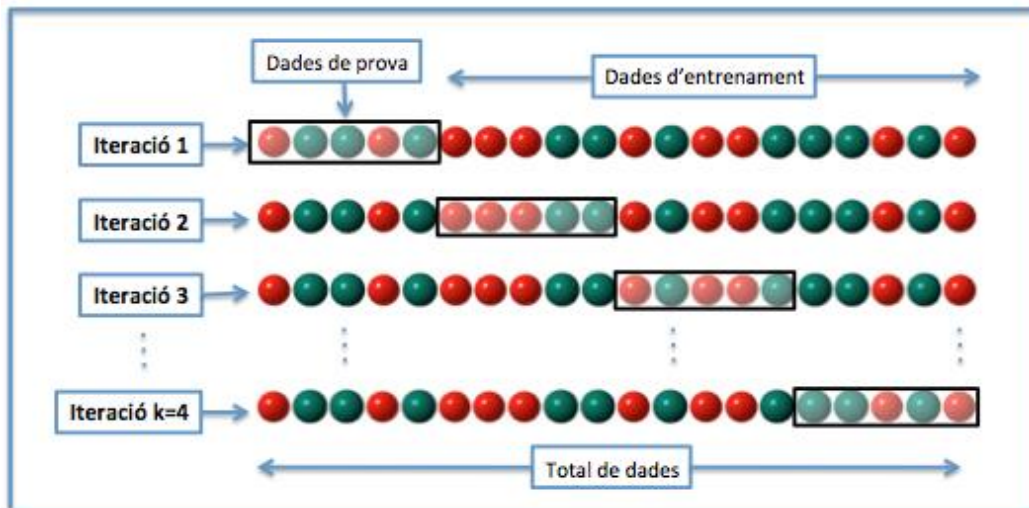


Fig. 5. Funcionament de la validació creuada de 10 iteracions

3.4.3. Classificació

3.4.3.1. Classificació amb J48

La classificació efectuada per l'algorithm J48 presenta un 65% d'instàncies correctament classificades en gairebé totes les seves fases de refinament d'atributs (v. Fig. 6). La lògica interna de l'arbre és, així mateix, invariable (v. Fig. 7): podem observar que el primer node fa referència a la variable *Clout*, la qual, si presenta una puntuació igual o menor a 54.74, és momentàniament assignada a la classe *Psico*; mentre que si és major d'aquesta xifra i presenta un nombre de pronoms menor a 20.31, és definitivament assignat a la classe *NoPsico*. El refinament òptim es troba als 3 atributs, els quals són *Clout* i *pron*. Amb 2 atributs ofereix una idèntica classificació, per bé que utilitza els atributs *Clout* i *Tone* (v. Fig. 8).

<i>J48</i>	<i>Percentatge d'encert</i>
<i>60 atributs</i>	60 %
<i>50 atributs</i>	65%
<i>40 atributs</i>	65%
<i>30 atributs</i>	65%
<i>20 atributs</i>	65%
<i>10 atributs</i>	65%
<i>7 atributs</i>	65%
<i>3 atributs</i>	70%
<i>2 atributs</i>	70%

Figura 6. Percentatges d'encert de l'algorithm *J4*

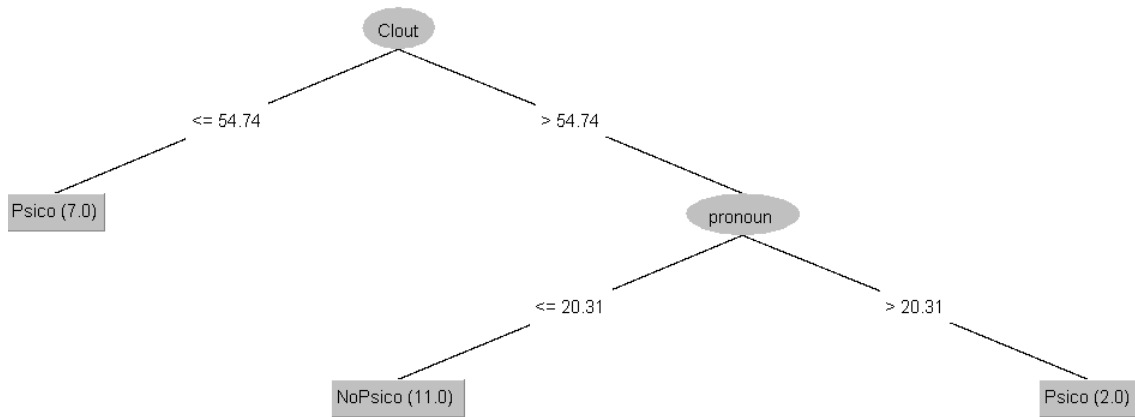


Figura 7. Arbre de decisió J48 amb 3 atributs (70%)

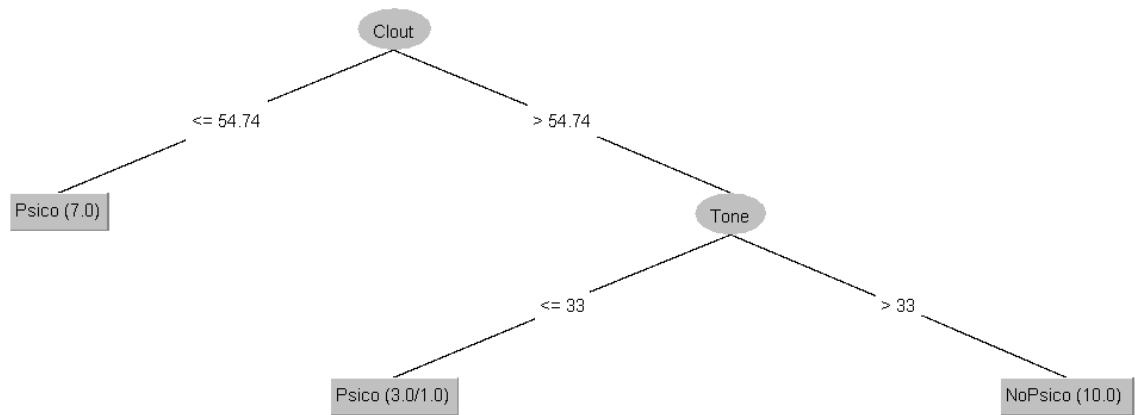


Figura 8. Arbre de decisió J48 amb 2 atributs (70%)

3.4.3.2. Classificació amb RandomTree

RandomTree no es manté uniforme com *J48* al llarg dels diversos refinaments; sinó que va millorant exponencialment la seva execució fins que, als 8 atributs, el percentatge d'encerts comença a baixar. Amb un total de 30 atributs, el classificador presentà un 60% d'encerts (v. Fig. 9), i amb un total de 20, un 70%. Veiem que, en aquests dos casos, l'algoritme selecciona com a categories altament predictores *death*, *Tone*, *adj* o *adverb*; també els pronoms *we* i *you* i, altre cop, la categoria *Clout*. L'arbre que presenta més percentatge d'encerts de tots és el que l'algoritme crea quan se l'entrena amb un total de 8 atributs – *Clout*, *Tone*, *pronoun*, *see*, *death*, *adj*, *verb*, *quant* (v. Fig. 11), el qual presenta exactament el mateix tipus d'*output* que s'aconsegueix amb *J48* en el seu millor refinament (v. Fig. 12).

<i>RandomTree</i>	<i>Percentatge d'encert</i>
60 atributs	50 %
50 atributs	50%
40 atributs	55%
30 atributs	60%
20 atributs	70%
8 atributs	75%

Fig. 9. Percentatges d'encert de l'algoritme *RandomTree*

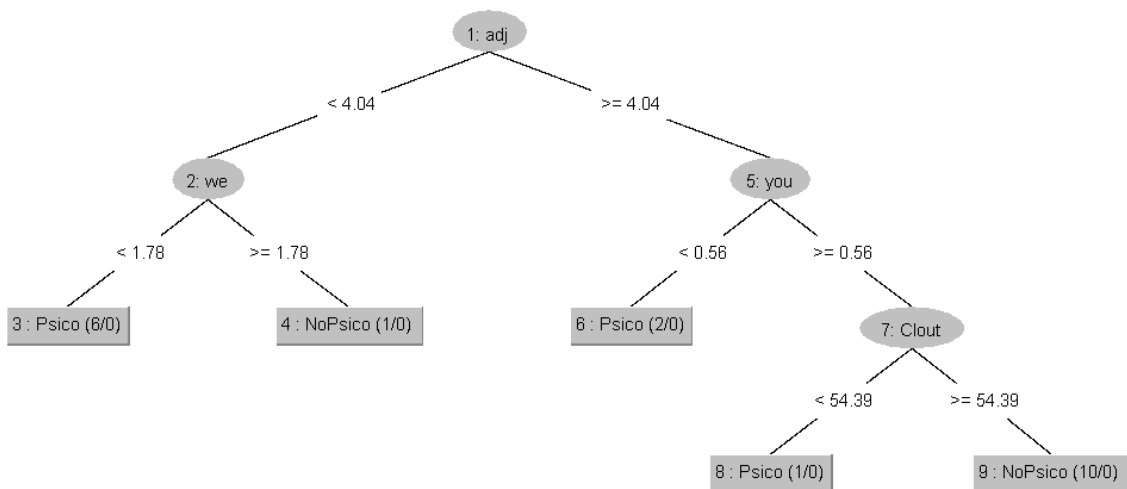


Fig. 10. Arbre de decisió *RandomTree* (60% d'encert)

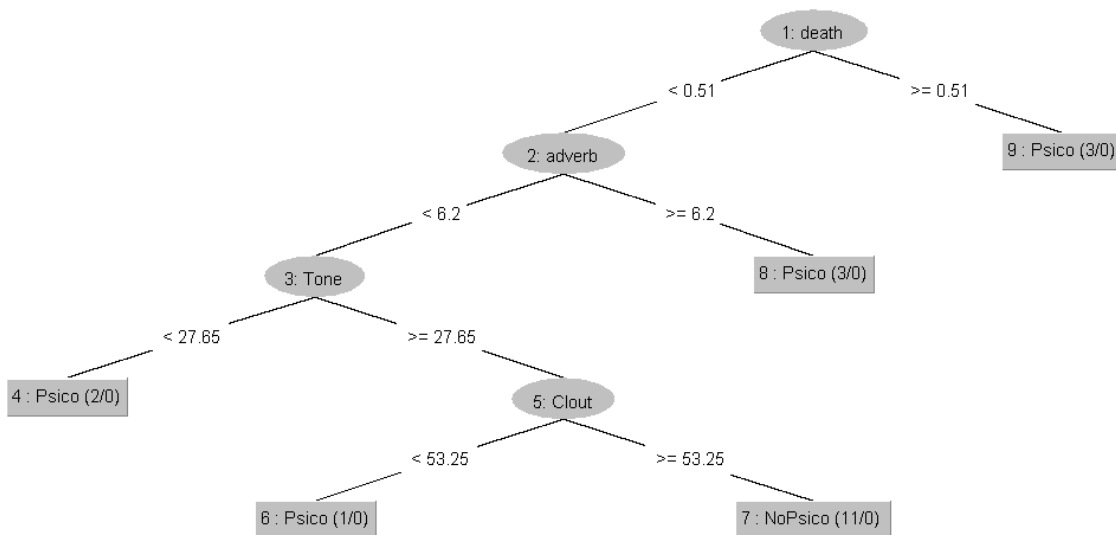


Fig. 11. Arbre de decisió *RandomTree* (70% d'encert)

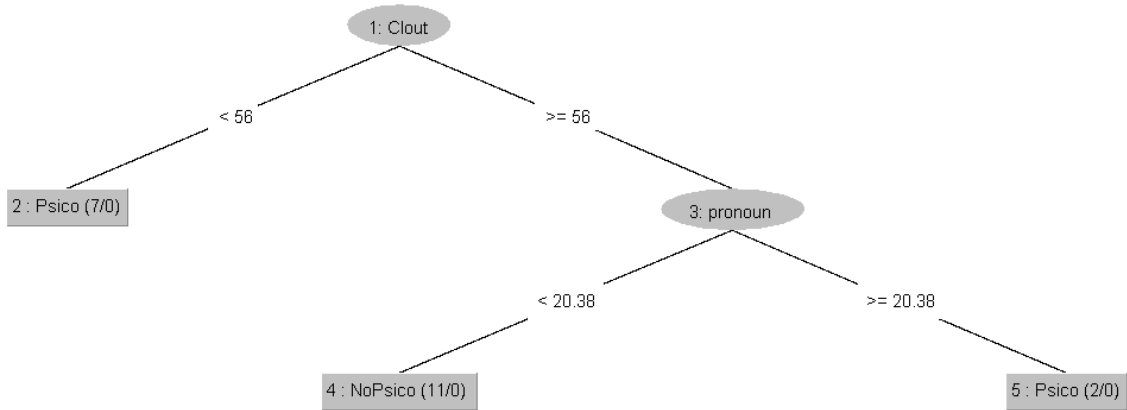


Fig. 12. Arbres de decisió *RandomTree* (75% d'encert)

3.4.3.3. Classificació amb SMO

Si bé aquest algoritme presenta classificacions realment exitoses, la seva lògica interna no és visible com la dels arbres de decisió. Tanmateix, el set d'atributs amb què aconseguix el seu èxit òptim de 90% de classificació sí que és ostensible (v. Figura 14).

<i>SMO</i>	<i>Percentatge d'encert</i>
60 atributs	85%
50 atributs	85%
40 atributs	80%
30 atributs	85%
20 atributs	90%
18 atributs	90%

Fig. 13. Percentatges d'encert de l'algoritme *SMO*

```

-1.8011 * (normalized) Clout
+ -0.6366 * (normalized) Tone
+ 0.2602 * (normalized) pronoun
+ -0.9296 * (normalized) see
+ 0.3851 * (normalized) death
+ -0.8284 * (normalized) adj
+ 0.2052 * (normalized) verb
+ 0.8307 * (normalized) quant
+ 0.2239 * (normalized) negate
+ -0.4415 * (normalized) compare
+ -0.1594 * (normalized) posemo
+ 0.0676 * (normalized) affect
+ 0.8829 * (normalized) anger
+ -0.3725 * (normalized) sad
+ -0.4552 * (normalized) anx
+ 0.0754 * (normalized) adverb
+ 0.3486 * (normalized) negemo
+ -0.0059 * (normalized) conj
  
```

Fig. 14. Atributs de l'algoritme *SMO*

4. RESULTATS I DISCUSSIÓ

Tal com indica la *prova U* de *Mann-Whitney*, el llenguatge emprat per individus amb trastorn de la personalitat psicopàtica difereix significativament del llenguatge dels individus que no presenten el trastorn pel que fa a les variables *Clout*, *Authentic*, *Tone*, *I*, *we*, *male*, *see*, *affiliation* i *money* del software LIWC2015. Per tal d'aconseguir una comprensió més profunda de l'*outcome*, es procedeix a descriure tant els resultats del test estadístic com cadascuna de les categories d'una manera més precisa. La informació sobre aquestes dimensions lingüístiques s'ha extret de de Pennebaker, Boyd, Jordan & Blackburn (2015); i els gràfics resultants de les proves estadístiques a què s'aludeixen es troben a l'Annex (7).

- i. *I*: aquesta categoria de LIWC és compresa per un total de 24 termes i recull els pronoms de primera persona del singular (*I*, *me*, *mine* en són exemples). El nivell de presència d'aquesta categoria en el discurs d'individus amb trastorn sociopàtic de la personalitat (*mean rank*=13,89) excedeix significativament del de les persones sense tal trastorn (*mean rank*=7,73).
- ii. *we*: aquesta categoria de LIWC és compresa per un total de 12 termes i recull els pronoms de primera persona del plural (*we*, *us*, *our* en són exemples). El nivell de presència d'aquesta categoria en el discurs d'individus amb trastorn sociopàtic de la personalitat (*mean rank*=7,00) és significativament menor del de les persones sense tal trastorn (*mean rank*=13,36).
- iii. *male*: aquesta categoria de LIWC és compresa per un total de 116 termes i recull mots que designen referents masculins (*boy*, *his*, *dad* en són exemples). El nivell de presència d'aquesta categoria en el discurs d'individus amb trastorn sociopàtic de la personalitat (*mean rank*=7,22) és significativament menor del de les persones sense tal trastorn (*mean rank*=13,18).
- iv. *see*: aquesta categoria de LIWC és compresa per un total de 126 termes i recull mots referents al sentit de la vista (*view*, *saw*, *seen* en són exemples). El nivell de presència d'aquesta categoria en el discurs d'individus amb trastorn sociopàtic de la personalitat

(*mean rank*=6,89) és significativament menor del de les persones sense tal trastorn (*mean rank*=13,45).

- v. *affiliation*: aquesta categoria de LIWC és compresa per un total de 248 termes i recull paraules referents l'afiliació, al vincle; a la relació personal (*ally, friend, social* en són exemples). El nivell de presència d'aquesta categoria en el discurs d'individus amb trastorn sociopàtic de la personalitat (*mean rank*=13,18) excedeix significativament el de les persones sense tal trastorn (*mean rank*=7,22).
- vi. *money*: aquesta categoria de LIWC és compresa per un total de 248 termes i recull paraules referents als diners (*audit, cash, owe* en són exemples). El nivell de presència d'aquesta categoria en el discurs d'individus amb trastorn sociopàtic de la personalitat (*mean rank*=7,00) és significativament més baix que el de les persones sense tal trastorn (*mean rank*=13,36).

Les variables *Authentic*, *Tone* i *Clout* són *variables resum*, una de les particularitats que incorpora la versió més novedosa de LIWC (2015). Cada variable resum ha estat originada a partir de descobriments publicats varis. El seu càlcul es fonamenta en percentils basats en diferents categories estandaritzades i no pas en recomptes de freqüències com a tal (Pennebaker, Boyd, Jordan & Blackburn, 2015). Si bé el software no és completament transparent respecte d'aquestes categories, és possible consultar els estudis de les quals parteixen per tal d'entendre com es caracteritzen. La puntuació que s'obté quant aquestes variables pot variar des de 0 a 100, essent 50 el punt mig.

- vii. *Authenticity*: aquesta categoria vol detectar honestat i personalitat en un text escrit (Cohn, Mehl & Pennebaker, 2004). La puntuació esdevé més alta en *Authenticity* si el discurs presenta percentatges alts en les categories *i, she/he, they* i *differ* i percentatges més baixos en *negative emotion* i *motion*. El nivell de presència d'aquesta categoria en el discurs d'individus amb trastorn sociopàtic de la personalitat (*mean rank*=13,78) és significativament més alt que el de les persones sense tal trastorn (*mean rank*=7,82).

- viii. *Tone*: aquesta categoria vol detectar el to d'un text escrit (Pennebaker et al., 2015). La puntuació esdevé més alta en *Tone* si el discurs presenta percentatges alts en la categoria *positive emotion* i percentatges baixos en la categoria *negative emotion*. El nivell de presència d'aquesta categoria en el discurs d'individus amb trastorn sociopàtic de la personalitat (*mean rank*=9) és significativament més baix que el de les persones sense tal trastorn (*mean rank*=11).
- ix. *Clout*: aquesta categoria vol detectar l'estatus social, la capacitat de lideratge i la seguretat en un text escrit (Cohn, Mehl & Pennebaker, 2004). La puntuació esdevé més alta en *Clout* si el discurs presenta percentatges alts en les categories *we*, *you* i *social* i percentatges més baixos en les categories *I*, *swear*, *negate* i *differe*. El nivell de presència d'aquesta categoria en el discurs d'individus amb trastorn sociopàtic de la personalitat (*mean rank*=5,89) és significativament més baix que el de les persones sense tal trastorn (*mean rank*=14,27).

Pel que fa als arbres de decisió, la majoria —i els més exitosos— es basen en algunes d'aquestes categories significatives (*Clout*, *Tone*, *we*) per a fer la classificació, per bé que també n'apareixen d'altres que en l'anàlisi estadístic no havien resultat significatives, com ara *pronoun*, *you*, *adj*, *death* o *adverb*. No obstant, els percentatges d'encert dels arbres de què formen part no són pas gaire alts. Mentre que de les categories *adj* i *adv* no se n'havia dit res en estudis previs, les categories *pronoun* i *you* sí que havien estat adreçades i són força reveladores.

Pel que fa al rànking d'atributs de l'algoritme SMO, dels que són estadísticament significatius només s'hi troben presents *Clout*, *Tone* i *see*. D'altra banda, podem veure atributs que no han resultat rellevants en el test estadístic però que en la literatura han estat destacats com a característics de la personalitat antisocial, com ara *pronoun*, *negate* o *quantitative*. El rànking d'SMO també emfasitza en variables referents a les emocions (*anger*, *sad*, *anxious*, *affect*, *positive emotion*, *negative emotion*).

Si analitzem els resultats obtinguts en relació als marcadors lingüístics que la literatura identifica com a caracteritzadors personalitat antisocial, és possible observar que la gran

majoria de marcadors de caràcter lèxic no han rebut suport empíric en aquest estudi, mentre que, pel que fa als relacionals, els resultats han estat més semblants.

- Al contrari del que s'especifica en la literatura, s'ha trobat que el discurs *no psicopàtic* presenta més termes referents a la categoria *power* (s'hi recullen els termes referents al poder) que no pas el discurs *psicopàtic*.
- Al contrari del que s'especifica en la literatura, s'ha trobat que el discurs *no psicopàtic* presenta més termes referents a la categoria *ingest* (s'hi recullen els termes referents al menjar i a la beguda) que no pas el discurs *psicopàtic*. Pel que fa a les altres categories referents a necessitats fisiològiques, per a les quals, recordem, es pronosticaven puntuacions altes; sí que s'ha detectat més presència de termes sexuals (categoria *sexual*) i referits al cos (categoria *body*), si bé no de manera significativa. La que sí ha resultat significativa és la categoria *money*, per bé que de la manera inversa a què s'especificava en estudis previs: inesperadament, el discurs *no psicopàtic* presenta més presència de termes referents als diners que el discurs *no psicopàtic*.
- Al contrari del que s'especifica en la literatura, s'ha trobat que la presència de termes referents a la categoria *see* (s'hi recullen els termes referents al sentit de la vista) en el discurs *no psicopàtic* és significativament més alta que no pas en el discurs *psicopàtic*. Pel que fa a les altres categories referents als sentits, per a les quals també es pronosticaven puntuacions altes (categories *hear* i *feel*) tampoc s'han complert les expectatives, per bé que les seves puntuacions no han resultat, en cap cas, significatives.
- No s'ha detectat diferències respecte a l'ús més extens de temps passat (categoria *focuspast*).
- Sí que s'ha detectat diferències respecte al major ús de termes de negació (categoria *negate*) en el discurs *psicopàtic* respecte al discurs *no psicopàtic*, per bé que no són significatives.
- Si bé la idea de que el discurs *psicopàtic* presenta menys termes emocionals no ha estat verificada amb una categoria en concret, les variables *Authenticity* i *Tone* (en les quals el llenguatge antisocial presenta puntuacions significatives) poden suggerir que la hipòtesi no va mal encaminada. La categoria *Authenticity* suggereix que el discurs *psicopàtic* incorpora un nombre reduït de *negative emotions*, i la puntuació en la categoria *Tone* mostraria que conté més *negative*

emotions que no pas *positive emotions*. En definitiva, tot sembla indicar que el discurs *psicopàtic* presenta, en general, pocs termes emocionals (i, dins d'aquests, encara més pocs referents a sentiments positius).

- S'ha complert —i de manera significativa— el pronòstic de que el discurs *no psicopàtic* incorpora una quantitat major de pronoms de primera persona singular (categoria *I*) que no pas el discurs *no psicopàtic*.
- Si bé no s'havia hipotitzat com a tal, l'anàlisi ha demostrat que el discurs *psicopàtic* presenta, i de manera significativa, menys termes referents al vincle humà (categoria *affiliation*) que no pas el discurs *no psicopàtic*.
- Si bé la hipòtesi de que el discurs *psicopàtic* presenta menys termes referents a la segona i tercera persona singular i plural no ha estat verificada, ja que les diferències varien segons el pronom i són mínimes; sí que s'ha observat que el discurs *psicopàtic* presenta una quantitat significativament menor de pronoms de primera persona del plural (categoria *we*) que no pas el discurs *no psicopàtic*.

Aquest contrast (alta presència de pronom exclusiu *I* i baixa presència de pronom inclusiu *we*) no només es reflecteix en la significació estadística d'ambdues categories, sinó que també es reflecteix en la baixa puntuació aconseguida en la variable resum *Clout*; la qual, recordem, es caracteritza per una alta presència del pronom *we* i una baixa presència del pronom *I*, entre d'altres. *Clout* és, de tots, l'atribut que ha rep un *valor p* més baix (és a dir, més significatiu) i que té més presència —i aconsegueix millors prediccions— en els algorismes de decisió. És, de lluny, el tret que diferencia més els dos discursos, i sembla basar-se en aquest contrast *inclusió-exclusió*.

De fet, Henning (2017) ja notà aquest contrast:

“[...] hauríem d'observar una barra prou densa en el diagrama de dispersió lèxica corresponent per als termes *I* i *me*, fet que és demostrat més endavant —així mateix, hauríem de poder notar una deficiència en mots *ego-exclusius* davant dels inclusius, fet que també és el cas. D'altra banda, l'anàlisi de dispersió lèxica pels subjectes no sociòpates mostra el contrari: una deficiència en termes *ego-inclusius*, i un pic en els termes *ego-exclusius*” (p.74).

Una manera d'aprofundir encara més en aquesta idea, així com en la naturalesa de la categoria *Clout*, és examinar l'estudi que donà origen a aquesta categoria (Kacewicz, Pennebaker, Davis, Jeon & Graesser; 2014). L'objectiu d'aquest estudi era “identificar

quines dimensions lingüístiques són les que aporten pistes més valuoses per a la identificació del lloc que ocupen les persones en una jerarquia social”, entenent aquesta jerarquia com una “ordenació d’individus o grups en una dimensió social de valors, en què la posició dins d’un grup ha estat definida per molts constructes que se sobreposen; inclòs l’estatus, el poder, la dominància i el prestigi” (Kacewicz et al., 2014).

L’estudi dona una importància cabdal als pronoms: sosté, com Chung & Pennebaker (2007) que les paraules funcionals poden reflectir estats psicològics i emocionals, personalitat, i d’altres trets referents a les relacions socials. Afirmar també que l’ús pronominal d’un text és valuós perquè subratlla si l’atenció està posada en els altres (*he/she, they*), en un mateix (*I*) o en un mateix imbricat en una relació social (*we*) (Zimmerman, Wolf, Bock, Peham & Benecke, 2013). També s’ha suggerit, en recerca prèvia, que un individu que se sent insegur de si mateix prestarà més atenció als seus propis pensaments, sentiments i comportaments (Duval & Wicklund, 1972) —en aquesta línia, s’ha demostrat que les persones que es troben en estat depressiu tendeixen a utilitzar més pronoms de primera persona singular que els individus que no pateixen depressió (Rude, Gortner & Pennebaker, 2004).

Kacewicz et al. (2013) apunta que els individus d’alt estatus són aquells que contribueixen a l’èxit grupal (Thibaut & Kelley, 1959), i just perquè l’estatus és conferit col·lectivament pel grup, aquelles persones que es mostren “orientades als altres” (cooperatives, justes, enfocades al col·lectiu, interessades en forjar relacions i en resoldre conflictes) obtenen un estatus més alt. Així, els líders carismàtics i influents fan menys referències a si mateixos i més referències al col·lectiu i, en conseqüència, utilitzen una quantitat major de termes inclusius com *we, us i our*” (Gardner & Avolio, 1998); i els individus amb un estatus baix utilitzen freqüentment el pronom *I* i menys sovint el pronom *you* (Dino, Reysen & Branscomb, 2008). Així mateix, l’alta presència de *I* ha estat negativament associada amb percepcions de dominança (Berry, Pennebaker, Mueller & Hiller, 1997).

5. CONCLUSIONS

El llenguatge emprat per individus amb trastorn de la personalitat antisocial difereix significativament del llenguatge dels individus que no presenten tal trastorn, de manera que podem dir que la hipòtesi nul·la a què ens acollim al començar l’estudi és descartable. Adoptem, doncs, la hipòtesi alternativa; la qual sosté que existeixen trets que diferencien el discurs *psicopàtic* del *no psicopàtic*. Amb tot, aquests trets semblen no ser de naturalesa lèxica (és possible que aquests tinguin més raó de ser en confessions

criminals, on es detallen els interessos que motiven les seves accions) —tot i que mancances lèxiques en el camp de l’afiliació social han estat també descobertes— sinó relacional.

Concretament, les característiques que diferencien els dos tipus de narratives són, pràcticament, reductibles a una categoria: gramaticalment parlant; la categoria pronominal; i, conceptualment parlant, la de l’enfocament en l’ego i el sentiment de no pertinença col·lectiva. Podem dir que la caracterització del llenguatge antisocial sembla basar-se en una sola característica que engloba dues tendències: una alta presència de pronoms de primera persona exclusius i una baixa presència de pronoms de primera persona inclusius. Aquest tipus de caracterització encaixa d’allò més, tal com Hare (2004) ja havia denotat, amb la teoria del desenvolupament de l’ego de Loevinger (1987).

De fet, la baixa presència, com dèiem, de termes lèxics pertanyents a la categoria *affiliation* (dimensió referent a l’afiliació social) gira també al voltant d’aquesta mateixa idea. Recordem que la definició d’*afiliació*, segons el DIEC, és “fer entrar (algú) com a membre en una societat, una congregació, un partit polític, etc.”, és a dir: incloure’l. En aquest sentit, la manca de termes d’aquest camp és del tot equiparable, conceptualment parlant, a la manca de pronoms de primera persona inclusius. Generalitzant, podem dir que l’individu sociòpata sembla sentir-se poc inclòs en les estructures humanes.

Per tal de provar la idea de la reductibilitat al concepte d’*inclusió* versus *exclusió*, es feu una classificació extra amb els algorismes utilitzats anteriorment; aquest cop refinant manualment els atributs de manera que els classificadors s’entrenessin només amb les variables que fan referència a aquesta idea (*Clout*, *I*, *we*, i *affiliation*). Tant l’arbre de decisió *J48* com el *RandomTree* milloraren la seva actuació en un 5% (amb un 80% i 75% d’èxit, respectivament) respecte les anteriors classificacions; fet que podria indicar que els atributs eliminats resultaven gairebé inútils per aquesta tasca de classificació.

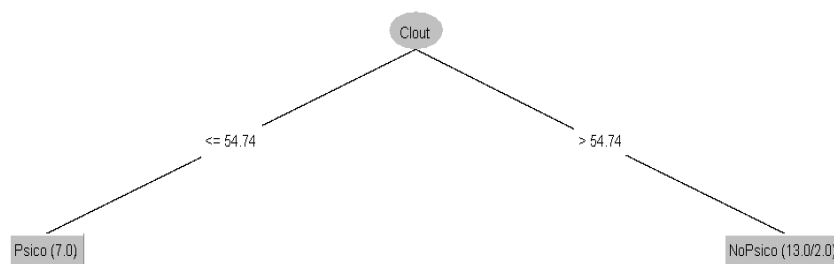


Fig. 15. Arbre de decisió *J48* (80% d’encert)

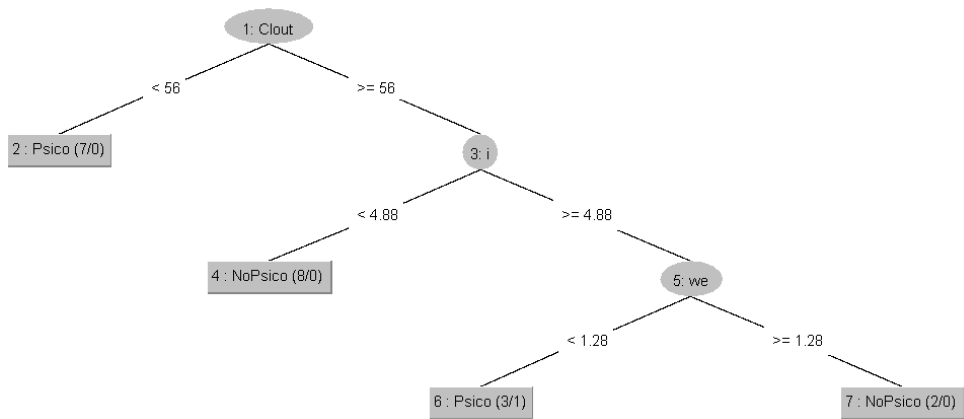


Fig. 16. Arbore de decisió *RandomTree* (75% d'encert)

6. REFERÈNCIES BIBLIOGRÀFIQUES

- Au, Randy (2008). *Uncovering psychopaths: an automated linguistic approach*. Cornell Theses and Dissertations.
- Berry, D. S., Pennebaker, J. W., Mueller, J. S., & Hiller, W. S. (1997). *Linguistic bases of social perception*. *Personality and Social Psychology Bulletin*, 23(5), 526–537.
- Bromberg, W. (1948). *Crime and the mind; an outline of psychiatric criminology*. Lippincott.
- Chung, C., & Pennebaker, J. (2007). *The Psychological Functions of Function Words*. In K. Fiedler (Ed.), *Frontiers of social psychology. Social communication* (p. 343–359). Psychology Press.
- Cleckley, H. (1941). *The mask of sanity; an attempt to reinterpret the so-called psychopathic personality*. Mosby.
- Day, R., & Wong, S. (1996). *Anomalous perceptual asymmetries for negative emotional stimuli in the psychopath*. *Journal of abnormal psychology*, 105(4), 648–652.
- de Oliveira-Souza, R., Hare, R. D., Bramati, I. E., Garrido, G. J., Azevedo Ignácio, F., Tovar-Moll, F., & Moll, J. (2008). *Psychopathy as a disorder of the moral brain: fronto-temporo-limbic grey matter reductions demonstrated by voxel-based morphometry*. *NeuroImage*, 40(3), 1202–1213.
- Dino, A., Reysen, S., & Branscombe, N. R. (2009). *Online interactions between group members who differ in status*. *Journal of Language and Social Psychology*, 28.
- Duval, S., & Wicklund, R. A. (1972). *A theory of objective self awareness*. Academic Press.
- Endres J. (2004). *The language of the psychopath: characteristics of prisoners' performance in a sentence completion test*. *Criminal behaviour and mental health : CBMH*, 14(3), 214–226.
- Gardner, W. L., & Avolio, B. J. (1998). *The charismatic relationship: A dramaturgical perspective*. *The Academy of Management Review*, 23(1), 32–58.
- Gnanambal, S., Thangaraj, M., Meenatchi, V.T., & Gayathri, V. (2018). *Classification Algorithms with Attribute Selection: An Evaluation Study using*

- WEKA. *International Journal of Advanced Networking and Applications*, 9, 3640-3644.
- Gottschalk, L. A., & Bechtel, R. (1995). *Computerized measurement of the content analysis of natural language for use in biomedical and neuropsychiatric research*. *Computer methods and programs in biomedicine*, 47(2), 123–130.
 - Hancock, J., Woodworth, M., & Porter, S. (2013). *Hungry like the wolf: A word-pattern analysis of the language of psychopaths*. *Legal and Criminological Psychology*, 18(1), 102-114.
 - Hancock, J.T., Woodworth, M., & Boochever, R. (2018). *Psychopaths Online: The Linguistic Traces of Psychopathy in Email, Text Messaging and Facebook*. *Media and Communication*, 6 (3), 83-92
 - Hare, R. D. (1996). *Psychopathy: A Clinical Construct Whose Time Has Come*. *Criminal Justice and Behavior*, 23(1), 25–54.
 - Henning, Andrew Stephen (2017). *Machine Learning and Natural Language Methods For Detecting Psychopathy In Textual Data*. *Electronic Theses and Dissertations*. 446.
 - Huprich, S. K., Gacono, C. B., Schneider, R. B., & Bridges, M. R. (2004). *Rorschach Oral Dependency in psychopaths, sexual homicide perpetrators, and nonviolent pedophiles*. *Behavioral Sciences & the Law*, 22(3), 345–356.
 - Ikonomakis K. E, Kotsiantis S., & Tampakas V. (2005). *Text Classification using Machine Learning techniques*. *WSEAS Transactions on Computers* 4(8):966-974.
 - Kacewicz, E., Pennebaker, J.W., Davis, M.J., Jeon, M., & Graesser, A.C. (2014). *Pronoun Use Reflects Standings in Social Hierarchies*.
 - Loevinger, J. (1976). *Ego Development*. San Francisco: Jossey-Bass.
 - Oberlander, J., & Gill, A. J. (2006). *Language With Character: A Stratified Corpus Comparison of Individual Differences in E-Mail Communication*. *Discourse Processes*, 42(3), 239–270.
 - Oxman, T. E., Rosenberg, S. D., Schnurr, P. P., Tucker, G. J., & Gala, G. (1988). *The language of altered states*. *Journal of Nervous and Mental Disease*, 176(7), 401–408.
 - Pennebaker, J. W. (2011). *The secret life of pronouns: What our words say about us*. Bloomsbury Press/Bloomsbury Publishing.

- Pennebaker, J. W., Boyd, R.L., Jordan, K., & Blackburn, K. (2015). *The development and psychiometric properties of LIWC2015*. Austin, TX: University of Texas at Austin.
- Pennebaker, J. W., & Graybeal, A. (2001). *Patterns of Natural Language Use: Disclosure, Personality, and Social Integration*. *Current Directions in Psychological Science*, 10(3), 90–93.
- Pennebaker, J. W., & King, L. A. (1999). *Linguistic styles: Language use as an individual difference*. *Journal of Personality and Social Psychology*, 77(6), 1296–1312.
- Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic inquiry and word count: LIWC*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Pennebaker, J.W., Mehl, M.R., & Niederhoffer, K. (2003). *Psychological aspects of natural language use: our words, our selves*. *Annual review of psychology*, 54, 547-77.
- Porter, S., & Woodworth, M. (2007). *"I'm sorry I did it... but he started it": a comparison of the official and self-reported homicide descriptions of psychopaths and non-psychopaths*. *Law and human behavior*, 31(1), 91–107.
- Raine A, Ishikawa SS, Arce E, et al. *Hippocampal structural asymmetry in unsuccessful psychopaths*. *Biol Psychiatry*. 2004;55(2):185-191.
- Rieber, R. W., & Vetter, H. (1994). *The language of the psychopath*. *Journal of psycholinguistic research*, 23(1), 1–28.
- Rude, S. S., Gortner, E.-M., & Pennebaker, J. W. (2004). *Language use of depressed and depression-vulnerable college students*. *Cognition and Emotion*, 18(8), 1121–1133.
- Thibaut, J. W., & Kelley, H. H. (1959). *The social psychology of groups*. John Wiley.
- Wolpert, D. H., Macready, W. G. (1997). *No free lunch theorems for optimization*. *EE Transactions on Evolutionary Computation*, vol. 1, no. 1, pp. 67-82.
- Zimmermann, J., Wolf, M., Bock, A., Peham, D., & Benecke, C. (2013). *The way we refer to ourselves reflects how we relate to others: Associations between first-person pronoun use and interpersonal problems*. *Journal of Research in Personality*, 47(3), 218–225.

7. ANNEX

Annex 1: Taula de dimensions lingüístiques (LIWC2015)

Category	Abbrev	Examples	Words in category	Internal Consistency (Uncorrected α)	Internal Consistency (Corrected α)
Word count	WC	-	-	-	-
Summary Language Variables					
Analytical thinking	Analytic	-	-	-	-
Clout	Clout	-	-	-	-
Authentic	Authentic	-	-	-	-
Emotional tone	Tone	-	-	-	-
Words/sentence	WPS	-	-	-	-
Words > 6 letters	Sixltr	-	-	-	-
Dictionary words	Dic	-	-	-	-
Linguistic Dimensions					
Total function words	funct	it, to, no, very	491	.05	.24
Total pronouns	pronoun	I, them, itself	153	.25	.67
Personal pronouns	ppron	I, them, her	93	.20	.61
1st pers singular	i	I, me, mine	24	.41	.81
1st pers plural	we	we, us, our	12	.43	.82
2nd person	you	you, your, thou	30	.28	.70
3rd pers singular	shehe	she, her, him	17	.49	.85
3rd pers plural	they	they, their, they'd	11	.37	.78
Impersonal pronouns	ipron	it, it's, those	59	.28	.71
Articles	article	a, an, the	3	.05	.23
Prepositions	prep	to, with, above	74	.04	.18
Auxiliary verbs	auxverb	am, will, have	141	.16	.54
Common Adverbs	adverb	very, really	140	.43	.82
Conjunctions	conj	and, but, whereas	43	.14	.50
Negations	negate	no, not, never	62	.29	.71
Other Grammar					
Common verbs	verb	eat, come, carry	1000	.05	.23
Common adjectives	adj	free, happy, long	764	.04	.19
Comparisons	compare	greater, best, after	317	.08	.35
Interrogatives	interrog	how, when, what	48	.18	.57
Numbers	number	second, thousand	36	.45	.83
Quantifiers	quant	few, many, much	77	.23	.64
Psychological Processes					
Affective processes	affect	happy, cried	1393	.18	.57
Positive emotion	posemo	love, nice, sweet	620	.23	.64
Negative emotion	negemo	hurt, ugly, nasty	744	.17	.55
Anxiety	anx	worried, fearful	116	.31	.73
Anger	anger	hate, kill, annoyed	230	.16	.53
Sadness	sad	crying, grief, sad	136	.28	.70
Social processes	social	mate, talk, they	756	.51	.86
Family	family	daughter, dad, aunt	118	.55	.88
Reward	reward	take, prize, benefit	120	.27	.69
Risk	risk	danger, doubt	103	.26	.68
Time orientations	TimeOrient				
Past focus	focuspast	ago, did, talked	341	.23	.64
Present focus	focuspresent	today, is, now	424	.24	.66
Future focus	focusfuture	may, will, soon	97	.26	.68
Relativity	relativ	area, bend, exit	974	.50	.86
Motion	motion	arrive, car, go	325	.36	.77
Space	space	down, in, thin	360	.45	.83
Time	time	end, until, season	310	.39	.79
Personal concerns					
Work	work	job, majors, xerox	444	.69	.93
Leisure	leisure	cook, chat, movie	296	.50	.86
Home	home	kitchen, landlord	100	.46	.83
Money	money	audit, cash, owe	226	.60	.90
Religion	relig	altar, church	174	.64	.91
Death	death	bury, coffin, kill	74	.39	.79
Informal language	informal		380	.46	.84
Swear words	swear	fuck, damn, shit	131	.45	.83
Netspeak	netspeak	btw, lol, thx	209	.42	.82
Assent	assent	agree, OK, yes	36	.10	.39
Nonfluencies	nonflu	er, hm, umm	19	.27	.69
Fillers	filler	Imean, youknow	14	.06	.27

Annex 2: Resum de contrast d'hipòtesis (prova u de Mann-Whitney)

Resumen de contrastes de hipótesis

	Hipótesis nula	Prueba	Sig.	Decisión
1	La distribución de Analytic es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,175 ^a	Conserve la hipótesis nula.
2	La distribución de Clout es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,001 ^a	Rechace la hipótesis nula.
3	La distribución de Authentic es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,025 ^a	Rechace la hipótesis nula.
4	La distribución de Tone es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,020 ^a	Rechace la hipótesis nula.
5	La distribución de Dic es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,295 ^a	Conserve la hipótesis nula.
6	La distribución de function es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,152 ^a	Conserve la hipótesis nula.
7	La distribución de pronoun es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,131 ^a	Conserve la hipótesis nula.
8	La distribución de ppron es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,295 ^a	Conserve la hipótesis nula.
9	La distribución de i es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,020 ^a	Rechace la hipótesis nula.
10	La distribución de we es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,016 ^a	Rechace la hipótesis nula.
11	La distribución de you es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	1,000 ^a	Conserve la hipótesis nula.
12	La distribución de shehe es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,131 ^a	Conserve la hipótesis nula.
13	La distribución de they es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,882 ^a	Conserve la hipótesis nula.
14	La distribución de ipron es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,230 ^a	Conserve la hipótesis nula.
15	La distribución de article es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,056 ^a	Conserve la hipótesis nula.
16	La distribución de prep es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,456 ^a	Conserve la hipótesis nula.
17	La distribución de auxverb es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,656 ^a	Conserve la hipótesis nula.
18	La distribución de adverb es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,370 ^a	Conserve la hipótesis nula.

19	La distribución de conj es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,941 ^a	Conserve la hipótesis nula.
20	La distribución de negate es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,552 ^a	Conserve la hipótesis nula.
21	La distribución de verb es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,456 ^a	Conserve la hipótesis nula.
22	La distribución de adj es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,095 ^a	Conserve la hipótesis nula.
23	La distribución de compare es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,230 ^a	Conserve la hipótesis nula.
24	La distribución de interrog es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,710 ^a	Conserve la hipótesis nula.
25	La distribución de number es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,230 ^a	Conserve la hipótesis nula.
26	La distribución de quant es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,656 ^a	Conserve la hipótesis nula.
27	La distribución de affect es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,882 ^a	Conserve la hipótesis nula.
28	La distribución de posemo es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,261 ^a	Conserve la hipótesis nula.
29	La distribución de negemo es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,882 ^a	Conserve la hipótesis nula.
30	La distribución de anx es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,552 ^a	Conserve la hipótesis nula.
31	La distribución de anger es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,230 ^a	Conserve la hipótesis nula.
32	La distribución de sad es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,456 ^a	Conserve la hipótesis nula.
33	La distribución de social es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,056 ^a	Conserve la hipótesis nula.
34	La distribución de family es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,067 ^a	Conserve la hipótesis nula.
35	La distribución de friend es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,261 ^a	Conserve la hipótesis nula.
36	La distribución de female es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,710 ^a	Conserve la hipótesis nula.
37	La distribución de male es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,025 ^a	Rechace la hipótesis nula.
38	La distribución de cogproc es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,552 ^a	Conserve la hipótesis nula.
39	La distribución de insight es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,230 ^a	Conserve la hipótesis nula.

40	La distribución de cause es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,766 ^a	Conserve la hipótesis nula.
41	La distribución de discrep es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,295 ^a	Conserve la hipótesis nula.
42	La distribución de tentat es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,201 ^a	Conserve la hipótesis nula.
43	La distribución de certain es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	1,000 ^a	Conserve la hipótesis nula.
44	La distribución de differ es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	1,000 ^a	Conserve la hipótesis nula.
45	La distribución de percept es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,201 ^a	Conserve la hipótesis nula.
46	La distribución de see es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,012 ^a	Rechace la hipótesis nula.
47	La distribución de hear es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,175 ^a	Conserve la hipótesis nula.
48	La distribución de feel es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,824 ^a	Conserve la hipótesis nula.
49	La distribución de bio es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,656 ^a	Conserve la hipótesis nula.
50	La distribución de body es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,456 ^a	Conserve la hipótesis nula.
51	La distribución de health es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,175 ^a	Conserve la hipótesis nula.
52	La distribución de sexual es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,175 ^a	Conserve la hipótesis nula.
53	La distribución de ingest es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,261 ^a	Conserve la hipótesis nula.
54	La distribución de drives es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,095 ^a	Conserve la hipótesis nula.
55	La distribución de affiliation es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,025 ^a	Rechace la hipótesis nula.
56	La distribución de achieve es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,095 ^a	Conserve la hipótesis nula.
57	La distribución de power es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,552 ^a	Conserve la hipótesis nula.
58	La distribución de reward es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	1,000 ^a	Conserve la hipótesis nula.
59	La distribución de risk es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,710 ^a	Conserve la hipótesis nula.
60	La distribución de focuspast es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,824 ^a	Conserve la hipótesis nula.

6 1	La distribución de focuspresent es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,370 ^a	Conserve la hipótesis nula.
6 2	La distribución de focusfuture es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,152 ^a	Conserve la hipótesis nula.
6 3	La distribución de work es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,295 ^a	Conserve la hipótesis nula.
6 4	La distribución de leisure es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	1,000 ^a	Conserve la hipótesis nula.
6 5	La distribución de home es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,552 ^a	Conserve la hipótesis nula.
6 6	La distribución de money es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,016 ^a	Rechace la hipótesis nula.
6 7	La distribución de relig es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,882 ^a	Conserve la hipótesis nula.
6 8	La distribución de death es la misma entre categorías de class.	Prueba U de Mann-Whitney para muestras independientes	,261 ^a	Conserve la hipótesis nula.

Annex 3: Gràfics de contrast d'hipòtesis (prova *u* de Mann-Whitney). Variables significatives

