

EL REpte DE MESURAR EL BIAIX IDEOLÒGIC DELS MITJANS DIGITALS ESCRITS

Ana Sofía Cardenal, Carolina Galais González, Joaquim Moré, Camilo Cristancho, Sílvia Majó.

ÍNDIX

1 INTRODUCCIÓ. PER QUÈ ESTUDIAR EL BIAIX DELS MITJANS DIGITALS..	3
2 MARC TEÒRIC. MESURAR EL BIAIX ALS MITJANS	11
2.1 Definicions i conceptes bàsics.	11
2.2. <i>Limitacions dels estudis previs sobre el biaix dels mitjans</i>	12
2.3 Una nova direcció	15
3. METODOLOGIA.....	15
3.1. Fase 1: Identificació del corpus per a detectar discursos ideologitzats.....	18
3.2. Fase 2: identificació de les relacions semàntiques característiques d'un discurs ideològic (<i>frames</i>)	22
Com es detecten els <i>frames</i> ?	23
Correspondència entre piulades i diaris segons la similitud de <i>frames</i>	29
Criteris per establir els ideology terms (IT)	23
3.3 Fase 3. <i>Comprovar les correspondències entre els frames d'una determinada ideologia i les notícies dels diaris</i>	24
4 RESULTATS	25
4.1. IT característiques del PP i Podemos.....	25
4.2. Correspondència entre piulades i diaris segons freqüència	26
4.3. Correspondència entre piulades i diaris segons similitud	28
4.4. Correspondència entre piulades i diaris segons focus t	28
5. CONCLUSIONS.	32
Bibliografia	34

EL REPTE DE MESURAR EL BIAIX IDEOLÒGIC EN ELS MITJANS ESCRITS DIGITALS.

1. INTRODUCCIÓ. PER QUÈ ESTUDIAR EL BIAIX DELS MITJANS DIGITALS.

Al nostre territori, com a la majoria dels països occidentals, la transformació en l'esfera dels mitjans de comunicació digital s'ha caracteritzat els darrers anys per l'aparició i popularització d'un nombre indeterminat de mitjans específicament digitals – publicacions, agregadors, blocs i d'altres formats híbrids que barregen informació amb serveis. El número de mitjans digitals creix constantment (Newman, Fletcher, Kalogeropoulos, Levy, & Nielsen, 2017) en part, degut a la reducció de costos que suposa la publicació online d'informació. Només a l'Estat Espanyol, l'any 2015 es van crear 579 nous mitjans, la majoria d'ells només amb versions en línia (APM, 2015) i encapçalats per periodistes procedents de mitjans tradicionals –aquells que ja existien abans d'Internet (Minder, 2015; Schoepp, 2016).

Aquesta creixent diversitat en la banda de l'oferta de mitjans de comunicació dibuixa un panorama fragmentat i, alhora, representa un repte per als ciutadans i per als investigadors en comunicació política. Aquest repte es manifesta en dos sentits. Primer, uns i altres desconeixem quin és el grau de pluralitat dels nostres mitjans digitals. És a dir s'han creat molts nous mitjans però desconeixem si aquest increment ha repercutit en una oferta més diversa des d'un punt de vista ideològic. Segon, perquè com la recerca ha demostrat, sabem que els mitjans incideixen en l'opinió pública. Ara bé, per saber quin és el possible efecte dels mitjans en l'opinió pública, és necessari conèixer en primer lloc quina és la inclinació política d'aquest mitjans utilitzats pels ciutadans per a informar-se.

És necessari destacar que el grau de pluralitat del sistema mediàtic d'un país, això és, la diversitat ideològica dels mitjans de comunicació, constitueix un criteri de valoració positiva d'aquell sistema de comunicació, segons el Consell d'Europa (1994). La diversitat ideològica s'ha utilitzat doncs com a element clau per a classificació dels sistemes de comunicació i la seva comparació internacional (veure (Hallin & Mancini, 2004). Per tant, identificar el biaix ideològic dels múltiples mitjans digitals ens permet alhora avaluar la diversitat d'un sistema mediàtic i, en definitiva, la seva contribució al procés democràtic. Això, alhora, ens permetria contribuir a debats rellevants a la literatura de la comunicació política, com ara respondre si certament la creixent oferta

de mitjans de comunicació comporta que els mitjans siguin cada vegada més partidistes i polaritzats (Stroud 2011).

La multiplicació de l'oferta de mitjans en línia també ha dificultat la identificació de les fonts d'informació per part de l'audiència. Establir la correspondència entre els contingut i l'autoria –els mitjans que produeixen les notícies- resulta cada cop complicat per a l'audiència (Kalogeropoulos & Newman, 2017). Aquest pas suposa un entrebanc més i és previ al reconeixement del biaix ideològic dels mitjans. En aquest sentit, proveir a l'audiència d'informació sobre el biaix dels nous mitjans contribuiria a la seva alfabetització mediàtica (Buckingham 2007, Gilster i Gilster 1997)i, per tant, repercutiria positivament en les seves competències cíviques, en el procés d'elaboració de les seves opinions, en la detecció de notícies falses i finalment, en un control més efectiu dels governants.

Pel que fa als efectes dels mitjans sobre l'opinió pública, la recerca ha demostrat que la seva influència està limitada pel biaix de confirmació i l'exposició selectiva, pels quals els individus cerquen informació coherent amb el què prèviament creuen (Lazarsfeld, Berelson i Gaudet 1948, Nickerson 1998) i eviten exposar-se a informació contrària a les seves actituds o creences, ja que aquest contrast genera incomoditat (Festinger, 1962; Olson i Stone, 2014). Tanmateix, la creixent diversitat i fragmentació en els mitjans de comunicació digitals constitueix un repte per al biaix de confirmació i l'exposició selectiva. La multiplicació de l'oferta informativa en línia fa difícil que els usuaris es facin una idea precisa del biaix ideològic de cada nou mitjà digital i, per tant, de la congruència entre aquests i les seves pròpies actituds polítiques. En definitiva, els ciutadans s'estarien exposant ara, a Internet, a estímuls i idees més diverses perquè no poden identificar el biaix de tots els mitjans digitals existents. Queda per saber quin és el grau de diversitat d'aquesta oferta, quins valors tracta de transmetre cadascun i, en conseqüència, quin és el sentit de la seva influència.

Així doncs, saber d'antuvi el biaix ideològic dels mitjans digitals té implicacions molt importants per al funcionament desitjable d'una democràcia. Ara bé, la qüestió clau és: Quin és el biaix individual dels mitjans que conformen l'oferta informativa d'un sistema mediàtic? Com s'assenyala des de la recerca històrica del periodisme, el naixement de la premsa a l'Europa occidental i als EEUU va estar caracteritzat per uns mitjans amb una clara identificació política. Entrat el segle XX, arran de la necessitat de legitimitzar l'activitat periodística, va emergir la imparcialitat i el correcte equilibri de fonts d'informació de diferent ideologies com a aspecte normatiu essencial per a la pràctica del periodisme (Carlson, 2017). No obstant, com han assenyalat Hallin i Maccini en els seus successius treballs sobre els sistemes mediàtics contemporanis (Hallin & Mancini, 2004, 2012), el biaix ideològic és un aspecte inherent de l'activitat periodística i cabdal per a la comprensió de la qualitat de l'oferta mediàtica. De fet, des d'una part de la literatura se subratlla clarament que no hi ha mitjans sense biaix ideològic (Fabos 2004). Aquí una excepció teòrica són els mitjans públics que per llei han de mantenir una posicionament plural i equilibrat. Però fins i tot en aquests darrers casos trobem, a

Europa, nivells diversos d'intervenció governamental i per tant també política en la seva organització i funcionament (Hallin & Mancini, 2012).

A l'hora de conèixer la diversitat ideològica del panorama mediàtic –i el “color” polític de cada mitjà- Catalunya i la resta del territori espanyol juguen amb un cert desavantatge si els comparem amb els Estats Units, on existeixen registres de biaixos ideològics pels principals mitjans (e.g. Editor & Publisher). Dins les fronteres de l'Estat, i en termes generals, és més freqüent referir-se a la ideologia dels mitjans de comunicació d'èpoques preconstitucionals (Jones 1998) que postconstitucionals. Entre les excepcions més destacades trobem les treballs d'Almirón, que ha analitzat l'estructura de propietat i les seves línies editorials per als mitjans tradicionals (2009) i pels diaris digitals sense referent imprès (2006). En una aproximació més recent, l'autora també s'ha ocupat de la diversitat ideològica d'aquests diaris analitzant en quins termes es refereixen a les ideologies més tradicionals, però sense atribuir a cada mitjà un biaix o etiqueta ideològica concreta, sinó representant el panorama conjunt que aquests mitjans ofereixen (Pineda i Almirón 2013). Altres intents de caracteritzar la premsa digital espanyola emfatitzen la importància de l'estructura d'enllaços o hipervincles entre els mitjans i com aquestes estructures creen xarxes d'afinitat que serveixen per a inferir la seva influència (Trillo 2008). Però continuem sense tenir una brúixola comunament acceptada a què referir-nos quan parlem dels biaixos ideològics dels nous mitjans digitals.

Confrontats amb el dilema d'inferir la ideologia d'un mitjà, alguns autors proposen identificar-los d'acord amb la ideologia de la seva audiència (Gentzkow & Shapiro, 2011; Newman et al., 2017). Alguns d'aquests treballs, però, presenten una limitació important, i és que es basen en el record dels individus (Newman et al., 2017) a l'hora de saber quins mitjans han visitat. Sobreestimar la memòria de l'audiència esbiaixa les anàlisis que es basen en aquest record (Prior, 2009). A més, el procediment assumeix que els individus saben quina és la inclinació de cada mitjà i que escullen informar-se a través d'un que els sigui afí. Aquest pressupost, com s'ha explicat abans, està entredit per la multiplicitat i novetat dels mitjans digitals.

Podem intentar una primera aproximació al fenomen de la ideologia dels mitjans digitals adoptant una de les estratègies més utilitzades pels investigadors quan es veuen en el repte de identificar la ideologia d'un partit: analitzar les percepcions de la ciutadania.¹ A continuació detallem els resultats d'aquesta metodologia que es basa en 3 enquestes sobre mostres diferents². En concret les figures 1 a 3 mostren la percepció de

¹ La literatura sobre identificació de la ideologia dels partits polítics constitueix un referent reiterat en la present recerca, i més endavant ens hi referirem més en detall. Val a dir, però, que els altres dos sistemes més utilitzats per a identificar la ideologia d'un partit són les enquestes a experts (per exemple, l'estudi Chapell Hill i sobre tot l'estudi de Benoit & Laver, 2006) i l'anàlisi (automatitzada o no) dels seus programes electorals.

² Aquestes enquestes es van efectuar entre 2015 i 2016. La primera és una enquesta del Grup de Recerca en eGovernança: administració i democràcia electrònica (GADE), de la Universitat Oberta de Catalunya (UOC) que es va efectuar en el marc del projecte Opinionet, finançat pel MEyC. La segona, és una enquesta del grup de recerca Democràcia, Eleccions i Ciutadania (DEC) de la Universitat Autònoma de

la ciutadania de la ideologia d'alguns dels principals mitjans digitals estatals- Obviant l'encert de l'audiència a l'hora de fer aquestes identificacions, el més destacable d'aquestes figures és el percentatge d'individus que no saben classificar els mitjans. En totes aquestes enquestes s'ha preguntat per la ideologia només si els individus afirmen conèixer un determinat mitjà. Per tant, entre el 23 i el 33% de persones no saben quina és la ideologia del *Huffington Post* (mitjà digital d'origen americà i amb implantació als principals sistemes mediàtics europeus) o de *20Minutos* mitjà gratuït amb versió també en paper. Aquesta informació és veu clarament a la Figura 1. Gairebé un terç de la població espanyola no coneix quina és la ideologia de mitjans tan populars com *Eldiario.es* o *ElConfidencial*, dos dels mitjans pioners en oferir versions exclusivament a Internet. Aquestes xifres creixen considerablement quan preguntem a estudiants (Figura 3). Cal destacar que el desconeixement de la ideologia d'aquests dos mitjans, es produeix per part d'una subpoblació, els estudiants, que compta amb un nivell d'alfabetització més alt que la mitjana de la població espanyola. La nostra mostra a més, només inclou estudiants "en línia" de la Universitat Oberta de Catalunya, el que encara permet pensar que els seu nivell d'alfabetització digital és superior. Tot i això, les xifres mostren com no poden identificar la ideologia de 3 dels mitjans que tenen una major penetració en el mercat periodístic espanyol pel que fa al nivell d'audiència a Internet. *ElConfidencial* es visitat per un 12,6% de l'audiència d'Internet a l'Estat, *Eldiario.es* per un 7,7% i el *Huffington Post* per un 5%³.

Tal i com mostren les figures d'1 a 3, entre aquest grup d'estudiants, gairebé la meitat no sap ubicar *Eldiario.es*, *ElConfidencial* o el *Huffington Post*; i entre un 63 i un 68% desconeix la ideologia de l'agregador de notícies *Reddit*. A primer cop de vista, sembla que no ser capaç d'identificar el biaix ideològic dels mitjans podria tenir a veure amb característiques individuals (joventut, estudis), i amb característiques dels propis mitjans (novetat, existència d'una versió tradicional en paper, popularitat...). Pel que fa al panorama dels mitjans, aquest sembla bastant divers, amb proporcions similars de mitjans de totes les ideologies, tot i que la primera enquesta (Figura 1) dóna més preponderància als mitjans d'esquerres i l'enquesta a estudiants (Figura 3) indica que aquests perceben els mitjans més a la dreta.

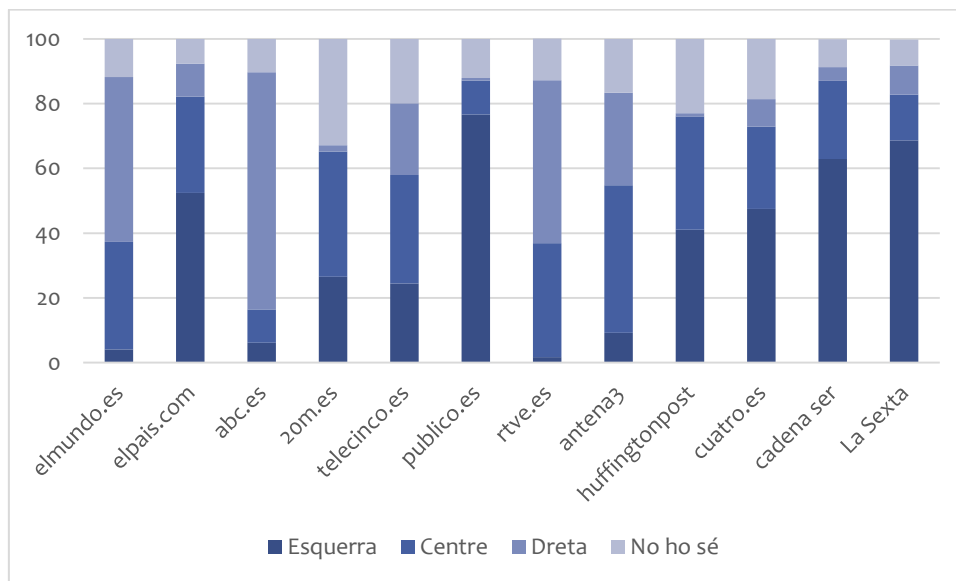
Un altre tret destacable és la variabilitat dels resultats en funció de la mostra. Efectivament, a l'enquesta de la UOC, administrada a una mostra dels seus estudiants, un 57% dels participants identifiquen *Libertad Digital* com un mitjà de dretes, però a l'enquesta DEC, realitzada sobre una mostra de la població general, només un 46% ho percep així. A l'enquesta d'Opinionet un 52% percep El País com un mitjà d'esquerres, però a l'enquesta DEC només un 40% pensa igual. Les variacions encara són més marcades per al *Huffington Post*, que es percebut com un mitjà d'esquerres per un 41%

Barcelona (UAB), on hi col·labora un dels membres de l'equip d'investigació. Es tracta de la 8^a onada d'una enquesta panel en línia. La tercera, es també una enquesta del grup GADE que es va passar als estudiants de l'UOC per conèixer les seves actituds respecte als nous mitjans digitals.

³ Dades proporcionades pel mesurador oficial d'audiències online a l'Estat, ComScore. Les xifres d'audiència corresponent a la mitjana dels mesos de maig, juny i juliol de 2015 i s'han obtingut mesurant el consum de notícies només en ordinadors, no inclouen per tan el consum des de dispositius mòbils.

de l'enquesta de maig de 2015, per un 33% per l'enquesta DEC i per només un 28% dels estudiants de la UOC. Aquesta variabilitat indica que preguntar les audiències per la seva percepció del biaix dels mitjans digitals no és un sistema gaire fiable per a classificar-los. Encara més rellevant és que aquests sistema d'aproximació de la ideologia dels mitjans constata la confusió i desconeixement del biaix ideològic per part de la ciutadania.

Figura 1. Percepció de distints mitjans digitals. Enquesta a la població espanyola OPINIONET, maig de 2015. N=753⁴



⁴ Aquesta enquesta inclou preguntes per als 42 mitjans recollits per Alexa, però aquí es presenten només els 12 més coneguts, identificats per un mínim de 100 respondents.

Figura 2. Percepció de distints mitjans digitals. Enquesta a la població espanyola grup de recerca DEC, maig de 2016. N=1040

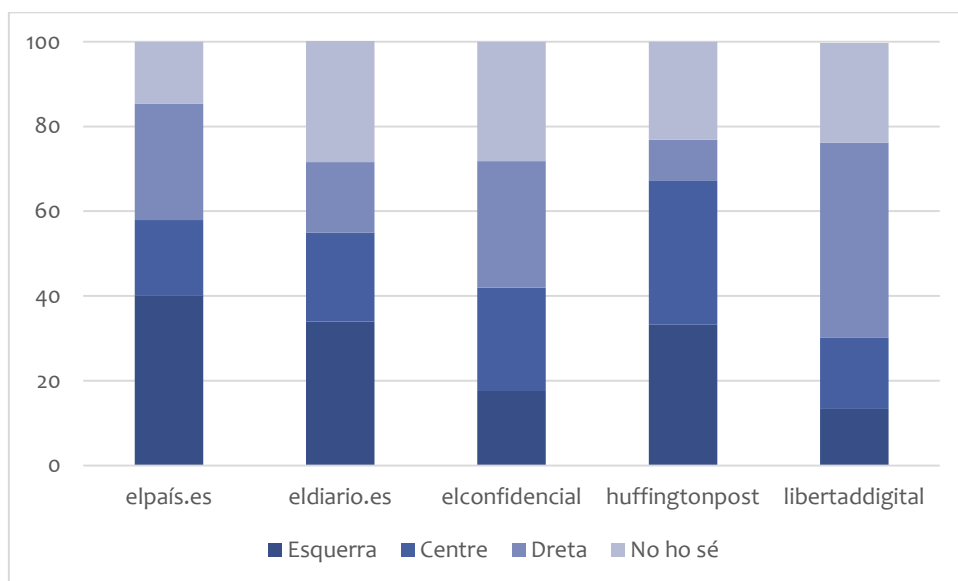
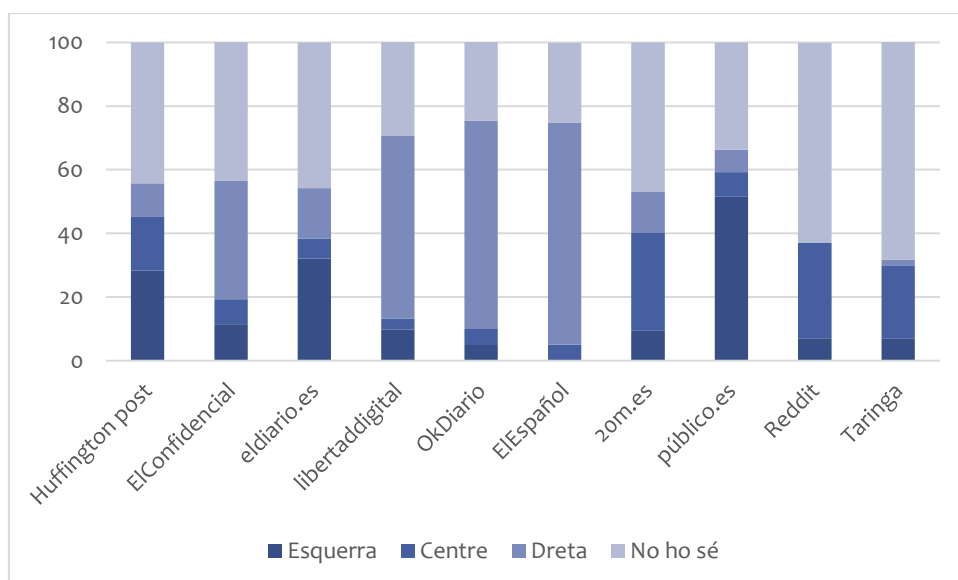


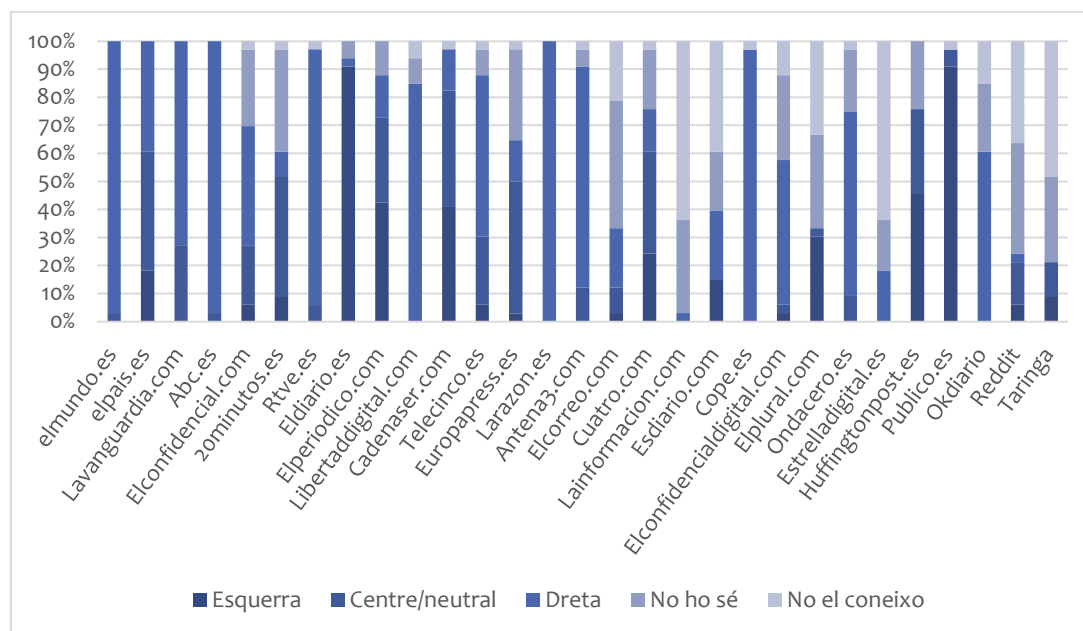
Figura 3: Percepció de distints mitjans digitals. Enquesta als estudiants de la UOC. Juny de 2016. N= 712.



Una estratègia alternativa, també inspirada en el estudis sobre ideologia de partits, consisteix en preguntar experts, en aquest cas en ideologia i comunicació. La figura 4 palesa els resultats d'una enquesta realitzada en setembre de 2017 a 33 experts en les

àrees de ciència política i ciències de la informació. Els experts van ser preguntats per la ideologia dels 30 mitjans més visitats l'any anterior, segons Alexa.

Figura 4: Percepció de distints mitjans digitals. Enquesta a experts. Setembre de 2017. N=33



Si exclouem les versions digitals de mitjans tradicionals com *El Mundo*, *ABC*, etc., trobem un percentatge sorprenentment alt de “no ho sé” i “no conec”, que arriba a ser de més del 50% per a *Lainformación* (3,6% de l'audiència digital a l'Estat visita aquest mitjà segons ComScore). Situar aquests mitjans en un mapa mental d'ideologies on cada mitjà ocupi una posició relativa respecte als altres resulta cada cop més complex per al ciutadà, fins i tot per a experts en mitjans i política.

Una comparació entre l'enquesta d'experts realitzada el 2017 (figura 4) i l'enquesta Opinionet realitzada al 2015 (Figura 1) també suggereix variacions temporals. Encara que les mostres no siguin comparables, s'apunta que el 2015 el País era percebut com a un mitjà d'esquerres, mentre que els experts, en 2017, el perceben com a un mitjà de centre. Aquesta possible variació temporal suggereix que la posició d'un mitjà pot canviar en un espai relativament curt de temps.

L'enquesta a experts desvetlla dos problemes addicionals que un sistema objectiu i automatitzat de classificació ideològica podria resoldre. En primer lloc, poder ubicar més o menys objectivament el mitjà en un espai ideològic sense haver de tenir-ne en compte la resta. És a dir, poder produir valoracions o classificacions de mitjans de manera absoluta, i no relativa, un problema que tindrem sempre que preguntem individus –siguin experts o no– i que els presentem una bateria de mitjans. En segon

lloc, ubicar nous mitjans segons sorgeixin, sense haver-ne de fer un seguiment exhaustiu durant un període de temps. En el mateix sentit, també permetria actualitzar la posició dels que ja existien en funció del que estiguin publicant en un període recent, i no en base al que hagin publicat des del seu naixement. En definitiva, no dependre de les percepcions dels individus sinó centrar-se en el contingut publicat permetria fer classificacions de mitjans segons la seva ideologia independentment del seu bagatge i història, a més de fer seguiments de la seva evolució d'any en any, per exemple.

La present recerca té l'objectiu de classificar els principals mitjans digitals al territori espanyol en funció del seu biaix ideològic, apostant per l'anàlisi de contingut automatitzat que permeti posar al dia la classificació resultant de manera eficient i el més objectiva possible. Aquesta informació serà d'utilitat no només en l'àmbit acadèmic per als debats ja presentats sobre exposició selectiva, sinó també de vital importància política –pel que fa a avaluar la pluralitat de mitjans- i per a millorar el grau d'alfabetització digital de la ciutadania, cosa que alhora hauria de repercutir positivament sobre la qualitat democràtica del sistema polític.⁵

El següent treball s'estructura de la següent manera. Primerament, analitzarem els antecedents als estudis de comunicació sobre ubicació ideològica dels mitjans i també en estudis sobre ideologia de partits. Seguidament, presentarem la metodologia de l'estudi, per acabar presentant els resultats més rellevants i conclusions assolides.

⁵ L'alfabetització mediàtica és el desenvolupament d'una comprensió raonada i crítica de la naturalesa dels mitjans de comunicació i els seus efectes, com creen significat i com organitzen la seva pròpia realitat (Gilster & Gilster 1997, Aparici 1996)

2. MARC TEÒRIC. MESURAR EL BIAIX ALS MITJANS.

2.1. Definicions i conceptes bàsics.

Una majoria de definicions sobre què és el biaix ideològic als mitjans coincideix en el fet que el biaix no és “prejudici” o “opinió”. L’existència d’una opinió no és ni suficient ni necessària per justificar la conclusió que les notícies produïdes contenen un biaix ideològic (Groeling, 2013). Aquestes definicions també coincideixen en que el biaix no implica un intent deshonest i deliberat de tergiversar la realitat (Groseclose i Milyo, 2005). Aquesta actitud estaria més a prop d’allò que avui es coneix com fake news o notícies falses (Allcott & Gentzkow, 2017; Nielsen & Graves, 2017). En canvi, en major o menor mesura, la literatura coincideix en que el biaix es pot entendre com una forma de descriure la realitat que està significativament i sistemàtica distorsionada (Groeling, 2013: 130).

Un pas previ per entendre aquesta distorsió es produiria a través de dues categories d’eleccions que poden fer aquells que elaboren les notícies: quina informació cobrir (què cobrir) i com cobrir-la (quin tractament donar-li). Groeling (2013) es refereix a les decisions en aquestes dues àrees com biaix de “selecció” i de “presentació” (p. 130). Una altra forma de referir-se a aquestes dues fonts de biaix en la literatura és parlar de “filtre” per referir-se a la selecció de temes (o issue filtering) i “d’enquadrament” pel que fa tractament dels mateixos (o issue framing) (Budak, Goel, & Rao, 2016).

Ara bé un pas previ necessari per abordar què entenem per biaix ideològic és la pròpia definició de la paraula ideologia. La ideologia s’ha definit com una distorsió d’una realitat objectiva (veure Diccionari d’Oxford). Alguns autors consideren que la ideologia és un conjunt de construccions mentals subjectives que en contacte amb les institucions i amb altres agents generen cognicions socials que descansen en percepcions de la realitat distorsionades (Benabou, 2008:1). És per això, que la ideologia es pot definir “com distorsions de la realitat col·lectivament sostingudes” (Benabou, 2008: 2). Un dels autors seminals en aquest debat, Converse (1964), defineix la ideologia com les parts (o subconjunts) d’un sistema de creences. Converse posa l’èmfasi en el que considera que defineix tot sistema de creences: la interdependència funcional o constricció entre els elements (o idees força) constitutius del sistema. Defineix un sistema de creences com: “una configuració d’idees i actituds on els elements estan lligats a través d’alguna forma de constricció o interdependència funcional” (p.207).

La idea que proposa Converse d’interdependència funcional entre els elements d’un sistema de creences té diverses implicacions. D’una banda, i des del punt de vista de l’agent, permet que una idea força o organitzadora permeti a aquest agent ubicar i donar sentit a una ampla varietat d’informació. Sense aquesta idea força aquesta informació

no es comprendria (Convers, 1964: 214). D'altra banda, com més dependència funcional hi hagi entre els elements d'un sistema de creences, menys recursos cognitius caldran per ser descrit o comprés (Convers, 1964). Des d'aquest punt vista, una de les idees força o dimensions de judici que més útil ha estat per organitzar i simplificar els esdeveniments en la política occidental, en el darrer segle i mig, ha estat la dimensió esquerra-dreta (o liberal-conservadora). Sobre aquesta dimensió els partits, els polítics, les polítiques, les decisions judicials i els altres objectes de la política han pogut ser – amb major o menor precisió – correctament ubicats (p.214).

La interdependència entre els elements que caracteritza un sistema de creences també explicaria, segons Converse, que la difusió social de les ideologies tendeixi a fer-se per 'paquets'. Uns conjunts de significats o conceptes que els ciutadans veuen com un tot 'natural' pel fet que siguin presentats d'aquesta manera. "Si creus això, llavors també creuràs allò ja que una cosa porta associada l'altra", explica Converse (1964: 211).

⁶Aquesta forma a partir de la qual es difonen les ideologies, per paquets, és rellevant perquè afecta el comportament, la interpretació i la comunicació de les mateixes ideologies polítiques. Els partits, per exemple, voten sobre diferents temes de forma connectada (Aldrich, 1995; Benoit i Laver, 2006, 2007) i quan ho fan presenten paquets d'alternatives als electors (Downs, 1957). Els electors utilitzen una dimensió de judici, normalment l'eix esquerra-dreta (o liberal-conservadora), per donar sentit a les decisions de vot dels partits i per prendre, ells mateixos, decisions sobre els paquets d'alternatives que els partits han presentat.

Aquesta seqüència del procés polític ens porta, finalment, a parlar dels actors que aquí són d'interès, els mitjans de comunicació. Ells també difonen les ideologies polítiques a través de paquets, en aquest cas de conjunt de paraules o termes que evoquen altres conceptes connectats ideològicament. Amb aquestes construccions verbals i substantives apel·len als diferents sistemes de creences i conceptes que els defineixen i estan funcionalment connectats.

Fets aquests aclariments, en aquest estudi entendrem que **el biaix ideològic als mitjans és l'ús d'una i/o les dues fonts de biaix esmentades més a d'alt: el biaix de filtre i el d'enquadrament**. Ambdós s'utilitzen per distorsionar el coneixement polític amb la finalitat de polaritzar l'opinió i/o d'influir en el seu comportament o per referir-se als actors de la política, com partits i/o polítics (Groeling, 2013).

2.2. Limitacions dels estudis previs sobre el biaix dels mitjans.

Estudis previs sobre biaix ideològic dels mitjans han fet servir bàsicament dues aproximacions per mesurar-lo: la primera, basada en l'estudi i la caracterització de l'audiència, i la segona, en el contingut publicat (veure també Budak et al., 2016). La primera aproximació ha utilitzat el perfil ideològic de l'audiència que llegeix, visita o segueix uns mitjans determinats per atribuir-los una ideologia. Seguint la literatura

⁶ Una altra conseqüència metodològica per a la nostra recerca de la idea d'ideologia com a interdependència de conceptes format paquets és que es correspon gairebé perfectament amb la noció de "frames" o marcs d'interpretació propi de les anàlisis semàntiques.

sobre exposició selectiva a la informació (Freedman i Sears, 1966; Sears i Freedman, 1967) i estudis empírics que mostren que els votants prefereixen consumir informació congruent (Iyengar i Hahn, 2009; Lazarsfeld, Berelson, i Gaudet, 1944; Stroud, 2011), s'assumeix que la audiència segueix mitjans que li són pròxims ideològicament. D'aquesta manera, coneixent la ideologia de la audiència es pot atribuir una ideologia als mitjans (Bakshy, Messing, i Adamic, 2015; Gentzkow i Shapiro, 2011; Newman, Fletcher, Kalogeropoulos, Levy, i Nielsen, 2017). Altres treballs, com el de Barberá i Sood (2014), es basen en una perspectiva similar per estimar la ideologia dels mitjans però en el seu cas utilitzen les decisions dels usuaris de xarxes socials, en particular de Twitter, de seguir a diferents mitjans per determinar el biaix ideològic d'aquest. El principi bàsic és el mateix: sobre la base del perfil ideològic dels seguidors dels mitjans de comunicació a Twitter s'estima la ideologia dels mitjans.

Aquesta aproximació té algunes avantatges. La principal és que és parsimoniosa i relativament simple: explota la informació que donen els individus amb les seves decisions de consumir, visitar o seguir un mitjà per assignar-los una ideologia. Té però alguns inconvenients. El primer, ja l'hem esmentat a la introducció d'aquest estudi: la proliferació de mitjans fa cada cop més difícil per a l'audiència conèixer el biaix ideològic dels mitjans. Un altre inconvenient d'aquesta aproximació és que proporciona mesures relatives i no objectives d'aquest biaix. Per tant si tenim en compte que els moviments de l'audiència poden ser molt sensibles a petites diferències en el biaix entre els mitjans, aquest mètode no ens permetria valorar bé les diferències existents (Budak et al. 2016).

La segona aproximació utilitzada a la literatura per identificar el biaix del mitjans es basa en el contingut que aquests elaboren. Aquesta aproximació pretén superar alguns dels problemes del mètode anterior basant-se en una mesura objectiva de biaix –el contingut publicat –i proporcionar diferències quantitatives mesurables del biaix entre els diferents mitjans existents en un sistema mediàtic. Tot i les avantatges potencials d'aquesta aproximació, també presenta reptes formidables. El més important es deriva del fet que una majoria de mitjans no pren posicions explícites sobre els temes que cobreix, i això afecta al gruix del contingut dels mitjans. És el que considerem informació factual o descripció ordenada i coherent de dades i d'accions, sense incorporar cap element interpretatiu que permeti donar-hi un significat més enllà del contingut que el propi text incorpora. Així doncs, és difícil estimar la ideologia dels mitjans a partir d'aquest tipus de continguts (Barbera i Sood, 2015). Davant aquesta limitació, les anàlisis s'han basat en mètodes indirectes i s'ha fet ús de diferents estratègies.

Podem distingir tres gran estratègies dins la literatura per superar aquesta limitació. La primera estratègia consisteix a limitar l'anàlisi a un conjunt reduït però altament informatiu del contingut publicat. Es tracta del contingut editorial, aquell que sí incorpora explícitament el posicionament dels mitjans sobre els fets d'actualitat que cobreix. Aquest contingut es troba ubicat en espais que mitjançant elements gràfics o visuals es diferencien dels espais que contenen contingut factual. Parlem de les

editorials o les pàgines d'opinió en els mitjans audiovisuals o els diaris respectivament. La segona gran estratègia es basa en l'aprenentatge automatitzat per detectar patrons (lingüístics) en un conjunt ampli i indiscriminat de notícies. La tercera i última, combina l'aprenentatge automatitzat i la codificació humana (o crowdsourcing).

Un exemple típic d'ús de la primera estratègia són els estudis que es basen precisament, en les pàgines editorials per estimar la ideologia dels mitjans. Per exemple, Ho i Quinn et al. (2008) utilitzen les posicions preses en casos del Tribunal Suprem d'EEUU en 1.500 editorials publicades per diferents mitjans per estimar-ne la seva ideologia. Altres exemples d'aquesta estratègia són utilitzar el recolzament a candidats electorals de diferents partits o les cites a grups d'interès d'ideologia coneguda. Per exemple, Groseclose i Mylio (2005) comparen el nombre de vegades que els mitjans de notícies citen grups d'interès amb la freqüència que ho fan els congressistes de coneguda ideologia per assignar una ideologia als mitjans. Per la seva banda, l'estudi de Puglisi i Snyder (2011) prova que la cobertura d'escàndols polítics varia sistemàticament segons el recolzament a candidats electorals que fan els mitjans estudiats.

És important destacar que una limitació d'aquests estudis és que se centren en un conjunt molt reduït d'articles o una selecció parcial del total del contingut total pels mitjans. En particular, s'ha criticat als estudis que fan ús dels editorials perquè mesuren només el biaix d'una part molt petita del contingut (la pàgina editorial), i no del gruix del contingut (les notícies). Conseqüentment, la ideologia estimada a partir de les editorials pot acabar sent una predictora gens fiable de la ideologia o del biaix de la globalitat del diari (Barberá i Sood, 2014). Així ho demostren Barberá i Sood (2014) els resultats dels quals evidencien que la pàgina editorial del Wall Street Journal és més conservadora que la resta de pàgines i per tant, de contingut del diari.

La segona gran estratègia es basa en fer servir l'aprenentatge automatitzat per detectar patrons lingüístics en el gruix del contingut publicat als mitjans. Uns patrons que després s'utilitzen per estimar la ideologia dels mitjans. La major part dels estudis dins d'aquesta línia segueixen treballs que escalen (o classifiquen) la ideologia política a partir de textos. Val a dir, que aquesta mateixa tècnica s'utilitza no només per estimar la ideologia dels mitjans sinó també la dels candidats electorals i els partits polítics (Monroe, Colaresi i Quinn, 2008). El procediment implica, primer, identificar un conjunt de documents (programes de partits, discursos de congressistes, piulades a Twitter...) a partir dels quals es procedirà a detectar paraules carregades ideològicament. Posteriorment, s'assigna una puntuació a cadascuna d'aquestes paraules, que després és utilitzada per estimar la ideologia del mitjà. Alguns exemples d'ús d'aquesta estratègia aplicades als mitjans són l'estudi de (Gentzkow & Shapiro, 2010) i de Wihbey, Coleman, Joseph, i Lazer (2017). El primer d'ells utilitza els discursos dels congressistes, de coneguda ideologia, per identificar frases –tècnicament conegudes com bigrams i trigrams- amb càrrega ideològica. Tot seguit estima la ideologia dels mitjans a partir de la relació entre la ideologia dels congressistes i freqüència d'ús d'aquestes frases per part dels mitjans. En aquesta mateixa línia, Wihbey et al., (2017) també es basen en declaracions públiques dels congressistes per extreure frases

ideològicament carregades que després utilitzen per estimar la ideologia de periodistes, en el seu cas, a partir dels articles que escriuen.

Els models que s'acaben d'explicar són bons exemples de la creixent sofisticació que caracteritza aquest àmbit d'estudi. Si bé la innovació metodològica amplia el recorregut i per tant la precisió amb la que s'identifiquen paraules càrrega ideològica, aquestes estratègies també tenen importants limitacions. La principal limitació és que per més sofisticats que siguin els models utilitzats per detectar i puntuar les paraules carregades ideològicament aquestes representen, en el millor dels casos, un percentatge encara molt petit del contingut total publicat pels mitjans i, per tant, les estimacions d'ideologia basades en la freqüència d'ús d'aquestes paraules solen produir un elevat volum de soroll (Gentzkow i Shapiro, 2010). La segona de les limitacions és que aquestes paraules o frases associades a una ideologia són utilitzades freqüentment per mitjans d'ideologia oposada en registres com l'humor, la ironia o el sarcasme en el context d'una estratègia de crítica adreçada als adversaris polítics. Clarament, aquest ús dificulta la classificació dels mitjans (Barberá i Sood, 2015:4). En el marc del nostre estudi, aquestes limitacions també són rellevants i com veurem més avall, condicionen els nostres resultats.

Finalment, la tercera estratègia és basa en una combinació d'aprenentatge automatitzat i codificació humana (o crowdsourcing) per superar algunes de les limitacions associades a l'estratègia basada únicament en l'aprenentatge automatitzat. La codificació humana per exemple, permet identificar la ironia i la broma i corregir per tant, el que en aquest camp es denominen falsos positius; aquesta estratègia també permet classificar la ideologia tenint en compte la propietat dels temes i no només el seu tractament (o enquadrament). Un exemple d'ús d'aquesta estratègia és l'estudi de Budak et al. (2016), on es fa servir l'aprenentatge automatitzat per seleccionar articles de contingut polític d'un ampli ventall de mitjans de notícies i la codificació humana o el crowdsourcing per classificar-los per ideologia a partir de les dues fonts de biaix esmentades: els temes que es tracten (el filtre) i com es tracten (l'enquadrament).

La principal limitació però d'una estratègia basada en la combinació d'aprenentatge automatitzat i codificació humana és que és molt costosa en recursos. L'estudi de Budak et al. (2016) per exemple fa servir cents de treballadors de la plataforma Amazon Mechanical Turk per classificar en diferents moments en el temps el tema, el tipus i el biaix ideològic de més de 10.000 articles.

2.3. Una nova direcció

En el present treball apostem per la segona estratègia, que es basa totalment en l'ús de l'aprenentatge automatitzat, per identificar o estimar la ideologia d'una mostra estratègica de mitjans. En concret, aquí seguim els treballs que classifiquen o escalen la ideologia política a partir de textos com ara els programes dels partits polítics, els discursos dels parlamentaris o els missatges, piulades, enviats a través de la xarxa social Twitter. La nostra proposta però presenta algunes millores i novetats respecte als treballs que hem revistat i discutit fins ara.

La primera de les novetats és que aquí anem una mica més enllà dels estudis previs, i no basem el nostre anàlisi en paraules (o frases curtes) carregades ideològicament sinó en un conjunt de sintagmes nominals connectats. D'aquesta manera ens assegurem que els termes pels què comencem tinguin significat per sí mateixos.

La segona de les decisions i novetats que caracteritzen el nostre treball, és que la nostra anàlisi utilitza com a text de referència per classificar la ideologia dels mitjans les piulades de polítics i partits a tuitar en comptes d'utilitzar, com han fet altres estudis esmentats, els programes electorals o els discursos parlamentaris. Com veurem, això ens garanteix l'accés a discursos més ideologitzats i compromesos que els presents als programes dels partits.

La tercera i darrera decisió que trenca amb la tradició en aquest camp és utilitzar Twitter com a font per detectar discursos ideologitzats (més endavant “frames”) que mirarem de trobar també als mitjans en una segona fase. Com hem vist, alguns estudis utilitzen els comptes de Twitter dels usuaris dels mitjans per a deduir-ne la ideologia i, en última instància, atribuir-la als mitjans. Però cap estudi fins a la data actual ha utilitzat els comptes de Twitter de polítics per a detectar quins termes i discursos són les típics d'una ideologia i, després, trobar-ne correspondències amb els distints mitjans.

Hi hauria, com a mínim, dos clars arguments per centrar l'anàlisi en les piulades dels polítics i per tant, utilitzar-les com a textos de referència. Alguns estudis defensen que Internet ha augmentat l'oferta de mitjans ideològics (o partisan), i amb ella, s'ha contribuït a una creixent polarització (Iyengar i Hahn, 2009; Bennett i Iyengar, 2008; Stroud, 2011). Aquesta polarització que com diem alguns autors atribueixen a l'augment de l'oferta de mitjans partidaris també afectaria les converses i discussions que es mantenen a les xarxes socials. Una part de la literatura que ha estudiat la homofília –comunitats definides per membres d'una mateixa ideologia- i la dinàmica de les converses a Twitter troba evidències de segregació en aquesta xarxa social (Colleoni, Rozza, i Arvidsson, 2014; Conover et al., 2011). Des d'aquesta perspectiva, esperem que a Twitter es faci servir un llenguatge amb més càrrega ideològica i divisiva que en altres textos (Toff i Kim, 2013). I per tant que ho facin també els polítics i els mitjans. Tanmateix, i reforçant aquesta expectativa, sabem que en les converses polítiques a Twitter els textos amb càrrega editorial o clar posicionament ideològic tenen més capacitat d'atracció de l'audiència (veure (Majo-Vazquez, Nurse, Simon, & Nielsen, 2017) i per tant, presenten un clar avantatge per ser més utilitzat per part dels actors polítics i els mitjans.

Tanmateix, també esperem que hi hagin més semblances entre el llenguatge utilitzat a Twitter i el que fan servir els mitjans o alguns d'ells per diferenciar-se de la competència en un escenari de creixent oferta (Mullainathan i Shleifer, 2005).

Un darrer argument per basar-se en un corpus de text creat a partir dels missatges de Twitter està relacionat amb un canvi en la conceptualització del que és un partit. Als estudis clàssics, un partit es defineix com una plataforma controlada per polítics per assolir objectius electorals. Estudis recents però, proposen una nova forma de

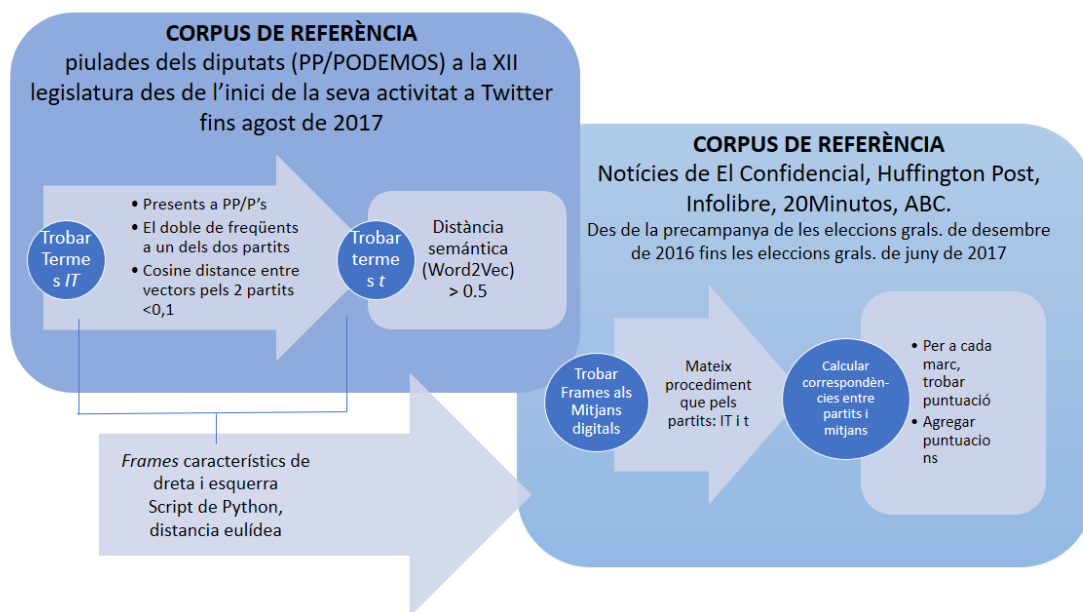
conceptualitzar els partits. Fruit dels canvis socials i tecnològics, en aquests estudis es proposa conceptualitzar els partits com coalicions laxes composades per grups d'interès, professionals de la comunicació, i simpatitzats que comparteixen una agenda i uns objectius comuns (Bawn et al., 2012). En el marc d'aquesta nova conceptualització del que és un partit assolirien molta importància els professionals de la comunicació i l'elecció i ús de paraules amb una càrrega ideològica orientada a la construcció d'un relat (Toff i Kim, 2013). El context o escenari on aquesta coalició d'interessos que són els partits posaria a prova aquest llenguatge amb càrrega ideològica no serien els programes electorals, que poca gent llegeix i són bastant neutres i formals, sinó les xarxes socials un espai molt més dinàmic i caracteritzat pel seu creixent ús en els darrers 5 anys, per part de la població en la majoria de democràcies occidentals (Newman et al., 2017)

3. METODOLOGIA

Aquest apartat presenta la metodologia emprada per a classificar els mitjans digitals segons la seva ideologia. Hem seguit 3 fases: 1) identificació del corpus per a detectar els discursos ideologitzats. 2) Identificar les relacions semàntiques que són característiques d'un discurs ideològic (*frames*). 3) Comprovar les correspondències entre els *frames* d'un discurs polític d'una determinada ideologia i les notícies dels diaris.

La Figura 5 resumeix sintèticament el procés seguit per a identificar els marcs d'interpretació dels partits i, finalment, per a trobar-ne correspondències amb els mitjans digitals.

Figura 5: resum de les fases seguides per identificar *frames* de partits als mitjans.



3.1. Fase 1: Identificació del corpus per a detectar discursos ideologitzats

El primer repte per a identificar la ideologia de les notícies dels mitjans i, des d'un punt de vista agregat, dels mateixos mitjans, era identificar el contingut definitori d'una ideologia. Donat que el contingut, temes i manera de tractar-los que defineixen una ideologia són dependents del context (tant de la societat com del moment en que es mesuri), podíem triar dues vies per a determinar quines eren les característiques definidores de les esquerres i les dretes. La primera, emprar criteris més o menys acordats per la literatura. La segona, determinar uns criteris "ex novo" que servissin per al context i conjuntura actual espanyols.

Vam provar la primera opció, considerant els fragments de programes electorals del PP, PSOE, Cs i Podemos, recollits i etiquetats pel Party Manifesto Project. Vam comprovar, per exemple, que temes vinculats a l'esquerra segons el Party Manifesto Project eren temes importants en el programa electoral dels partits de dreta (e.g: ampliació dels serveis socials) i a la inversa. Un fenomen que es podria explicar a partir de la teoria dels valence issues, termes consensuals que generen l'acord entre partits i electors (com ara les virtuts del creixement econòmic o la condemna de la corrupció (Stokes 1963, Enelow i Hinich 1982). El recurs a aquests termes per part dels partits els permetria intentar atreure el votant mitjà evitant el conflicte.

La segona via, que té en compte el context i la conjuntura amb criteris "ex novo", ens va semblar més adequada. Ara bé, agafar aquesta via implica un treball més inductiu pel qual hem emprat procediments automatitzats de processament del llenguatge natural. Això també significa que l'exercici que fem idealment s'hauria de repetir periòdicament i que la categorització que en resulti no és estàtica. Els continguts que identificaran una notícia i mitjà com de dretes o esquerres seran els denominats "*frames*" o marcs d'interpretació. Més endavant explicarem com identificarem les paraules clau d'aquests marcs d'interpretació.

La segona qüestió era triar els corpus o textos de referència per tal d'identificar *frames* que ens permetessin destriar entre textos d'esquerres i dretes. Les possibilitats eren diverses: utilitzar els discursos dels principals líders polítics (per exemple, en el debat sobre l'estat de la nació), els argumentaris dels partits (sovint disponibles en línia), els programes polítics o les piulades dels principals actors de cada partit a Twitter. Vam decantar-nos per aquesta darrera font després de comprovar, com hem dit abans, que els termes emprats pels partits en els seus discursos i documents oficials són molt semblants.

Twitter és una eina caracteritzada per la immediatesa, que no dona temps per a la correcció política i on hi afloren amb més cruesa concepcions i prejudicis ideològics que no es manifesten tan clarament als programes electorals. També és més probable trobar notícies als diaris que presenten el fet amb el mateix biaix ideològic en què es presenta a la piulada d'un polític. D'altra banda, Twitter té recursos retòrics que són comuns amb els diaris. L'estil perifràstic, carregat de sobreentesos, que sostenen la interpretació d'un missatge de 140 caràcters, el fa semblant als titulars de diaris, i sovint la referència directa a les declaracions d'un representant de la ideologia contrària, tant a Twitter com en un diari, és suficient com perquè el lector ho interpreti com una referència plena d'ironia.

Ara bé, aquestes característiques- perífrasi i ironia- comuns entre les declaracions dels polítics i els mitjans de comunicació són tot un repte per al processament del llenguatge natural. Per això ens sembla adient la nostra proposta de *frames*, que aborda els sobreentesos que hi ha en l'ús de certes paraules per detectar la ideologia que hi ha al darrera, tot i no fer-se explícits. Fins i tot no cal que les paraules tinguin per sí mateixes una càrrega ideològica per a que el titular o la piulada en tingui. Ho podem veure en la

piulada (1) i el titular (2), ambdós carregats de fort contingut ideològic, tot i que no hi ha cap paraula amb una càrrega ideològica per sí mateixa. Més que les paraules “Grecia” i “Venezuela”, és el que es considera sobreentès i no es diu el que marca el biaix ideològic.

1. @xxx pregúntele a los griegos quién hundió grecia el señor al que fue apoyar vuestro presidente
2. Rajoy en Sevilla a Pablo Iglesias: «España no es Venezuela»

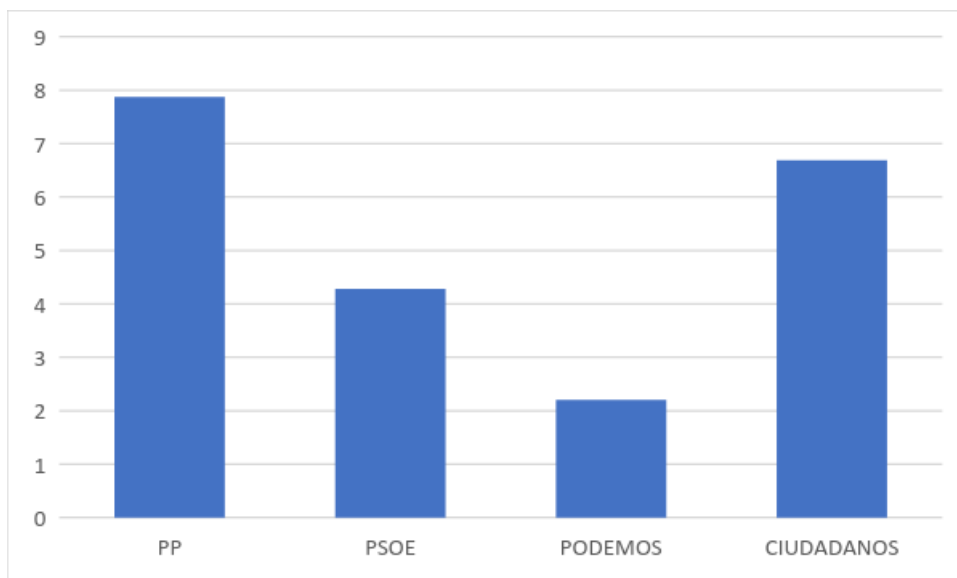
Així doncs, vam optar per triar un corpus de referència menys subjecte a les línies oficials dels partits: els comptes de Twitter de 296 parlamentaris espanyols en la XII legislatura.⁷

Per a explotar el màxim nivell de contrast, i optimitzar la tasca d’atribució d’ideologia als diputats, ens hem limitat en aquesta recerca als dos partits amb una ideologia més extrema i clara a l’eix esquerra/dreta: La coalició Unidos Podemos (simplement Podemos o P’s, d’ara en endavant) i Partido Popular (PP), respectivament.

La decisió de triar els dos actuals principals partits d’àmbit estatal té a veure amb l’univers de l’estudi, els mitjans digitals espanyols. La referència ideològica havia de ser la d’actors polítics suficientment importants com per afectar amb el seu discurs la ideologia d’un mitjà. En el context espanyol això és tradueix en partits polítics d’àmbit estatal (PAES) amb suficient representació parlamentària. A més, aquests dos partits reflecteixen prou bé les dues principals etiquetes ideològiques (esquerra i dreta). Tot i que identificar la ideologia d’un partit també és una qüestió teòrica i empírica de certa dificultat, en aquesta recerca fonamentem la representativitat dels partits respecte a l’eix ideològic en la identificació que en fan els votants. Així, l’enquesta del grup DEC/UAB de desembre de 2015 identifica clarament al PP i a Podemos com els més propers als pols de la dreta i l’esquerra, respectivament (veure la Figura 6).

Figura 6: ubicació dels principals partits a l’eix ideològic pels ciutadans.

⁷ D’entre els 350 diputats, només tenen un compte actiu de Twitter 296.



Font: vuitena onada del PANEL DEC/UAB. Desembre de 2015. N=1040.

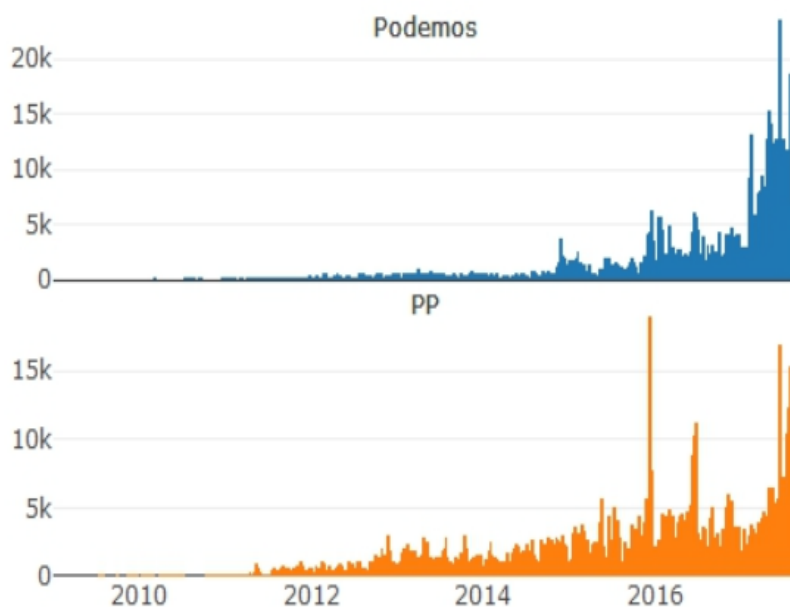
El conjunt de dades analitzades consisteix en gairebé mig milió de piulades dels diputats de Podemos i PP al congrès dels diputats.⁸ La distribució del nombre de les piulades per partit es presenta a la Taula 1 i a la Figura 6.

Taula 1: distribució en el temps de les piulades dels diputats espanyols a la XII legislatura des de l'inici de la seva activitat a la xarxa social.

Any	Usuaris PP	Usuaris Podemos	Piulades PP	Piulades Podemos
2009	6	7	1214	270
2010	15	10	2993	1492
2011	35	17	21324	7377
2012	38	19	48498	20362
2013	50	25	77700	27010
2014	60	32	94667	35147
2015	76	48	166789	77927
2016	88	56	203838	156512
2017	102	62	173722	298474

Figura 6. Representació longitudinal de les piulades utilitzades, per partit.

⁸ Hi hem accedit fent "scraping" (recerca i buidatge automatitzat) dels continguts de cerca de Twitter.



Per als 164 diputats de Podemos i del PP que tenien un compte de Twitter actiu a l'inici del treball de camp (agost 2017), es van rastrejar les publicacions fins que la seva activitat a la xarxa social comença, això és, fins 2009.⁹ Queda coberta doncs una part important del discurs a la política electoral a Espanya en el període d'estudi. Només 13 diputats eren actius a la xarxa el 2009. El nombre d'usuaris augmenta fins als 164 usuaris actuals, com també ho fa la seva activitat a la xarxa. De les 114 piulades per diputat i any que trobem de mitjana el 2009 passem a les més de 2800 per diputat en el que portem de 2017.

3.2. Fase 2: identificació de les relacions semàntiques que són característiques d'un discurs ideològic (frames).

El concepte de frames

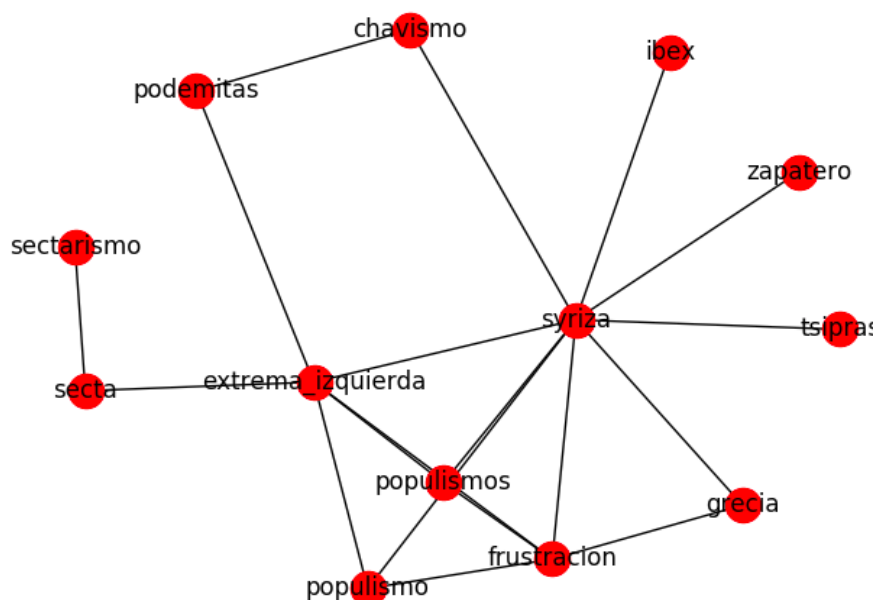
Metodològicament, un *frame* és una relació de proximitat semàntica entre un terme *IT* (*ideology term*, que també podem entendre com a paraula clau) del discurs i uns termes *t* del mateix discurs.¹⁰ La conjunció d'un terme IT amb una sèrie de termes *t* indiquen una determinada visió de les coses referides pel terme IT. Tot seguit s'explica el procediment per a l'obtenció dels termes IT i *t*, però comencem amb un exemple per

⁹ Això no vol dir que aquestes persones fossin diputats el 2009. Alguns sí que ho eren, sobretot del PP, però dins de la coalició Unidos Podemos, només algunes persones ja eren diputats per Izquierda Unida abans de 2016. La resta s'estrenen com a parlamentaris després de les eleccions de desembre de 2016. Entenem, però, que amb les seves piulades abans d'aquesta data estan difonent missatges i valors consonants amb el que serà la línia programàtica d'aquest partit.

¹⁰ Entenem aquí proximitat semàntica com a coocurrència, o aparèixer en posicions adjacents al mateix text. Es tracta d'un concepte propi de l'anàlisi quantitatiu de textos. L'algoritme utilitzat per determinar-la (Wor2Vec) recull aquesta proximitat física de les paraules mantenint les propietats gramaticals dels textos de què s'extreuen.

fer-ho més entenedor. Mentre que, per al PP, el IT “populismos” té associats els termes t “populismo, frustración, Syriza, extrema_izquierda, Grecia”. Alhora, el terme IT “Syriza” presenta associats els termes “Zapatero, frustración, Grecia”. El terme IT “nacionalismo” té associats els termes “ruptura, farsa, sociedad_catalana, independentismo”. Així, durant el període en què es van publicar les piulades, els representants del PP relacionaven Grècia amb el populisme i la frustració, el nacionalisme amb la ruptura, etc. Com que cadascun dels termes t poden esdevenir també termes IT depenent del seu pes al discurs, podem observar xarxes de termes i també de *frames*. La Figura 7 representa una xarxa que relaciona els termes al voltant de l'IT *populismo*. Xarxes com aquestes poden ajudar a entendre l'argumentari ideològic que sosté un discurs breu i comprimit com és el d'una piulada. Així, no ens sorprendria trobar una piulada o el titular d'un editorial que digués que Zapatero és un populista i que ha estat el Tsipras d'Espanya. La piulada o el titular concentra una tesi, un missatge i uns determinats valors del partit X (en aquest cas, el PP) expressats amb uns determinats termes que conformen un discurs, que és el que recullen els *frames*.

Figura 7. Representació gràfica del *frame* sorgit del IT “populismos” per al PP.



Cal tenir en compte, però, que la freqüència no és l'únic criteri utilitzat per identificar *frames* propis del PP i de Podemos. Donat que Twitter és també un espai de crítica, controvèrsia i expressió d'opinions, hem de tenir en compte també que un diputat (i també un mitjà) pot fer referència al *frame* de l'opció ideològica contrària per criticar-lo i convertir-lo en tema de discussió. Per tant, es pot donar el cas de que un terme amb un *frame* propi d'un partit sigui més referenciat precisament en les piulades i els diaris de tendència contrària. Tractarem aquestes qüestions a l'apartat dedicat a les correspondències entre els *frames* dels partits i els dels diaris, i sobre els resultats obtinguts.

Com es detecten els frames?

Per detectar els *frames*, primer calculem els sintagmes nominals de les piulades dels representants d'una determinada ideologia. Per fer això, cal fer prèviament un procés de

‘neteja’ de les piulades, traient-ne les emoticones, mencions a altres comptes, i caràcters estranys.

Per calcular els sintagmes nominals, hem fet servir l’eina *Parse Tree* del paquet *pattern.es* del projecte CLIPS¹¹. El càlcul de sintagmes nominals i no de paraules soles o combinacions de paraules consecutives (*ngrams*, que podem trobar també en la literatura sobre anàlisi de textos), s’explica perquè el sintagma nominal és -juntament amb els verbs- l’element bàsic que estructura una oració, la principal seu del significat lèxic i, en definitiva, la manera en com es denominen els conceptes. Així podem recollir denominacions com *Tribunal Superior de Justicia*, mentre que si partíssim de bigrams obtindríem el IT “Tribunal Superior” i si ho féssim de paraules úniques tindríem “Tribunal”, “Superior” i “Justicia”, tres ITs que en realitat expressen una única idea. Amb un corpus considerable de piulades, el nombre de sintagmes nominals obtinguts és molt gran, per la qual cosa s’han considerat aquells que apareixen al corpus de referència (conjunt de piulades per a cada partit) més de 5 vegades.

Un cop obtinguts els sintagmes nominals, es busquen els seus termes *t*; és a dir, els termes més propers semànticament en el conjunt de totes les piulades. Per obtenir-los apliquem el mètode Word2Vec, en el mode Skip-Gram, implementat mitjançant un mòdul de Python. Explicat de manera molt resumida, el mètode Word2Vec indica que dues paraules *p* i *p'* són properes si apareixen en contextos similars. És a dir, les paraules que solen estar al voltant de *p* solen també estar al voltant de *p'*. Aplicat a la detecció dels termes *t*, l’explicació de que populismos i extrema_izquierda són propers és que les paraules que envolten populismos solen aparèixer també propers a extrema_izquierda’.

Word2Vec utilitza un algoritme que calcula, per cada sintagma nominal, els sintagmes nominals més propers. La proximitat és un valor que va del 0 al 1 (de menys proper a més proper). Per aquest projecte hem considerat com termes *t* els que superen el valor de la mediana (0.5).

Criteris per establir els ideology terms (IT)

D’entre tots els sintagmes nominals, vam haver d’establir un(s) criteri(s) per identificar aquells que són un IT (ideology term). Hem fet servir tres criteris. Em primer lloc, el sintagma nominal ha d’aparèixer tant a les piulades del PP com de Podemos. Sense aquesta condició no podem decidir si hi ha discrepància en els *frames* entre els dos partits (atès que només un el fa servir) i tampoc podem pròpiament fer una atribució del *frame* a un partit (el *frame* no és prou important ja que no hi ha posicionament sobre ell). En segon lloc, l’IT ha d’aparèixer amb més freqüència en les piulades d’un partit que de l’altre. En tercer lloc, els *frames* dels partits (és a dir, els termes *t* associats al IT) han de ser diferents. Vegem aquests dos últims criteris més en detall.

El segon criteri apunta a què un sintagma nominal es considera IT si un terme és força present al seu discurs (corpus de referència, piulades dels seus diputats). Com més elevada sigui aquesta freqüència en relació amb l’altre partit, més característic del partit serà el terme. Considerem aquí que un criteri raonable és que un terme “propi” d’un partit ha d’aparèixer a les piulades dels seus diputats més del doble de vegades que al corpus de referència de l’altre partit.

¹¹ <https://www.clips.uantwerpen.be/pages/pattern-es>

El tercer criteri indica que el vector que es genera amb les piulades d'un partit té una distància considerable respecte al seu vector en les piulades del partit ideològicament oposat. És la manera de detectar que un mateix sintagma nominal (SN) té un biaix interpretatiu segons el discurs polític en el que es trobi. Per a detectar aquestes diferències representem els SNs d'un partit com vectors de trets. Hi ha tants trets com termes t té el SN a les piulades d'un partit i de l'altre. Els valors de cada tret del vector d'un SN del PP, posem per cas, són els valors de distància semàntica (mesurada, recordem, mitjançant Word2Vec) que un terme t té a les piulades del PP. Un cop es creen els vectors per l'SN del PP i de Podemos, es calcula la cosine similarity entre ambdós. Ens quedarem com a candidats a IT aquells que tinguin una cosine similarity inferior a 0.1, apuntant, per tant, a una gran diferència.

Veiem un exemple:

SN piulades PP: rupturista

Termes t: rupturistas, proceso independentista, soberanistas

Valors de similaritat semàntica dels termes t: [0.794, 0.539, 0.594]

SN piulades Podemos: rupturista

Termes t: jefatura, jefatura del estado, coyuntura, fractura, ruptura democrática, ruptura, postura

Valors de similaritat semàntica dels termes t: [0.506, 0.594, 0.578, 0.629, 0.510, 0.522, 0.516]

Distància entre tots dos SN: 0.0, ja que no hi ha cap terme t del PP que estigui en el *frame* del Podemos i viceversa. Es tracta de dos vectors absolutament diferents.

Així doncs, un terme es considerarà IT quan a) estigui present al discurs de tots dos partits, b) la seva freqüència superi el doble de la freqüència del mateix IT en les piulades de l'altre partit i c) el seu vector sigui molt distant del vector en les piulades de l'altre partit (Cosine Similarity <0.1).

3.3. Fase 3. Comprovar les correspondències entre els frames d'un discurs polític d'una determinada ideologia i les notícies dels diaris

A l'hora d'aplicar el mètode triat i trobar correspondències entre *frames* de partits i *frames* de diaris hem decidit centrar-nos en alguns dels mitjans que s'ha detectat (veure introducció) que generaven més confusió per part de l'audiència, sense perjudici que el mètode pugui ser estès a més mitjans. Aquests mitjans són: Huffington Post, *ElConfidencial*, *Infolibre* i *20Minutos*. A més, hem inclòs l'ABC pel fet de ser el mitjà més clarament situat a la dreta en totes les enquestes analitzades en la present recerca, el que ens pot servir de punt de referència.

Hem obtingut els textos de la base de dades de premsa FACTIVA, i hem acotat la recerca entre inicis de desembre de 2016 (precampanya eleccions generals 2016) i finals

de juny de 2017 (recollint així els dies posteriors a les eleccions del 26 de Juny de 2017, i l'inici de la XII legislatura).

Per fer la comprovació de les correspondències hem considerat diferents opcions:

- a. Detectar una correspondència directa entre la freqüència dels IT d'una determinada ideologia i el diari ideològicament afí. Per exemple, un diari més afí al PP farà servir més IT-PP que un diari de línia ideològica d'esquerres.
- b. Els vectors que descriuen els IT en les piulades i els vectors que descriuen els *frames* d'aquests ITs als diaris són semblants. Per exemple, els diaris afins a Podemos solen vincular la UE amb l'austeritat i amb Àngela Merkel que no pas els diaris afins al PP
- c. La correspondència no està tant en els IT com en els termes t i la seva capacitat de formar part de *frames* d'ITs de diferent ideologia.

Per provar aquestes tres possibilitats ha calgut calcular automàticament els sintagmes nominals dels titulars i els cossos de notícia dels diaris. A l'apartat dedicat als resultats explicarem amb més detall altres processos que hem realitzat per determinar l'ideologia dels mitjans segons aquestes tres possibles variants.

4. RESULTATS.

En aquest apartat mostrarem exemples dels IT1 característics del PP i de Podemos. Seguidament presentarem resultats sobre les correspondències entre la ideologia de les piulades i les notícies dels diaris.

4.1. IT característics del PP i de Podemos

Hem obtingut 327 IT característics del PP (IT-PP), seguint els criteris explicats a l'apartat metodològic. És a dir, sintagmes nominals també presents als discursos del partit oposat, amb una freqüència superior al doble que en les piulades del partit ideològicament oposat, i que el seu vector de t_s tingui una distància (*cosine distance*) inferior a 0.1 respecte al vector del mateix sintagma nominal a les piulades del partit oposat.

A continuació en presentem un exemple d'alguns ITs del PP, amb el seu valor de distància respecte al mateix IT a Podemos

proceso independentista	0.09782474131	gran_pacto	0.07272315273
sede parlamentaria	0.095738877	tranquilidad	0.07031200271
grupo socialista	0.09536627573	reforma_electoral	0.06996984863
centralidad	0.09404395815	abismo	0.06942207419
extremismos	0.09095882373	sentencia del tc	0.05383532882
recuperación económica	0.08946776756	crisis_económica	0.04498589957
comunidades autónomas	0.08613694577	anterior_legislatura	0.02661601968
gobierno estable	0.0842600721	rupturista	0.0
elecciones europeas	0.08331973245	inestabilidad	0.0
empresas españolas	0.08118751714	separatismo	0.0
inversiones	0.08081599019	sistema_educativo	0.0

Això vol dir que tant el PP com Podemos parlen del “proceso independentista”, però que el PP parla més del doble de vegades que Podemos d'aquesta idea, i que els termes amb què s'hi refereixen (els termes t , no representats aquí) són extremadament diferents (valor de la cosine distance entre el vector “proceso independentista” del PP respecte al vector generat per al mateix terme IT de Podemos = 0.0978). La mateixa lògica s'aplica a la resta de termes IT representats a l'exemple i als altres 305 que no s'hi representen.¹²

D'altra banda, hem obtingut 111 IT característics de Podemos. També en presentem un exemple, amb els corresponents valors de distància respecte al mateix IT al PP.

¹² Els IT-PP giren al voltant de la necessitat que tenia el PP en aquell moment de fer un pacte amb el PSOE per frenar Podemos. D'aquí que s'apel·lés a la centralitat, estabilitat política, etc. i s'advertís dels perills dels extremismos, atribuïts a Podemos i a l'esquerra (referències al govern socialista anterior), que posarien en perill la recuperació econòmica. També hi ha esment especial a la societat catalana, incidint en l'independentisme, vinculat amb el rupturisme. D'altra banda, s'esmenten altres qüestions que estan relacionades amb l'independentisme, tals com el sistema educatiu, i esments al tribunal constitucional.

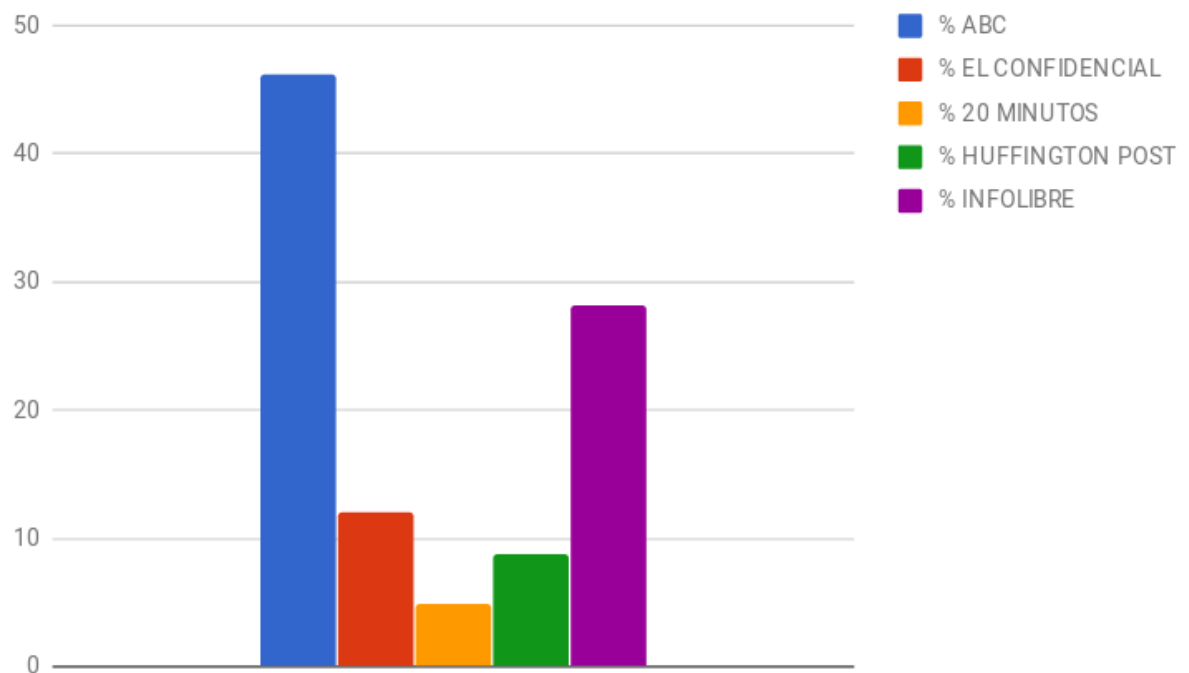
emergencia_social	0.09822456084	animalistas	0.0
agenda política	0.09689463198	renta básica	0.0
redistribución	0.08827979275	confluencias	0.0
capitalismo	0.07074209785	círculos	0.0
sistema electoral	0.0	coleta	0.0
código ético	0.0	feministas	0.0
banca pública	0.0	mayoría social	0.0
proceso constituyente	0.0	república	0.0
oligarquía	0.0	movimientos sociales	0.0
populismo	0.0	ibex	0.0
tacticismo	0.0	proetarras	0.0

A banda de presentar propostes econòmiques (renta bàsica) i fer referències a moviments socials com el 15-M, els IT-PODEMOS giren al voltant d'una crítica al tacticisme dels partits davant la situació de cerca del pacte PP-PSOE i la posició del PSOE d'abstenir-se. També giren al voltant del capitalisme i als seus vicis i defectes (puertas giratorias), a més que reclamen una política amb valors ètics (código ético). D'altra banda, presenten conceptes lligats a la sociologia i la ciència política (mayoría social, movimientos sociales, oligarquía), reflectint la formació i trajectòria professional dels seus principals líders (molts d'ells politòlegs). També es distingeixen del discurs polític del PP incidint en la plurinacionalitat, la república, el feminisme i, fins i tot amb l'animalisme. Termes propis del partit com *casta*, *confluencias*, *círculos* o *marea* també estan en aquests IT. Ara bé, el que crida l'atenció és la presència de *populismo*, *proetarras* o *coleta*, que són termes que la dreta utilitza per desacreditar Podemos. Això apunta a que *les piulades* de Podemos tenen una bona càrrega de referencialitat del discurs del partit contrari ideològicament. Com això incideix en la detecció automàtica de la correspondència entre partits i diaris es tractarà en els propers apartats.

4.2. Correspondència entre piulades i diaris segons la freqüència dels IT

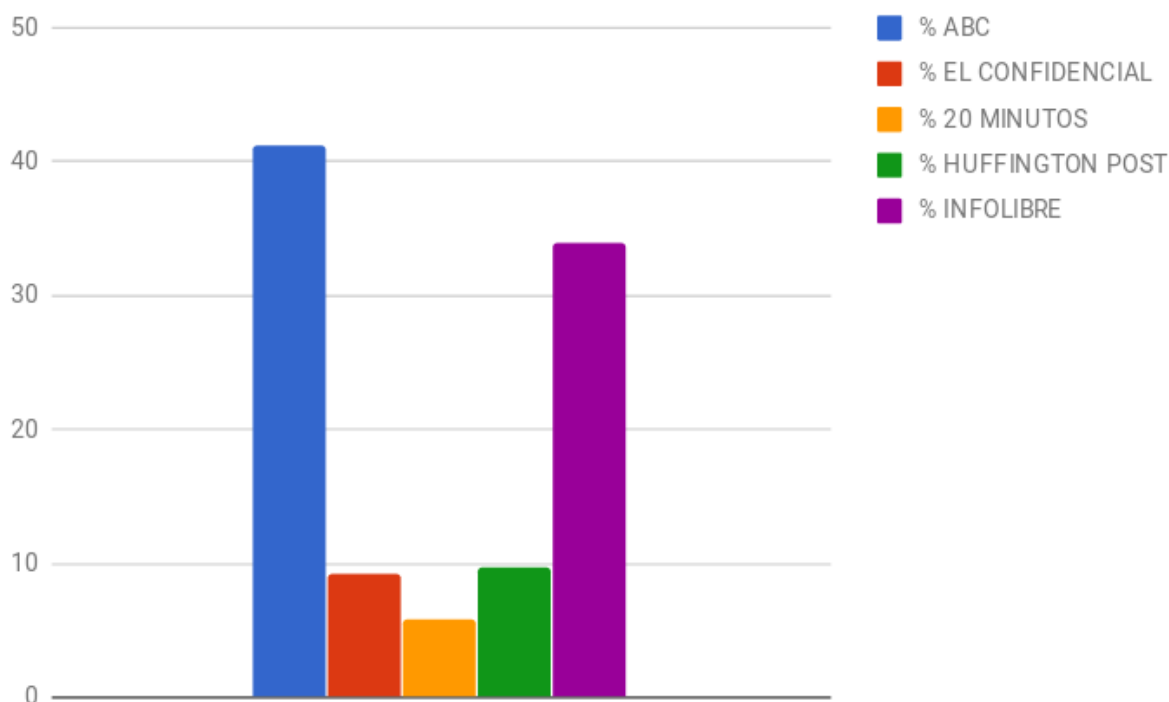
La primera opció per comprovar la correspondència entre les piulades i els diaris va ser comprovar la correspondència directa entre la freqüència dels IT d'una determinada ideologia i el diari ideològicament afí. Per exemple, un diari més afí al PP farà servir més IT-PP que un diari de línia ideològica d'esquerres. Hem de tenir en compte que les piulades i les notícies són espais on pot haver moltes referències als IT ideològicament contraris per criticar-los i polemitzar sobre ells. Per aquesta raó, comprovem si aquest fet influeix a l'hora de trobar-hi correspondències.

Figura 8: Percentatge de IT-PP per als diaris analitzats.



A la Figura 8, veiem el percentatge de IT-PP distribuïts pel diaris que hem analitzat. Es pot veure que un poc més del 40% de les aparicions de IT-PP estan a l'ABC. Li segueix el diari *Infolibre* i *ElConfidencial*. Pel que mostra la gràfica, sembla que hi ha un alineament bastant clar entre les piulades del PP i les notícies de l'ABC, i que *20Minutos* tindria l'alineament menor. Què passa, però, si observem la correspondència entre les IT de Podemos i els mateixos diaris?

Figura 9: Percentatge de IT-PODEMOS per als diaris analitzats.



A la Figura 9 veiem el percentatge de IT-PODEMOS distribuïts pel diaris analitzats. Segons la gràfica, l'ABC és també el diari on es concentren més IT-PODEMOS, encara que menys acusadament que a l'exemple anterior. La distribució relativa de la resta de diaris és molt similar, del que podríem deduir, erròniament, que ABC i Infolibre són els diaris més afins a Podemos i la resta els més hostils o llunyans. Per tant, no sembla que la distribució per freqüència dels IT segons la ideologia serveixi per detectar alineaments clars entre les piulades dels polítics i els diaris. Això es pot deure a l'ús ja comentat de *frames* derivats de ITs com *populismo*, *proetarras* o *coletas*, termes sorgits de la dreta per a criticar a Podemos que s'han trobat més freqüentment als piulades de Podemos per a minimitzar o ridiculitzar les crítiques que rebien. La referència als termes característics de la ideologia contrària i, sobretot, en diaris que semblen els més vehements en utilitzar ITs com són l'ABC i Infolibre poden explicar aquests resultats.

4.3. Correspondència entre piulades i diaris segons la similitud de frames

El següent pas va ser comprovar si els vectors que descriuen els IT en les piulades i els vectors que descriuen els *frames* d'aquests ITs als diaris són semblants. Per exemple, volíem comprovar si els diaris afins a Podemos solen vincular més sovint la UE amb l'austeritat i amb Àngela Merkel que no pas els diaris afins al PP.

Vam processar els diaris i extreure els seus sintagmes nominals. De la mateixa manera que havíem fet amb les piulades dels diputats, vam convertir els sintagmes de cada diari en vectors, les dimensions dels quals eren els termes t ; és a dir, els termes més relacionats semànticament, obtinguts amb el Word2Vect. Els vectors dels IT-PP i IT-PODEMOS es van comparar- via cosine similarity- amb els vectors dels mateixos sintagmes nominals dels diaris. Tot i que en alguns sintagmes nominals les diferències eren significatives, vam comprovar que la referencialitat als ITs del partit oposat

ideològicament era també una característica dels diaris, per la qual cosa vam obtenir resultats semblants als de la freqüència d'IT.

4.4. Correspondència entre piulades i diaris segons els focus en les t

L'última passa de l'anàlisi va ser comprovar si la correspondència no estava tant en els IT com en els termes t i la seva capacitat de relacionar-se amb ITs de diferent ideologia. Com havíem dit, els partits i els diaris incideixen sobre uns aspectes i no sobre uns altres en parlar de determinats temes. En termes de frames, és com dir que, donat un IT, els diaris afins a un partit coincideixen en posar uns termes t en el seu focus. Encara que l'IT sigui característic del partit o ideologia contraris.

Per comprovar-ho, vam recollir els termes t relacionats semànticament amb els IT de les piulades del PP. Vam fer el mateix, treient els termes t relacionats amb els IT de les piulades de Podemos. Anomenarem els primers termes t com t_{pp} i els segons t_{pod} . Llavors, vam comprovar quants t_{pp} coapareixien amb un IT als titulars i al cos de les notícies d'un diari. També ho vam fer per cada t_{pod} . Per exemple, imaginem que un IT de les piulades del PP té 10 t_{pp} relacionats. Llavors, per cada diari, es va calcular quants d'aquests t_{pp} coapareixien amb l'IT en un titular o cos de diari. Il·lustrant-ho amb un exemple concret, es va calcular quants t_{pp} relacionats amb *extremismos* coapareixien amb *extremismos* als titulars i als cossos de notícia de l'ABC, El Confidencial, etc. Així vam crear, per cada IT un vector amb el nombre de t_{pp} i t_{pod} coocorrents per cada diari.

La taula següent il·lustra aquests vectors amb els t_{pp} relacionats amb centralismo, abismo, coleta i populismo. Per exemple, centralidad i coleta té 19 i 1 t_{pp} coocorrents al diari ABC respectivament però cap t_{pp} a Infolibre. Populismo té dos t_{pp} a l'ABC i un a El Confidencial, però cap a 20 Minutos i al Huffington Post.

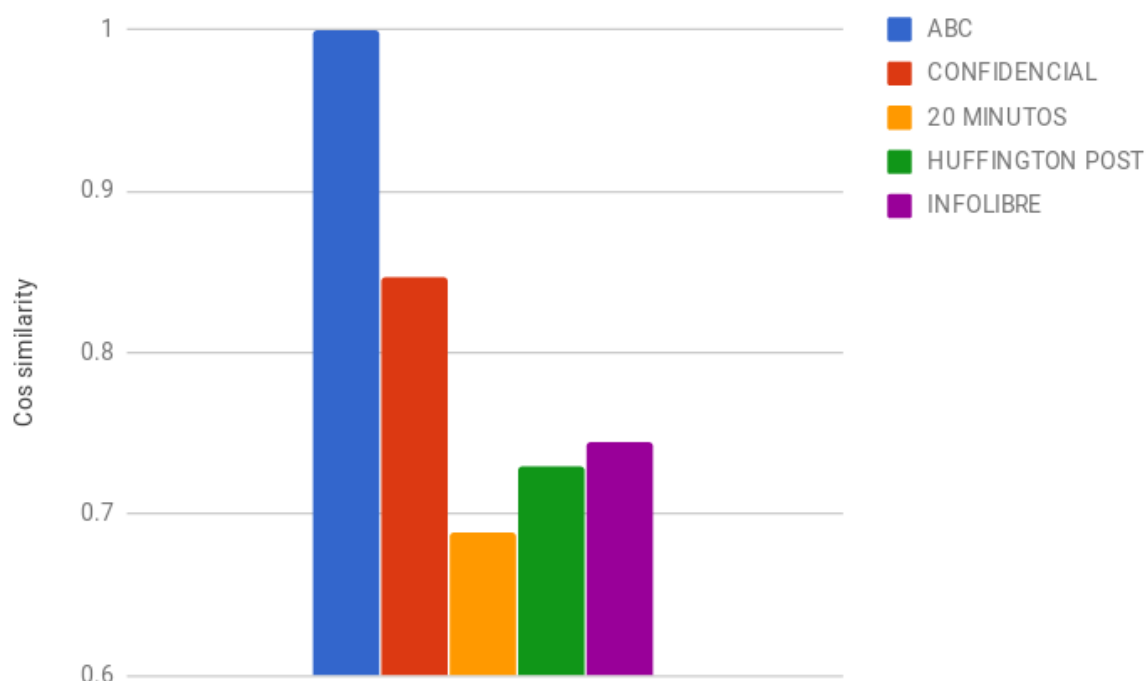
IT	ABC	EL CONFIDENCIAL	20 MINUTOS	HUFFINGTON POST	INFOLIBRE
centralidad	19	17	0	13	0
abismo	8	3	0	0	0
coleta	1	0	0	2	0
populismo	2	1	0	0	1

La mètrica utilitzada per mesurar la co-ocurrència ha estat la Normalized Google Distance (NGD). Adoptem la NGD perquè és una mesura de distància semàntica segons el grau de co-aparició de dos termes, en el nostre cas entre el IT i el seu t , al titular i al cos de la notícia. El seu càlcul no requereix de corpus d'entrenament de gran mida, com és el cas de mètriques de proximitat semàntica com el Word2Vect. La NGD és una mètrica que va del 0 al 1 (de cap proximitat a més proximitat).

Creats els vectors per cada diari, agafem un diari de referència. Tenint en compte, com hem vist abans, que l'ABC està més alineat amb els IT del PP, el prenem com a referència per veure en quina mesura els altres diaris s'hi acosten o s'allunyen.

A la següent gràfica es pot veure la distància dels diaris respecte a l'ABC, quan es calcula la co-ocurrència de IT amb termes t en les piulades del PP.

Figura 10: Proximitat dels diaris respecte al ABC respecte als *frames* del PP.



Com es pot veure, *ElConfidencial* és el diari més proper a l'ABC quan tenim en compte la freqüència d'aparició dels IT amb els tpp. *20Minutos*, *Huffington Post* i *Infolibre* estan més allunyats, essent *20Minutos* el que ho està més.

El fet de que no hàgim trobat un alineament ideològic clar a través dels IT de Podemos, produït en gran part per les moltes referències als termes característics de la dreta, fa que no puguem agafar un diari de referència per mesurar la tendència dels diaris a adoptar el discurs de Podemos. Amb un processament sense intervenció humana és difícil esbrinar que proetarras és un terme característic de la dreta si durant el període de temps en què es van publicar les piulades, aquest terme ha estat un terme molt més discutit pels partits d'esquerra que pels de dreta. Per aquest motiu, com s'ha dit a l'apartat de metodologia, la categorització no és estàtica i caldria anar repetint la classificació d'IT periòdicament fins arribar a una classificació més estable al cap d'un temps.

Ara bé, els vectors de co-ocurrència en els diaris, explicats abans, poden ser útils per trobar falsos IT d'esquerres. Per exemple, aquest és el vector corresponent a proetarras.

IT	ABC	EL CONFIDENCIAL	20 MINUTOS	HUFFINGTON POST	INFOLIBRE
proetarras	3	0	0	0	0

Proetarras té 3 termes t amb els que coapareix en un diari; de fet en un sol diari, que és l'ABC. Tenint en compte que aquests t són Otegui, Bildu i ETA, s'haurà de pensar si la co-aparició d'un IT amb uns t determinats en un diari ja alineat ideològicament (com ja hem fet amb l'ABC), és un criteri per classificar un IT nou, malgrat ser molt utilitzat pel partit ideològicament contrari.

5. CONCLUSIONS.

Hem justificat la necessitat d'aproximar-nos a la mesura del biaix dels mitjans digitals escrits espanyols per un seguit de raons. En primer lloc, perquè necessitem conèixer el seu biaix per a poder saber quin és l'abast del seu efecte (persuasió). En segon lloc, perquè necessitem saber quina és la ideologia que professa cada mitjà per avaluar la pluralitat del panorama informatiu. En tercer i darrer lloc, per a poder millorar el grau d'alfabetització digital de la ciutadania -en facilitar-los aquesta informació-, cosa que alhora hauria de repercutir positivament sobre la qualitat democràtica del sistema polític.

La revisió de la literatura existent sobre mesura del biaix ideològic dels mitjans ha revelat que els distints mitjans emprats fins l'actualitat tenen un seguit de limitacions. Atribuir a cada mitjà la ideologia de la seva audiència, per exemple, assumeix que els ciutadans coneixen el biaix dels mitjans i s'hi exposen selectivament; però ni el primer supòsit ni el segon són sempre veritat. La segona aproximació utilitza el contingut publicat, seguint tres possibles variants. La primera és limitar-se a una petita quantitat de text molt indicativa del contingut (editorials), la segona consisteix en detectar automàticament patrons lingüístics i la darrera en combinar aquests procediments automatitzats amb codificació humana. Estudis previs suggereixen que la primera estratègia pot presentar una ideologia més extremista de la que té realment el mitjà, i que la darrera és força costosa en termes de recursos. Hem adoptat, doncs, la segona estratègia. A més, hi hem afegit dues novetats rellevants. Primer, la nostra unitat d'anàlisi no és una llista (més o menys amplia) de paraules o frases carregades ideològicament sinó un conjunt de sintagmes nominals connectats i carregats ideològicament. Segon, la mesura utilitzada per assignar una ideologia no s'estableix únicament a partir de la freqüència d'ús d'aquestes cadenes de paraules sinó sobretot, en la discrepància entre elles.

Un altre aspecte que distingeix la nostra recerca té a veure no amb el cos de text analitzat relatiu als mitjans sinó amb el cos de text que utilitzem de referent per a identificar ideologies o, més concretament, frames ideològics. Com hem vist, utilitzem piulades de líders polítics en comptes de programes electorals o discursos parlamentaris. Val a dir que després de diferents proves exploratòries, s'han desestimat els discursos parlamentaris perquè no era possible construir un corpus de text suficientment gran per extreure'n termes o conjunts de termes carregats ideològicament. En aquest sentit, s'ha optat també per prescindir dels programes electorals perquè en les anàlisis preliminars realitzades, primer, no s'han detectat discrepàncies significatives en el *frames* dels diferents partits a partir dels programes electorals. Segon, els programes electorals (i les codificacions proposades pel Projecte Party Manifesto que parteixen d'aquests mateixos programes i que hem explorat en aquest estudi) ja no es fan servir per estimar les posicions ideològiques dels partits (Benoit i Laver, 2006, 2007). Per últim, els programes dels partits utilitzen un llenguatge molt formal que s'allunya del llenguatge

més informal i amb càrrega ideològica que sí que es fa servir als titulars i als corpus dels mitjans de notícies com a estratègia de diferenciació davant l'augment de la oferta i la conseqüent competència.

Així doncs, hem centrat els nostres esforços en identificar “frames” (conjunts de paraules semànticament properes al voltant d'un terme IT) propis dels dos partits d'àmbit estatal més polaritzats d'acord amb les percepcions de l'opinió pública espanyola: el PP i Podemos. Mitjançant l'algoritme de Word2Vec hem detectat un seguit de termes comuns a tots dos partits però més presents en les piulades dels diputats d'un partit que de l'altre. Hem verificat que els termes t de què s'acompanyen siguin força distints abans d'identificar frames, una diferència que hem detectat amb la cosine distance per a cada parell de vectors (PP i Podemos) generat pels mateixos ITs.

Durant aquest procés hem topat amb distintes vies mortes o camins poc productius. Com ja hem esmentat, utilitzar els programes dels partits com a corpus de referència per a identificar discursos ideologitzats va ser un dels primers. Però n'hi ha hagut més. Comptar les correspondències dels *frames* de cada partit als diaris ha estat un altre, degut, probablement, a l'apropiació, sobretot de Podemos, de *frames* crítics amb ells sorgits a la dreta. De manera similar, comparar la distància entre frames de partits i de diaris ens duu al mateix punt: els resultats semblen tenir sentit si atenem només als frames propis del PP i a les similituds entre aquest partit i els mitjans, però el mateix no s'aplica a Podemos.

Propers desenvolupaments d'aquesta recerca hauran d'intentar solucionar el problema de les referències iròniques als marcs d'interpretació de l'adversari. Aquest ha estat apuntat com un dels principals problemes de la metodologia adoptada en aquest estudi – l'anàlisi del contingut a través de l'aprenentatge automatitzat – per atribuir una ideologia als mitjans (Barbera & Good, 2014) i nosaltres ho podem verificar. El principal problema l'hem trobat amb l'apropiació de Podemos de les crítiques sorgides des de la dreta a les seves actituds i argumentacions suposadament “populistes”. Els líders de Podemos en fan befa tan sovint, i utilitzant els mateixos termes que el propi PP, que és impossible assignar per procediments d'anàlisi quantitatiu de text aquest terme i els *frames* que ens sorgeixen al PP (i independentment de les tècniques utilitzades per a decidir la “proximitat” dels termes al discurs, ja sigui Word2Vec o la Normalized Google Distance). A més de pensar sistemes per a detectar la ironia, es podria incloure una dimensió temporal, per a donar més pes als termes que apareguin primer en el temps com a elements identificadors del *frame* d'un partit. Una altra aproximació que podria diluir aquest error de mesura és ampliar el corpus de referència a la resta de partits d'àmbit estatal; amb què aquest fenomen que sembla típic de Podemos quedaria diluït entre les piulades del bloc d'esquerres, de què formaria també part el PP. Tinguem en compte, a més, que Podemos és un agent força recent a la política espanyola. Per bé que una part dels seus diputats provinguin d'un partit de llarga tradició (Izquierda Unida), tant aquest partit tradicional com la coalició resultant són o han estat en el millor dels casos tercera força parlamentària. Té sentit, doncs, que els mitjans

reproduceixin preferentment el discurs dels dos principals partits que sintetitzarien les pugnes entre l'esquerra i la dreta.

Una altra solució en la mateixa línia (modificar/ampliar el corpus de referència) tindria a veure amb utilitzar textos no tan estereotipats com els programes (susceptibles, com hem vist, de tractar termes poc conflictius i de convergir en els interessos del votant mitjà) . Una possibilitat podrien ser entrevistes a líders polítics i/o els discursos parlamentaris recollits al diari de sessions del Congrés dels Diputats.

Altres vies no explorades en aquesta recerca podrien limitar-se a com cadascun dels principals agents polítics veu l'adversari. És a dir, com els diputats de dretes parlen de les esquerres o dels partits d'esquerres, i com els diputats d'esquerres parlen de la dreta i dels partits de dreta. Tindríem doncs dos *frames* fonamentals (en comptes dels gairebé 500 amb què hem treballat en aquesta recerca) que miraríem de veure fins a quin punt es reproduceixen en els textos dels diaris digitals.

Finalment, una darrera via podria ser abandonar, al menys parcialment, l'estratègia utilitzada per adoptar la de combinar l'aprenentatge automatitzat amb la codificació humana. Aquesta estratègia tot i ser més costosa, com ja hem apuntat, ens permetria destriar el que és ironia i sarcasme en l'ús del llenguatge del que no i poder assignar amb més garanties que una estratègia completament automatitzada una ideologia als mitjans.

Bibliografia

Aldrich, J. H. (1995). *Why parties?: The origin and transformation of political parties in America*. University of Chicago Press.

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. National Bureau of Economic Research. Recuperado a partir de <http://www.nber.org/papers/w23089>

APM. (2015). *Informe Anual de la Profesion Periodística*. Madrid.

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130–1132.

Barberá, P., & Sood, G. (2014). *Follow Your Ideology: A Measure of Ideological Location of Media Sources*. Unpublished manuscript.

Bawn, K., Cohen, M., Karol, D., Masket, S., Noel, H., & Zaller, J. (2012). A theory of political parties: Groups, policy demands and nominations in American politics. *Perspectives on Politics*, 10(3), 571–597.

Benabou, R. (2008). *IDEOLOGY WORKING PAPER 13907. NBER WORKING PAPER SERIES*, 13907.

Bennett, W. L., & Iyengar, S. (2008). A new era of minimal effects? The changing foundations of political communication. *Journal of Communication*, 58(4), 707-731.

Benoit, K., & Laver, M. (2006). *Party policy in modern democracies*. Routledge.

Benoit, K., & Laver, M. (2007). Estimating party policy positions: Comparing expert surveys and hand-coded content analysis. *Electoral Studies*, 26(1), 90–107.

Buckingham, D. (2007). Media education goes digital: an introduction. *Learning, Media and Technology*, 32(2), 111-119.

Budak, C., Goel, S., & Rao, J. M. (2016). Fair and balanced? quantifying media bias through crowdsourced content analysis. *Public Opinion Quarterly*, 80(S1), 250–271.

Carlson, M. (2017). *Journalistic Authority: Legitimizing News in the Digital Era*. Columbia University Press.

Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal of Communication*, 64(2), 317–332.

Conover, M., Ratkiewicz, J., Francisco, M. R., Gonçalves, B., Menczer, F., & Flammini, A. (2011). Political Polarization on Twitter. *ICWSM*, 133, 89–96.

Converse, P. E. (1964). *The nature of mass opinion beliefs. Ideology and Discontent*. New York, NY: The Free Press of Glencoe.

- Downs, A. (s. f.). 1957 *An economic theory of democracy*. New York: Harper and Row.
- Enelow, J. M., & Hinich, M. J. (1982). Ideology, issues, and the spatial theory of elections. *American Political Science Review*, 76(3), 493-501.
- Fabos, B. (2004). *Wrong Turn on the Information Superhighway: Education and the Commercialization of the Internet*. New York: Teachers College Press.
- Festinger, L. (1962). *A theory of cognitive dissonance* (Vol. 2). Stanford university press.
- Freedman, J. L., & Sears, D. O. (1966). II SELECTIVE EXPOSURE1. *ADV EXPERIMENTAL SOCIAL PSYCHOLOGY*, 2, 57.
- Gentzkow, M., & Shapiro, J. M. (2010). What drives media slant? Evidence from US daily newspapers. *Econometrica*, 78(1), 35–71.
- Gentzkow, M., & Shapiro, J. M. (2011). Ideological Segregation Online and Offline. *The Quarterly Journal of Economics*, 126(4), 1799–1839.
<http://doi.org/10.1093/qje/qjr044>
- Gilster, P., & Glister, P. (1997). *Digital literacy*. New York: Wiley Computer Pub..
- Groeling, T. (2013). Media bias by the numbers: Challenges and opportunities in the empirical study of partisan news. *Annual Review of Political Science*, 16.
- Groseclose, T., & Milyo, J. (2005). A measure of media bias. *The Quarterly Journal of Economics*, 120(4), 1191–1237.
- Hallin, D. C., & Mancini, P. (2004). *Comparing media systems: Three models of media and politics*. Cambridge University Press.
- Hallin, D. C., & Mancini, P. (2012). *Comparing media systems beyond the Western world*. Cambridge University Press.
- Ho, D. E., & Quinn, K. M. (2008). Measuring explicit political positions of media. *Quarterly Journal of Political Science*, 3(4), 353-377.
- Iyengar, S., & Hahn, K. S. (2009). Red media, blue media: Evidence of ideological selectivity in media use. *Journal of Communication*, 59(1), 19–39.
- Jones, D. E. (1998). Investigación sobre comunicación en España: evolución y perspectivas. *Zer: Revista de estudios de comunicación= Komunikazio ikasketen aldizkaria*, (5), 1.
- Kalogeropoulos, A., & Newman, N. (2017). “I saw the news on Facebook”. Brand attribution when accessing news from distributed environments. Oxford, UK. Retrieved from [http://reutersinstitute.politics.ox.ac.uk/sites/default/files/Brand attributions report.pdf](http://reutersinstitute.politics.ox.ac.uk/sites/default/files/Brand%20attributions%20report.pdf)

- Lazarsfeld, P. F., Berelson, B., & Gaudet, H. (1944). *The people's choice: How the voter makes up his mind in a presidential election*. New York: Duell, Sloan and Pearce.
- Majó-Vázquez, S., Nurse, J. R., Simon, F. M., & Nielsen, R. K. (2017). Digital-Born and Legacy News Media on Twitter during the German Federal Election.
- Minder, R. (2015, November 5). Spain's News Media Are Squeezed by Government and Debt. *The New York Times*. New York. Retrieved from http://www.nytimes.com/2015/11/06/world/europe/as-spains-media-industry-changes-rapidly-some-worry-about-objectivity.html?_r=0
- Monroe, B. L., Colaresi, M. P., & Quinn, K. M. (2008). Fightin' words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis*, 16(4), 372–403.
- Mullainathan, S., & Shleifer, A. (2005). The market for news. *American Economic Review*, 1031–1053.
- Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D. A., & Nielsen, R. K. (2017). *Digital News Report 2017*. Oxford, UK. Retrieved from <http://www.digitalnewsreport.org/>
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, 2(2), 175.
- Nielsen, R. K., & Graves, L. (2017). “News you don't believe”: Audience perspectives on fake news. Oxford, UK.
- Olson, J. M., & Stone, J. (2014). The Influence of Behavior. *The handbook of attitudes*, 223.
- Pineda, A., & Almiron, N. (2013). Ideology, Politics, and Opinion Journalism: A Content Analysis of Spanish Online-Only Newspapers. *tripleC: Communication, Capitalism & Critique*. Open Access Journal for a Global Sustainable Information Society, 11(2), 558-574.
- Prior, M. (2009). The Immensely Inflated News Audience: Assessing Bias in Self-Reported News Exposure. *Public Opinion Quarterly*, 73(1), 130–143.
- Puglisi, R., & Snyder Jr, J. M. (2011). Newspaper coverage of political scandals. *The Journal of Politics*, 73(3), 931-950.
- Schoepp, V. S. (2016, March 1). Reporter im Wohnzimmer. *Süddeutsche Zeitung*. Munich. Retrieved from <http://www.sueddeutsche.de/medien/neue-medien-reporter-im-wohnzimmer-1.2886624?reduced=true>
- Sears, D. O., & Freedman, J. L. (1967). Selective exposure to information: A critical review. *Public Opinion Quarterly*, 31(2), 194–213.

Stokes, D. E. (1963). Spatial models of party competition. *American political science review*, 57(2), 368-377.

Stroud, N. J. (2011). *Niche news: The politics of news choice*. Oxford University Press on Demand.

Toff, B. J., & Kim, Y. M. (2013). Words That Matter: Twitter and Partisan Polarization. En UWMadison's Political Behavior Research Group meeting, November (Vol. 13, p. 2013).

Trillo-Domínguez, M. (2008). Análisis cibernético de la prensa digital española. Ranking de calidad web y mapa de influencia mediática.

Wihbey, J., Coleman, T. D., Joseph, K., & Lazer, D. (2017). Exploring the Ideological Nature of Journalists' Social Networks on Twitter and Associations with News Story Content. arXiv preprint arXiv:1708.06727.