TREBALL DE FI DE GRAU

Doble grau en Economia i Estadística

Títol: Anàlisi de patrons de parla i estratègies comunicatives dels mitjans segons la seva ideologia política

Autor: Carlota Castellano Escuder

Directora: Laura Marquès

Data: 27-06-2022

Departament: Econometria, Estadística i Economia

Convocatòria: Primera

UNIVERSITAT DE BARCELONA

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat de Matemàtiques i Estadística

RESUM

Anàlisi dels patrons de parla i expressions de les diferents ideologies polítiques en les que s'identifiquen els mitjans nacionals. S'examina si hi ha diferències de parla, expressions o conceptes entre els mitjans nacionals, separats segons la seva ideologia política, al explicar un mateix fet internacional, com és la guerra entre Ucraïna i Rússia.

Altrament, saber quin és l'impacte econòmic que ha causat aquesta guerra. S'estudia aquest efecte a través d'indicadors econòmics concrets i a través de les notícies publicades per els tres grans diaris econòmics espanyols: *El Economista, Cinco Días* i *Expansión*.

ABSTRACT

This work analyzes the speech patterns and expressions of the different political ideologies in which the Spanish national media identify themselves. This work has been examined whether there are differences in speech, expressions or concepts between the national media when explaining the same international event, in this case, the war between Ukraine and Russia.

On the other hand, this work analyzes the economic impact of this war and the way the media talk about it. This effect is analyzed through specific economic indicators and the news published by the three major Spanish economic newspapers*: El Economista, Cinco Días,* and *Expansión.*

PARAULES CLAU

- Valor econòmic: Total del valor econòmic publicitari (VEP) o valor publicitari equivalent (VPE) associat a les notícies.

- Número de notícies: Total de notícies diferents que fan referència a un mitjà.

- Freqüència absoluta d'un concepte: Número de vegades que apareix una paraula o expressió en la mostra de notícies analitzades.

- Audiència: número de persones que llegeix una notícia.

- *Data frame*: estructura de dades de dues dimensions

KEY WORDS

- Economic value: Total economic advertising value (VEP) or equivalent advertising value (VPE) associated with the news.

- Number of news items: Total number of different news items referring to a medium.

- Absolute frequency of a concept: Number of times a word or expression appears in the sample of news items analyzed.

- Audience: Number of people who read a news item.

- Data frame: two-dimensional data structure.

CLASSIFICACIÓ AMS

Per falta de classificació específica de mineria de text, es classifica a: *62-07 Data analysis*

ÍNDEX

LLISTA DE FIGURES

LLISTA DE TAULES

I. INTRODUCCIÓ

El dia 24 de febrer de 2022 tots els mitjans de comunicació van fer ressò de la invasió russa al seu país veí. El president de Rússia, Vladimir Putin, va ordenar a les seves tropes bombardejar i envair Ucraïna, en la que està sent la primera gran agressió d'aquest tipus a Europa des de la Segona Guerra Mundial el 1945 i la fi de la dictadura dels nazis a Alemanya.

Durant bona part del segle XX, Rússia formava part de la Unió de Repúbliques Socialistes Soviètiques (URSS), estat que va néixer després de la Revolució del 1917, quan l'imperi dels tsars de Rússia es va desfer i es va crear un nombre important de països com Polònia, Estònia, Letònia o Lituània, Ucraïna, Geòrgia i evidentment, Rússia.

Rússia es va convertir en el primer estat socialista del món que un cop consolidat va començar a atacar als països que la rodejaven, va envair i recuperar molts dels territoris que havia perdut i els va incorporar a l'URSS, un d'ells va ser Ucraïna. Rússia es va convertir en un imperi i quasi tot Ucraïna va quedar en les seves mans, la cultura i idioma d'Ucraïna van ser menyspreats i perseguits durant molts anys.

Durant la Segona Guerra Mundial, els nazis van envair l'URSS i van provocar una gran destrucció a Ucraïna, però amb el temps es va recuperar i va passar a ser un dels territoris més importants de l'URSS, considerada un dels centres agraris i industrials. El col·lapse de l'URSS el 1991 va provocar que els països membres s'independitzessin i en aquell moment, Ucraïna va votar en un referèndum per crear un estat propi.

Des del 2014, Ucraïna es debat entre el seu desig de pertànyer a Europa i el seu històric i cultural vincle amb Rússia. El govern ha mostrat intencions d'unir-se a l'aliança militar occidental, l'OTAN i sumar-se a la Unió Europea, idees inacceptables per Rússia que conceben el país com a part de la seva identitat i el seu espai d'influència.

L'any 2014 manifestants ucraïnesos van prendre els carrers massivament i van acabar derrotant i expulsant el, fins aleshores president prorús Víktor Yanukovich. Crisi política que Putin va aprofitar per envair la península ucraïnesa de Crimea i recolzar a separatistes prorussos en el Donbass, regió d'Ucraïna que fa frontera amb Rússia.

Entre el 2014 i el 2015, Rússia i Ucraïna van firmar acords d'alto el foc amb l'ajuda de França i Alemanya, anomenats Protocols de Minsk, els quals van acabar quan el passat 24 de febrer de 2022 Putin va llençar la invasió en diversos punts de la frontera.

A partir d'aquest moment, la majoria de les notícies dels mitjans de comunicació estan focalitzades en la guerra esdevinguda entre Ucraïna i Rússia.

Els mitjans de comunicació són els encarregats d'explicar i informar a la població sobre tot el que passa al seu voltant de manera objectiva, és a dir, sense donar l'opinió. Així i tot, és difícil que un mitjà no rebel·li mai la seva opinió sobre un tema, la seva ideologia política o les seves preferències.

En aquest treball, ens preguntem si depenent de la ideologia política, els mitjans detallen el mateix fet internacional, com és la guerra entre Ucraïna i Rússia, d'una manera diferent o deixen a banda les seves creences per parlar de manera totalment objectiva.

La ideologia dels mitjans nacionals és pública, però així i tot, tots defensen la seva imparcialitat en el contingut de les seves notícies.

L'objectiu principal és l'anàlisi dels patrons de parla i expressions de les diferents ideologies polítiques en les quals s'identifiquen els mitjans nacionals espanyols. És a dir, observar quines diferències de parla, expressions o paraules hi ha entre els mitjans nacionals, separats per la seva ideologia política, a l'explicar un mateix fet internacional, com és la guerra entra Ucraïna i Rússia. Es tracta de trobar tendències i semblances entre la tipologia d'expressions dels mitjans nacionals espanyols.

Aquesta anàlisi de patrons es basa en dues hipòtesis:

- Hipòtesis 1: Els mitjans de la mateixa ideologia utilitzen el mateix vocabulari, llenguatge i expressions per expressar una mateixa idea.

- Hipòtesis 2: Les ideologies d'extrems fan servir vocabulari, llenguatge i expressions diferents per expressar la mateixa idea.

Un altre objectiu és analitzar, també a través de notícies, quin és l'impacte econòmic que ha causat aquesta guerra. Estudiar aquest efecte a través d'indicadors econòmics concrets i a través del contingut de les publicacions de tres dels grans diaris econòmics espanyols: *El Economista*, *Cinco Días* i *Expansión*.

El programa usat per a tot el treball és l'R, la lectura de les notícies, la neteja del text, l'anàlisi i els resultats s'han obtingut completament a partir d'aquest.

Pel que fa a les parts del treball, en primer lloc es troba la metodologia, en la qual s'expliquen les tècniques estadístiques emprades, la procedència de les dades i els recursos informàtics.

En segon lloc es desenvolupa el cos del treball, el qual està dividit en dues parts, corresponents als dos objectius del treball:

- La primera part exposa l'anàlisi de patrons de parla i expressions dels mitjans i de les ideologies, la secció queda dividia en dues parts, una per l'anàlisi exploratòria de les dades i l'altre per l'anàlisi semàntica.
- La segona part exposa l'anàlisi de l'impacte econòmic de la guerra. Per acabar, es duu a terme una discussió dels resultats obtinguts i es detallen les conclusions del treball.

II. METODOLOGIA

Per tal de contrastar com les diferents ideologies polítiques cobreixen el mateix fet internacional, és a dir, com enfronten els mitjans espanyols el mateix conflicte mundial, s'han escollit 7 mitjans de premsa.

Si bé és cert, els mitjans es classifiquen segons la ideologia política a la qual són més fidels, però no existeixen el mateix nombre de mitjans per a cada ideologia. Existeixen 6 mitjans nacionals espanyols de premsa impresa amb les mateixes característiques d'audiència i difusió, o si més no, comparables entre si, comparteixen idioma, territori i cobertura.

En els mitjans escollits amb aquestes característiques trobem: *El Periódico* d'esquerra, de centre *La Vanguardia* i *El País* i de dreta *El Mundo*, *ABC* i *La Razón*.

Per tal de no focalitzar l'opinió de la ideologia d'esquerra en un sol diari, s'ha escollit un diari espanyol digital que ha publicat un nombre de notícies similar a la resta en el període escollit, aquest ha estat el diari *Público.es*.

D'aquesta manera, l'anàlisi es farà sobre 7 mitjans, classificats per ideologies, 6 d'ells nacionals espanyols de premsa impresa i un espanyol digital. La classificació queda de la següent manera:

ESQUERRA	**CENTRE**	**DRETA**
El Periódico (944)	El País (957)	El Mundo (1080)
Público.es (969)	La Vanguardia (1004)	La Razón (907)
		ABC (939)

Taula 2.1: Mitjans analitzats en l'anàlisi de patrons

L'anàlisi de patrons ha estat dividit en dos grans blocs. Per una banda, s'ha fet una anàlisi exploratòria de les dades obtingudes per tal de comparar a través d'estadístics bàsics les principals diferències d'audiència i valor publicitari equivalent entre els mitjans nacionals. Per altra banda, s'ha fet l'anàlisi semàntica de les notícies.

Pel que fa a l'anàlisi de l'impacte econòmic en el primer mes de guerra, s'han escollit tres mitjans econòmics espanyols a partir dels quals s'ha fet l'anàlisi semàntica dels conceptes econòmics i s'han explicat els efectes de la guerra:

Expansión	Cinco Días	El Economista

Taula 2.2: Mitjans analitzats en l'anàlisi econòmic

S'expliquen les tècniques estadístiques, la procedència de les dades, els programes utilitzats i alguns dels paquets.

Per obtenir les dades s'ha usat l'hemeroteca[1] de *Mynews: MyNews S.L.* es tracta d'una empresa que va néixer el 1995 per desenvolupar el projecte *MyNews Online*, el primer diari electrònic personalitzat creat a Europa.

Actualment, *MyNews* es dedica al *press clipping*[2], la gestió del coneixement i l'anàlisi de dades. Cobreix les necessitats dels professionals de la informació relacionades amb l'obtenció i explotació dels continguts publicats a premsa. Permet als seus clients monitoritzar tot el que els interessi i crear tants seguiments com vulguin, els prepara dossiers personalitzats i obté les dades d'audiència i valor publicitari equivalent de les seves notícies.

MyNews té una cobertura de més de 500 fonts de premsa impresa, 200.000 fonts en línia, 50 canals i emissores de televisió i ràdio i monitorització de les principals xarxes socials per cobrir qualsevol necessitat.

L'empresa disposa de l'única hemeroteca digital de premsa impresa moderna d'Espanya que permet recuperar qualsevol notícia publicada en premsa des de 1996.

La cerca executada ha estat realitzada deu vegades, una vegada per mitjà, en cadascun d'ells s'ha seleccionat el període que abasta des del 24 de febrer del 2022 fins al 24 de març del 2022. El període escollit fa referència a un mes exacte a partir del dia en què Putin va anunciar que Rússia estava iniciant una operació militar especial en el Donbass i va llençar una invasió a gran escala a Ucraïna.

Totes les notícies estudiades estan escrites en castellà degut a pertànyer a diaris nacionals espanyols, és per això que tant els resultats de l'anàlisi com la cerca de les notícies han estat elaborades en l'idioma castellà.

[1] Hemeroteca: dipòsit digital de notícies des del 1995, amb cercador específic i aplicador de filtres per a la cerca.

[2] *Press clipping*: activitat que selecciona articles o retalls de premsa en els que una empresa determinada apareix en forma de notícia

Per tal d'escollir exclusivament aquelles notícies que feien referència a la guerra entre Ucraïna i Rússia, en la cerca s'ha obligat l'existència de certes paraules, una del grup: *guerra*, *crisis*, *invasión*, o *conflicto* i una del grup: *Ucraina*, *Ucrania*, *Ucraïna*, *Rússia* o *Rusia*. S'ha fet d'aquesta manera, ja que, si s'hagués fet només per a algunes de les paraules clau hagueren entrat notícies no pertinents al fet analitzat, en canvi, si s'obliga la contenció d'una paraula de cada grup, s'assegura la veracitat de les notícies que s'escolliran.

Les dades de cada mitjà s'obtenen en format Excel (xls) i en format XML, arxius a partir dels quals s'executarà tota l'anàlisi a través d'R. Aquests arxius ens els proporciona la pròpia hemeroteca de *Mynews* directament, des de la seva plataforma és des d'on es fa la cerca i es descarreguen les notícies amb el format desitjat.

El treball es basa en la mineria de text, la qual es pot entendre com a l'anàlisi del text que té com a objectiu extreure informació a partir d'aquest, però sense la informació que ell mateix rebel·la. S'obté sempre informació nova a partir de grans quantitats de text.

El desenvolupament del treball ha estat dut a terme completament amb R amb paquets que permeten llegir, netejar i analitzar text com ara *dplyr*, *NLP*, *tm* i *stringi*.

Les tècniques de visualització de dades es basen en diagrames de barres, núvols de paraules, taules de freqüència, diagrames de correlació, un diagrama de Venn i un PCA. Totes les taules i gràfics del treball són d'elaboració pròpia, igual com el codi, el qual es troba completament a l'annex.

III. ANÀLISI DE PATRONS DE PARLA I EXPRESSIONS

1. Anàlisi exploratòria de les dades

Un cop obtingudes les dades, es desenvolupa l'anàlisi exploratòria d'aquestes per tal d'entendre, conèixer i familiaritzar-se amb les dades què es treballarà. Es disposa d'una base de dades per a cada mitjà, totes elles amb el mateix format i mateixes variables.

En el cas d'aquesta anàlisi, moltes de les variables no tenen nivells, ja que, s'unifiquen en un sol tipus, és a dir, el nom de la publicació, per exemple, sempre serà el mateix, ja que, a cada base de dades hi haurà el nom del seu mitjà.

Les variables que trobem a la base de dades són les següents:

- *Publicación:* variable qualitativa que indica el nom del mitjà que ha fet la publicació.
- *Edición:* variable qualitativa que indica l'edició en què s'ha publicat la notícia.
- *Sección:* variable qualitativa que marca la secció del diari en què surt la notícia.
- *Página:* variable numèrica que indica la pàgina de la notícia.
- *Fecha:* variable numèrica que indica la data de publicació.
- *Título:* variable qualitativa que indica el títol de la notícia.
- *Subtítulo:* variable qualitativa que indica el subtítol de la notícia.
- *Audiencia:* variable numèrica que mostra el nombre de persones que llegeix aquesta publicació.
- *Difusión:* variable numèrica que quantifica el nombre de persones al que arriba aquesta publicació.
- *Valor:* variable numèrica que representa el valor publicitari equivalent de la notícia, és a dir, el cost d'aquesta.
- *IdDocument:* variable qualitativa que dona un identificador a cada document.
- *Palabra Clave:* variable qualitativa que indica les paraules exactes per les quals la notícia ha estat seleccionada.
- *Género Publicación:* variable qualitativa que indica en gènere de publicació de la notícia.
- *Cobertura:* variable qualitativa que mostra la cobertura que té la publicació, en aquest cas, sempre nacional.
- *Territorio:* variable qualitativa que mostra el territori en què s'ha publicat la notícia, en aquest cas, sempre nacional.

- *Tipo:* variable qualitativa que expressa el tipus de premsa de la notícia, en aquest cas, sempre premsa impresa.
- *Contenido:* variable qualitativa on s'emmagatzema el contingut de la notícia.

En l'anàlisi exploratòria de les dades únicament s'analitza el comportament de les variables audiència, valor publicitari equivalent i la data de publicació.

L'audiència i el valor publicitari es defineixen per l'EGM (Estudi General de Mitjans), L'OJD (Oficina de la Justificació de la Difusió) i el SimilarWeb, proporcionen serveis d'analítica web, mineria de dades i intel·ligència empresarial.

- EGM és un estudi sobre el consum dels mitjans de comunicació a Espanya realitzat per l'Associació per a la Investigació de Mitjans de comunicació (AIMC), aporta informació sobre el nombre de lectors de cada mitjà i, per tant, sobre l'audiència.
- L'OJD s'encarrega del control de tiratge i difusió de diversos tipus de mitjans de comunicació a Espanya, revistes, diaris i mitjans de comunicació per Internet, aporta informació sobre els exemplars publicats.
- SimilarWeb utilitza tecnologies de *big data* per a recollir, mesurar, analitzar i proporcionar estadístiques d'interacció d'usuaris a pàgines web i aplicacions mòbils, els valors acostumen a variar cada mes.

L'audiència depèn de la repercussió de cada mitjà en les seves notícies i el valor, del cost d'aquella publicació, o tarifa publicitària, definit per l'espai que ocupa cada notícia en el paper o pel tipus de mitjà. La revisió d'aquest valor es du a terme mensualment, de manera que en alguns mitjans no hi ha variacions en el transcurs d'un mes.

Per a cada mitjà es calcula el número de notícies, la mitjana, la mediana, la desviació típica, la variància, els quartils i el màxim i el mínim del valor publicitari equivalent i de l'audiència. Aquesta anàlisi es divideix per ideologies de mitjans i es compara entre aquestes. Per començar, s'exploren els mitjans d'ideologia d'esquerra.

Valor publicitari equivalent

Publicacion	n	Mean	Sd	Var	Q1	Q2	Q3	Min	Max
El Periódico	944	5896	4425.754	19587298	2930.5	5104.5	7789.75	17	34881
publico.es	968	6796	0.000	0	6796.0	6796.0	6796.00	6796	6796

Taula 3.1: Estadístics bàsics del VPE d'esquerra

Audiència

Publicacion	n	Mean	Sd	Var	Q1	Q2	Q3	Min	Max
El Periódico	944	361704	4917.834	24185092	361000	361000	361000	361000	396000
publico.es	968	5481321	0.000	0	5481321	5481321	5481321	5481321	5481321

Taula 3.2: Estadístics bàsics de l'audiència d'esquerra

Com es pot comprovar, en el mitjà digital *Públic*o.es no és possible calcular els estadístics bàsics, ja que, el seu VPE i la seva audiència no varien segons la seva definició en el període analitzat. S'observa que el valor publicitari de la ideologia d'esquerra es troba al voltant de 6.345€ i l'audiència al voltant de 2.921.512 lectors.

En segon lloc, es procedeix amb els mitjans de centre i s'observa que en el període analitzat, *La Vanguardia* no té variacions d'audiència, com s'ha explicat, és degut a les revisions mensuals d'aquests valors i a la poca variància d'aquests en períodes curts. El VPE d'ideologia central és molt diferent segons els mitjans però sempre superior al d'ideologia d'esquerra. Dels vists fins ara, *La Vanguardia* és el primer mitjà en superar les 1.000 publicacions mensuals.

Valor publicitari equivalent

Publicacion	n	Mean	Sd	Var	Q1	Q2	Q3	Min	Max
El País	957	28210	18685.179	349135926	14627	25362	36220.0	649	119185
La Vanguardia	1004	8306	8889.356	79020643	4238	7010	9657.5	58	87129

Taula 3.3: Estadístics bàsics del VPE de centre

Audiència

Publicacion	n	Mean	Sd	Var	Q1	Q2	Q3	Min	Max
El País	957	911312	12555.3	157635546	913000	913000	913000	818000	913000
La Vanguardia	1004	489000	0.0	0	489000	489000	489000	489000	489000

Taula 3.4: Estadístics bàsics de l'audiència de centre

En últim lloc, els mitjans de dreta. Observem que *ABC* es troba en la mateixa situació que *La Vanguardia* respecte no tenir variacions d'audiència. El VPE mitjà de la ideologia és de 11.493€, molt elevat, però inferior al de centre, tot i comptar amb un mitjà més que podria provocar l'augment del valor. Aquesta diferència es pot veure clarament en la mitjana de les taules.

Valor publicitari equivalent

Publicacion	n	Mean	Sd	Var	Q1	Q2	Q3	Min	Max
ABC	939	11048	8846.134	78254092	4962.50	8137	15492.5	32	50096
El Mundo	1080	14657	11799.541	139229158	6209.75	12278	20132.0	37	76630
La Razón	907	8777	6435.599	41416933	4484.50	7712	11466.5	99	54436

Taula 3.5: Estadístics bàsics del VPE de dreta

Audiència

Publicacion	n	Mean	Sd	Var	Q1	Q2	Q3	Min	Max
ABC	939	430000	0.000	0	430000	430000	430000	430000	430000
El Mundo	1080	681919	58849.652	3463281584	687000	687000	687000	1000	687000
La Razón	907	227417	3366.578	11333847	228000	228000	228000	178000	228000

Taula 3.6: Estadístics bàsics de l'audiència de dreta

Parlant de mitjans individuals, pel que fa a l'audiència, la de *Público.es* és clarament la més elevada degut probablement a ser un mitjà digital, el segueix *El País* amb una mitjana de 911.312 seguidors al mes, el qual també guanya a la resta de mitjans amb mitjana de VPE.

Si fem una representació gràfica amb el nombre de notícies publicat per cada mitjà, podem veure que és força similar entre els 7 mitjans. Tanmateix, *El Mundo* encapçala el rànquing amb 1.080 publicacions i *La Razón* el tanca amb 907. Vegem-ho:

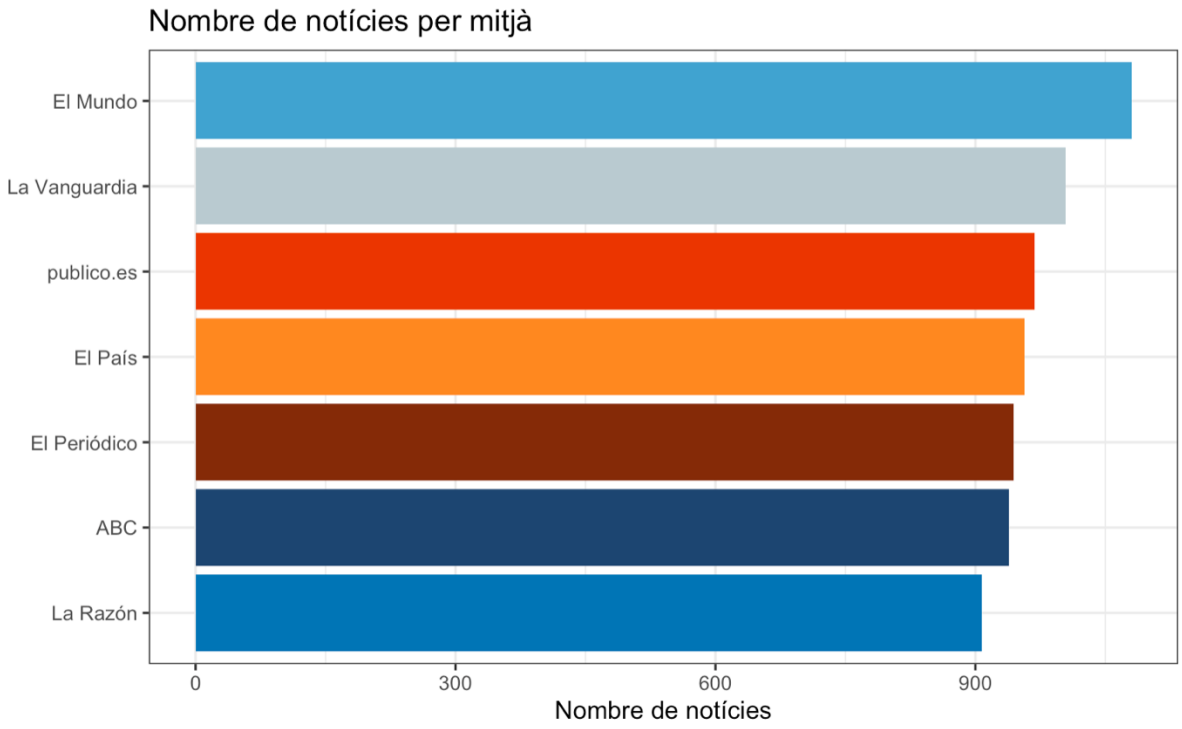

Figura 3.1: Diagrama de barres del nombre de notícies de cada mitjà en el període analitzat

Seguidament, es fa una anàlisi de l'evolució de les notícies cronològicament i dels titulars a nivell diari. Durant el mes analitzat, hi ha pics de publicacions que comparteixen la majoria dels mitjans i que s'expliquen pels fets ocorreguts en la guerra en aquelles dates. Tot i això, el ritme de publicacions segueix una dinàmica bastant constant.

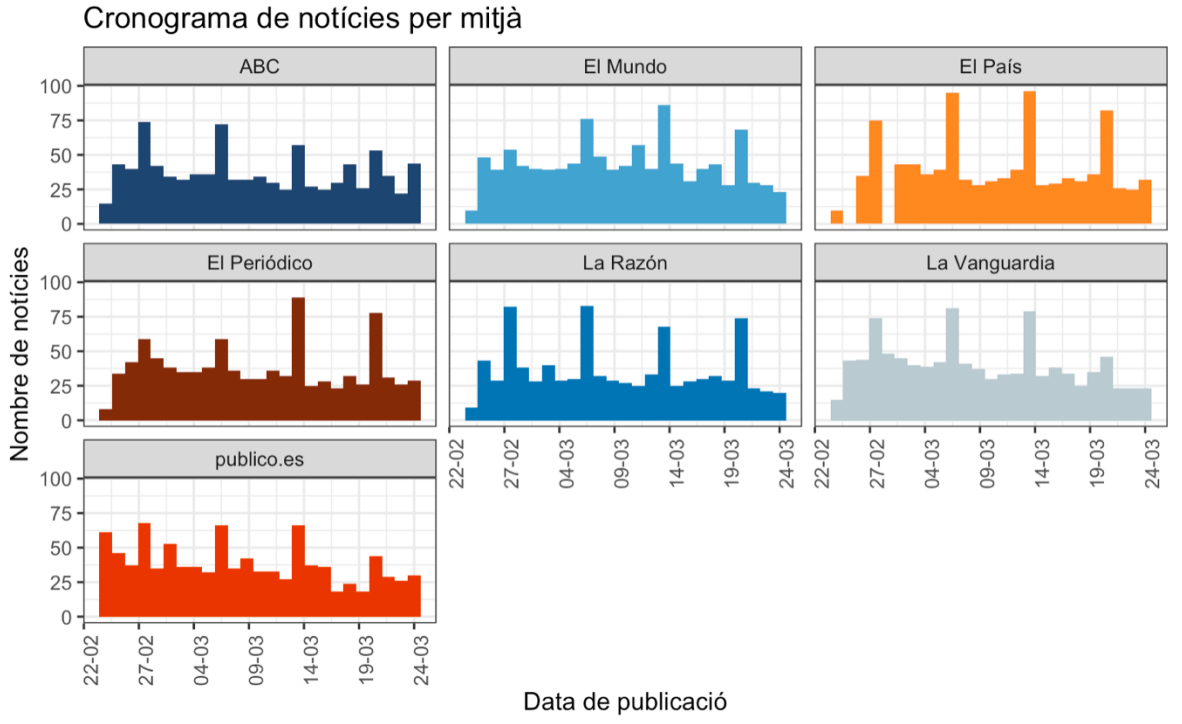

Figura 3.2: Histogrames de la cronologia de notícies per a cada mitjà

El dia 6 de març *La Razón*, *El País* i *La Vanguardia* tenen un repunt de notícies possiblement degut a ser el dia en què les forces russes van bombardejar una intersecció en una carretera que utilitzaven els civils per fugit de l'avenç de l'exèrcit rus en el nord d'Ucraïna cap a Kiev. A més a més, Ucraïna va acusar Rússia d'atacar un reactor nuclear experimental a Járkiv.

L'onze de març les publicacions de *El Mundo,* mitjà espanyol de premsa impresa, augmenten mentre les dels altres mitjans es mantenen. Destaquen titulars com:

- *"Una indiscreción de Borrell impidió la transferencia de cazas de Polonia a Ucrania, según la prensa de EEUU"*
- *"El FMI advierte de bancarrota en Rusia y la "recesión profunda" de su economía por las sanciones*"
- *"Aterrizan en Madrid 25 niños ucranianos con cáncer para seguir su tratamiento en España"*
- "EEUU iguala el régimen comercial de Rusia al de Cuba y Corea del Norte y prohíbe importaciones".

S'observa que la majoria d'ells són de caràcter econòmic.

El dia tretze de març es torna a percebre un repunt genèric de notícies que es pot suposar que és degut a la mort del periodista Brent Renaud al ser tirotejat a Ucraïna o a la destrucció de la base militar de Yavoriv.

L'últim repunt té lloc gairebé al final del període analitzat, el diumenge vint de març per a tots els mitjans menys per *ABC* i *Público.es*, la suposició apunta a ser degut a l'ultimàtum de Rússia a Mariupol per rendir-se en les pròximes hores.

Un fet internacional sempre és capaç de revolucionar la població i els mitjans i si és te en compte que l'anàlisi pertany al primer més d'aquesta guerra es pot dir que els sentiments eren més a flor de pell i qualsevol fet podia remoure les notícies dels mitjans.

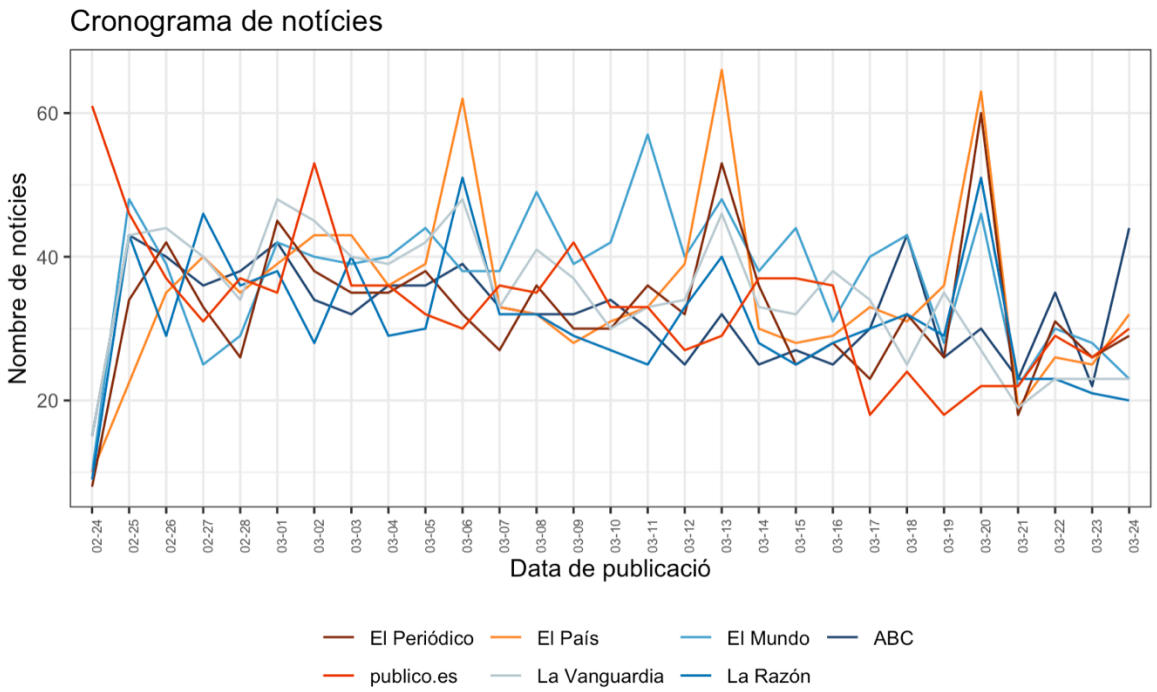

Figura 3.3: Cronograma de l'evolució del nombre de notícies

2. Anàlisi semàntica de les dades

Un cop explorades i conegudes les dades, es procedeix a l'anàlisi de cada una de les notícies quantificades per a cada mitjà. Es tracta d'observar les tendències en el llenguatge dels mitjans.

L'anàlisi semàntica de les dades explica la relació entre un mitjà i els conceptes associats a aquest mitjà. L'objectiu de l'anàlisi semàntic és establir relacions entre 2 variables: mitjans i paraules amb valor semàntic.

En l'anàlisi semàntica de les notícies dels mitjans s'han llegit els conjunts de notícies de cada mitjà en format *XML,* extrets de l'hemeroteca i s'han ajuntat en un sol *data frame*, a partir d'aquí el que s'ha treballat són els textos per tal que fossin analitzables. En aquest punt és on entren les tècniques de mineria de text, que consten de diverses fases.

Primer, s'ha procedit a la neteja del text, s'ha passat tot el text a minúscules, s'han eliminat les pàgines web o enllaços, s'han eliminat els signes de puntuació, els números, els múltiples espais en blanc i els accents. Després, s'han eliminat les *stopwords*, un conjunt de paraules definit pel mateix R, que no tenen significat per elles mateixes sinó que modifiquen o n'acompanyen d'altres, acostumen a ser paraules com articles, pronoms, preposicions, adverbis o alguns verbs. A més a més de l'extracció d'aquest tipus de paraules, també s'ha fet

una tria manual de paraules no útils per a l'anàlisi i s'han extret d'aquest. En últim lloc, s'ha procedit a eliminar les paraules amb 3 o menys lletres, per la suposició de poca rellevància d'aquestes en el contingut de les notícies.

Un cop netejat el text, s'ha procedit a la tokenització d'aquest. Tokenitzar és el procés de separar un text en les unitats que el formen, sent una unitat l'element més senzill amb significat propi, en aquest cas, les paraules. S'ha fet un comptatge de les vegades que es repeteix una mateixa paraula en cada mitjà per obtenir el *data frame* final, el qual consta de 3 columnes:

- Mitjà (nom del mitjà que menciona la paraula)
- Paraula (paraula de la què parlem)
- N (nombre de vegades que el mitjà repeteix la paraula)

Un exemple del procés realitzat seria aquest:

- Partim d'una petita frase de l'anàlisi: "*La guerra de Ucrania ha dejado la pandemia de la covid en un ensayo de pesadilla, pues no hay vacuna a la vista capaz de frenar a Putin."*
- Es passa tot a minúscules, s'eliminen els signes de puntuació, els múltiples estpais en blanc i els accents: "*la guerra de ucrania ha dejado la pandemia de la covid en un ensayo de pesadilla pues no hay vacuna a la vista capaz de frenar a putin"*
- S'eliminen les *stopwords* i altres paraules irrellevants per l'anàlisi*:* "*guerra ucrania pandemia covid pesadilla vacuna frenar putin"*
- Es procedeix a la tokenització*:* "*guerra", "ucrania", "pandemia", "covid", "pesadilla", " vacuna", "frenar", "putin"*

Les paraules que tenen la mateixa rellevància semàntica per aquesta anàlisi, com ara "kyiv" i "kiev" s'han agrupat en una sola paraula, això es pot fer, ja que, la importància de l'anàlisi no està en la semàntica de la paraula, sinó en el contingut de la notícia. També s'han eliminat totes aquelles paraules que s'havien mencionat 15 o menys vegades, ja que, si és té en compte que el període analitzat és de 30 dies, aquestes paraules només corresponien a un quart de les notícies de l'anàlisi i no es podien considerar representatives d'aquest.

Un cop treballats els textos i aconseguida la neteja d'aquests, s'ha executat tota l'anàlisi.

Degut a la neteja del text, les paraules dels resultats de l'anàlisi no contenen accents ni signes de puntuació i, per tant, s'han d'interpretar com si no se'ls hi haguessin extret.

Fins ara els elements de l'anàlisi, és a dir, les observacions, eren les notícies i se situaven una a cada fila complint amb la condició de *tidy data* (una observació, una fila). A partir d'aquí, amb la tokenització feta, l'element d'estudi han passat a ser els tokens, és a dir, les paraules, es continua complint la condició *tidy data*, però ara les paraules estan repetides per a cada mitjà.

2.1. Freqüència absoluta

La freqüència absoluta d'un concepte, és el número de vegades que apareix una paraula o expressió en la mostra de notícies analitzades.

En primer lloc, es mostren les 20 paraules amb més freqüència de tota l'anàlisi i es grafiquen per a cada mitjà. En el gràfic global s'observen les paraules ordenades per freqüència absoluta general, és a dir, no es mostra ordenat cada gràfic de mitjà sinó el global total.

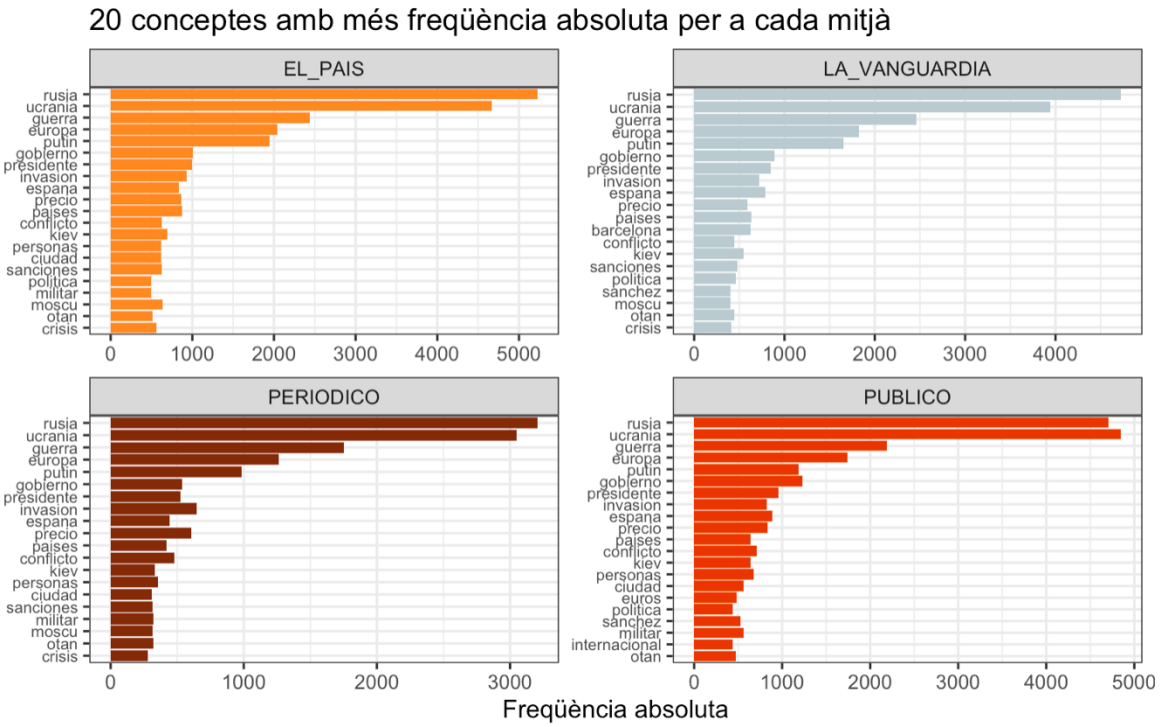

Figura 3.4: Diagrama de barres dels 20 conceptes amb més freqüència absoluta separats per mitjans

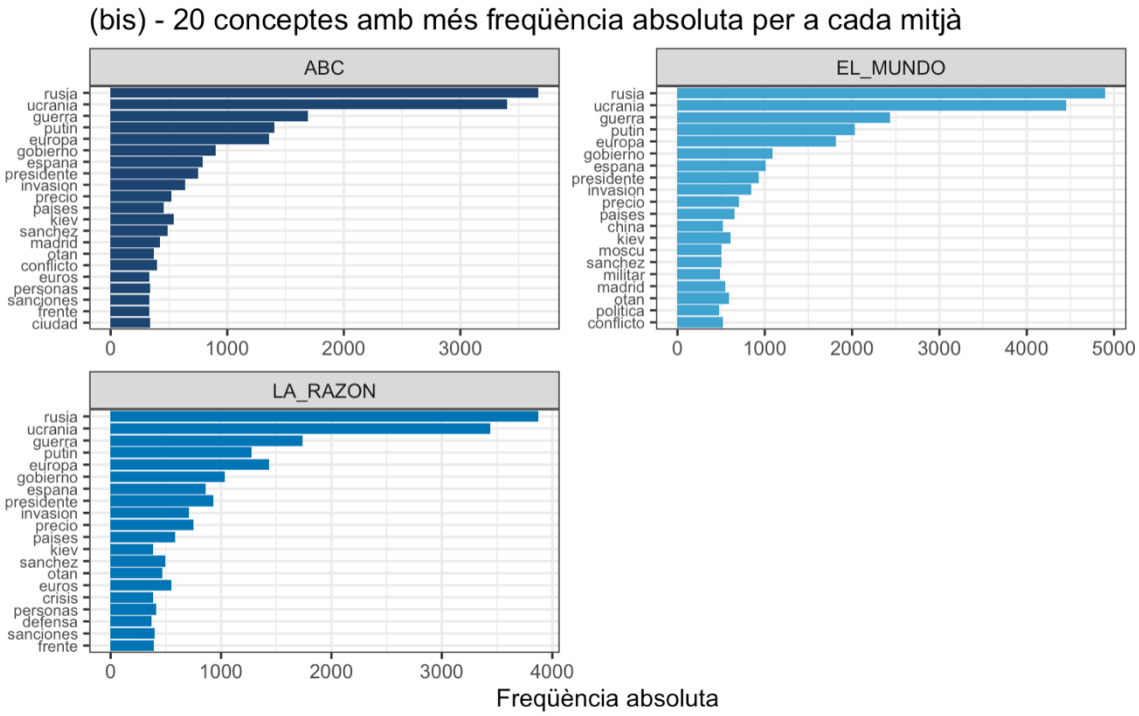

Figura 3.5: (bis) - Diagrama de barres dels 20 conceptes amb més freqüència absoluta separats per mitjans

La majoria de mitjans comparteixen l'ordre de les paraules més usades, sent Rússia, Ucraïna i Guerra les tres que encapçalen el rànquing.

Putin també es troba entre les paraules més usades, però curiosament Zelenski no hi és, això denota que la població està concebent una disputa entre dos països clarament identificats però associats a una sola persona. Els mitjans no parlen de l'altra cara de la moneda.

Entre els termes més usats abunden els relacionats amb l'economia, com ara *euros*, *precio* o *crisis*, també es veuen molts termes relacionats amb el propi país dels mitjans, que revelen la preocupació d'un fet internacional envers el país concret i estableixen una comparació, possiblement per sensibilitzar al lector i apropar-lo al conflicte: *Sánchez*, *Madrid*, *España* o *Europa*.

Exceptuant *La Razón*, es pot veure que tots els mitjans utilitzen els conceptes *invasión* i *conflicto* com a sinònims de guerra i ho fan aproximadament amb la mateixa freqüència, sempre prevalent *invasión*.

En segon lloc, es mostren les 20 paraules amb més freqüència de cada mitjà de manera individual però agrupats per ideologies polítiques. Els gràfics de barres no expressen resultats

gaire diferents als del gràfic anterior global però són més entenedors a l'hora de comprar mitjans d'ideologies de manera directa.

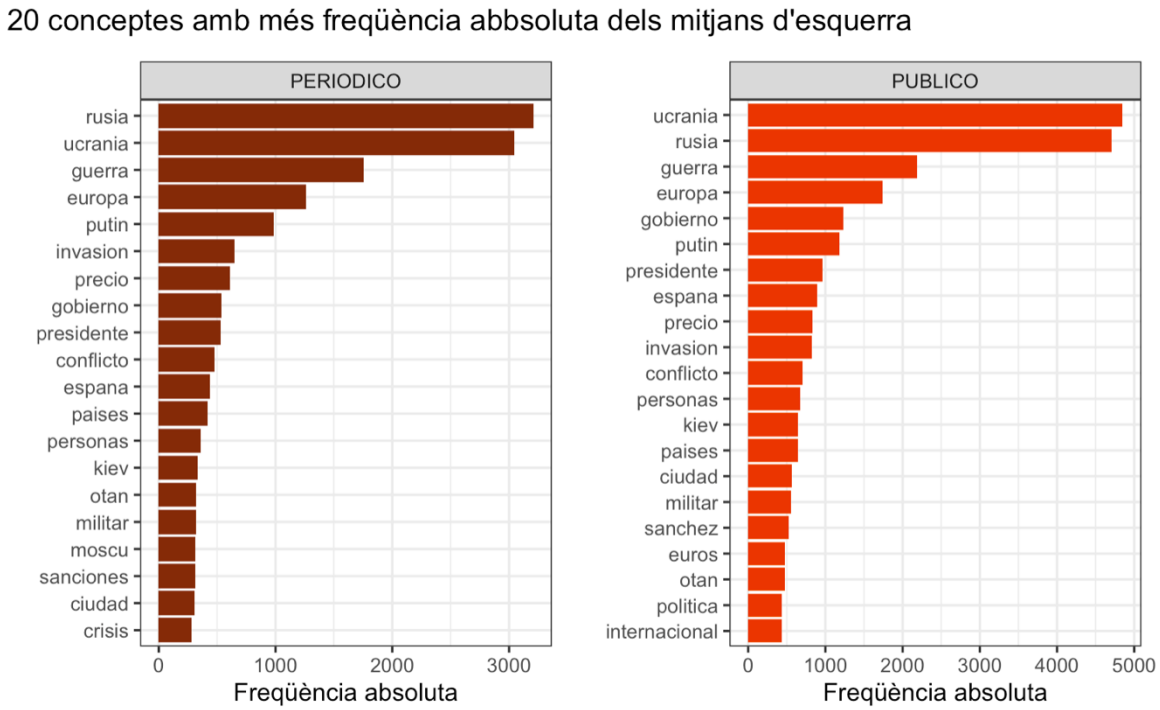

Figura 3.6: Diagrama de barres de freqüència absoluta de les paraules en els mitjans d'esquerra

Alhora s'observa un núvol de paraules per a cada ideologia. Els núvols de paraules, ens ajuden a identificar d'una manera més visual quina és la paraula amb més freqüència, la qual és proporcional a la mida de la paraula.

Figura 3.7: Wordcloud de la ideologia d'esquerra

Observem que els mitjans d'esquerra divergeixen entre ells en 6 paraules del total del rànquing de freqüència absoluta. Les paraules que destaquen són pràcticament les mateixes que en el global de tots els mitjans.

Catalunya apareix com a concepte en el núvol de paraules d'esquerra i, en canvi, no hi apareix en els de centre o dreta, Figures *3.9* i *3.11*.

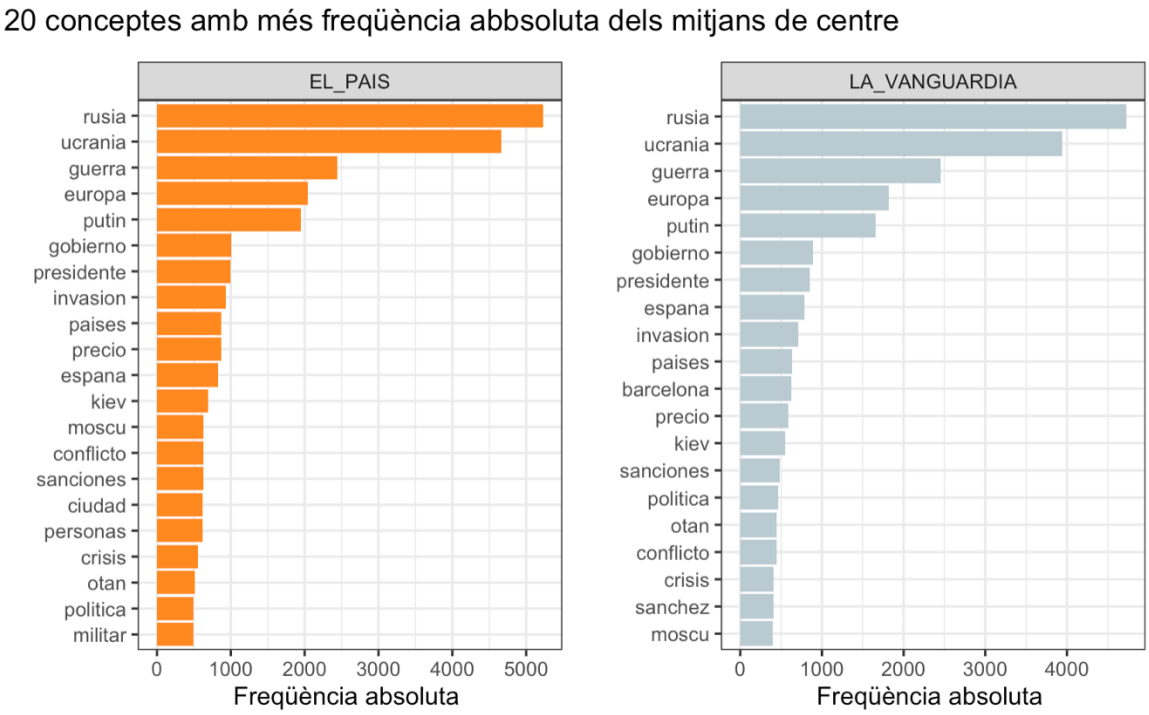

Figura 3.8: Diagrama de barres de freqüència absoluta de les paraules en els mitjans de centre

Figura 3.9: Wordcloud de la ideologia de centre

En les dues figues anteriors, podem veure que els mitjans de centre divergeixen en 5 paraules del seu rànquing de freqüència absoluta. Aparentment no s'observen clares diferències a la resta de mitjans.

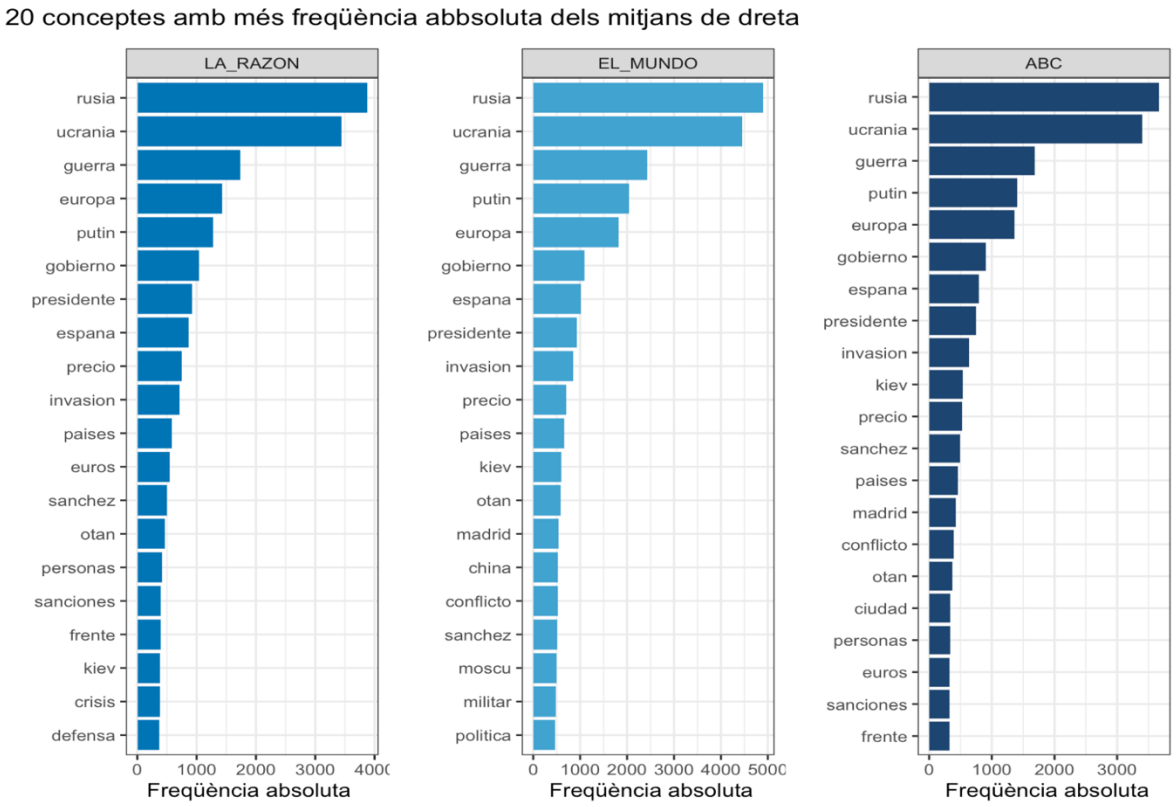

Figura 3.10: Diagrama de barres de freqüència absoluta de les paraules en els mitjans de dreta

Figura 3.11: Wordcloud de la ideologia de dreta

Els mitjans de dreta, igual com els de centre, divergeixen en 5 conceptes. No obstant, igual com en els mitjans d'ideologia central, no s'observen clares diferències entre els mitjans.

Per a la freqüència absoluta els mitjans de la mateixa ideologia sembla que utilitzen els mateixos patrons, llenguatge i expressions per a explicar un mateix fet internacional.

2.2. Rellevància dels termes

El *tf* és la freqüència amb què apareix un terme dins un document concret.

$$tf = \frac{n_{terme}}{longitud\ document}$$

Aquesta aproximació té la limitació de donar molta importància a aquelles paraules que surten moltes vegades tot i no aportar informació selectiva. Per tal de solucionar el problema, es poden ponderar els valors *tf* multiplicats per la inversa de la freqüència amb la qual el terme apareix a la resta de documents (*idf*).

$$idf = \log\left(\frac{n_{documents}}{n_{documents\ amb\ el\ terme}}\right)$$

D'aquesta manera, es redueix el valor d'aquells termes que apareixen a molts documents i que, per tant, no aporten informació selectiva. El *tf-idf* és la freqüència d'ocurrència del terme en un document concret en relació amb la presència que el terme té en el conjunt de documents analitzats.

$$tf - idf = \frac{n_{terme}}{longitud\ document} * \log\left(\frac{n_{documents}}{n_{documents\ amb\ el\ terme}}\right)$$

Es grafiquen les 20 paraules que defineixen més cada mitjà, segons el *tf-idf*. Es veuen paraules molt diferents de les de freqüència absoluta i molt més interessants en sentit semàntic. Es deixen de banda els conceptes obvis i lògics de les notícies i l'enfocament es focalitza en els detalls d'aquestes.

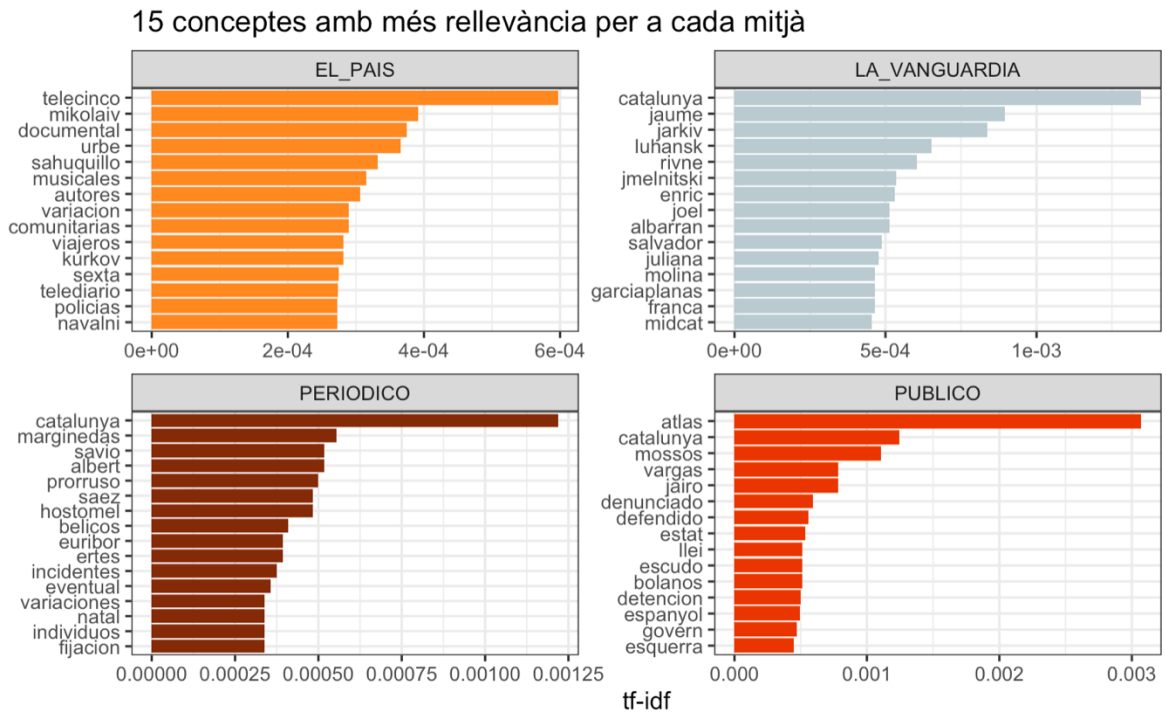

Figura 3.12: Diagrama de barres dels 15 conceptes amb més rellevància per a cada mitjà

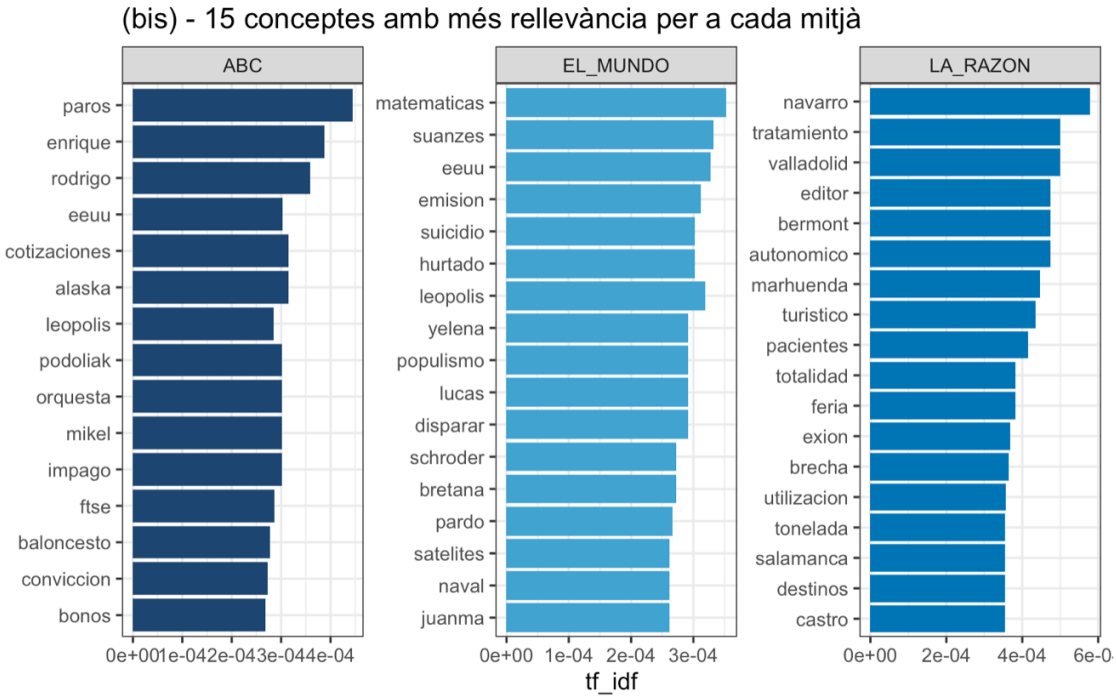

Figura 3.13: (bis) - Diagrama de barres dels 15 conceptes amb més rellevància per a cada mitjà

Apareixen noms concrets de persones implicades en la guerra, noms de ciutats, cadenes de televisió, conceptes econòmics concrets, etc.

A la **ideologia de dretes** es troben conceptes com **atur, suïcidi, cotitzacions** o **pacients.** D'altres com emissió, que es pot suposar que es tracta d'emissions de gas, estats units com a país clau del conflicte, o noms de personatges implicats en els mitjans com ara Enrique Serbeto, corresponsal del diari *ABC*, Francisco Marhuenda, director del diari *La Razón* o Pablo.R Suanzes, corresponsal de *El Mundo*.

En el diari *ABC* hi destaquen l'atur, les cotitzacions, els impagaments o l'índex FTSE 100, índex borsari publicat pel Financial Times compost pels 100 principals valors de la borsa de Londres.

En la **ideologia d'esquerra** destaca la menció a Catalunya, així com al mitjà centralista La Vanguardia. Noms de persones implicades en els mitjans com ara Marc Marginedas, periodista català destacat per la seva activitat com a corresponsal de guerra, o Jairo Vargas, periodista de *Público.es*. Altres conceptes econòmics molt rellevant i interessants són l'EURIBOR, taxa d'interès interbancària de referència de la zona euro que es mou sense una tendència clara fins que el BCE confirmi si lluitarà contra la inflació sense perjudicar el creixement, o les ERTOS de guerra per evitar la fallida de les empreses per la crisi d'Ucraïna.

2.3. Correlació entre mitjans i ideologies

Un diagrama de correlació és una representació gràfica que mostra la relació d'una variable respecte una altra sense necessitat que aquesta correlació sigui causa-efecte. S'ha calculat la freqüència de cada paraula segons cada mitjà i s'han executat diagrames de correlació per als conjunts de notícies de cada ideologia per tal de contrastar les següents hipòtesis:

Hipòtesis 1: Els mitjans de la mateixa ideologia utilitzen un vocabulari similar per expressar una mateixa idea.
Hipòtesis 2: Les ideologies d'extrems usen vocabulari diferent per expressar la mateixa idea.

La idea principal és que si dos mitjans escriuen de forma similar, tendiran a emprar les mateixes paraules i amb freqüències similars.

En el següent gràfic es mostra la correlació entre els dos mitjans de la ideologia d'esquerra, s'observa que la correlació és positiva i directa i que el gruix més gran de paraules compartides es troba entre l'1% i el 10%.

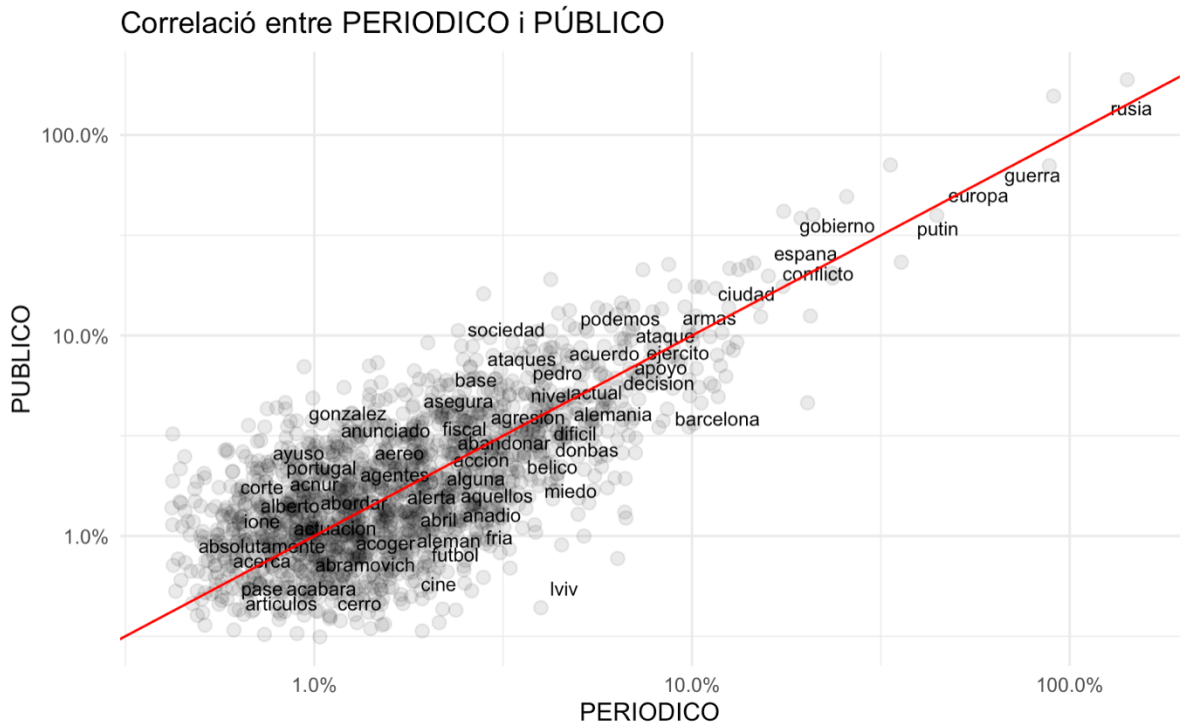

Figura 3.14: Diagrama de correlació entre els conceptes de *Público.es* i *El Periódico*

A continuació vegem la correlació dels mitjans de centre. De nou, s'observa una correlació positiva i amb cap paraula extremadament allunyada de la línia vermella de correlació lineal.

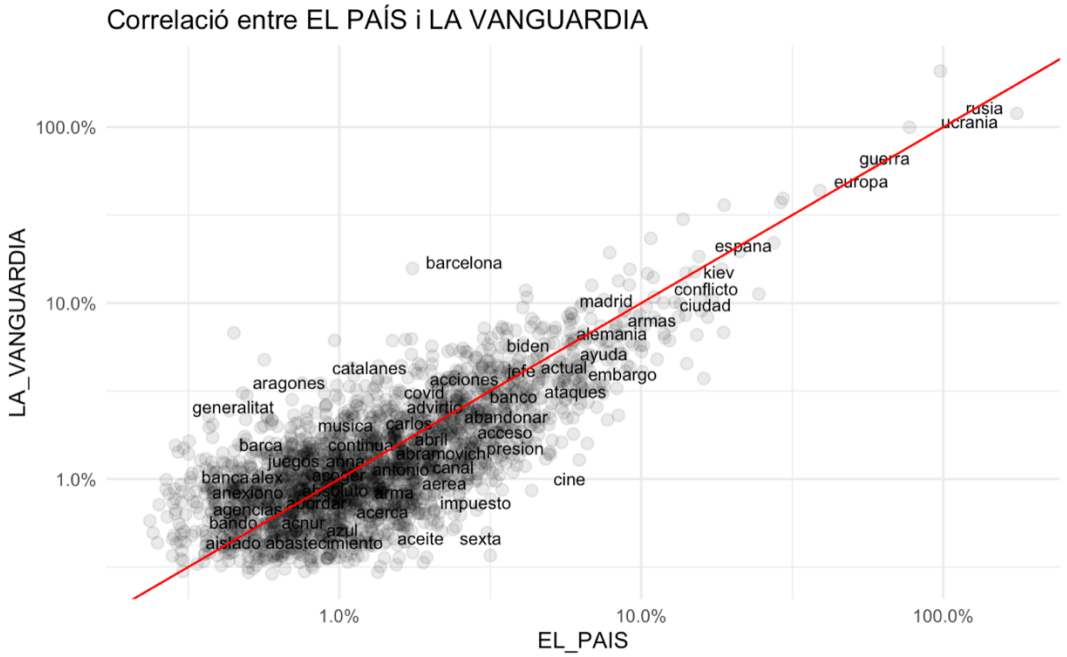

Figura 3.15: Diagrama de correlació entre els conceptes de *La Vanguardia* i El País

Per últim, per a les correlacions dels mitjans de dreta no s'hi veuen diferències de comportament respecte les altres ideologies. Les tres correlacions són lineals i el gruix de paraules compartides està entre l'1 i el 10%. A més a més, es corrobora el que s'ha dit en l'apartat de freqüència absoluta que *La Razón* és el mitjà que menys usa la paraula *conflicto* per referir-se a la guerra.

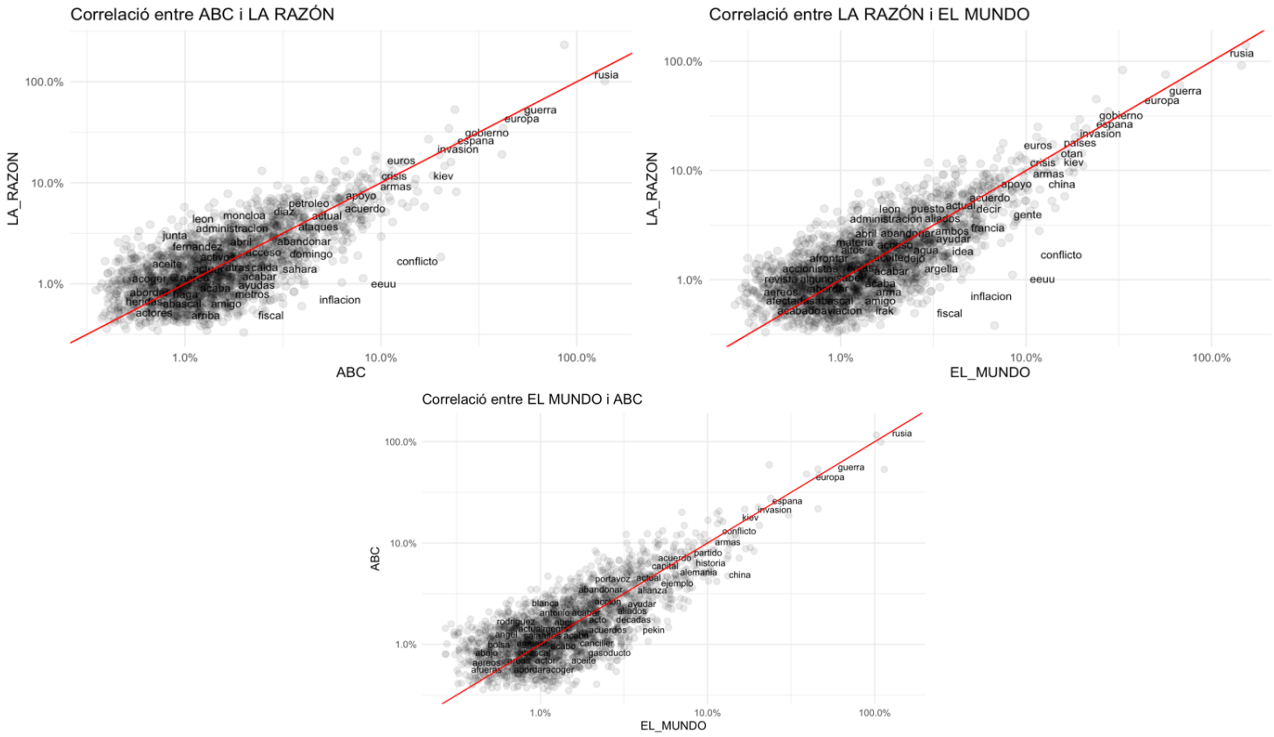

Figura 3.16: Diagrama de correlació entre els conceptes de *La Razón*, ABC i *El Mundo*

Es pot dir, que segons les correlacions entre mitjans, la hipòtesi 1 sembla ser certa i que els mitjans de la mateixa ideologia utilitzen un vocabulari similar per expressar una mateixa idea.

L'interès principal de l'anàlisi és comprar els extrems, per tant, la hipòtesi a contrastar és la que es qüestiona si les ideologies de dreta i esquerra usen termes diferents per expressar les mateixes idees.

Es calcula el nombre de paraules comunes que tenen les dues ideologies. Individualment esquerra té un total de 3.263 termes i dreta un total de 3.950. Conjuntament, comparteixen 2.721 termes.

Com es veu en la Figura 3.17, contra tot pronòstic i suposicions inicials, les ideologies de dreta i esquerra, igual com les ideologies entre elles mateixes, mostren una correlació lineal positiva i directa.

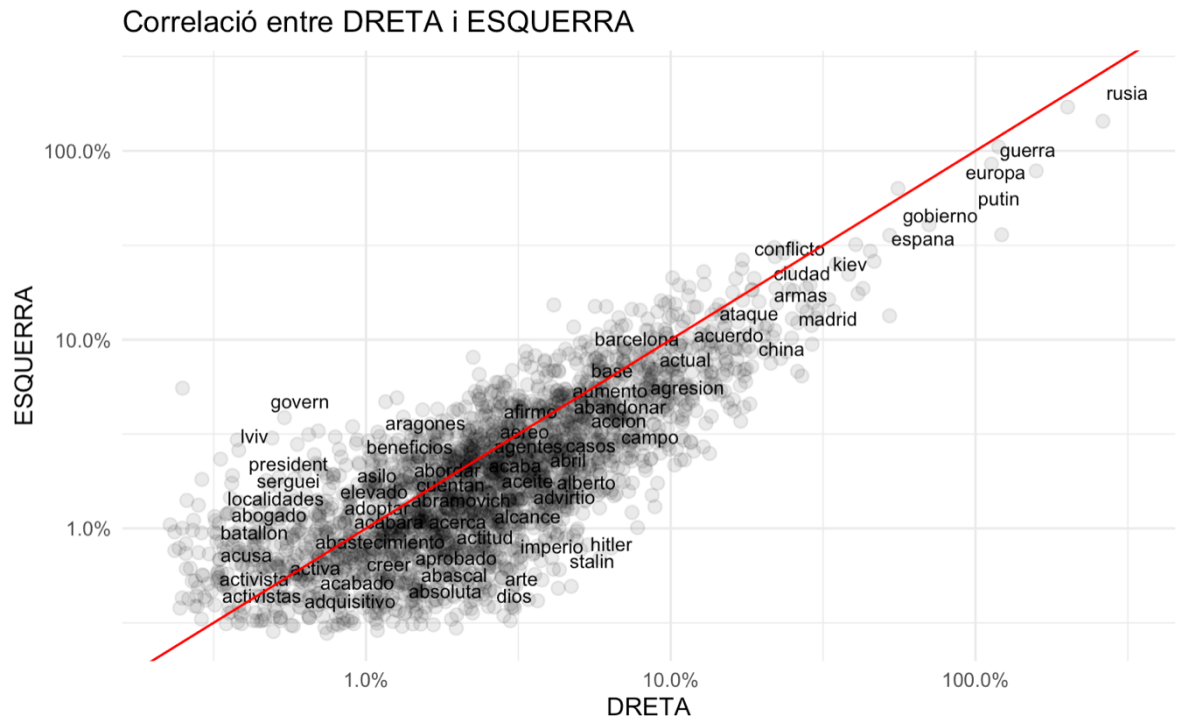

Figura 3.17: Diagrama de correlació entre els conceptes de les ideologies d'extrems

No es pot dir que el llenguatge de cada ideologia és diferent quan parlen d'un mateix fet internacional i, per tant, segons la correlació, caldria rebutjar la hipòtesi 2.

2.4. Diferència de termes

A continuació, s'observa quines són les paraules que s'utilitzen de forma més diferenciada per a cada mitjà, si un mitjà utilitza molt una paraula que un altre no utilitza gens. Per a fer-ho, s'usen els *log of odds ratio* de les freqüències de cada paraula. Un *odds ratio* (OR) és una estadística que quantifica la força de l'associació entre dos esdeveniments, A i B, es defineix com la relació entre les probabilitats d'A en presència de B i les probabilitats d'A en absència de B, o equivalent.

Dos esdeveniments són independents si i només si l'OR és igual a 1, és a dir, les probabilitats d'un esdeveniment són les mateixes en presència o absència de l'altre esdeveniment. Si l'OR és superior a 1, aleshores A i B s'associen positivament. Per contra, si l'OR és menor que 1, aleshores A i B estan correlacionats negativament, i la presència d'un esdeveniment redueix les probabilitats de l'altre esdeveniment.

El *log of odds ratio* és el mateix que el OR, però expressat amb el logaritme de la probabilitat.

$$\log \text{ of odds ratio} = \log\left(\frac{\left(\frac{n_k+1}{N+1}\right)_{MITJÀ_1}}{\left(\frac{n_k+1}{N+1}\right)_{MITJÀ_2}}\right)$$

n_k és el nombre de vegades que apareix el terme en les notícies de cada mitjà i N el nombre total de paraules de cada mitjà.

Tenint en compte que la correlació entre els mitjans de la mateixa ideologia és positiva, les paraules que poden sortir en els *log of odds ratio* de cada ideologia no haurien de ser gaire rellevants parlant des del context semàntic de la paraula.

En atenció a això, seguidament es mostren els *log of odds ratio* de cada ideologia, però, s'observa que les paraules que no comparteixen els mitjans d'aquestes no tenen una rellevància per fer que siguin diferents, són conceptes amb poca força per poder diferenciar el llenguatge dins la pròpia ideologia.

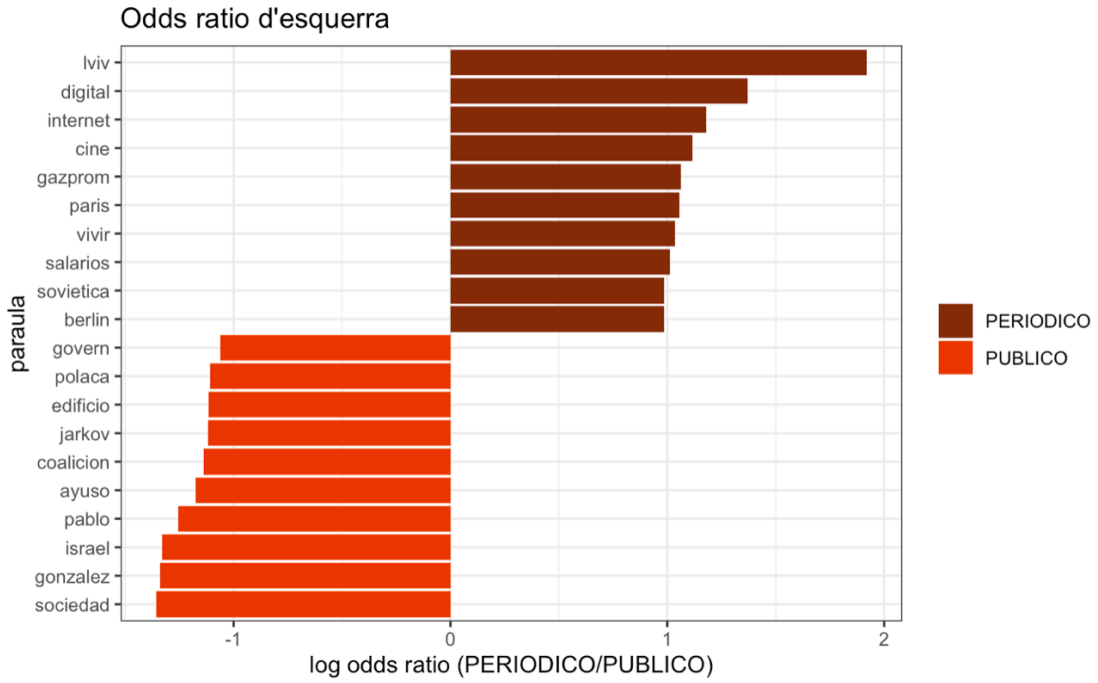

Figura 3.18: Log odds ratio dels conceptes de *El Periódico* i de *Público.es*

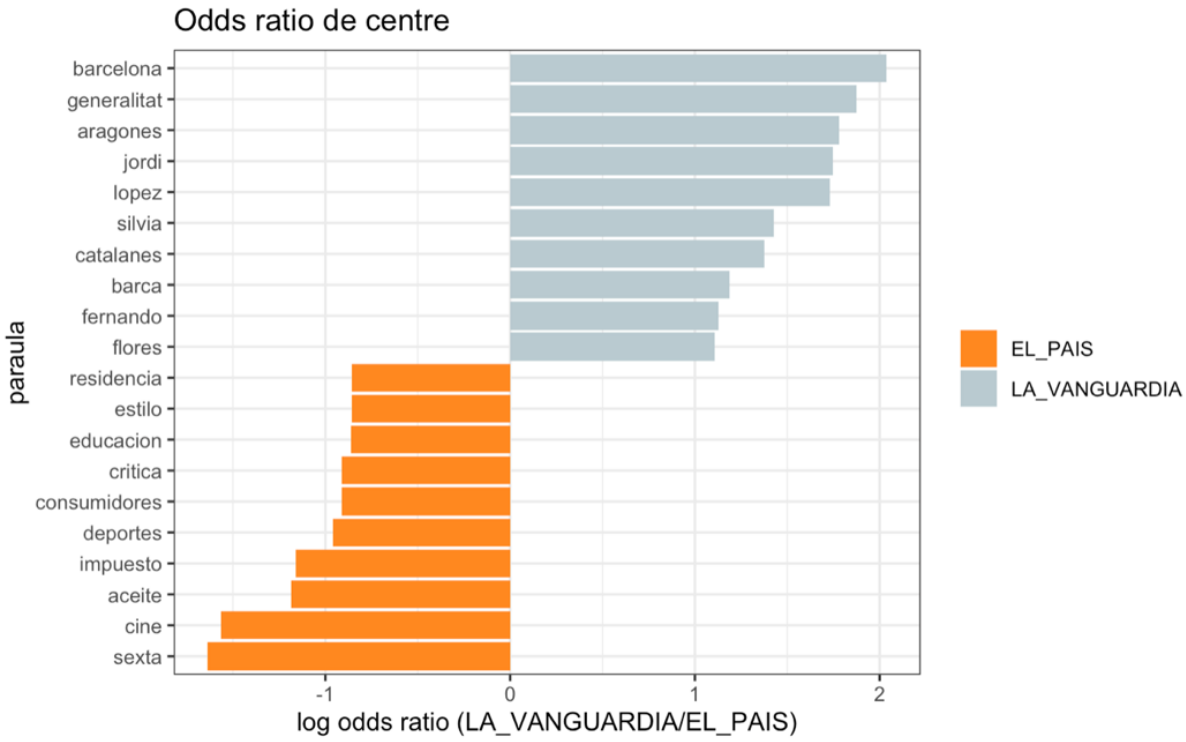

Figura 3.19: Log odds ratio dels conceptes de *El País* i de *La Vanguardia*

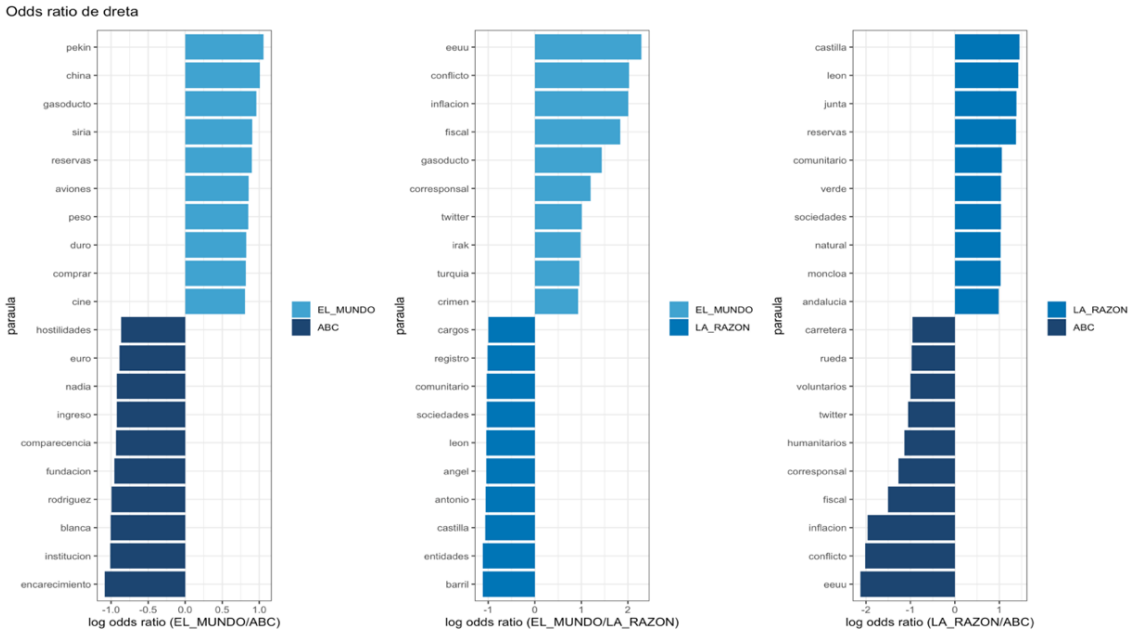

Figura 3.20: Log odds ratio dels conceptes de *El Mundo, ABC* i *La Razón*

El més interessant d'aquesta comparació de paraules més diferenciades, igual com en les correlacions, és el contrast entre ideologies d'extrems.

En la Figura 3.21, s'observa que la ideologia de dreta utilitza conceptes com ara *Hitler*, *nación*, *dictador*, *Stalin* o *dios* i la ideologia d'esquerra aposta més per conceptes com *Sputnik*[3], *física*, *Rufian* o *president*.

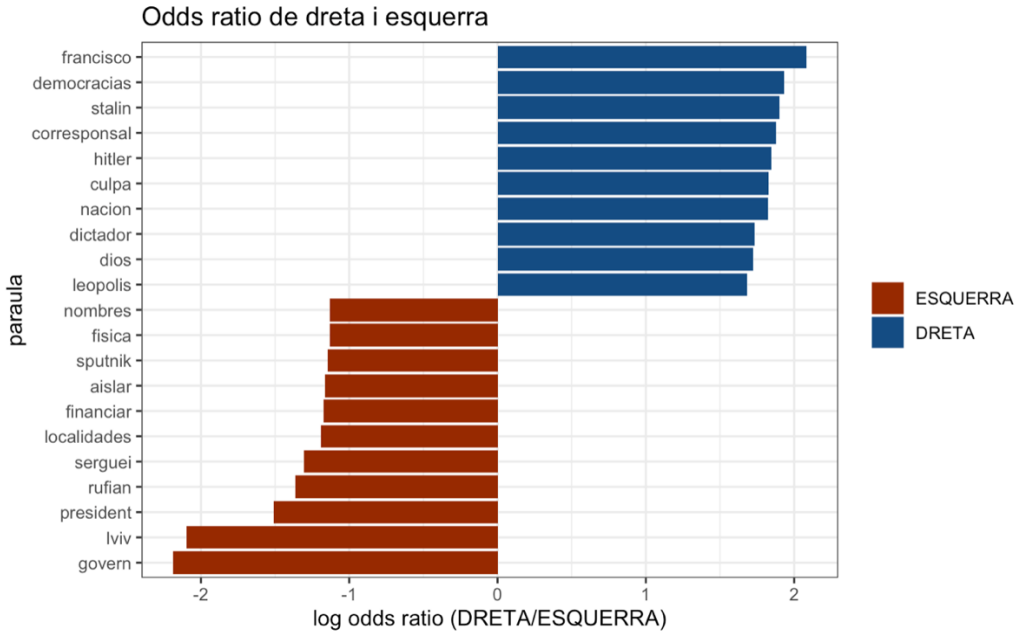

Figura 3.21: Log odds ratio dels conceptes de les ideologies d'extrems

Com a conclusió, es pot dir que els conceptes que ressalten en la ideologia de dreta semblen encarats a criticar l'esquerra, s'observa que els mitjans de dreta semblen tenir tendència a ressaltar conceptes clarament partidistes, remarcant la pròpia ideologia en les seves notícies. Els mitjans d'esquerra, en canvi, tot i optar per alguns conceptes claus en la seva ideologia, semblen més generalistes i parcials a l'hora d'exposar les seves idees.

2.5. PCA

L'anàlisi de components principals o PCA és una de les tècniques d'aprenentatge no supervisat que permet extreure informació d'un número concret de variables. Un dels seus usos és el de visualització de dades de "n" observacions amb "x" variables, el PCA fa una rotació dels eixos de coordenades de les variables originals a uns nous eixos ortogonals, de manera que aquests nous coincideixin amb la direcció de màxima variància de les dades. A més a més, el PCA no requereix la suposició de normalitat de les dades.

[3] Sputnik: Agencia de notícies, lloc web de notícies i servei de radiodifusió del govern rus, el qual va ser creat per l'agència estatal russa *Rossiya Segodnya* el 10 de novembre del 2014.

A partir de les dades usades per l'execució de tota l'anàlisi prèvia, es prepara un nou *data frame* per tal d'aplicar el PCA, en el qual cada paraula és una columna, és a dir, una de les variables, i els set mitjans són cadascuna de les files.

El PCA resultant n'és el següent:

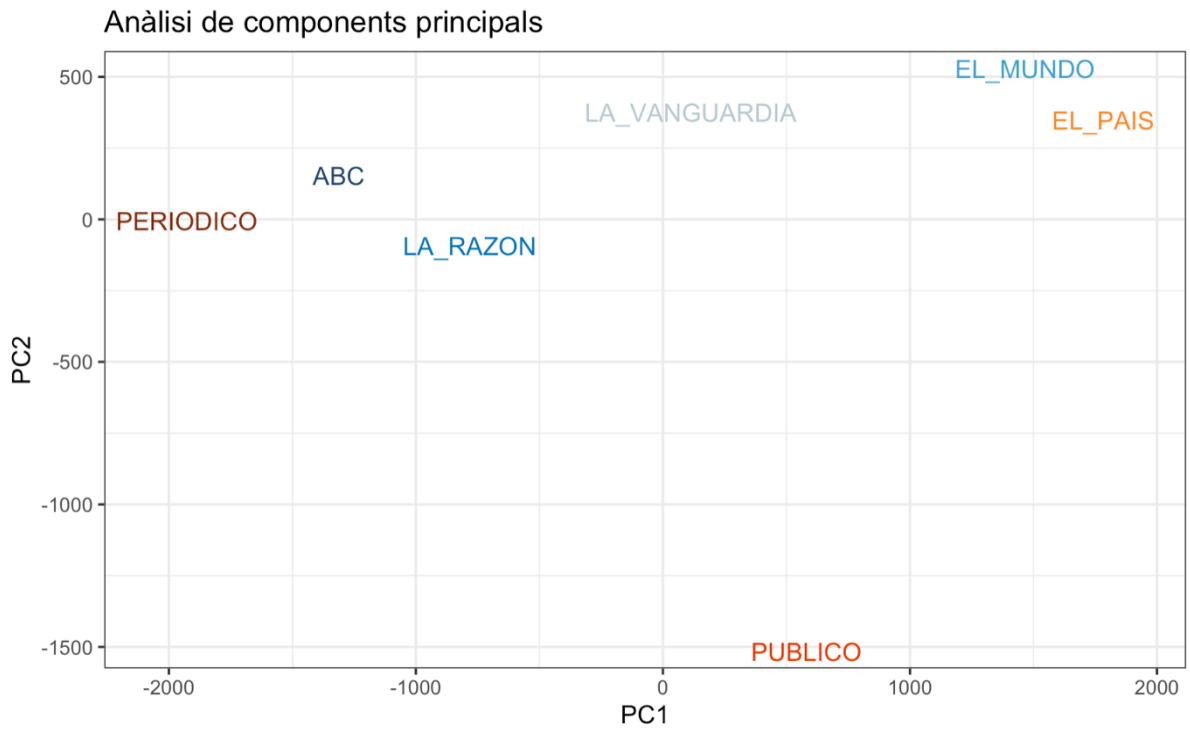

Figura 3.22: Anàlisi de components principals

Sis dels mitjans queden repartits a la part superior del PCA i només *Público.es* es situa a la part inferior, a molta distància de la resta.

Tres dels mitjans de premsa impresa es troben a l'esquerra del PCA, *La Vanguardia* força al centre i *El Mundo* i *El País* a la part dreta.

Contrari al que hagués corroborat les ideologies del mitjans, aquests no s'agrupen en relació a elles. Els mitjans s'agrupen, però no ho fan per ideologies. Tanmateix, sí que es remarca la diferència dels mitjans de premsa impresa amb l'únic mitja digital de l'anàlisi.

Es pot dir que el mitjà centralista *El País* comparteix idees amb un mitjà de dretes com és *El Mundo* i que el mitjà d'ideologia d'esquerra El Periódico, s'apropa molt a idees de dos mitjans d'ideologia de dreta com són *ABC* o *La Razón.*

2.6. Diagrama de Venn

Un diagrama de Venn és una representació gràfica que utilitza cercles superposats per il·lustrar la relació lògica entre dos o més grups d'elements.

En aquest tipus de diagrama es pot veure com de similars o afins són els grups que s'estudien a partir dels elements que tenen en comú.

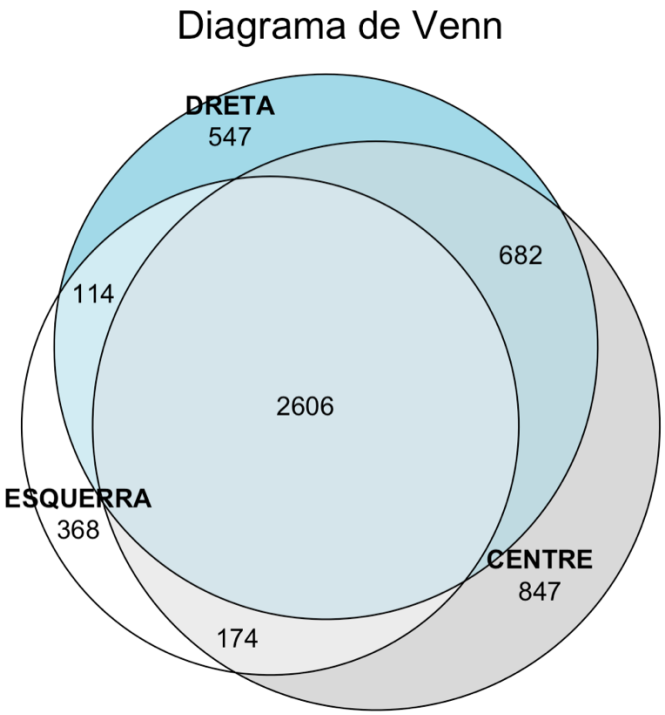

Figura 3.23: Diagrama de Venn de les 3 ideologies analitzades

En el diagrama s'observa que la intersecció més gran és la de dreta i centre, donant a entendre que el centre és més afí a una ideologia més de dretes que no pas d'esquerra. La intersecció més petita és la que formen els extrems, corroborant així que els termes usats i el llenguatge de cada ideologia és diferent tot i parlar del mateix fet internacional.

IV. ANÀLISI DE L'IMPACTE ECONÒMIC DE LA GUERRA

Pel que fa a la repercussió econòmica en el primer mes de guerra, s'ha seguit el mateix procés que en l'anàlisi semàntica de les paraules, però en mitjans diferents. Aquest cop els mitjans analitzats han estat els tres grans mitjans econòmics espanyols *Cinco Días*, *Expansión* i *El Economista.*

S'han escollit els tres mitjans econòmics espanyols de premsa impresa i s'han recollit de l'Hemeroteca de *MyNews* totes les notícies del període del 24 de febrer de 2022 fins al 24 de març de 2022 que tenien relació amb la guerra entre Ucraïna i Rússia.

El filtratge de les notícies i la neteja del text s'ha fet amb el mateix procés que el primer anàlisi. Aquest cop, però, els termes s'han separat en bigrames de dos, és a dir, amb dues unitats de text.

Es considera que els conceptes econòmics acostumen a anomenar-se a partir de grups de paraules i no pas amb una paraula concreta. Tanmateix, abans d'unir les paraules en bigrames de dos termes s'ha fet una petita observació per identificar tres conceptes econòmics d'un sol terme que s'han inclòs a l'anàlisi de manera individual: *iva*, *or* i *ley*.

És a dir, en primer lloc es va realitzar l'anàlisi amb unigrames, però els resultats no eren representatius econòmicament i no se'n podien extreure unes conclusions clares. En conseqüència, es va decidir procedir amb l'anàlisi a través de bigrames, però sense excloure els únics 3 conceptes econòmics que es van identificar com a conceptes individuals, els quals són: *iva, oro* i *ley.*

A partir d'aquesta anàlisi i de les següents representacions gràfiques, es pretén fer una fotografia de les conseqüències econòmiques que ha portat la guerra en el seu primer mes i desenvolupar-ne l'explicació.

Es grafiquen els 20 bigrames amb freqüències absolutes més elevades de cada mitjà econòmic espanyol analitzat.

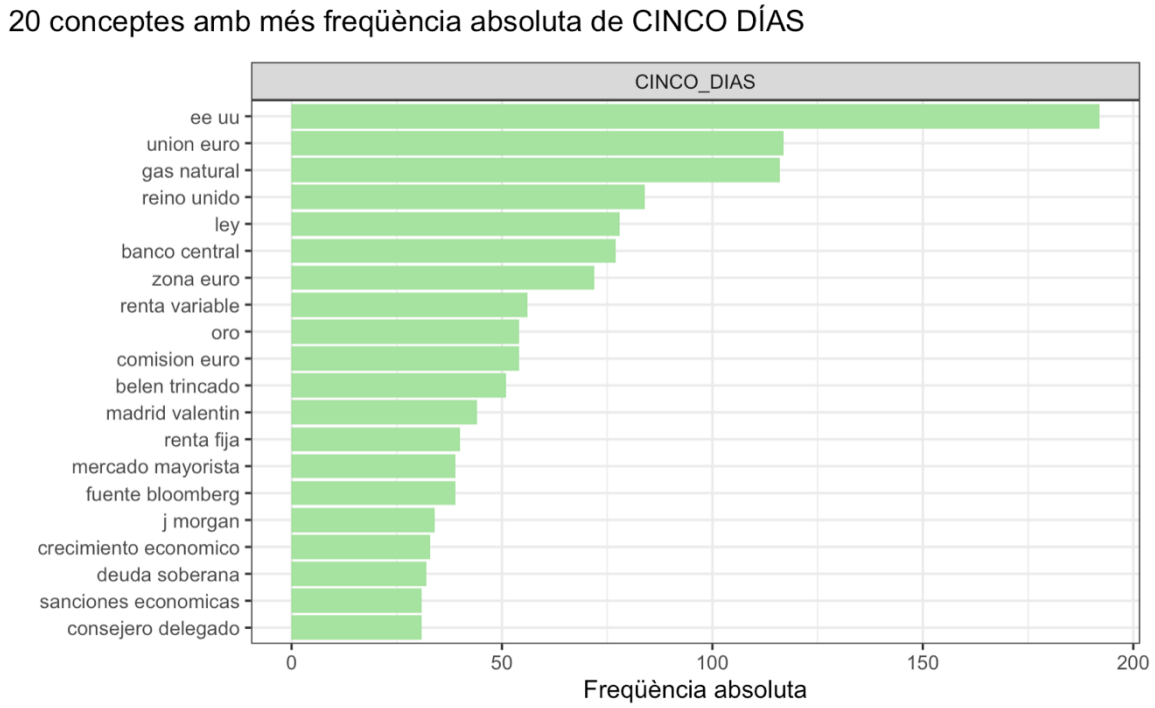

Figura 4.1: Diagrama de barres de freqüència absoluta dels bigrames de Cinco Días

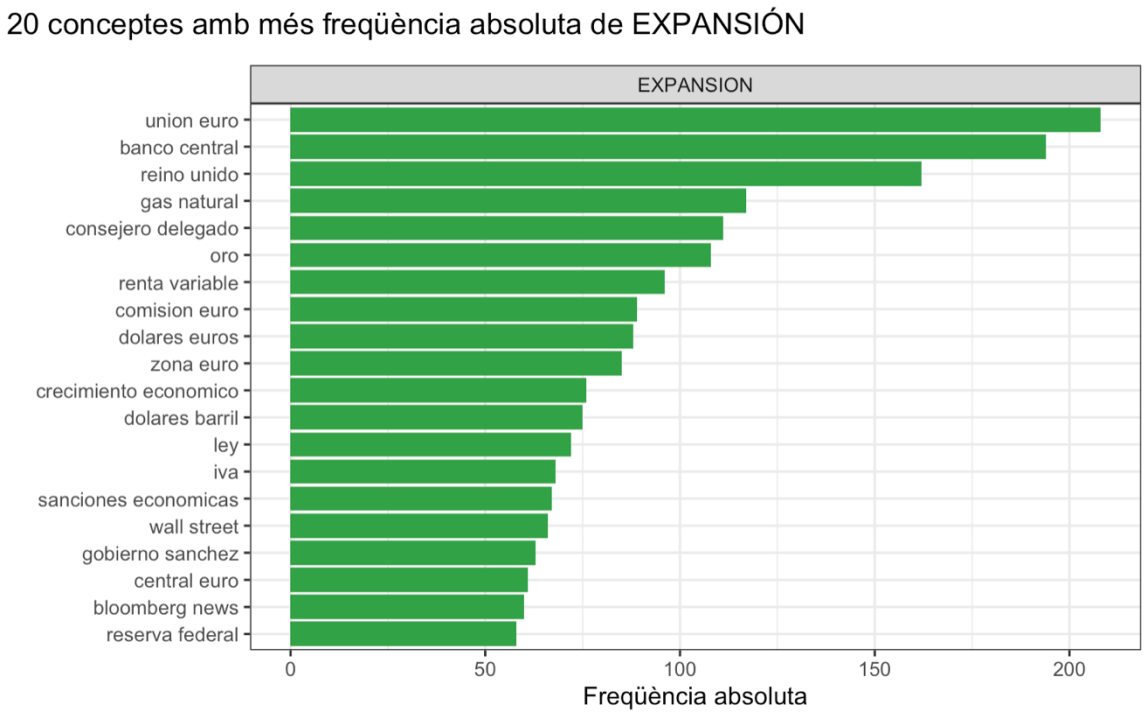

Figura 4.2: Diagrama de barres de freqüència absoluta dels bigrames de Expansión

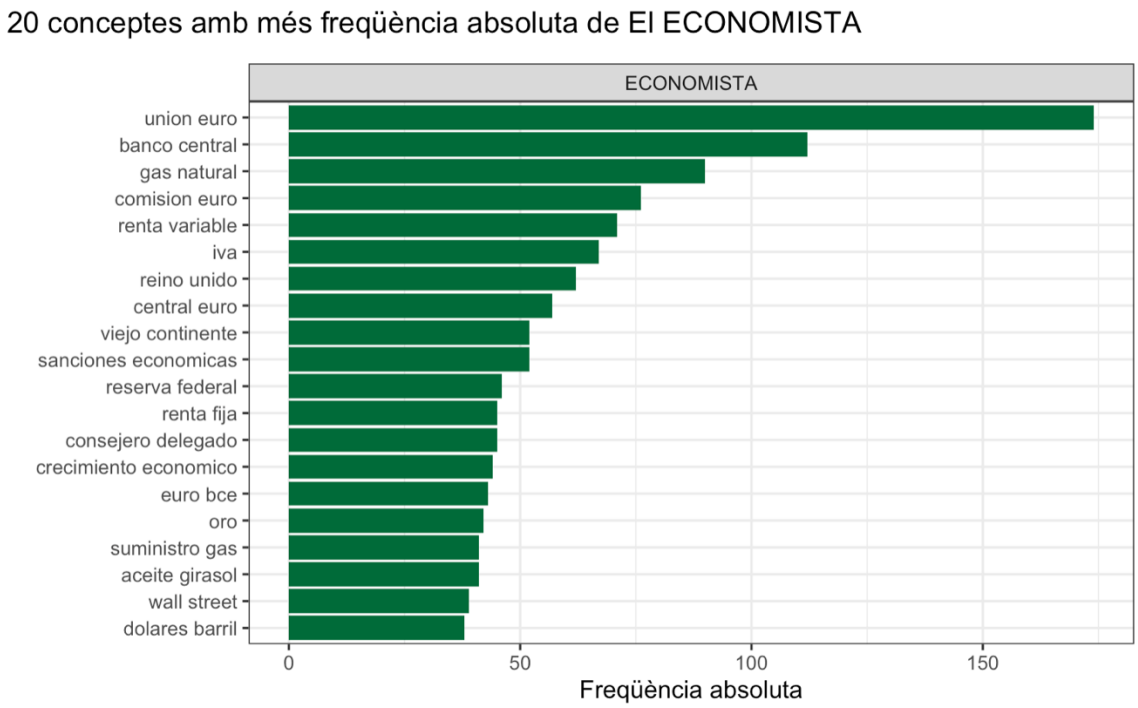

Figura 4.3: Diagrama de barres de freqüència absoluta dels bigrames de El Economista

S'observa que els mitjans parlen de conceptes generals com la unió europea, el banc central, el regne unit, els estats units, el gas natural o la renta variable.

S'obté el gràfic de les freqüències absolutes totals sense tenir en compte el mitjà econòmic en el qual han aparegut els termes, per tal, de poder decidir sobre quins conceptes parlar envers la repercussió econòmica de la guerra.

En la figura, els conceptes més rellevants són *unión europea, banco central, gas natural, reino unido* i *renta variable*.

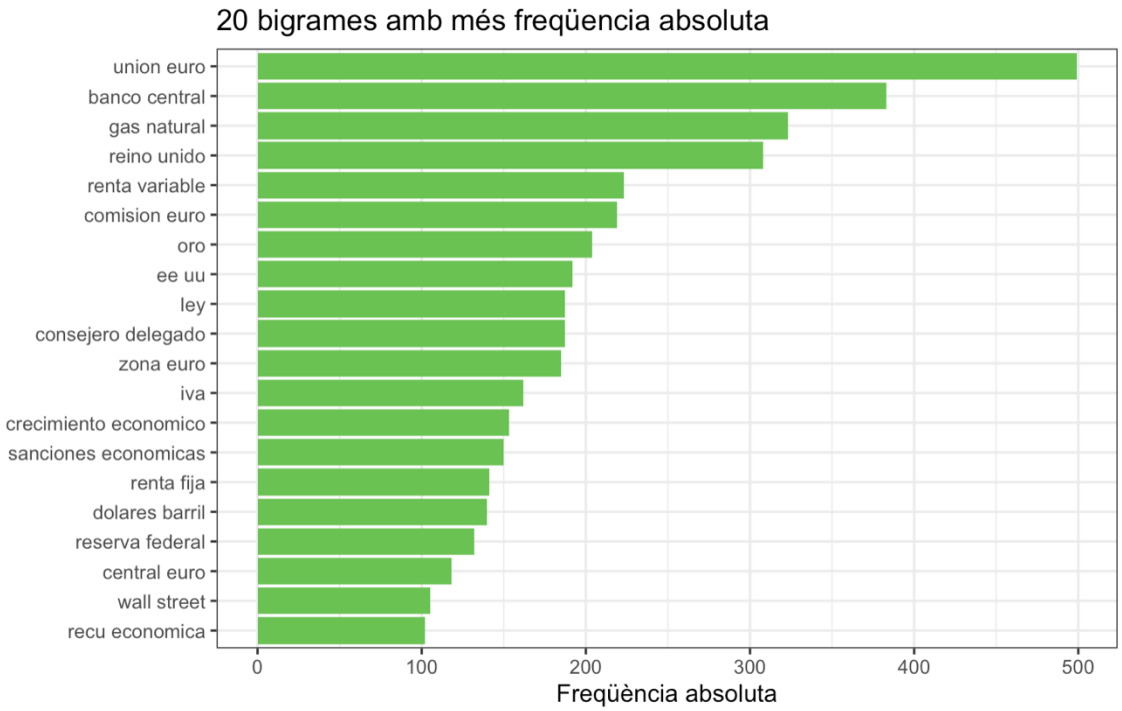

Figura 4.4: Diagrama de barres de freqüències absolutes dels bigrames dels mitjans econòmics

1. Conseqüències de la guerra

Les guerres afecten de forma significativa a les economies dels països; els efectes del conflicte sempre van més enllà de les morts o la despesa militar i acostumen a tenir repercussions profundes i persistents al llarg dels anys. Una d'aquestes repercussions té a veure amb les conseqüències econòmiques de la guerra, les quals s'exposaran en els següents paràgrafs a través de la interpretació dels conceptes del diagrama de barres anterior, Figura 4.4.

La guerra ha provocat una escalada dels preus molt important. A Espanya, la inflació se situa al voltant del 10% i, durant el mes de febrer, els preus han pujat un 7,6%, taxa que no es veia des del 1986. A la zona euro, la inflació s'ha situat a un 5,8%, el percentatge més alt de tota la seva història. De fet, el Banc Central Europeu preveu que els preus escalaran fins al 5,1%, de mitjana el 2022 a la zona euro, quasi dos punts per sobre de la seva última estimació, ja que, al desembre, es creia que se situaria en el 3,2%. Per tant, tot i saber que la inflació ja era alta abans de la invasió russa, es fa evident que aquesta ha provocat que encara incrementi més ràpidament.

Pel que fa a l'or, els primers dies de guerra, aquest ha pujat un 1,5%, arribant a tocar els 2.050,76 dòlars. La llum també ha incrementat el seu preu; el megawatt hora (MWh) ha superat per primera vegada els 400 euros en diferents zones i franges horàries. No només l'or i la llum, la electricitat en general, sinó que durant aquest mes el mercat de matèries primeres

també ha augmentat, provocant que el petroli i el gas natural cotitzin a la zona de màxims històrics, igual que els metalls i els productes agrícoles. Metalls com el níquel, el coure o l'alumini registren els preus més alts des del 2007. Els experts preveuen que el preu del petroli de referència a Europa, el Brent, oscil·larà entre els 150 i els 250 dòlars, i adverteixen d'una possible crisi en l'oferta si es tanca l'exportació de petroli rus.

Segons Putin, no té sentit subministrar els seus productes a la Unió Europea o als Estats Units i rebre els pagaments en dòlars, euros o altres monedes. Aquesta amenaça ha disparat el Brent fins als 121,7 dòlars per barril, mentre que el gas natural holandès ha pujat un 30%. Pel que fa a aliments bàsics, Rússia i Ucraïna sumen conjuntament el 29% de les exportacions mundials de blat, el 19% de les de blat de moro i el 80% de l'oli de girasol. Espanya per exemple, importa el 80% d'oli de girasol d'Ucraïna. Això provoca que la majoria d'importacions europees topin amb dificultats a causa de la guerra.

Respecte a les divises, el ruble rus ha caigut un 30% en el seu canvi contra el dòlar quan el Banc Central Rus va anunciar que pujava els tipus al 20%. L'euro també ha perdut en el seu canvi contra el dòlar; al principi del conflicte la divisa va caure un 1%, però des de l'inici de la guerra ja ha perdut un 2,7%. L'Ibex-35, índex borsari de referència espanyol, va tancar la segona setmana del conflicte amb la major caiguda setmanal des de l'inici de la pandèmia el març de 2022.

A banda d'aquestes clares repercussions econòmiques, durant el període del primer mes de guerra, s'ha vist un gran moviment de criptoactius. Segons Christine Lagarde (2022), presidenta del BCE, "l'inici de la guerra ha augmentat el canvi de rubles a criptomonedes per evitar les sancions que ha implementat Rússia".

En definitiva, mentre s'espera que Rússia i Ucraïna pactin un alto el foc, la Comissió Europea treballa en mesures per revertir aquest augment dels preus de l'energia a Europa i, així, evitar una major escalada de la inflació i donar aire al sector industrial.

V. DISCUSSIÓ

En moltes ocasions, les notícies dels mitjans de comunicació són completament diferents entre elles tot i voler explicar el mateix fet. Això és degut a la manera que té cadascun dels mitjans d'intentar persuadir i apropar el lector cap al seu mitjà. A més a més, tal com s'ha esmentat al llarg del treball, la ideologia de cada mitjà és pública i en alguns casos es veu reflectida en les notícies i publicacions que fan. Tanmateix, això no hauria de passar, ja que, els mitjans idealment haurien de parlar sempre de manera objectiva.

El present treball ha posat el focus en el cas de la guerra entre Rússia i Ucraïna per comprovar si la ideologia dels mitjans es veia reflectida en les seves publicacions. Així, els resultats esperats es basaven en pensar que els mitjans de la mateixa ideologia seguien els mateixos patrons de llenguatge i que els d'ideologies oposades en seguien uns de divergents.

Posant èmfasi en els resultats obtinguts, es pot considerar que la hipòtesi inicial es confirma parcialment atenent que s'ha pogut comprovar que els mitjans que comparteixen ideologies, expressions, llenguatge, i la majoria de conceptes. No obstant això, aquest fenomen es repeteix amb ideologies oposades; si bé és cert que en aquests mitjans destaquen alguns conceptes que descobreixen la seva ideologia, quan es parla de manera generalitzada, no es detecta una gran diferència d'expressió entre els mitjans d'ideologia oposada.

S'ha de tenir en compte que pel cas de l'anàlisi de components principals, el resultat esperat no ha concordat amb l'obtingut. Així mateix, s'esperava que els mitjans s'agrupessin segons els grups establerts en el treball, és a dir, els tres grups de cada ideologia i que, a més a més, aquests grups quedessin separats entre ells. El PCA, en canvi, ha mostrat una clara separació entre els mitjans de premsa impresa i l'únic mitjà digital de l'anàlisi, i ha agrupat el mitjà de dretes *El Mundo* amb el mitjà de centre *El País. El Periódico* també ha quedat relativament agrupat amb *ABC* i *La Razón,* i *La Vanguardia* ha complert les expectatives, posicionant-se al centre del PCA i sense quedar agrupada amb cap altre mitjà.

Finalment, i pel que fa a l'anàlisi econòmica, en aquest cas els resultats obtinguts concorden amb la realitat de les conseqüències de la guerra. Tot i que l'argumentació de les repercussions econòmiques derivades de la guerra permet explicar-se d'una manera subjectiva, tendint al posicionament ideològic, la situació econòmica, en termes numèrics, s'explicarà sempre d'igual manera i la notícia, independentment del mitjà, transmetrà la mateixa informació.

VI. CONCLUSIONS

En aquest treball s'ha donat resposta a la qüestió sobre si en funció de la ideologia política, els mitjans expliquen el mateix fet internacional – en aquest cas la guerra entre Ucraïna i Rússia – d'una manera diferent o deixen de banda les seves ideologies per parlar de manera totalment objectiva. L'objectiu principal ha estat l'anàlisi dels patrons de parla i expressions de les diferents ideologies polítiques en les quals s'identifiquen els mitjans nacionals espanyols, i les hipòtesis plantejades han estat les següents:

- Hipòtesis 1: Els mitjans de la mateixa ideologia utilitzen el mateix vocabulari, llenguatge i expressions per expressar una mateixa idea.

- Hipòtesis 2: Les ideologies d'extrems fan servir vocabulari, llenguatge i expressions diferents per expressar la mateixa idea.

S'ha observat que la primera hipòtesi es confirma i que, per tant, quan els mitjans de la mateixa ideologies exposen una idea, ho fan de manera similar, amb una correlació positiva. Així, de les paraules que són rellevants semànticament, no en mostren cap de diferent i estan en concordança amb les seves idees.

Pel que fa a la segona hipòtesis, s'ha conclòs que no es pot confirmar i que, per tant, es rebutja. El vocabulari usat per les ideologies oposades quan expressen un mateix fet és similar i amb correlació positiva. Tanmateix, amb els *log of odds ratio* s'ha observat que, per una banda, els mitjans d'ideologia de dreta mencionen conceptes extremistes acord amb les seves idees. I, d'altra banda, s'ha observat que els mitjans d'ideologia d'esquerra són més discrets que els de dretes a l'hora d'expressar la seva ideologia. A més a més, s'ha contemplat que els mitjans de centre sempre són més afins a la ideologia de dreta que no pas a la d'esquerra, compartint més conceptes a l'hora de redactar les notícies.

No només això, sinó que el treball també s'ha centrat en les conseqüències econòmiques de la guerra. Per aquesta part, l'objectiu ha estat analitzar, també a través de notícies, quin és l'impacte econòmic que ha causat la guerra entre Rússia i Ucraïna.

Així, com a conclusió, aquesta guerra ha provocat un fort augment de la inflació, que ha afectat més bruscament al preu de les matèries primeres com el gas natural, el petroli i alguns metalls com el níquel, el coure o l'alumini, i també al preu de productes d'alimentació bàsics, com és el cas del blat.

En aquest treball s'ha tractat als mitjans d'ideologies oposades com a contraris, ja que, així ho requeria l'anàlisi segons les hipòtesis plantejades. Tanmateix, si l'objectiu és exclusivament saber com parlen els mitjans i si depèn de la seva ideologia la manera d'explicar les idees, es podria repetir l'anàlisi sense focalitzar-lo en un fet concret i obtenint totes les notícies dels mitjans durant un període concret.

VII. BIBLIOGRAFIA

ABC.es. (2022). *ABC - Tu diario en español - ABC.es*. abc. https://www.abc.es

24/02/2022 - 24/03/2022

Cinco Días. (2022). *Cinco Días: economía y mercados*. Cinco Días.

https://cincodias.elpais.com

24/02/2022 - 24/03/2022

Coder, R. (2022). *Paletas de colores en R*. R CHARTS. https://r-charts.com/es/paletas-colores/

elEconomista. (2022). *Líder en noticias de economía, bolsa y finanzas. -elEconomista.es*. elEconomista.es. https://www.eleconomista.es

24/02/2022 - 24/03/2022

Expansión - Diario Económico e información de mercados. (2022). EXPANSION.

https://www.expansion.com

24/02/2022 - 24/03/2022

EL MUNDO - Diario online líder de información en español. (2022). ELMUNDO.

https://www.elmundo.es

24/02/2022 - 24/03/2022

La Vanguardia – Últimas noticias, actualidad y última hora en Cataluña, España y el mundo. (2022). La Vanguardia. https://www.lavanguardia.com

24/02/2022 - 24/03/2022

My News SL. (2022). *MyNews: La hemeroteca digital de prensa escrita moderna de los periódicos españoles.* Hemeroteca MyNews. https://hemeroteca.mynews.es

MyNews, Hemeroteca Digital y Seguimiento de Medios Online. (2022, 22 marzo). My News.

https://mynews.es

Padinger, G. (2022, 24 mayo). *¿Por qué Rusia atacó e invadió Ucrania? ¿Cuáles son los motivos y el origen del conflicto?* CNN. https://cnnespanol.cnn.com/2022/05/24/por-que-rusia-ucrania-guerra-invasion-motivos-orix/

País, E. (2022). *EL PAÍS: el periódico global*. El País. https://elpais.com
24/02/2022 - 24/03/2022

Pereira, G. (2022, 3 junio). *A 100 días de la invasión, por qué es la guerra entre Rusia y Ucrania: qué pasó en 2014*. ECC. https://www.cronista.com/internacionales/a-100-dias-de-la-invasion-por-que-es-la-guerra-entre-rusia-y-ucrania-que-paso-en-2014/

Periódico, E. (2022). *El Periódico - Noticias y última hora de Catalunya, España y el mundo*. elperiodico. https://www.elperiodico.com/es/
24/02/2022 - 24/03/2022

Público. (2022). Público.es. https://www.publico.es
24/02/2022 - 24/03/2022

La Razón - Diario de Noticias de España y Actualidad. (2022). La Razón. https://www.larazon.es
24/02/2022 - 24/03/2022

Silge, J., & Robinson, D. (2017). *Text Mining with R: A Tidy Approach*. O'Reilly Media.

Team, V. (2021a, julio 28). *Introducción Al Text Mining Con R (parte 2)*. Visionarios. https://blogvisionarios.com/articulos-data/introduccion-al-text-mining-con-r-parte-2/

Team, V. (2021b, julio 28). *Introducción Al Text Mining Con R: Parte I*. Visionarios. https://blogvisionarios.com/articulos-data/introduccion-al-text-mining-con-r-parte-i/

Universidad de Sevilla. (2020, marzo). Utilidades para documentos R Markdown. https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwiv_KKxqbT4AhWD0YUKHcHrCLgQFnoECAQQAQ&url=http%3A%2F%2Fd

estio.us.es%2Fcalvo%2Fmemoriatfe%2FMemoriaTFE_v1.2_paraweb.pdf&usg=AOvVaw08xXnp13dIJs92gRBc61mS

CODI

S'ha penjat al codi i una mostra de les dades a la plataforma GitHub per tal de poder-lo executar i examinar.

A continuació es troba l'enllaç per accedir-hi:

https://github.com/ccastellano/TFG2022

ANNEX

III. ANÀLISI DE PATRONS DE PARLA I EXPRESSIONS

1. Anàlisi exploratòria de les dades

```
## Ideologia Esquerra
el_periodico <- read_excel("periodico.xls", col_names = TRUE)
publico <- read_excel("publico.xls", col_names = TRUE)

esquerra <- rbind(el_periodico, publico)

## Ideologia Centre
el_pais <- read_excel("pais.xls", col_names = TRUE)
la_vanguardia <- read_excel("vanguardia.xls", col_names = TRUE)

centre <- rbind(el_pais, la_vanguardia)

## Ideologia Dreta
el_mundo <- read_excel("mundo.xls", col_names = TRUE)
la_razon <- read_excel("razon.xls", col_names = TRUE)
abc <- read_excel("abc.xls", col_names = TRUE)

dreta <- rbind(el_mundo, la_razon,abc)
tots <- rbind(esquerra,centre,dreta) #TOTS JUNTS

cols <- c("El Periódico" = "#7B3014", "publico.es" = "#D94602", "El País"
= "#FD8E3F", "La Vanguardia" = "#BCCACF", "El Mundo" = "#5BA2CC", "La Razó
n"= "#1F74B1", "ABC"="#26456E")

paleta <- c("#7B3014", "#8B310E", "#9C3206", "#AF3602","#C53E01","#D94602"
,
  "#E35408", "#ED620F", "#F4711F", "#FA7F2E", "#FD8E3F", "#FE9C52",
  "#FDAA65","#EEB78D","#D8C4B6", "#BCCACF", "#9EC9D9","#7AC7E2",
  "#72BCDC","#6AB1D6","#5BA2CC","#4993C0","#3885B6","#2D7DB4","#1F74B1",
  "#1C6AA8","#1C5F9E","#1F5591","#244D7F","#26456E")

#mitjana del valor economic i l'audiencia de les publicacions de la ideolo
gia esquerra
dt1 <- summarise(
    group_by(esquerra, Publicacion),
    n = n(),
    Mean = round(mean(Valor),0),
    Sd = sd(Valor),
    Var = var(Valor),
    Q1 = quantile(Valor, .25),
    Q2 = quantile(Valor, .50),
    Q3 = quantile(Valor, .75),
    Min = min(Valor),
    Max = max(Valor))

dt2 <- summarise(
```

```
  group_by(esquerra, Publicacion),
  n = n(),
  Mean = round(mean(Audiencia),0),
  Sd = sd(Audiencia),
  Var = var(Audiencia),
  Q1 = quantile(Audiencia, .25),
  Q2 = quantile(Audiencia, .50),
  Q3 = quantile(Audiencia, .75),
  Min = min(Audiencia),
  Max = max(Audiencia))

dt1 %>%
  kable() %>%
  kable_styling(latex_options = c("condensed","hold_position"),
                position = "center", full_width = FALSE) %>%
                row_spec(0, bold = T)  %>%
                add_header_above(c("Valor publicitari equivalent"=10))

kable(dt2) %>%
  kable_styling(latex_options = c("condensed","hold_position"),
                position = "center", full_width = FALSE) %>%
                row_spec(0, bold = T) %>%
                add_header_above(c("Audiència"=10))

#mitjana del valor economic i l'audiencia de les publicacions de la ideolo
gia centre
dt3 <- summarise(
  group_by(centre, Publicacion),
  n = n(),
  Mean = round(mean(Valor),0),
  Sd = sd(Valor),
  Var = var(Valor),
  Q1 = quantile(Valor, .25),
  Q2 = quantile(Valor, .50),
  Q3 = quantile(Valor, .75),
  Min = min(Valor),
  Max = max(Valor))

dt4 <- summarise(
  group_by(centre, Publicacion),
  n = n(),
  Mean = round(mean(Audiencia),0),
  Sd = sd(Audiencia),
  Var = var(Audiencia),
  Q1 = quantile(Audiencia, .25),
  Q2 = quantile(Audiencia, .50),
  Q3 = quantile(Audiencia, .75),
  Min = min(Audiencia),
  Max = max(Audiencia))

kable(dt3) %>%
  kable_styling(latex_options = c("condensed","hold_position"),
                position = "center", full_width = FALSE) %>%
                row_spec(0, bold = T) %>%
                add_header_above(c("Valor publicitari equivalent"=10))
```

```r
kable(dt4) %>%
  kable_styling(latex_options = c("condensed","hold_position"),
                position = "center", full_width = FALSE) %>%
                row_spec(0, bold = T) %>%
                add_header_above(c("Audiència"=10))

#mitjana del valor economic i l'audiencia de les publicacions de la ideolo
gia dreta
dt5<- summarise(
  group_by(dreta, Publicacion),
  n = n(),
  Mean = round(mean(Valor),0),
  Sd = sd(Valor),
  Var = var(Valor),
  Q1 = quantile(Valor, .25),
  Q2 = quantile(Valor, .50),
  Q3 = quantile(Valor, .75),
  Min = min(Valor),
  Max = max(Valor))

dt6 <- summarise(
  group_by(dreta, Publicacion),
  n = n(),
  Mean = round(mean(Audiencia),0),
  Sd = sd(Audiencia),
  Var = var(Audiencia),
  Q1 = quantile(Audiencia, .25),
  Q2 = quantile(Audiencia, .50),
  Q3 = quantile(Audiencia, .75),
  Min = min(Audiencia),
  Max = max(Audiencia))

kable(dt5) %>%
  kable_styling(latex_options = c("condensed","hold_position"),
                position = "center", full_width = FALSE) %>%
                row_spec(0, bold = T) %>%
                add_header_above(c("Valor publicitari equivalent"=10))

kable(dt6) %>%
  kable_styling(latex_options = c("condensed","hold_position"),
                position = "center", full_width = FALSE) %>%
                row_spec(0, bold = T) %>%
                add_header_above(c("Audiència"=10))

#Nombre de notícies per mitjà
tots %>%
  group_by(Publicacion) %>%
  summarize(news = n_distinct(IdDocument)) %>%
  ggplot(aes(news, reorder(Publicacion, news), fill= Publicacion)) +
  geom_col() +
  ggtitle("Nombre de notícies per mitjà") +
  scale_colour_manual(values = cols, aesthetics = c("fill")) +
  theme_bw() +
  theme(legend.position="none") +
```

```
  labs(x = "Nombre de notícies",
       y = NULL)

# grafic de les publicacions cronologieques per mitja
ggplot(tots, aes(x = as.Date(Fecha), fill = Publicacion)) +
  geom_histogram(position = "identity", bins = 25, show.legend = FALSE) +
  scale_fill_manual(values = cols, aesthetics = c("fill")) +
  scale_x_date(date_labels = "%d-%m", date_breaks = "5 days") +
  labs(x = "Data de publicació", y = "Nombre de notícies") +
  facet_wrap(~ Publicacion, ncol = 3) +
  ggtitle("Cronograma de notícies per mitjà") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 90))

#Cronograma del nombre de notícies publicades segons la data de publicació
tots %>%
  mutate(mes_any = format(Fecha, "%m-%d")) %>%
  group_by(Publicacion, mes_any) %>%
  summarise(n = n()) %>%
  ggplot(aes(x = mes_any, y = n, color = Publicacion)) +
  geom_line(aes(group = Publicacion)) +
  scale_fill_manual(values = cols, aesthetics = c("color")) +
  labs(x = "Data de publicació", y = "Nombre de notícies") +
  ggtitle("Cronograma de notícies") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 90, size = 6, hjust = 1), legen
d.position = "bottom", legend.title = element_blank())
```

2. Anàlisi semàntica de les dades

```
## Ideologia Esquerra
per <- xmlToDataFrame("periodico.xml")
pub <- xmlToDataFrame("publico.xml")

## Ideologia Centre
pai <- xmlToDataFrame("pais.xml")
van <- xmlToDataFrame("vanguardia.xml")

## Ideologia Dreta
mun <- xmlToDataFrame("mundo.xml")
raz <- xmlToDataFrame("razon.xml")
ab <- xmlToDataFrame("abc.xml")

m <- c(rep("PERIODICO",length(per[,1])),
       rep("PUBLICO",length(pub[,1])),
       rep("EL_PAIS",length(pai[,1])),
       rep("LA_VANGUARDIA",length(van[,1])),
       rep("EL_MUNDO",length(mun[,1])),
       rep("LA_RAZON",length(raz[,1])),
       rep("ABC",length(ab[,1])))

dades <- rbind(per,pub,pai,van,mun,raz,ab)
dades <- cbind(dades,m)
```

```r
#Neteja del text
text <- dades[,"body"]

#Conviertim el text a minúsculas
text <- tolower(text)
# Eliminem pàgines web
text <- str_replace_all(text,"http\\S*", "")
text <- str_replace_all(text,"www\\S*", "")
# Eliminem signes de puntuació
text <- removePunctuation(text)
# Eliminem els números
text <- removeNumbers(text)
# Eliminem de espais múltiples en blanc
text <- stripWhitespace(text)
text <- stri_trans_general(text, "Latin-ASCII")
#Eliminem les stopwords
text <- removeWords(text, words = stopwords("spanish"))
#Elimnare paraules que no tenen significat o no ens interessen
fora <- read.csv("paraules.csv", header=F)
vector <- c(fora$V1)
text <- removeWords(text, vector)

dades <- data.frame(mitja = dades$m, text = text)

#Separem el text en paraules
mitja_paraules <- dades %>%
  unnest_tokens(word, text) %>%
  filter(str_detect(word, "[a-z']$"))

colnames(mitja_paraules) <- c("mitja", "paraula")

# Eliminación de tokens con una longitud < 3
mitja_paraules <- mitja_paraules %>%
  filter(nchar(paraula) > 3 )

# Nombre de vegades que cada marca utilitza certes paraules
paraules_per_marca <- mitja_paraules %>%
  count(mitja, paraula, sort = TRUE) %>%
  ungroup()

#ajuntar paraules similars
canvi <- read.csv("canvi.csv", header = T, sep = ";")

canvi <- canvi %>%
  pivot_longer(cols = -DEFINITIVA) %>%
  dplyr::select(-name) %>%
  drop_na()

for (i in 1:nrow(paraules_per_marca)) {
  if(paraules_per_marca$paraula[i] %in% canvi$value) {
    paraules_per_marca$paraula[i] <- canvi$DEFINITIVA[canvi$value == parau
les_per_marca$paraula[i]]
  } else {
```

```
    paraules_per_marca$paraula[i] <- paraules_per_marca$paraula[i]
  }
}

saveRDS(paraules_per_marca, file = "paraules_per_marca.Rds") #desem les da
des

paraules_per_marca <- readRDS("paraules_per_marca.Rds") #llegim les dades

# Nombre de vegades que cada marca utilitza certes paraules
paraules_per_marca <- paraules_per_marca %>%
 group_by(mitja, paraula) %>%
 summarise_at(vars(n), list(n = sum)) %>%
 filter(n > 15)

colors <- c("PERIODICO" = "#7B3014", "PUBLICO" = "#D94602", "EL_PAIS" = "#
FD8E3F",
           "LA_VANGUARDIA" = "#BCCACF", "EL_MUNDO" = "#5BA2CC",
           "LA_RAZON"= "#1F74B1", "ABC"="#26456E")

paleta <- c("#7B3014", "#8B310E", "#9C3206", "#AF3602","#C53E01","#D94602"
,
  "#E35408", "#ED620F", "#F4711F", "#FA7F2E", "#FD8E3F", "#FE9C52",
  "#FDAA65","#EEB78D","#D8C4B6", "#BCCACF", "#9EC9D9","#7AC7E2",
  "#72BCDC","#6AB1D6","#5BA2CC","#4993C0","#3885B6","#2D7DB4","#1F74B1",
  "#1C6AA8","#1C5F9E","#1F5591","#244D7F","#26456E")

p1 <- paleta[1:8]
p2 <- paleta[9:18]
p3 <- paleta[19:30]

ES <- paraules_per_marca %>%
   filter(mitja %in% c("PERIODICO", "PUBLICO"))%>%
   select(paraula,n) %>%
   group_by(paraula) %>%
   summarise_at(vars(n), list(n = sum))

esquerra <- as.vector(ES$paraula)
esquerra <- as.character(esquerra)

CEN <-paraules_per_marca %>%
   filter(mitja %in% c("EL_PAIS", "LA_VANGUARDIA"))%>%
   select(paraula,n) %>%
   group_by(paraula) %>%
   summarise_at(vars(n), list(n = sum))

centre <- as.vector(CEN$paraula)
centre <- as.character(centre)

DRE <- paraules_per_marca %>%
   filter(mitja %in% c("EL_MUNDO", "ABC", "LA_RAZON")) %>%
   select(paraula,n) %>%
   group_by(paraula) %>%
```

```
   summarise_at(vars(n), list(n = sum))

dreta <- as.vector(DRE$paraula)
dreta <- as.character(dreta)

i <- c(rep("ESQUERRA",nrow(ES)),
       rep("DRETA",nrow(DRE)))

esdre <- rbind(ES,DRE)
esdre <- cbind(esdre,i)
```

2.1. Freqüència absoluta

```
# Les visualitzem per les marques
paraules_per_marca %>%
  filter(mitja %in% c("PERIODICO","PUBLICO", "EL_PAIS", "LA_VANGUARDIA"))
%>%
   group_by(mitja) %>%
   top_n(20, n) %>%
   arrange(mitja, desc(n)) %>%
   ggplot(aes(x = reorder(paraula,n), y = n, fill = mitja)) +
   scale_fill_manual(values = colors, aesthetics = c("fill")) +
   geom_col() +
   theme_bw() +
   labs(y = "Freqüència absoluta", x = "") +
   theme(legend.position = "none", axis.text.y = element_text(size = 7)
         ) +
   coord_flip() +
   ggtitle("20 conceptes amb més freqüència absoluta per a cada mitjà")+
   facet_wrap(~ mitja, scales = "free", nrow = 2)

paraules_per_marca %>%
  filter(mitja %in% c("EL_MUNDO", "LA_RAZON", "ABC")) %>%
   group_by(mitja) %>%
   top_n(20, n) %>%
   arrange(mitja, desc(n)) %>%
   ggplot(aes(x = reorder(paraula,n), y = n, fill = mitja)) +
   scale_fill_manual(values = colors, aesthetics = c("fill")) +
   geom_col() +
   theme_bw() +
   labs(y = "Freqüència absoluta", x = "") +
   theme(legend.position = "none", axis.text.y = element_text(size = 7)
         ) +
   coord_flip() +
     ggtitle("(bis) - 20 conceptes amb més freqüència absoluta per a cada
mitjà")+
   facet_wrap(~ mitja, scales = "free", nrow = 2)

perio <- paraules_per_marca %>%
  filter(mitja %in% c("PERIODICO")) %>%
   group_by(mitja) %>%
   top_n(20, n) %>%
   arrange(mitja, desc(n)) %>%
   ggplot(aes(x = reorder(paraula,n), y = n, fill = mitja)) +
```

```
   scale_fill_manual(values = colors, aesthetics = c("fill")) +
   geom_col() +
   theme_bw() +
   labs(y = "Freqüència absoluta", x = "") +
   theme(legend.position = "none") +
   coord_flip() +
   facet_wrap(~ mitja, scales = "free", drop = TRUE)

publi <- paraules_per_marca %>%
  filter(mitja %in% c("PUBLICO")) %>%
   group_by(mitja) %>%
   top_n(20, n) %>%
   arrange(mitja, desc(n)) %>%
   ggplot(aes(x = reorder(paraula,n), y = n, fill = mitja)) +
   scale_fill_manual(values = colors, aesthetics = c("fill")) +
   geom_col() +
   theme_bw() +
   labs(y = "Freqüència absoluta", x = "") +
   theme(legend.position = "none") +
   coord_flip() +
   facet_wrap(~ mitja, scales = "free", drop = TRUE)

p <- perio + publi
p +plot_annotation(title = '20 conceptes amb més freqüència abbsoluta dels
mitjans d\'esquerra')

wordcloud(words = ES$paraula, freq = ES$n, scale = , min.freq = 20,
          max.words=80, random.order=FALSE, rot.per=0.35, colors = p1)

pais <- paraules_per_marca %>%
  filter(mitja %in% c("EL_PAIS")) %>%
   group_by(mitja) %>%
   top_n(20, n) %>%
   arrange(mitja, desc(n)) %>%
   ggplot(aes(x = reorder(paraula,n), y = n, fill = mitja)) +
   scale_fill_manual(values = colors, aesthetics = c("fill")) +
   geom_col() +
   theme_bw() +
   labs(y = "Freqüència absoluta ", x = "") +
   theme(legend.position = "none") +
   coord_flip() +
   facet_wrap(~ mitja, scales = "free", drop = TRUE)

vang <- paraules_per_marca %>%
  filter(mitja %in% c("LA_VANGUARDIA")) %>%
   group_by(mitja) %>%
   top_n(20, n) %>%
   arrange(mitja, desc(n)) %>%
   ggplot(aes(x = reorder(paraula,n), y = n, fill = mitja)) +
   scale_fill_manual(values = colors, aesthetics = c("fill")) +
   geom_col() +
   theme_bw() +
   labs(y = "Freqüència absoluta ", x = "") +
   theme(legend.position = "none") +
```

```
   coord_flip() +
   facet_wrap(~ mitja, scales = "free", drop = TRUE)

t <- pais + vang
t +plot_annotation(title = '20 conceptes amb més freqüència abbsoluta dels
mitjans de centre')

wordcloud(words = CEN$paraula, freq = CEN$n, scale = , min.freq = 20,
          max.words=80, random.order=FALSE,rot.per=0.35, colors = p2)

raz <-paraules_per_marca %>%
  filter(mitja %in% c("LA_RAZON")) %>%
   group_by(mitja) %>%
   top_n(20, n) %>%
   arrange(mitja, desc(n)) %>%
   ggplot(aes(x = reorder(paraula,n), y = n, fill = mitja)) +
   scale_fill_manual(values = colors, aesthetics = c("fill")) +
   geom_col() +
   theme_bw() +
   labs(y = "Freqüència absoluta ", x = "") +
   theme(legend.position = "none") +
   coord_flip() +
   facet_wrap(~ mitja, scales = "free", drop = TRUE)

mund <- paraules_per_marca %>%
  filter(mitja %in% c("EL_MUNDO")) %>%
   group_by(mitja) %>%
   top_n(20, n) %>%
   arrange(mitja, desc(n)) %>%
   ggplot(aes(x = reorder(paraula,n), y = n, fill = mitja)) +
   scale_fill_manual(values = colors, aesthetics = c("fill")) +
   geom_col() +
   theme_bw() +
   labs(y = " Freqüència absoluta", x = "") +
   theme(legend.position = "none") +
   coord_flip() +
   facet_wrap(~ mitja, scales = "free", drop = TRUE)

ab <-paraules_per_marca %>%
  filter(mitja %in% c("ABC")) %>%
   group_by(mitja) %>%
   top_n(20, n) %>%
   arrange(mitja, desc(n)) %>%
   ggplot(aes(x = reorder(paraula,n), y = n, fill = mitja)) +
   scale_fill_manual(values = colors, aesthetics = c("fill")) +
   geom_col() +
   theme_bw() +
   labs(y = " Freqüència absoluta", x = "") +
   theme(legend.position = "none") +
   coord_flip() +
   facet_wrap(~ mitja, scales = "free", drop = TRUE)

z <- raz + mund + ab
```

```
z +plot_annotation(title = '20 conceptes amb més freqüència abbsoluta dels
mitjans de dreta')

wordcloud(words = DRE$paraula, freq = DRE$n, scale = , min.freq = 20,
          max.words=80, random.order=FALSE, rot.per=0.35, colors = p3, tit
le ="Wordcloud de dreta")
```

2.2. Rellevància dels termes

```
# Busquem les paraules que més defineixen els mitjans
tf_idf <- paraules_per_marca %>%
   bind_tf_idf(paraula, mitja, n) %>%
    arrange(desc(tf_idf))

# Les visualitzem per els mitjans
tf_idf %>%
  filter(mitja %in% c("PERIODICO","PUBLICO", "EL_PAIS", "LA_VANGUARDIA"))
%>%
  group_by(mitja) %>%
  slice_max(tf_idf, n = 15) %>%
  ungroup() %>%
  mutate(word = reorder(paraula, tf_idf)) %>%
  ggplot(aes(tf_idf, word, fill = mitja)) +
  scale_fill_manual(values = colors, aesthetics = c("fill")) +
  geom_col(show.legend = FALSE) +
  theme_bw() +
  ggtitle("15 conceptes amb més rellevància per a cada mitjà") +
  facet_wrap(~ mitja, scales = "free") +
  labs(x = "tf-idf", y = NULL)

# Les visualitzem per els mitjans
tf_idf %>%
  filter(mitja %in% c("EL_MUNDO", "LA_RAZON", "ABC")) %>%
  group_by(mitja) %>%
  slice_max(tf_idf, n = 15) %>%
  ungroup() %>%
  mutate(word = reorder(paraula, tf_idf)) %>%
  ggplot(aes(tf_idf, word, fill = mitja)) +
  scale_fill_manual(values = colors, aesthetics = c("fill")) +
  geom_col(show.legend = FALSE) +
  theme_bw() +
  facet_wrap(~ mitja, scales = "free") +
  ggtitle("(bis) - 15 conceptes amb més rellevància per a cada mitjà")
  labs(x = "tf-idf", y = NULL)
```

2.3. Correlació entre mitjans i ideologies

```
## GRAFICS DE CORRELACIÓ ENTRE 2 MITJANS
paraules_per_marca1 <- paraules_per_marca %>%
  group_by(mitja) %>%
  mutate(freq=n/n())

## mitjans d'ESQUERRA
paraules_per_marca1 %>%
  filter(mitja %in% c("PERIODICO", "PUBLICO")) %>%
```

```
  select(-n) %>%
  pivot_wider(names_from = mitja, values_from = freq) %>%
  ggplot(aes(PERIODICO, PUBLICO)) +
  geom_jitter(alpha = 0.1, size = 2.5, width = 0.25, height = 0.25) +
  geom_text(aes(label = paraula), check_overlap = TRUE, vjust = 1.5, size
= 3) +
  scale_x_log10(labels = percent_format()) +
  scale_y_log10(labels = percent_format()) +
  ggtitle("Correlació entre PERIODICO i PÚBLICO") +
  theme_minimal() +
  geom_abline(color = "red")

## mitjans de centre
paraules_per_marca1 %>%
  filter(mitja %in% c("EL_PAIS", "LA_VANGUARDIA")) %>%
  select(-n) %>%
  pivot_wider(names_from = mitja, values_from = freq) %>%
  ggplot(aes(EL_PAIS, LA_VANGUARDIA)) +
  geom_jitter(alpha = 0.1, size = 2.5, width = 0.25, height = 0.25) +
  geom_text(aes(label = paraula), check_overlap = TRUE, vjust = 1.5, size
= 3) +
  scale_x_log10(labels = percent_format()) +
  scale_y_log10(labels = percent_format()) +
  ggtitle("Correlació entre EL PAÍS i LA VANGUARDIA") +
  theme_minimal() +
  geom_abline(color = "red")

## mitjans de dreta
g1 <- paraules_per_marca1 %>%
  filter(mitja %in% c("ABC", "LA_RAZON")) %>%
  select(-n) %>%
  pivot_wider(names_from = mitja, values_from = freq) %>%
  ggplot(aes(ABC, LA_RAZON)) +
  geom_jitter(alpha = 0.1, size = 2.5, width = 0.25, height = 0.25) +
  geom_text(aes(label = paraula), check_overlap = TRUE, vjust = 1.5, size
= 3) +
  scale_x_log10(labels = percent_format()) +
  scale_y_log10(labels = percent_format()) +
  ggtitle("Correlació entre ABC i LA RAZÓN") +
  theme_minimal() +
  geom_abline(color = "red")

g2 <- paraules_per_marca1 %>%
  filter(mitja %in% c("EL_MUNDO", "LA_RAZON")) %>%
  select(-n) %>%
  pivot_wider(names_from = mitja, values_from = freq) %>%
  ggplot(aes(EL_MUNDO, LA_RAZON)) +
  geom_jitter(alpha = 0.1, size = 2.5, width = 0.25, height = 0.25) +
  geom_text(aes(label = paraula), check_overlap = TRUE, vjust = 1.5, size
= 3) +
  scale_x_log10(labels = percent_format()) +
  scale_y_log10(labels = percent_format()) +
  ggtitle("Correlació entre LA RAZÓN i EL MUNDO") +
  theme_minimal() +
```

```
  geom_abline(color = "red")

g3 <- paraules_per_marca1 %>%
  filter(mitja %in% c("EL_MUNDO", "ABC")) %>%
  select(-n) %>%
  pivot_wider(names_from = mitja, values_from = freq) %>%
  ggplot(aes(EL_MUNDO, ABC)) +
  geom_jitter(alpha = 0.1, size = 2.5, width = 0.25, height = 0.25) +
  geom_text(aes(label = paraula), check_overlap = TRUE, vjust = 1.5, size
= 3) +
  scale_x_log10(labels = percent_format()) +
  scale_y_log10(labels = percent_format()) +
  ggtitle("Correlació entre EL MUNDO i ABC") +
  theme_minimal() +
  geom_abline(color = "red")

# ddaes per la correlació dreta contra esquerra

esdre <- esdre %>%
  group_by(i) %>%
  mutate(freq=n/n())

un <- esdre %>% filter(i=="DRETA") %>% select(paraula)
dos <- esdre %>% filter(i=="ESQUERRA") %>% select(paraula)
un <- data_frame(un$paraula)
dos <- data_frame(dos$paraula)
colnames(dos) <- c("V1")
colnames(un) <- c("V1")

comunes <- nrow(intersect(un,dos))

#Correlació dreta contra esquerra
esdre %>%
  filter(i %in% c("DRETA","ESQUERRA")) %>%
  select(-n) %>%
  pivot_wider(names_from = i, values_from = freq) %>%
  ggplot(aes(DRETA, ESQUERRA)) +
  geom_jitter(alpha = 0.1, size = 2.5, width = 0.25, height = 0.25) +
  geom_text(aes(label = paraula), check_overlap = TRUE, vjust = 1.5, size
= 3) +
  scale_x_log10(labels = percent_format()) +
  scale_y_log10(labels = percent_format()) +
  ggtitle("Correlació entre DRETA i ESQUERRA") +
  theme_minimal() +
  geom_abline(color = "red")
```

2.4. Diferències de termes

```
#Paraules que diferenecien més els mitjans
# Cálcul de odds i log of odds de cada paraula (esquerra)
logo <- paraules_per_marca1 %>%
   filter(mitja %in% c("PERIODICO", "PUBLICO")) %>%
   group_by(mitja) %>%
   left_join(paraules_per_marca1 %>%
```

```
                group_by(mitja) %>%
                summarise(N = n(), by= "mitja")) %>%
   mutate(odds = (n + 1) / (N + 1)) %>%
   select(mitja, paraula, odds) %>%
   spread(key = mitja, value = odds) %>%
   mutate(log_odds = log(PERIODICO/PUBLICO), abs_log_odds = abs(log_odds))
%>%
   mutate(freq_mitja = if_else(log_odds > 0,"PERIODICO", "PUBLICO")) %>%
   arrange(desc(abs_log_odds))

primera <- c("PERIODICO" = "#7B3014", "PUBLICO" = "#D94602")

logo %>%
   group_by(freq_mitja) %>%
   top_n(10, abs_log_odds) %>%
   ggplot(aes(x = reorder(paraula, log_odds), y = log_odds,
              fill = freq_mitja)) +
   scale_fill_manual(values = primera, aesthetics = c("fill")) +
   geom_col() +
   labs(x = "paraula", y = "log odds ratio (PERIODICO/PUBLICO)") +
   coord_flip() +
   ggtitle("Odds ratio d'esquerra")+
   theme_bw() +
   theme(legend.title = element_blank())

# Cálcul de odds i log of odds de cada paraula (centre)
logo <- paraules_per_marca1 %>%
   filter(mitja %in% c("LA_VANGUARDIA", "EL_PAIS")) %>%
   group_by(mitja) %>%
   left_join(paraules_per_marca1 %>%
                group_by(mitja) %>%
                summarise(N = n(), by= "mitja")) %>%
   mutate(odds = (n + 1) / (N + 1)) %>%
   select(mitja, paraula, odds) %>%
   spread(key = mitja, value = odds) %>%
   mutate(log_odds = log(LA_VANGUARDIA/EL_PAIS),
          abs_log_odds = abs(log_odds)) %>%
   mutate(freq_mitja = if_else(log_odds > 0,"LA_VANGUARDIA", "EL_PAIS")) %
>%
   arrange(desc(abs_log_odds))

segona <- c("EL_PAIS" = "#FD8E3F", "LA_VANGUARDIA" = "#BCCACF")

logo %>%
   group_by(freq_mitja) %>%
   top_n(10, abs_log_odds) %>%
   ggplot(aes(x = reorder(paraula, log_odds), y = log_odds,
              fill = freq_mitja)) +
   scale_fill_manual(values = segona, aesthetics = c("fill")) +
   geom_col() +
   labs(x = "paraula", y = "log odds ratio (LA_VANGUARDIA/EL_PAIS)") +
   coord_flip() +
   ggtitle("Odds ratio de centre")+
```

```
   theme_bw() +
   theme(legend.title = element_blank())

# Cálcul de odds i log of odds de cada paraula (dreta)
logo <- paraules_per_marca1 %>%
   filter(mitja %in% c("EL_MUNDO", "ABC")) %>%
   group_by(mitja) %>%
   left_join(paraules_per_marca1 %>%
                group_by(mitja) %>%
                summarise(N = n(), by= "mitja")) %>%
   mutate(odds = (n + 1) / (N + 1)) %>%
   select(mitja, paraula, odds) %>%
   spread(key = mitja, value = odds) %>%
   mutate(log_odds = log(EL_MUNDO/ABC), abs_log_odds = abs(log_odds)) %>%
   mutate(freq_mitja = if_else(log_odds > 0,"EL_MUNDO", "ABC")) %>%
   arrange(desc(abs_log_odds))

tercera <- c("EL_MUNDO" = "#5BA2CC", "ABC"="#26456E")

d1 <-logo %>%
   group_by(freq_mitja) %>%
   top_n(10, abs_log_odds) %>%
   ggplot(aes(x = reorder(paraula, log_odds), y = log_odds,
              fill = freq_mitja)) +
   scale_fill_manual(values = tercera, aesthetics = c("fill")) +
   geom_col() +
   labs(x = "paraula", y = "log odds ratio (EL_MUNDO/ABC)") +
   coord_flip() +
   theme_bw() +
   theme(legend.title = element_blank())

logo <- paraules_per_marca1 %>%
   filter(mitja %in% c("EL_MUNDO", "LA_RAZON")) %>%
   group_by(mitja) %>%
   left_join(paraules_per_marca1 %>%
                group_by(mitja) %>%
                summarise(N = n(), by= "mitja")) %>%
   mutate(odds = (n + 1) / (N + 1)) %>%
   select(mitja, paraula, odds) %>%
   spread(key = mitja, value = odds) %>%
   mutate(log_odds = log(EL_MUNDO/LA_RAZON), abs_log_odds = abs(log_odds))
%>%
   mutate(freq_mitja = if_else(log_odds > 0,"EL_MUNDO", "LA_RAZON")) %>%
   arrange(desc(abs_log_odds))

quarta <- c("EL_MUNDO" = "#5BA2CC", "LA_RAZON"= "#1F74B1")

d2 <-logo %>%
   group_by(freq_mitja) %>%
   top_n(10, abs_log_odds) %>%
   ggplot(aes(x = reorder(paraula, log_odds), y = log_odds,
              fill = freq_mitja)) +
   scale_fill_manual(values = quarta, aesthetics = c("fill")) +
   geom_col() +
```

```
   labs(x = "paraula", y = "log odds ratio (EL_MUNDO/LA_RAZON)") +
   coord_flip() +
   theme_bw() +
   theme(legend.title = element_blank())

logo <- paraules_per_marca1 %>%
   filter(mitja %in% c("LA_RAZON", "ABC")) %>%
   group_by(mitja) %>%
   left_join(paraules_per_marca1 %>%
                group_by(mitja) %>%
                summarise(N = n(), by= "mitja")) %>%
   mutate(odds = (n + 1) / (N + 1)) %>%
   select(mitja, paraula, odds) %>%
   spread(key = mitja, value = odds) %>%
   mutate(log_odds = log(LA_RAZON/ABC), abs_log_odds = abs(log_odds)) %>%
   mutate(freq_mitja = if_else(log_odds > 0,"LA_RAZON", "ABC")) %>%
   arrange(desc(abs_log_odds))

cinquena <- c("LA_RAZON"= "#1F74B1", "ABC"="#26456E")

d3 <-logo %>%
   group_by(freq_mitja) %>%
   top_n(10, abs_log_odds) %>%
   ggplot(aes(x = reorder(paraula, log_odds), y = log_odds,
              fill = freq_mitja)) +
   scale_fill_manual(values = cinquena, aesthetics = c("fill")) +
   geom_col() +
   labs(x = "paraula", y = "log odds ratio (LA_RAZON/ABC)") +
   coord_flip() +
   theme_bw() +
   theme(legend.title = element_blank())

s <- d1 + d2 + d3
s + plot_annotation(title = "Odds ratio de dreta")

# Cálcul de odds i log of odds de cada paraula (esquerra contra dreta)
logo <- esdre %>%
   filter(i %in% c("DRETA", "ESQUERRA")) %>%
   group_by(i) %>%
   left_join(esdre %>%
                group_by(i) %>%
                summarise(N = n(), by= "i")) %>%
   mutate(odds = (n + 1) / (N + 1)) %>%
   select(i, paraula, odds) %>%
   spread(key = i, value = odds) %>%
   mutate(log_odds = log(DRETA/ESQUERRA), abs_log_odds = abs(log_odds)) %>
%
   mutate(freq_mitja = if_else(log_odds > 0,"DRETA", "ESQUERRA")) %>%
   arrange(desc(abs_log_odds))

ultima <- c("ESQUERRA" = "#8B310E", "DRETA" = "#244D7F")

logo %>%
   group_by(freq_mitja) %>%
```

```
   top_n(10, abs_log_odds) %>%
   ggplot(aes(x = reorder(paraula, log_odds), y = log_odds,
              fill = freq_mitja)) +
   scale_fill_manual(values = ultima, aesthetics = c("fill")) +
   geom_col() +
   labs(x = "paraula", y = "log odds ratio (DRETA/ESQUERRA)") +
   coord_flip() +
   ggtitle("Odds ratio de dreta i esquerra") +
   theme_bw() +
   theme(legend.title = element_blank())
```

2.5. PCA

```
#PCA

dadesPCA <- paraules_per_marca %>%
   pivot_wider(names_from = paraula, values_from = n)

dadesPCA[is.na(dadesPCA)] <- 0

pca_res <- prcomp(dadesPCA[, 2:ncol(dadesPCA)])

pca_res$x %>%
   as.data.frame() %>%
   dplyr::select(PC1, PC2) %>%
   mutate(mitja = dadesPCA$mitja) %>%
   ggplot(aes(PC1, PC2, label = mitja, color = mitja)) +
  ggtitle("Anàlisi de components principals") +
   ggrepel::geom_text_repel(check_overlap = TRUE, show.legend = FALSE) +
  scale_color_manual(values = colors) +
   theme_bw()
```

2.6. Diagrama de Venn

```
#Diagrama de Venn

overlap <- data.frame(paraules = unique(c(esquerra, centre, dreta))) %>%
  mutate(ESQUERRA = paraules %in% esquerra,
         CENTRE = paraules %in% centre,
         DRETA = paraules %in% dreta)

overlap %>%
  dplyr::select(ESQUERRA, CENTRE, DRETA) %>%
  euler() %>%
  eulerr:::plot.euler(counts = TRUE, quantities = TRUE, main="Diagrama de
Venn")
```

IV. ANÀLISI DE L'IMPACTE ECONÒMIC DE LA GUERRA

```
## Economia
el_economista <- read_excel("economista.xls", col_names = TRUE)
eco <- xmlToDataFrame("economista.xml")
```

```
cinco_dias <- read_excel("cinco_dias.xls", col_names = TRUE)
cin <- xmlToDataFrame("cinco_dias.xml")
expansion <- read_excel("expansion.xls", col_names = TRUE)
exp <- xmlToDataFrame("expansion.xml")

economia <- rbind(el_economista, cinco_dias, expansion)

m2 <- c(rep("ECONOMISTA",length(eco[,1])),
       rep("CINCO_DIAS",length(cin[,1])),
       rep("EXPANSION",length(exp[,1])))

dades2 <- rbind(eco,cin,exp)
dades2 <- cbind(dades2,m2)

#Neteja del text

text2 <- dades2[,"body"]

#Conviertim el text a minúsculas
text2 <- tolower(text2)
# Eliminem pàgines web
text2 <- str_replace_all(text2,"http\\S*", "")
text2 <- str_replace_all(text2,"www\\S*", "")
text2 <- str_replace_all(text2,"p\\S*", "")
# Eliminem signes de puntuació
text2 <- removePunctuation(text2)
# Eliminem els números
text2 <- removeNumbers(text2)
# Eliminem de espais múltiples en blanc
text2 <- stripWhitespace(text2)
text2 <- stri_trans_general(text2, "Latin-ASCII")
#Eliminem les stopwords
text2 <- removeWords(text2, words = stopwords("spanish"))
#Elimnare paraules que no tenen significat o no ens interessen
fora2 <- read.csv("paraules2.csv", header=F)
vector2 <- c(fora2$V1)
text2 <- removeWords(text2, vector2)

dades2 <- data.frame(mitja = dades2$m2, text = text2)

# Contatge d'aparició de cada bigrama
bigrama <- dades2 %>%
  unnest_tokens(input = text, output = "word", token = "ngrams", n = 2, dr
op = TRUE) %>%
  count(mitja, word, sort = TRUE) %>%
  ungroup()

bigrama <- bigrama %>%
 group_by(mitja, word) %>%
 summarise_at(vars(n), list(n = sum)) %>%
 filter(n > 25)

unigrama <- dades2 %>%
  unnest_tokens(word, text) %>%
```

```
  count(mitja, word, sort = TRUE) %>%
  ungroup() %>%
  filter(word %in% c("iva", "oro", "ley"))

paraules_per_marca2 <- rbind(unigrama, bigrama)

saveRDS(paraules_per_marca2, file = "paraules_per_marca2.Rds")

paraules_per_marca2 <- readRDS("paraules_per_marca2.Rds")

paraules_per_marca2 <- paraules_per_marca2 %>%
   group_by(mitja, word) %>%
   summarise_at(vars(n), list(n = sum)) %>%
   filter(n > 15)


 cols1 <- c("CINCO_DIAS" = "#B3E0A6", "EXPANSION" = "#559F52", "ECONOMISTA
" = "#24693D")

ci <- paraules_per_marca2 %>%
  filter(mitja %in% c("CINCO_DIAS")) %>%
   group_by(mitja) %>%
   top_n(20, n) %>%
   arrange(mitja, desc(n)) %>%
   ggplot(aes(x = reorder(word,n), y = n, fill = mitja)) +
   scale_fill_manual(values = cols1, aesthetics = c("fill")) +
   geom_col() +
   theme_bw() +
   labs(y = "Freqüència absoluta", x = "") +
   theme(legend.position = "none") +
   coord_flip() +
   facet_wrap(~ mitja, scales = "free", drop = TRUE)

ex <- paraules_per_marca2 %>%
  filter(mitja %in% c("EXPANSION")) %>%
   group_by(mitja) %>%
   top_n(20, n) %>%
   arrange(mitja, desc(n)) %>%
   ggplot(aes(x = reorder(word,n), y = n, fill = mitja)) +
   scale_fill_manual(values = cols1, aesthetics = c("fill")) +
   geom_col() +
   theme_bw() +
   labs(y = "Freqüència absoluta", x = "") +
   theme(legend.position = "none") +
   coord_flip() +
   facet_wrap(~ mitja, scales = "free", drop = TRUE)

ec <- paraules_per_marca2 %>%
  filter(mitja %in% c("ECONOMISTA")) %>%
   group_by(mitja) %>%
   top_n(20, n) %>%
   arrange(mitja, desc(n)) %>%
   ggplot(aes(x = reorder(word,n), y = n, fill = mitja)) +
   scale_fill_manual(values = cols1, aesthetics = c("fill")) +
```

```
   geom_col() +
   theme_bw() +
   labs(y = "Freqüència absoluta", x = "") +
   theme(legend.position = "none") +
   coord_flip() +
   facet_wrap(~ mitja, scales = "free", drop = TRUE)

ex + plot_annotation(title = "20 conceptes amb més freqüència absoluta de
EXPANSIÓN")

paraules_per_marca2 %>%
   select(word, n) %>%
   group_by(word) %>%
   summarise_at(vars(n), list(n = sum)) %>%
   top_n(20, n) %>%
   ggplot(aes(x = reorder(word,n), y = n)) +
   geom_col(fill="#80BF61") +
   theme_bw() +
   labs(y = "Freqüència absoluta", x = "") +
   ggtitle("20 bigrames amb més freqüencia absoluta") +
   theme(legend.position = "none") +
   coord_flip()
```