

MÁSTER UNIVERSITARIO EN TECNOLOGÍAS DEL LENGUAJE

Curso 2020 – 2021

Trabajo Fin de Máster

Aproximación al análisis pragmático en redes sociales mediante el uso de técnicas de enmarcado y Transformers

Alumno: Raul Alvarez Prieto



2021

TUTOR: Dr. Víctor Fresno Fernández

If you can't annoy somebody with what you write, I think there's little point in writing.

[Sir Kingsley William Amis](#). English novelist, poet, critic and teacher.

ABSTRACT

This research project aim is to advance in the field of Natural Language Processing (NLP) and pragmatic analysis by using Framing Theory and Transformers (BERT in our case).

Building on some features of word embeddings models, text scaling, sentiment analysis and “thought vectors”, we argue that cognitive frames are a useful tool in NLP for the study of culture, [social](#) and [psychological constructs](#), in particular political and mind traits. As word embeddings are thought to represent semantic relations between words as relationships between vectors in a high-dimensional space, we argue that frame embeddings allow us to reduce the complexity of the human mind by locating the frames used by an entity (person, group, text) on a multidimensional vector space defined in terms of frames.

We have based our analysis on several dimensions making it possible to locate each entity (whether a person, a political party or a text) on a map of political variations and show how their “political views” change.

Key Words: BERT, word embedding, political communication, frame embedding, framing, text scaling, NLP, pragmatic analysis, Artificial General Intelligence.

Si no puedes molestar a alguien con lo que escribes, creo que no tiene mucho sentido escribir.

Sir Kingsley William Amis. Novelista, poeta, crítico y docente inglés.

RESUMEN

El objetivo de este proyecto de investigación es avanzar en el campo del Procesamiento del Lenguaje Natural (PLN) y el análisis pragmático mediante el uso de la teoría de marcos y Transformers (BERT en nuestro caso).

Sobre la base de algunas características de los modelos de inserción de palabras, escalado de texto, análisis de sentimientos y vectores de pensamiento, argumentamos que los marcos cognitivos son una herramienta útil en PLN para el estudio de construcciones culturales, sociales y psicológicas, en particular también los rasgos políticos y mentales. Al igual que los embeddings representan relaciones semánticas entre palabras como relaciones entre vectores en un espacio de alta dimensión, argumentamos que los embeddings basados en marcos nos permiten reducir la complejidad de la mente humana al ubicar los marcos utilizados por una entidad (persona, grupo, texto) en un espacio vectorial multidimensional definido en términos de marcos.

Palabras clave: BERT, embedding, comunicación política, embedding de marcos, enmarcado, escalado de texto, PLN, análisis pragmático, Inteligencia General Artificial.

TABLA DE CONTENIDOS

1.	Introducción y justificación del tema seleccionado	13
2.	Objetivos del proyecto de investigación	18
3.	Formulación del Problema y preguntas de Investigación	21
4.	Fundamentación: Antecedentes y estado del arte	23
4.1	Estado del arte en NLP	23
4.1.1	CoreNLP	28
4.1.1	GPT-3	30
4.1.2	BERT y ALBERT	32
4.1.3	Traducción automática neuronal (NMT)	33
4.2	Estado del arte en el Espacios Vectoriales en Aprendizaje Automático	34
4.2.1	Espacios vectoriales en entornos de “Word Embedding”	35
4.2.2	Word2Vec	39
4.2.3	GloVe	46
4.2.4	Autocodificadores y Espacios latentes	48
4.3	Procesamiento de lenguaje natural y análisis político	52
4.3.1	La Brújula Política (The Political Compass)	52
4.3.2	Escalado, Word Embeddings y análisis político	56
4.4	Psicografía	60
4.5	Vectores del Pensamiento	66
4.6	Teoría de Marcos	67
5.	Desarrollo del trabajo	68
5.1	Fase I: Análisis Semántico	75
5.1.1	Selección de datos y Corpus utilizado	75
5.1.2	Fase de enmarcado	77

5.1.3	Estudio y diseño del espacio vectorial de marcos	80
5.1.4	Marcado de polaridad de marcos	82
5.2	Fase II: Operaciones de Aprendizaje Automático (MLOps)	85
5.2.1	Dataset de Entrenamiento.....	86
5.2.2	Búsqueda automatizada de marcos	88
5.2.3	Ubicación de entidades políticas en espacio vectorial	93
5.2.4	Entrenamiento del clasificador.....	94
5.2.5	Test y evaluación de la solución (BERT + Clasificador).....	99
5.2.6	Solución tecnológica	103
5.3	Resultados	110
5.4	Evaluación de resultados.....	114
5.4.1	Análisis de resultados desde la Ciencia Informática.....	114
5.4.2	Análisis de resultados desde la Ciencia Política	118
6.	Discusión y Conclusiones	129
7.	Conocimientos adquiridos, limitaciones del trabajo y futuras líneas de investigación ..	131
8.	Referencias.....	133
9.	Anexo I, la teoría de Marcos (Framing).....	139
9.1	Los marcos y el lenguaje.....	139
9.2	Marcos y Cultura.....	140
9.3	Marcos cognitivos utilizados en la comunicación política.....	143
9.4	Identificando Marcos en los medios de comunicación	147
9.4.1	Fase I. Estudio Inductivo.....	149
9.4.2	Fase II. Estudio Deductivo	151
9.5	Método de búsqueda de marcos en los medios y redes sociales	154
9.5.1	Metodología de búsqueda de marcos en la fase Inductiva.....	154
9.5.2	Metodología de búsqueda de marcos en la fase Deductiva.....	156
9.5.3	Metodología de Codificación de datos Cualitativos	156
9.5.4	Esquema gráfico de la Metodología de búsqueda de marcos.....	157

9.6	Selección de fuerzas políticas utilizadas en el estudio comparativo	162
9.7	Recogida de información y selección de herramientas	163
9.7.1	Recogida de documentación ideológica y programática de las formaciones políticas	163
9.7.2	Recogida de información en redes sociales.....	164
9.7.3	Herramientas de gestión de información cualitativa utilizadas.....	166
9.8	Marcos Cognitivos utilizados por PODEMOS	168
9.8.1	Matriz de Marcos Cualitativa (fase inductiva).....	168
9.8.2	Matriz de Marcos Cuantitativa (fase deductiva)	173
9.8.3	Discusión de resultados	177
10.	Anexo II.....	178
11.	Anexo III, Etiquetado de datos.....	183
11.1	España Plurinacional	183
11.2	España Uninacional.....	185
11.3	España Monárquica	187
11.4	España Republicana	190
12.	Anexo IV. Tablas y gráficos	192

ÍNDICE DE ILUSTRACIONES

Ilustración 1 Niveles de análisis lingüístico. Fuente: (Zhou & Hripcsak, April 2007).....	13
Ilustración 2 El efecto de "enmarcado". Fuente: (The decision Lab, 2021)	14
Ilustración 3 Técnicas de Análisis de Sentimientos (Aydogan & Akcayol, 2016).....	15
Ilustración 4 Niveles de procesado Lingüístico. Fuente: NLTK.....	16
Ilustración 5 Reducción de dimensiones: de una mente compleja a un espacio vectorial	19
Ilustración 6 Esquema de alto nivel para utilizar la "Brújula de sentimiento político". Fuente: (Falck & Marstaller, 2018).....	20
Ilustración 7 "Mente política" y “marcos cognitivos”.....	20
Ilustración 8 Técnicas de Análisis de Sentimientos (Liu, Bi, & Fan , 2017).....	24
Ilustración 9 Diccionarios utilizados en el análisis de sentimientos (Reagan, Danforth, Tivnan, & et al., 2017).....	25
Ilustración 10 Fuentes utilizadas en minería de textos utilizando técnicas semánticas	25
Ilustración 11 Modelos avanzados en NLP y número de parámetros (Rosset, 2020).....	27
Ilustración 12 Número de parámetros en diversos modelos de NLP (Saifee, 2020).....	27
Ilustración 13 GPT-3. Fuente: How GPT3 Works.....	31
Ilustración 14 Precisión de diversos algoritmos en el reto RACE. Fuente: Google Research.	32
Ilustración 15 GNMT (Google).....	33
Ilustración 16 Espacio vectorial de ejemplo (Desagulier, 2018)	35
Ilustración 17 Similitud entre palabras mediante uso del coseno (Fuente: (Chadha, 2020))...	36
Ilustración 18 Red neuronal con una capa oculta.....	37
Ilustración 19 Visualización 3D de vectores de palabra (Word vector) después de reducción de dimensiones (Fuente: Eric Kim)	40
Ilustración 20 Modelo CBOW	41
Ilustración 21 Modelo Skip-gram	42
Ilustración 22 Estructura general de la red en Word2Vec. Fuente: Chris McCormick	43
Ilustración 23 Visualización de la red neuronal en Word2Vec. Fuente: Sanket Gupta	43
Ilustración 24 Topología de los modelos CBOW y Skip-gram	45
Ilustración 25 Modelo de GloVe.....	46
Ilustración 26 Arquitectura de un códec convolucional. Fuente: Hackernoon	48
Ilustración 27 El "cuello de botella" de un autocodificador.....	50
Ilustración 28 Forma en la que trabaja un autocodificador variacional	51
Ilustración 29 Esquema de alto nivel para utilizar la "Brújula de sentimiento político". Fuente: (Falck & Marstaller, 2018).....	52

Ilustración 30 Espacio vectorial bidimensional utilizado por la brújula política.....	53
Ilustración 31 Cambio en la posición en el espacio vectorial de partidos en el Reino Unido. Fuente: Political Compass.....	54
Ilustración 32 Brújula de sentimiento político". Fuente: (Falck & Marstaller, 2018)	54
Ilustración 33 Diagrama de Nolan	55
Ilustración 34 Dimensiones políticas utilizadas por “RightValues”	55
Ilustración 35 Modelo de generación y de análisis de textos políticos (Egerod, B. C. K. & Klemmensen, R., 2020).....	56
Ilustración 36 Posiciones de partidos alemanes (1999-2005) en varias dimensiones (Slapin & Proksch, 2008, pág. 714).....	57
Ilustración 37 Ejemplo de dimensión en espacio político (Egerod, B. C. K. & Klemmensen, R., 2020).....	58
Ilustración 38 Utilización de Word embeddings en el análisis político.....	59
Ilustración 39 Demografía vs Psicografía. Fuente: Branding Compass.....	60
Ilustración 40 Ruleta de las emociones	63
Ilustración 41 La mente política. Fuente: Amazon	67
Ilustración 42 Fases de la solución de análisis de sentimiento político (frame embedding) ...	68
Ilustración 43 Fases I del sistema a desarrollar y ejemplo de lenguaje metafórico	69
Ilustración 44 Metodología seguida en el proyecto	70
Ilustración 45 Búsqueda automatizada de marcos en textos	72
Ilustración 46 De una mente compleja a un espacio vectorial reducido	72
Ilustración 47 Marcos cognitivos como herramienta para reducir la complejidad del mundo	74
Ilustración 48 Tareas de la Fase I de análisis	75
Ilustración 49 Tarea de etiquetado y creación del corpus de marcos cognitivos	77
Ilustración 50 Ejemplo de metáforas utilizadas por el independentismo catalán	78
Ilustración 51 Ejemplo de marcos utilizados por PODEMOS. Fuente (Alvarez Prieto, 2018)78	
Ilustración 52 n-gramas utilizados con mayor probabilidad por liberales y conservadores (Iyyer, Enns, Boyd-Graber, & Resnik, 2014, pág. 1121)	79
Ilustración 53 Selección de dimensiones del espacio vectorial.....	80
Ilustración 54 Ubicación ideológica de senadores (Congreso 114): Fuente: (Rheault & Cochrane, 2020, pág. 26)	81
Ilustración 55 Mercado de polaridad en marcos para PODEMOS	83
Ilustración 56 Mercado de polaridad en marcos para VOX.....	83
Ilustración 57 Fase II de aprendizaje automático.....	85

Ilustración 58 Tareas a realizar en la Fase II.....	85
Ilustración 59 Dataset "Rotten Tomatoes" con su escala de sentimiento.....	86
Ilustración 60 Tarea de búsqueda automatizada de marcos en textos.....	88
Ilustración 61 Marcos a encontrar en la búsqueda automatizada de marcos en textos	89
Ilustración 62 Conjunto de etiquetas utilizados en dataset de entrada (Fuente: Amber Ellen Boydston)	92
Ilustración 63 Ubicación de tres líderes políticos en el espacio vectorial pragmático.....	93
Ilustración 64 Entrenamiento del clasificador para la detección automatizada de marcos	94
Ilustración 65 Clasificación multiclase (“Multiclass classification”)	94
Ilustración 66 BERT en clasificación de textos. Fuente: Swatimeena.....	95
Ilustración 67 Preentrenamiento y ajuste fino en BERT. Fuente: BERT paper.....	95
Ilustración 68 Ajuste fino de un transformer BERT para clasificación de textos.....	96
Ilustración 69 BERT for Sequence Classification.....	97
Ilustración 70 Etapa final de clasificación de frases	98
Ilustración 71 Generalización del coeficiente de correlación de Mathews para un caso multiclase (Jurman, Giuseppe et al. “, 2021)	99
Ilustración 72 Visión global del sistema a desarrollar	103
Ilustración 73 Utilización de BERT para clasificación de textos.....	104
Ilustración 74 Fases en operaciones de aprendizaje automático (MLOps)	105
Ilustración 75 Fases de diseño, desarrollo del modelo y operaciones en aprendizaje automático	106
Ilustración 76 Plataforma de ejecución GPU utilizada	106
Ilustración 77 Búsqueda automatizada de marcos es Twitter	107
Ilustración 78 MongoDB Atlas 4.2.3	108
Ilustración 79 Tareas en los diferentes niveles de procesado lingüístico. Fuente: (Lensu, 2002)	109
Ilustración 80 Entrenamiento del clasificador para la detección automatizada de marcos	110
Ilustración 81 Resultados del clasificador BERT para dos dimensiones	111
Ilustración 82 Evaluación mediante clasificador BERT de diferentes entidades políticas en el eje 4 (Organización Territorial del Estado).....	112
Ilustración 83 Evaluación mediante clasificador BERT de diferentes entidades políticas en el eje 5 (Modelo Político del Estado).....	113
Ilustración 84 Métricas de evaluación utilizados por (Ma, 2019).....	115
Ilustración 85 Modelos BERT utilizados por (Ma, 2019).....	116

Ilustración 86 Ubicación de usuarios según Polaridad Neta (PN) en ejes 4 y 5 por tipo de usuario (tamaño del círculo = Polaridad Bruta, ejes invertidos).....	119
Ilustración 87 Ubicación de entidades políticas en espacio vectorial de marcos.....	120
Ilustración 88 Ubicación de partidos políticos en el cuadrante definido por los dos ejes bajo estudio	121
Ilustración 89 Detalle de la ubicación de partidos políticos en ejes.....	122
Ilustración 90 Polaridad Neta en el Eje 1 para partidos políticos	122
Ilustración 91 Polaridad Neta en el Eje 2 para partidos políticos	123
Ilustración 92 Ubicación de entidades en el cuadrante Republicano-Uninacional	124
Ilustración 93 Servicio Amazon Mechanical Turk". Fuente: Amazon	125
Ilustración 94 Ubicación de partidos políticos en el cuadrante definido por los dos ejes bajo estudio	130
Ilustración 95 Modelo cascada de marcos (Suau Gomila, 2015, pág. 83).....	146
Ilustración 96 Actores en los procesos de comunicación política (Matthes, 2012).....	147
Ilustración 97 Recursos de enmarcado y de razonamiento en un artículo de periódico acerca de la pobreza (D'Angelo & Kuypers, 2010).....	150
Ilustración 98 Transformar esquemas en categorías del libro de códigos (D'Angelo & Kuypers, 2010, pág. 127).....	153
Ilustración 99 Matriz de marcos con sus recursos de enmarcado y razonamiento (Van Gorp, 2005, pág. 491).....	153
Ilustración 100 Matriz de marcos con recursos de enmarcado y de razonamiento representativos	155
Ilustración 101 Matriz de Marcos con sus "Recursos de Enmarcado".....	155
Ilustración 102 Matriz de Marcos con sus "Recursos de Razonamiento"	156
Ilustración 103 Metodología de Búsqueda de Marcos	157
Ilustración 104 Metodología y herramientas utilizadas en la investigación	158
Ilustración 105 Metodología en la Fase Cualitativa – Inductiva.....	158
Ilustración 106 Metodología en la Fase Cuantitativa – Deductiva	159
Ilustración 107 Categorías encontradas (Figura 12 (Meijer, 2014, pág. 41))	160
Ilustración 108 Codificación Axial (Meijer, 2014, pág. 42)	160
Ilustración 109 Marcos dominantes encontrados (Meijer, 2014, pág. 43).....	160
Ilustración 110 Recogida de datos en Twitter mediante la herramienta Capture-NVivo	165
Ilustración 111 Importación de datos procedentes de Twitter en la herramienta NVivo.....	166
Ilustración 112 Pantalla de inicio del proyecto PODEMOS en la plataforma NVIVO	167

Ilustración 113 Ejemplo de Codificación de documento recogido en internet relativo a Podemos y los Perdedores de la "Globalización"	167
Ilustración 114 Nodos comparados por el número de referencias en la codificación	167
Ilustración 115 Marcos utilizados por PODEMOS (Fase Inductiva).....	168
Ilustración 116 Marco "Élites culpables de la crisis" utilizado por PODEMOS	169
Ilustración 117 Marco "El Pueblo como los de Abajo" utilizado por PODEMOS	170
Ilustración 118 Marco "Final del Estado del 78" utilizado por PODEMOS	171
Ilustración 119 Marco "España Plurinacional" utilizado por PODEMOS	172
Ilustración 120 Comparativa entre fuerzas políticas en Twitter del uso de "Pueblo" y "Globalización"	173
Ilustración 121 Porcentaje de tuits referidos a "Pueblo" y "Globalización" para diversas formaciones y dirigentes políticos.....	174
Ilustración 122 Utilización de los términos "Pueblo" y "Globalización en Twitter por PODEMOS.....	174
Ilustración 123 Utilización de los términos "Pueblo" y "Globalización en Twitter por ERC	175
Ilustración 124 Marco fase cuantitativa "Final del Estado del 78" utilizado por ERC	176
Ilustración 125 Resultados detallados en dos ejes bajo estudio	193
Ilustración 126 Ubicación de entidades políticas en los cuatro cuadrantes	194
Ilustración 127 Ubicación de entidades políticas en el cuadrante República-Uninacional....	195
Ilustración 128 Resultados eje "Monarquía-República"	196
Ilustración 129 Resultados eje "Plurinacionalidad-Nación única"	197
Ilustración 130 Gráfico "Polaridad Neta"	198
Ilustración 131 Gráfico "Polaridad Relativa"	199
Ilustración 132 Tabla definición dimensiones del espacio vectorial pragmático.....	200
Ilustración 133 Ubicación de los marcos utilizados por PODEMOS en el espacio vectorial pragmático	201
Ilustración 134 Ubicación de los marcos utilizados por VOX en el espacio vectorial pragmático	202

ÍNDICE DE TABLAS

Tabla 1 Cronología de aparición de diferentes sistemas avanzados NLP	26
Tabla 2 Dimensiones del Espacio Vectorial Pragmático	81
Tabla 3 Recursos de razonamiento y de enmarcado	90
Tabla 4 Correlación de resultados entre ambos ejes	120
Tabla 5 Marcos utilizados en USA por Demócratas y Republicanos	145

1. Introducción y justificación del tema seleccionado

Este trabajo de investigación pretende sentar las bases para la creación de sistemas de Minería de Información Social centrados en el **análisis de la comunicación política** mediante el uso de **técnicas de enmarcado** o [Framing](#) en lengua inglesa, términos que utilizaremos indistintamente a lo largo de este trabajo, unas técnicas que podrían ubicarse en el **nivel pragmático de análisis lingüístico** al tomar en cuenta aspectos complejos del discurso y la utilización de lenguaje figurado, como las metáforas por ejemplo, junto con sistemas basados en **Transformers** que será **BERT** en nuestro caso.

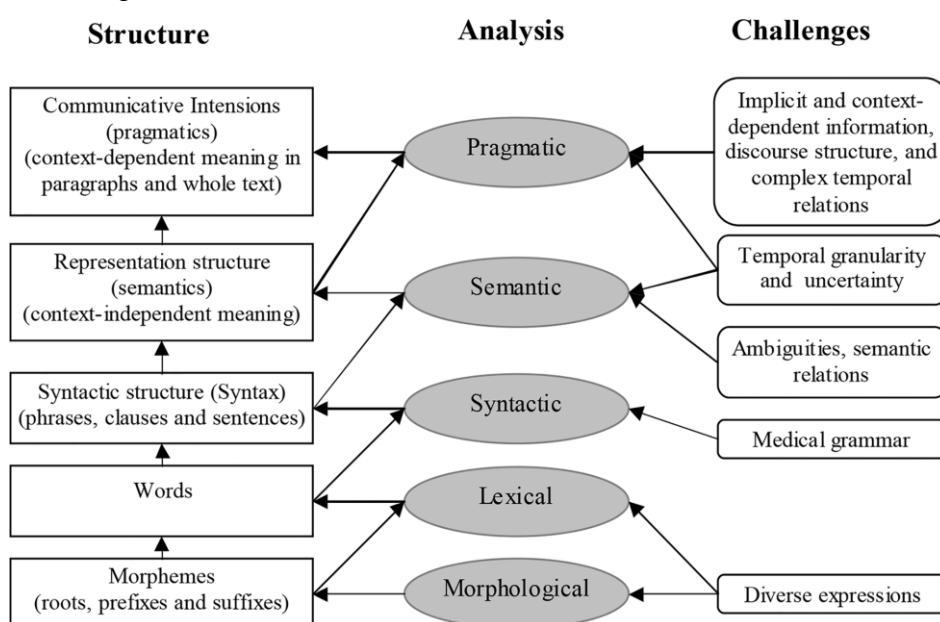


Ilustración 1 Niveles de análisis lingüístico. Fuente: (Zhou & Hripcsak, April 2007)

Los sesgos ideológicos son difíciles de detectar incluso para humanos ya que la tarea se basa no sólo en disponer de los conocimientos “políticos” necesarios, sino también en la capacidad del anotador/anotadores para captar elementos sutiles del uso del lenguaje (Iyyer, Enns, Boyd-Graber, & Resnik, 2014, pág. 1113). Los enfoques existentes en la detección de sesgos (clasificadores del tipo "bolsa de palabras" BoW por ejemplo) ignoran contextos lingüísticos “enriquecidos” y complejos de este tipo, operando a menudo a nivel de documentos completos.

El **efecto marco** es un sesgo cognitivo por el que las preferencias de una persona ante un mensaje o problema de decisión dependen de cómo se presente éste, esto es, de su “marco” (The decision Lab, 2021). Por ejemplo, en el anuncio presentado en la Ilustración 2, es muy diferente el efecto en el público objetivo al publicarlo como un anuncio centrado en la existencia

de grasa en el yogur (yogur con un 20% de grasa) que si lo presentamos como un yogur sin grasa (yogur donde se ha eliminado un 80% de la grasa). **La forma en la que se presenta un tema o producto (el yogur en este caso) activa diferentes “experiencias”, “sesgos”, y “emociones” en la mente de cada persona**, e incluso se activan diferentes áreas del cerebro si estudiáramos este efecto desde el área de la neuropsicología.

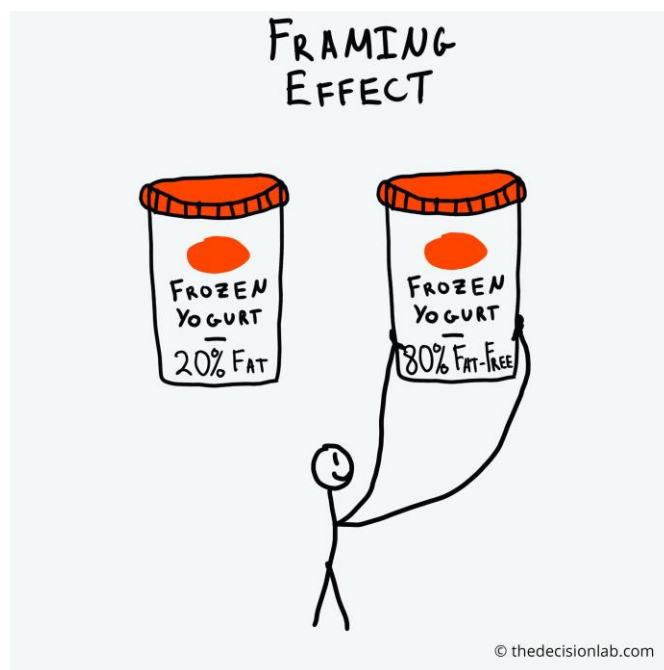


Ilustración 2 El efecto de "enmarcado". Fuente: (The decision Lab, 2021)

Este trabajo de investigación está centrado en el **estudio del sentimiento político a través de las redes sociales**, por lo que una parte importante de los datos serán recogidos de dichas redes sociales, en particular de **Twitter**. El [análisis de sentimientos](#) es un campo dentro del Procesamiento de Lenguaje Natural (PLN) que construye sistemas que tratan de **identificar y extraer opiniones dentro del texto** siendo una tarea dentro del NLP que podríamos ubicar en un nivel alto y más cercano al pragmático. Normalmente, además de identificar la opinión, estos sistemas extraen atributos como la polaridad (opinión positiva o negativa), el asunto sobre el que se está hablando y la persona o que expresa la opinión. **Dado que creemos que muchas de las técnicas utilizadas en el análisis de sentimientos se pueden aplicar en nuestro trabajo centrado en el análisis de sentimiento político, profundizaremos en las técnicas utilizadas en dicho campo.** Como posteriormente veremos, en este trabajo evaluaremos marcos contrapuestos para ubicar a una entidad política en un espacio vectorial político, lo que podría definirse como una medida de polaridad política sobre varios ejes.

La mayoría de las aplicaciones de análisis de sentimientos utilizan dos enfoques diferentes: aproximaciones semánticas y técnicas de aprendizaje automático. Las **aproximaciones semánticas** destacan por el uso de lexicones o diccionarios de términos combinados con una orientación semántica de polaridad y se caracterizan por el procesamiento y segmentación en palabras de un texto para así asignar un valor de polaridad a dicho texto mediante la suma de los valores de polaridad de los términos lematizados que lo componen. Las **técnicas de aprendizaje automático** por su parte suelen utilizar un algoritmo de aprendizaje supervisado para crear un clasificador a partir de un entrenamiento con una colección de textos anotados con su polaridad. Este tipo de representaciones permite, a través de espacios de representación de conceptos, calcular **similitudes semánticas** entre textos y entre los algoritmos más utilizados están los basados en “máquinas de vectores de soporte” (SVM), el cálculo de probabilidades de “Naïve Bayes” y KNN (k-vecinos más parecidos). Existen otras técnicas más avanzadas como el Análisis de la Semántica Latente (LSA) que permite encontrar "conceptos" a partir de asociaciones de términos y BERT (*Bidirectional Encoder Representations from Transformers*), una técnica basada en redes neuronales que permite **representar una palabra en el contexto en el que aparece por medio de un proceso de entrenamiento basado en enmascaramiento, es decir, entrenando a la red neuronal para que aprenda la siguiente palabra a un conjunto de palabras dado. Esta será la técnica que utilizemos en este trabajo.**

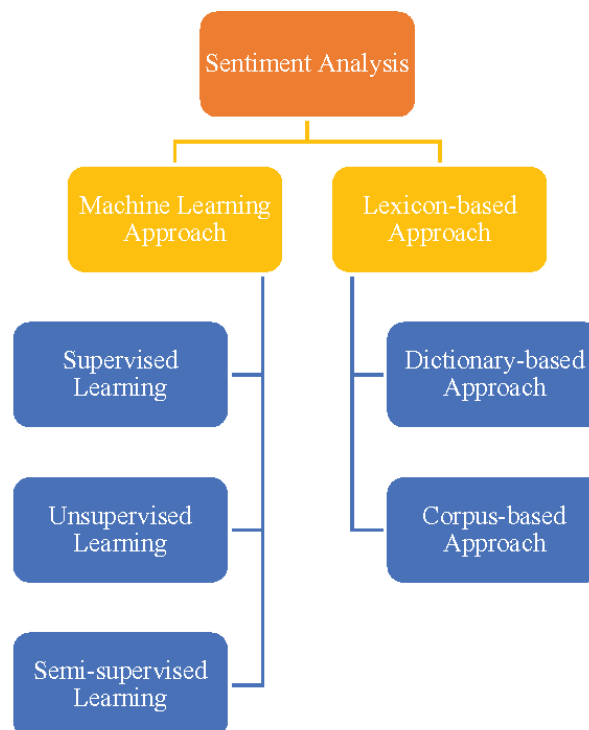


Ilustración 3 Técnicas de Análisis de Sentimientos (Aydogan & Akcayol, 2016)

El nivel de análisis morfológico, sintáctico, semántico o pragmático de los textos bajo estudio se aplica dependiendo del objetivo de cada aplicación. En nuestro caso, un sistema como el que queremos construir requiere información detallada del contexto y del dominio temático ya que queremos llevar a cabo un **sistema de análisis lingüístico que llegue hasta el nivel de análisis pragmático**. Por lo tanto, la realización de este proyecto implicará, la utilización de componentes del procesamiento del lenguaje natural en varios niveles:

- **Análisis morfológico y léxico:** Análisis interno de las palabras que forman oraciones para extraer lemas, rasgos flexivos y unidades léxicas compuestas.
- **Análisis sintáctico:** Análisis de la estructura de las oraciones de acuerdo con el modelo gramatical empleado.
- **Análisis semántico:** Proporciona la interpretación de las oraciones una vez eliminadas las ambigüedades morfosintácticas presentes.
- **Análisis pragmático:** Incorpora el análisis del contexto de uso a la interpretación final. Aquí se incluye tanto el tratamiento del *lenguaje figurado (metáfora e ironía)* como el conocimiento del campo específico (comunicación y tendencias políticas) necesario para entender un texto.

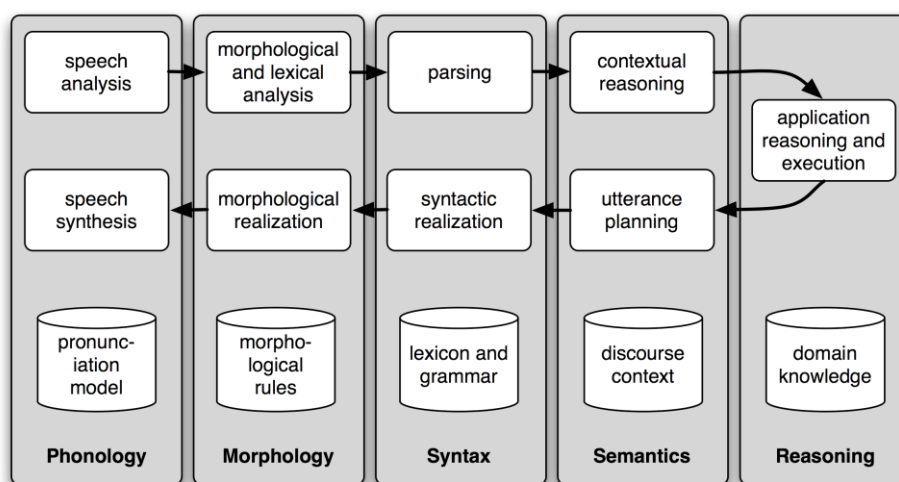


Ilustración 4 Niveles de procesado Lingüístico. Fuente: NLTK

Los motivos por los que creemos que este tema de investigación tiene interés y se puede justificar una investigación sobre el mismo son:

1. Los *algoritmos utilizados en la actualidad basados principalmente en técnicas de "Word Embedding"* (ELMo, BERT, Word2vec, GloVe, etc.) dónde las palabras o frases del lenguaje natural son representadas como vectores de números reales, cuantifican y categorizan las semejanzas semánticas entre elementos lingüísticos

basándose en sus propiedades distribucionales en grandes muestras de textos incluyendo en algunos casos información acerca del contexto de dichas palabras, como por ejemplo en los [Transformers](#). Por ejemplo, **Word2Vec** es un algoritmo de aprendizaje no supervisado que asume la hipótesis distribucional (palabras que aparecen en contextos similares tendrán significado similar) o el segundo principio de Wittgenstein en el que se establece que la semántica de una palabra se infiere desde su uso (es decir, a partir de las palabras con las que coaparece, etc..). Modelos como **GPT-3**, que utiliza un espacio vectorial con [12.288 dimensiones](#), **tratan de capturar información más allá de las coocurrencias** y, a partir de este otro tipo de información, podemos esperar ir subiendo en los niveles de análisis lingüístico.

2. Creemos que incluyendo *información procedente del nivel pragmático* mediante el uso de marcos cognitivos, podremos conseguir *ahondar en la relación entre lenguaje, psicología, neuropsicología y otras ramas de la ciencia* a la hora de detectar relaciones entre contenidos en textos y a la hora de identificar y analizar sentimientos políticos.
3. El objeto de estudio tiene aplicabilidad práctica en el **campo de la minería de información social y de la comunicación política** ya que pretendemos encontrar métodos para descubrir los principales mensajes (o marcos cognitivos) utilizados en la comunicación política, la forma en que se organizan estos mensajes y cómo se distribuyen a través de las redes sociales.
4. El **tema es complejo y podemos decir que no ha sido investigado suficientemente** en la literatura, tanto a nivel nacional como internacional. A pesar de que los estudios sobre comunicación política y redes sociales han pasado a ser objeto de investigación y del debate académico, *la utilización del análisis de marcos en redes sociales y su relación con el procesamiento del lenguaje natural en el nivel pragmático y el* no han alcanzado un grado de desarrollo suficiente.
5. Diseñar modelos eficientes que sean capaces de capturar el significado de unidades del lenguaje complejas es un reto importante que el procesamiento del lenguaje natural debe resolver en el futuro.

Por ello consideramos que el tema propuesto en este trabajo tiene interés y relevancia suficiente como para plantear el desarrollo de una investigación que busque un nivel de abstracción en NLP más profundo que el existente.

2. Objetivos del proyecto de investigación

Como se ha mencionado en el apartado anterior, la mayoría de las técnicas y algoritmos utilizados en la actualidad en NLP carecen de "comprensión del mundo" ya que solo analizan la relación entre palabras sin realizar un análisis básico del significado de las palabras (semántica). Creemos que se deben explorar otros caminos y para ello nuestros principales objetivos serán:

1. Definir métodos para **encontrar los principales mensajes o marcos presentes en textos y redes sociales** que son utilizados en un dominio de conocimiento dado, en nuestro caso la comunicación política en España y la forma en que se organizan y se distribuyen estos mensajes a través de las redes sociales. Para ello en una primera fase de este trabajo nos centraremos en la recopilación de recursos de enmarcado (metáforas, eslóganes y narrativas principalmente) utilizados en la comunicación política en España tomando como entrada un corpus de textos.
2. **Definir y generar un espacio vectorial multidimensional**, inicialmente en el rango de decenas de dimensiones pero que en un escenario comercial puede ser de varios cientos de ellas, *donde cada "marco" único en el corpus* (y no cada palabra o conjunto de palabras como en otras técnicas) *tiene asignado un vector en el espacio vectorial*. Los vectores marco se representan en el espacio vectorial de forma que marcos que comparten un "contexto" común están ubicados cercanos unos a otros en dicho espacio.
3. **Ubicar textos, partidos políticos y usuarios de redes sociales**, a partir de los mensajes compartidos en dichas redes, en este **espacio vectorial** multidimensional construido no a partir de palabras, términos o contenido semántico de los mismos (como en el caso de Word2Vec), sino **diseñado para ubicar "marcos cognitivos"**, que están a su vez definidos por los elementos que definen dichos marcos como por ejemplo metáforas y eslóganes. *Usuarios que comparten marcos comunes* obviamente estarán ubicados en áreas cercanas del espacio vectorial (con lo que será posible utilizar medidas para tratar de estimar su similitud en lo que a ideas políticas se refiere).
4. **Desarrollar las técnicas y algoritmos necesarios para implementar un primer prototipo de "Sistema de análisis de sentimiento Político"** en Twitter utilizando técnicas de aprendizaje automático y procesado lingüístico **mediante la utilización de técnicas de enmarcado y Transformers**. Este prototipo estará basado en la utilización de técnicas estándar como redes neuronales y Transformers y allá donde sea posible, la

reutilización de algoritmos, soluciones y librerías existentes en Python (Pandas, Transformers, Spacy, etc.).

5. Sentar las bases para desarrollar **sistemas de análisis que utilicen métodos de análisis a nivel pragmático** como medio para acercarnos a métodos de Inteligencia Artificial General o fuerte (*Artificial general Intelligence*, [AGI](#)). Este tipo de métodos están enfocados a que una máquina tenga la capacidad para entender una tarea intelectual de una forma similar a como lo hace un humano.

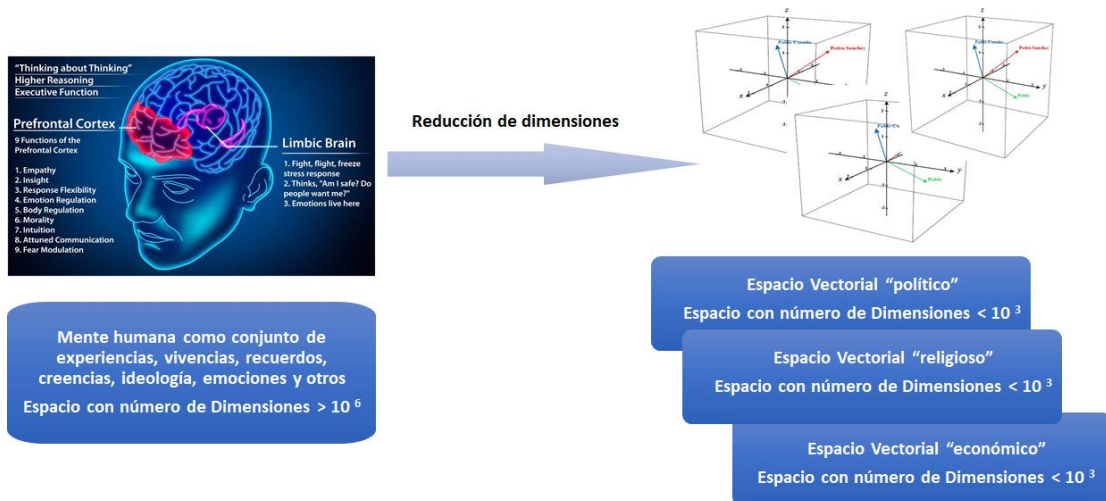


Ilustración 5 Reducción de dimensiones: de una mente compleja a un espacio vectorial

En cuanto al **análisis de sentimiento político**, no nos limitaremos a un análisis semántico de los textos como el realizado por (Falck & Marstaller, 2018) que nos lleve a una extracción de entidades y al análisis del sentimiento político de dichos textos, sino que utilizaremos los resultados de este trabajo para ubicar las entidades bajo estudio (como partidos políticos y medios de comunicación) en un espacio político multidimensional.

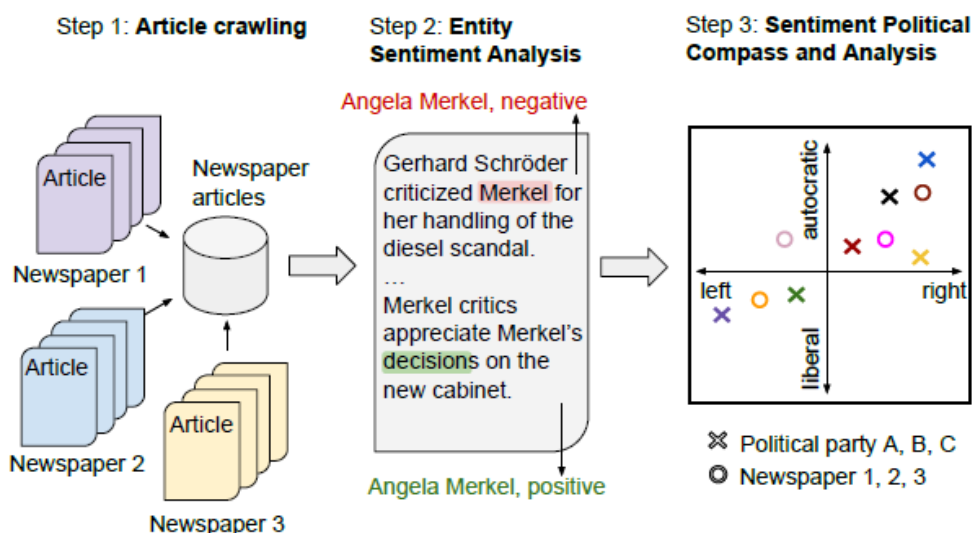


Ilustración 6 Esquema de alto nivel para utilizar la "Briújula de sentimiento político". Fuente: (Falck & Marstaller, 2018)

Es decir, en nuestro caso **nuestro objetivo es un análisis de textos para poder ubicar a los partidos políticos, a sus dirigentes políticos o a cualquier entidad política en un espacio vectorial multidimensional que permita ubicar marcos (frames) culturales o cognitivos.**



Ilustración 7 "Mente política" y "marcos cognitivos"

Presentamos en los siguientes apartados las preguntas de investigación que se pretenden abordar y el desarrollo de este trabajo, incluyendo la selección de datos de información social que utilizaremos, el corpus de datos sobre el que vamos a trabajar y el conjunto de procesos que vamos a aplicar para poder llegar a nuestro objetivo, es decir, qué conocimiento se pretende extraer de dicha información. Finalmente definiremos la forma en la que vamos a evaluar el sistema a implementar.

3. Formulación del Problema y preguntas de Investigación

Las líneas de investigación generales planteadas en el desarrollo de presente estudio, y que están mutuamente interrelacionadas del caso son:

1. Uso de técnicas propias de niveles superiores de procesamiento de lenguaje natural (nivel pragmático) como podría ser la **Teoría de Marcos o Encuadre** (framing):
 - a. Marcos y sociedad: Partimos de la premisa (compartida con varios autores como Van Gorp) de que **los marcos son parte de la cultura de una sociedad**, son generados, utilizados y amplificadas por dicha sociedad y por los medios de comunicación.
 - b. **Marcos y redes sociales**: pretendemos investigar cómo los marcos son generados, utilizados y distribuidos por una formación política a través de las redes sociales y su impacto en los usuarios de dichas redes.
 - c. **Marcos y comunicación política**: pretendemos investigar el conjunto de mensajes centrales (marcos) que una formación política emite y que resuenan entre sus seguidores y que son capaces de permanecer en el tiempo (las ideas principales o lo que algunos denominan ideología).
2. Estudio de **espacios vectoriales** diseñados para ubicar **marcos** en lugar de por palabras o conjuntos de palabras.
3. **Redes sociales** como elemento para recopilar corpus de datos utilizados para encontrar y ubicar sentimientos políticos.
4. **Modelos de lenguaje novedosos basados en Transformers (BERT en nuestro caso particular).**

Se incluye en este trabajo como anexo I un extracto del trabajo del mismo autor (Alvarez Prieto, 2018) en la Facultad de Ciencias Políticas y Sociología de la UNED que puede servir al lector para contextualizar las preguntas de investigación (no siendo necesaria su lectura). En dicho trabajo se sentaron las bases para el análisis del discurso político en España a través del uso de los marcos utilizando las técnicas y herramientas propias de la Sociología, la Ciencia Política y de la Comunicación. *En el presente TFM se han utilizado las técnicas propias del procesamiento del lenguaje natural para automatizar la mayoría de las tareas que un equipo trabajando en la comunicación política (sociólogos, politólogos, expertos en comunicación) emplearían en el estudio de la comunicación política en las redes sociales* (por ejemplo, la

adquisición de datos de Twitter y su procesamiento posterior para extraer “conocimiento” político. Así pues, las **preguntas de investigación** sobre las que queremos trabajar son:

PREGUNTAS DE INVESTIGACIÓN

1. *¿ Es posible avanzar en la teoría de marcos a partir del uso de técnicas de procesamiento de lenguaje natural?*
2. *¿Son de utilidad las técnicas de procesamiento del lenguaje natural basadas en Transformers para una tarea de detección de marcos en una red de microblogging?*

Como hipótesis iniciales y respuestas tentativas a las preguntas proponemos:

1. **Creemos que la utilización de técnicas de PLN** y la utilización de conceptos propios del nivel pragmático de la lingüística, nos permitirán avanzar en el campo de la teoría de marcos al permitir, entre otras razones, la creación de herramientas automatizadas de detección de marcos en textos.
2. **Intuimos que la utilización de Transformers**, con la posibilidad de captar información de niveles superiores en textos, nos permitirá dotar de herramientas adicionales de análisis a las Ciencias Sociales.

Exploraremos adicionalmente la forma en que el resultado de este trabajo se puede relacionar con trabajos como el de [Geoffrey Hinton](#) (uno de los primeros investigadores que demostró el uso del algoritmo de backpropagation para entrenar redes neuronales) acerca de los “[Vectores de Pensamiento](#)” (DL4J, 2020). Para este autor (véase el apartado 4.4) un “Vector de Pensamiento” es un “*pensamiento vectorizado*” que nos permite relacionar unos pensamientos con otros al igual que en Word2Vec se puede relacionar una palabra con otras.

4. Fundamentación: Antecedentes y estado del arte

Presentamos en este apartado el estado del arte en el campo del Procesamiento del Lenguaje Natural desde un punto de vista de computación neuronal (autocodificadores, Word Embeddings, transformers, etc.), así como otras aproximaciones a niveles altos de análisis lingüístico (en particular, análisis de sentimientos), del procesamiento de lenguaje natural en el ámbito político (escalado de textos), así como en el uso de otras técnicas innovadoras utilizadas en este proyecto de investigación como es la teoría de marcos (framing) o los “Vectores del Pensamiento”. No pretendemos describir completamente el funcionamiento de cada técnica o algoritmo sino mostrar las principales características de estos y sobre todo aquellas que puedan tener relación con el objetivo de este trabajo.

4.1 Estado del arte en NLP

Este apartado presenta el estado del arte en NLP desde un punto de vista de computación neuronal y poniendo el foco en niveles elevados de análisis lingüístico. Entre las técnicas más destacadas estarían la utilización de modelos de lenguaje, el uso de transformers, y en general todas las técnicas de aprendizaje automático utilizadas en este campo y que como veremos a lo largo de este trabajo encajan perfectamente en el estudio de sentimiento político mediante el uso de técnicas de enmarcado.

Como comentamos en la introducción de este trabajo, las aplicaciones NLP que encontramos en la literatura que han abordado niveles altos de procesamiento lingüístico se encuentran dentro del área del análisis de sentimientos y utilizan enfoques basados en aproximaciones semánticas o en técnicas de aprendizaje automático:

- Las **aproximaciones semánticas** están basadas en el uso de lexicones o diccionarios de términos combinados con una orientación semántica de polaridad. En ellas se segmentan los textos en palabras para asignar un valor de polaridad a dichos textos mediante la suma de los valores de polaridad de los términos lematizados que lo componen.
- Las **técnicas de aprendizaje automático** están basadas en técnicas **estadísticas** y suelen utilizar un algoritmo de aprendizaje supervisado para crear un clasificador a

partir de un entrenamiento con una colección de textos anotados con su polaridad. Este tipo de representaciones permite calcular similitudes semánticas entre textos. Entre los algoritmos más utilizados están los basados en “máquinas de vectores de soporte” (SVM), el cálculo de probabilidades de “Naïve Bayes” y KNN (k-vecinos más parecidos).

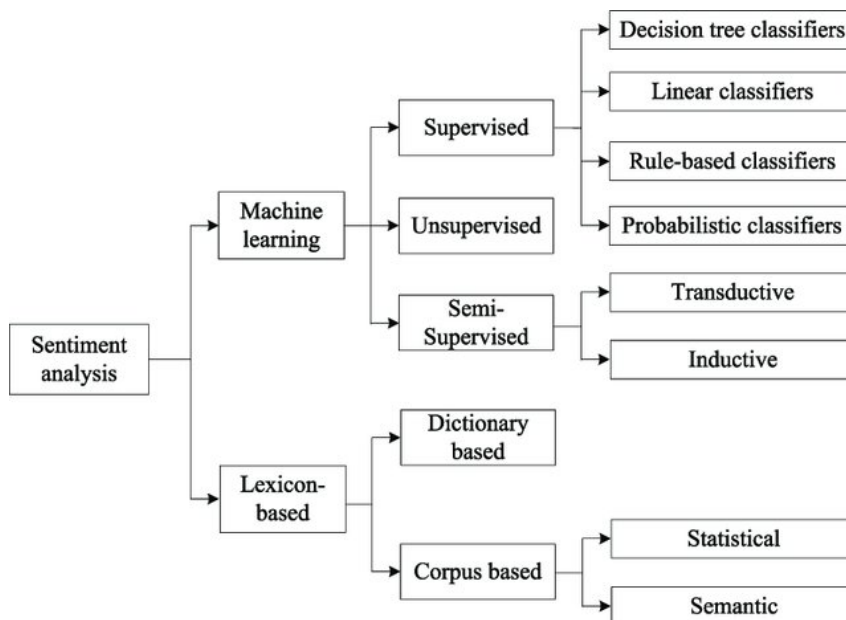


Ilustración 8 Técnicas de Análisis de Sentimientos (Liu, Bi, & Fan, 2017)

En las *técnicas basadas en aproximaciones semánticas* (“lexicon-based” en la Ilustración 8), generalmente se utiliza un lexicon de sentimientos para evaluar la polaridad (en lo que a sentimientos se refiere) de un texto a base de evaluar la polaridad de cada palabra presente en dicho documento. Este es el caso de las técnicas basadas en el uso de un diccionario donde se almacena la polaridad de cada palabra. Estos diccionarios pueden también contener sinónimos y antónimos de cada palabra. En las técnicas basadas en corpus, la similitud semántica está basada en las asociaciones entre palabras aprendidas a partir de grandes colecciones de textos (corpus). Si por ejemplo una palabra aparece intermitentemente entre textos positivos, se le asigna una polaridad positiva y si aparece entre textos negativos su polaridad es considerada negativa. En el caso de que aparezca tanto en textos considerados positivos como considerados negativos, su polaridad será neutra. ***Muchos sistemas utilizados están basados en asumir que palabras que muestran opiniones similares aparecen mayoritariamente juntas dentro de un corpus.*** Así, si dos palabras aparecen con frecuencia en un mismo contexto existe una alta probabilidad de que tengan la misma polaridad (y de la misma manera podemos calcular la polaridad de una palabra desconocida calculado la frecuencia de coocurrencia con otras palabras).

Dictionary	# Entries	Range	Construction	License
labMT	10,222	1.3–8.5	Survey: MT, 50 ratings	CC
ANEW	1,034	1.2–8.8	Survey: UF Intro Psych	Free for research
LIWC07	4,483	[–1,0,1]	Manual	Paid, commercial
MPQA	7,192	[–1,0,1]	Manual + ML	GNU GPL
OL	6,782	[–1,1]	Dictionary propagation	Free
WK	13,915	1.3–8.5	Survey: MT, 14-18 ratings	CC
LIWC01	2,322	[–1,0,1]	Manual	Paid, commercial
LIWC15	6,549	[–1,0,1]	Manual	Paid, commercial
PANAS-X	20	[–1,1]	Manual	Copyrighted paper
Pattern	1,528	–1.0–1.0	Unspecified	BSD
SentiWordNet	147,700	–1.0–1.0	Synset synonyms	CC BY-SA 3.0
AFINN	2,477	[–5,–4,...,4,5]	Manual	ODbL v1.0
GI	3,629	[–1,1]	Harvard-IV-4	Unspecified
WDAL	8,743	0.0–3.0	Survey: Columbia students	Unspecified
EmoLex	14,182	[–1,0,1]	Survey: MT	Free for research
MaxDiff	1,515	–1.0–1.0	Survey: MT, MaxDiff	Free for research
HashtagSent	54,129	–6.9–7.5	PMI with hashtags	Free for research
Sent140Lex	62,468	–5.0–5.0	PMI with emoticons	Free for research
SOCAL	7,494	–30.2–30.7	Manual	GNU GPL
SenticNet	30,000	–1.0–1.0	Label propagation	Citation requested
Emoticons	132	[–1,0,1]	Manual	Open source code
SentiStrength	2,615	[–5,–4,...,4,5]	LIWC + GI	Free for research
VADER	7,502	–3.9–3.4	MT survey, 10 ratings	Freely available
Umigon	927	[–1,1]	Manual	Public Domain
USent	592	[–1,1]	Manual	CC
EmoSenticNet	13,188	[–10,–2,–1,0,1,10]	Bootstrapped extension	Non-commercial

Ilustración 9 Diccionarios utilizados en el análisis de sentimientos (Reagan, Danforth, Tivnan, & et al., 2017)

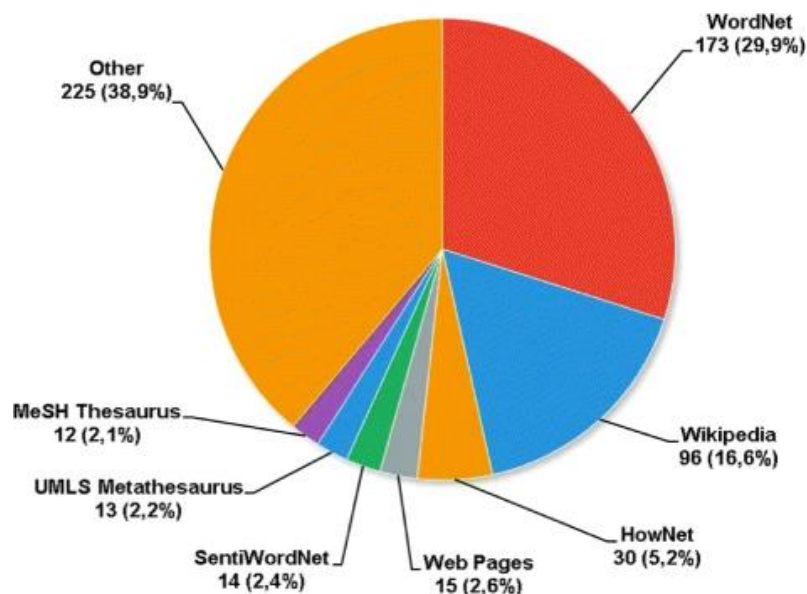


Ilustración 10 Fuentes utilizadas en minería de textos utilizando técnicas semánticas

Entre las diferentes técnicas y algoritmos existentes, las **técnicas de aprendizaje automático**, y dentro de ellas los [transformers](#), se han convertido rápidamente en los **más utilizados** en los problemas de PLN reemplazando a otros modelos basados en redes neuronales recurrentes ([RNN](#)) como [LSTM](#). El modelo de los Transformers ha facilitado la *paralelización durante la*

fase de entrenamiento, lo que a su vez ha permitido la utilización de mayores conjuntos de datos de entrenamiento ([training dataset](#)). Esto a su vez ha impulsado el desarrollo de **sistemas pre-entrenados** como [BERT \(Bidirectional Encoder Representations from Transformers\)](#) y [GPT \(Generative Pre-trained Transformer\)](#) que han sido entrenados con extensos conjuntos de textos y que pueden ser ajustados a tareas específicas del lenguaje.

Una cronología en la aparición de los sistemas avanzados de NLP durante la última década sería:

Modelo	Año	Empresa	Técnica
Word2vec	2013	Google	Word Embedding
GloVe	2014	Stanford University	Word Embedding
FastText	2016	Facebook	Word Embedding
Transformers	2017	Google	Transformers
ELMO	2018	AI2 (Allen Institute for AI)	Word Embedding
GPT	2018	OpenAI	Transformers
BERT	2018	Google	Transformers
GPT-2	2019	OpenAI	Transformers
Turing-NLP	2020	Microsoft	Transformers
GPT-3	2020	OpenAI	Transformers

Tabla 1 Cronología de aparición de diferentes sistemas avanzados NLP

Es importante hacer notar que la tendencia en NLP hacia los sistemas utilizando Transformers también ha ido acompañada de un aumento del número de parámetros que dichos algoritmos utilizan (véase Ilustración 11 e Ilustración 12), y por lo tanto un aumento de las necesidades de procesamiento (que afortunadamente han ido acompañadas de un notable aumento de soluciones HW y SW de mayor potencia disponibles en el mercado).

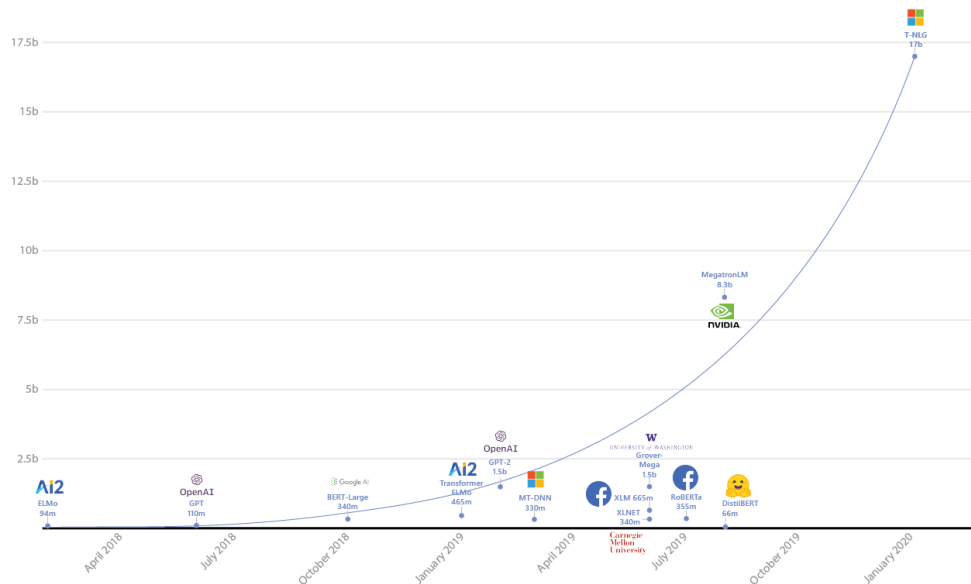


Ilustración 11 Modelos avanzados en NLP y número de parámetros (Rosset, 2020)

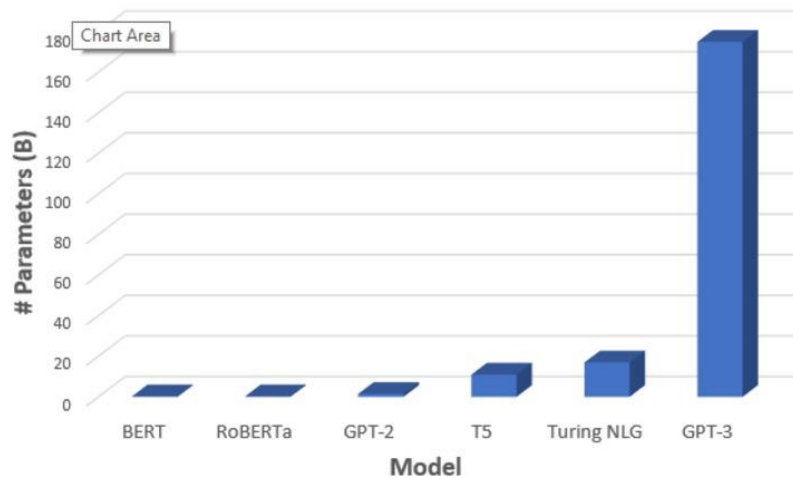


Ilustración 12 Número de parámetros en diversos modelos de NLP (Saifee, 2020)

Model	PF/S-Days	TPU V3-Days	Cost low	Cost high
BERT-Large	6,16	14,67	\$ 1.689,60	\$ 5.632,00
RoBERTa-Large	49,3	117,38	\$ 13.522,29	\$ 45.074,29
AlphaGoZero	1.850	4.404,76	\$ 507.428,57	\$ 1.691.428,57
GPT-3 175B	3.640	8.666,67	\$ 998.400,00	\$ 3.328.000,00
GPT-4 1,75T estimation	36.400	86.666,67	\$ 9.984.000,00	\$ 33.280.000,00

Figure 1 Coste estimado para entrenar diversos modelos (Seele, 2021)

En cualquier caso debemos tomar en cuenta que la principal fuente de información disponible para todos los métodos no supervisados para aprender representaciones de palabras, son las estadísticas de apariciones de palabras en un corpus, y la pregunta que todo investigador se debe seguir haciendo es *cómo se puede generar “significado” a partir de estas estadísticas* (si es que se genera realmente algún significado semántico) y *cómo los vectores resultantes pueden representar ese significado* (si es que lo hacen).

4.1.1 CoreNLP

El modelo de aprendizaje profundo CoreNLP de Stanford construye una representación de oraciones completas basándose en la estructura de dichas oraciones. Calcula el sentimiento basándose en cómo las palabras crean el significado de frases más largas, estando basada en un tipo de **red neuronal recursiva** operando sobre las estructuras gramaticales según sus investigadores. El conjunto de datos en el que se entrenó este modelo es el “[Stanford Sentiment Treebank](#)”. Este modelo, como cualquier otro en realidad, mejora su aprendizaje al alimentarlo con más oraciones etiquetadas. Un ejemplo de estudio de sentimientos se puede encontrar en [este enlace](#) y varias son las operaciones que se realizan sobre los textos como: [etiquetado POS](#), [NER](#), [parseado de componentes \(constituency parsing\)](#), [parseado de dependencias](#), [resolución de correferencias](#), estudio de [sentimiento](#), o [Open IE](#).

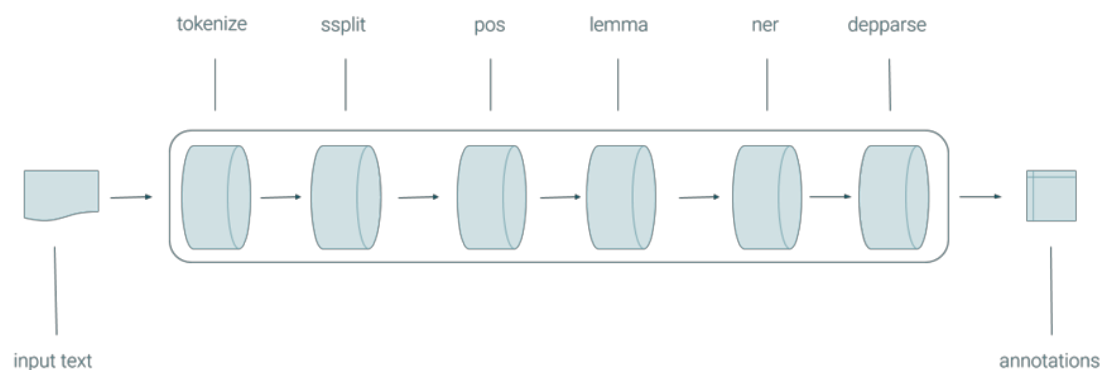


Figure 2 Pipeline de CoreNLP. Fuente: [Stanford CoreNLP](#)

Los espacios semánticos de palabras han sido muy útiles pero no pueden expresar el significado de frases más largas. Un mayor progreso hacia la comprensión de tareas como la detección de sentimientos requiere de mayores recursos de etiquetado, evaluación y modelos de composición más potentes. Para remediar esto los investigadores han utilizado en este caso el dataset “Sentiment Treebank” mencionado anteriormente que incluye etiquetas de sentimiento (de grano fino) para 215.154 frases en los árboles de análisis de 11.855 oraciones. Para abordar estos problemas, se ha utilizado una red de tensor neuronal recursivo (Recursive Neural Tensor Network) que, al entrenarse en el nuevo conjunto de datos de entrenamiento supera a todos los métodos anteriores en varias métricas consiguiendo mejorar la clasificación positiva / negativa en oraciones simples del 80% al 85,4%. La precisión de la predicción de sentimientos para todas las frases alcanza el 80,7%, una mejora del 9,7% con respecto al

baseline del conjunto de funciones. Por último, *los investigadores indican que este es el único modelo que puede capturar con precisión el efecto de conjunciones contrastivas, así como la negación en su alcance que es de varios niveles de árbol para frases tanto positivas como negativas.*

4.1.1 GPT-3

GPT-3 es un modelo de lenguaje, cuyo objetivo principal es predecir qué es lo siguiente que debe seguir a un texto en función de los datos previos. Los **Transformers generativos pre-entrenados** (“[Generative pre-trained transformers](#)”) son modelos de lenguaje que producen texto similar al que un humano escribiría. GPT-3 es el modelo de lenguaje más potente desarrollado hasta ahora. Su predecesor, GPT-2, ya podía generar flujos de texto convincentes en una variedad de estilos diferentes cuando se le alimenta con una frase o texto inicial. Pero GPT-3 es un gran avance ya que el modelo tiene 175.000 millones de parámetros (los valores que una red neuronal intenta optimizar durante el entrenamiento), en comparación con los ya enormes 1.500 millones de GPT-2. Y con los modelos de lenguaje, el tamaño realmente importa. Podríamos decir que GPT-3 está esencialmente preentrenado con todos los textos disponibles en Internet.

El Transformer es una arquitectura de red neuronal que utiliza capas de atención como su componente principal. “Todo lo que necesita es atención”, se refiere a una *arquitectura de secuencia a secuencia (o [Seq2Seq](#)) de red neuronal que transforma una secuencia particular de elementos (por ejemplo las palabras de una oración determinada) en una secuencia diferente*. Los modelos Seq2Seq consisten en una combinación de un codificador y un decodificador con el codificador tomando secuencias como entradas y mapeando lo que se conoce como un vector n-dimensional o una dimensión espacial superior. Luego, el vector se coloca en un decodificador que genera una secuencia de salida que, por ejemplo, puede estar en un idioma diferente. Una parte importante del mecanismo de atención es que permite que el codificador se centre en las palabras clave que son relevantes semánticamente en una oración determinada. Esto, a su vez, permite que el decodificador comprenda las partes de la secuencia en la oración que son significativas y proporcionan sentido a la oración.

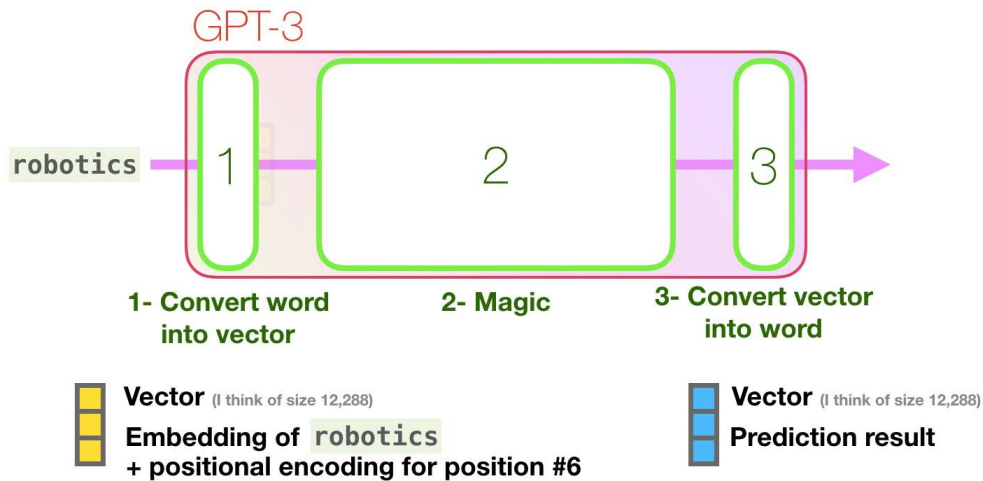


Ilustración 13GPT-3. Fuente: [How GPT3 Works](#)

Los “word embeddings” de GPT-3 no se basan en la percepción del mundo lo que explica por qué a menudo tiene dificultades para responder preguntas de sentido común. También *carece de intenciones, emociones, metas, creencias y deseos que impulsan y modulan el uso del lenguaje en los seres humanos*. Sus expresiones no tienen "propósito", no presenta empatía ni “piensa” antes de hablar en la medida en que esto implica encajar una idea y palabras como componentes de una oración o pensamiento que se desea expresar. Sin embargo, *las representaciones internas de GPT-3 (basadas en el aprendizaje a partir de un corpus de textos inmenso) le permiten componer oraciones de una manera que a menudo parece natural y crear relaciones entre palabras dentro de párrafos completos*.

4.1.2 BERT y ALBERT

Aunque técnicas como [word2vec](#) o [GloVe](#) generan una representación para cada palabra del vocabulario de una forma no dependiente del contexto (aunque sí que tomen en cuenta el contexto), otros sistemas basados en [Transformers](#) como [BERT](#) *sí generan vectores contextuales (vectores diferentes según el contexto en el que aparezca la palabra)*. En nuestro caso, la utilización del contexto podría ser de interés para *buscar una representación que tenga en cuenta el contexto de cada ocurrencia de un marco*, es decir, podríamos buscar representaciones de marcos que sean útiles para predecir otros marcos presentes en un documento alrededor del anterior.

Así como **Word2VEC codifica la coaparición de palabras**, BERT relaciona todas las palabras, codificando más información como la cercanía en contextos más amplios. Ya no codificamos únicamente la coaparición de palabras si no que BERT codifica, podríamos decir, algo de sintaxis, algo de semántica, etc. y los vectores generados por este tipo de sistemas contienen mucha más información que aquellos generados por Word2Vec.

[ALBERT](#) (A Lite BERT) es un sistema mejorado basado en BERT diseñado por Google AI que utiliza un número mucho menor de parámetros (hasta un 89%) con una pérdida mínima de precisión.

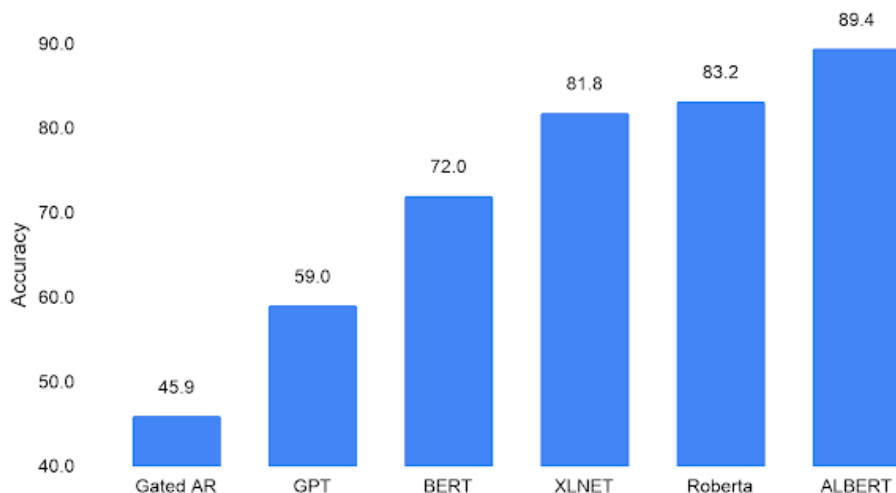


Ilustración 14 Precisión de diversos algoritmos en el reto RACE. Fuente: [Google Research](#)

4.1.3 Traducción automática neuronal (NMT)

La traducción automática neuronal (Neural Machine Translation, [GNMT](#)) es un campo de estudio utilizado entre otros por Google que utiliza técnicas de aprendizaje automático profundo (“deep learning”) para producir mejores traducciones entre diferentes lenguajes.

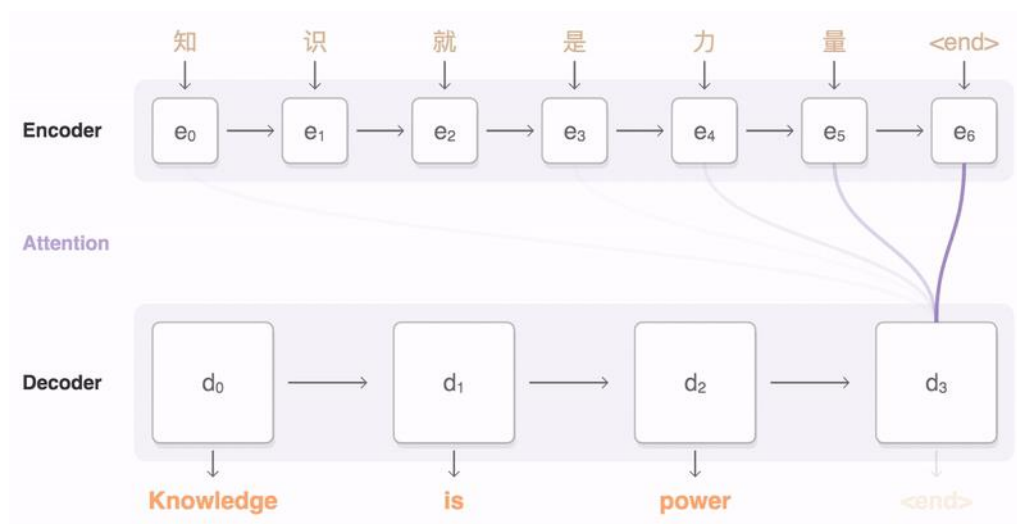


Ilustración 15 GNMT (Google)

Inicialmente (2016) Google utilizó en GNMT un modelo basado en [redes neuronales recurrentes \(RNN\)](#) que mejoró enormemente la calidad de las traducciones en múltiples idiomas. Sin embargo, en 2020 Google anunció que [modificó el sistema GNMT original para entrenar modelos con un codificador transformer](#) y un **decodificador RNN**, (implementados en [Lingvo](#), un framework para construir redes neuronales en [TensorFlow](#)) al demostrarse que los modelos de transformer son generalmente más efectivos en la traducción automática que los modelos RNN, aunque Google sugirió que la mayoría de estas ganancias de calidad provienen del codificador del transformer, y que el decodificador del transformer no era significativamente mejor que el decodificador RNN. Dado que el decodificador RNN es mucho más rápido en el tiempo de inferencia, lo han mantenido aplicando una variedad de optimizaciones antes de acoplarlo con el codificador del transformer. Los modelos híbridos resultantes según Google son de mayor calidad, más estables en el entrenamiento y de una menor latencia.

4.2 Estado del arte en el Espacios Vectoriales en Aprendizaje Automático

En este apartado vamos a centrarnos en las diferentes teorías y técnicas utilizadas en el campo del procesamiento del lenguaje natural que utilizan espacios vectoriales como forma de representar palabras y estudiar su similitud semántica. Esto nos permitirá construir la base sobre la que investigar si podemos utilizar técnicas similares para ubicar “marcos” en un espacio vectorial para representar “entidades” (personas, grupos políticos, documentos) así como para medir su similitud (ideológica, política, de personalidad, etc.). Así como mediante técnicas de “word embedding” podemos “reducir” o “insertar” un vocabulario completo dentro de un espacio lineal de baja dimensionalidad donde las dimensiones son las principales características o rasgos (“features”) latentes en dicho vocabulario, ***creemos que mediante “frame embedding” podemos “reducir” dentro de un espacio vectorial las principales características o rasgos políticos (“features”) latentes en una entidad (persona, partido o texto político).***

Un espacio vectorial es una colección de vectores caracterizados por su dimensión. Para medir la similitud entre dos palabras de destino (v y w), necesitamos una métrica que tome dos vectores de la misma dimensión y nos dé una medida de su similitud. La métrica de similitud más común es utilizar el coseno del ángulo entre los vectores. La [similitud coseno](#), como la mayoría de las medidas de similitud de vectores utilizadas en PNL, se basa en el [producto escalar](#) (el operador del producto escalar de álgebra lineal), también llamado producto interno.

Las dos familias de modelos principales para el aprendizaje de vectores de palabras son:

- 1) Métodos de factorización de matrices globales, como el análisis semántico latente ([LSA](#)) que asume que las palabras que tienen un significado cercano aparecerán en fragmentos de texto similares (hipótesis de distribución o “[distributional hypothesis](#)”).
- 2) Métodos de ventana de contexto local, como el modelo skip-gram de Mikolov mencionado en el apartado 4.2.2 (Word2Vec).

4.2.1 Espacios vectoriales en entornos de “Word Embedding”

La intuición detrás de un sistema como word2vec es que en lugar de contar la frecuencia con la que cada palabra “w” aparece cerca de otra palabra (por ejemplo “albaricoque”), entrenaremos a un clasificador en una tarea de predicción binaria: “¿Es probable que la palabra “w” aparezca cerca de albaricoque?” Una vez realizada la predicción tomaremos **los pesos del clasificador** aprendidos como las inserciones o “**word embeddings**”.

Un corpus no es una simplemente una bolsa de palabras ya que tiene faceta interesante que es su naturaleza distributiva. La distribución adopta muchas formas, por ejemplo, las palabras de un corpus tienden a concentrarse en determinados contextos y a estar más dispersas en otros (dispersión). Tienden a coocurrir con otras palabras específicas o en contextos gramaticales específicos. Por las razones anteriores, la lingüística de corpus es principalmente una ciencia distributiva. Una fuente de inspiración para la lingüística de corpus es la **Hipótesis distributiva** (“*Distributional Hypothesis*”, DH) que en su forma más simple establece que “*los lexemas con una distribución similar tienen «significados» similares*”.

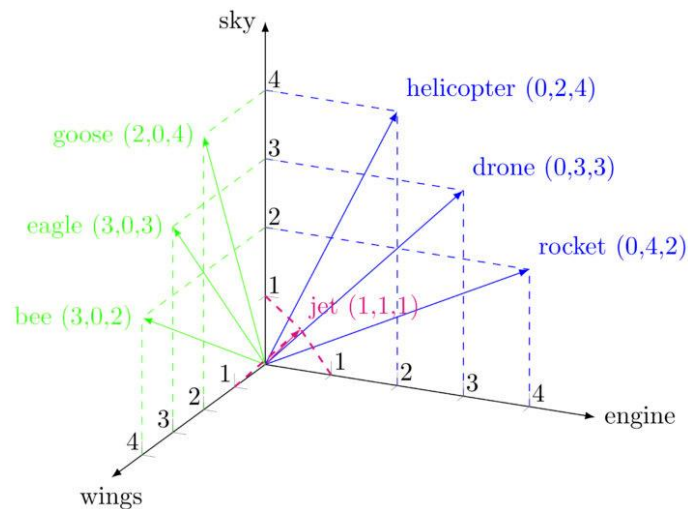


Ilustración 16 Espacio vectorial de ejemplo (Desagulier, 2018)

Los “word embeddings” son una implementación computacional de la hipótesis distributiva. Supongamos que tenemos un mini corpus con siete palabras (abeja, águila, ganso, helicóptero, dron, cohete y jet) y tres contextos (alas, motor y cielo). Cada palabra se caracteriza por tres coordenadas que corresponden al número de veces que se encuentra la palabra en cada contexto.

Por ejemplo en la figura anterior la palabra “helicóptero” no se encuentra en el contexto de las alas y ocurre dos y cuatro veces en los contextos motor y cielo, respectivamente. Por tanto, sus coordenadas en este espacio vectorial simple serían (0,2,4). Cada palabra ocupa una posición específica en el espacio vectorial y el vector que representa a una palabra es la flecha desde el punto donde los tres ejes se cruzan hasta el punto final definido por las coordenadas. La suposición subyacente es que las similitudes semánticas están relacionadas con las afinidades contextuales. Por ejemplo, helicóptero y dron están cerca porque ocurren en contextos similares, tienen perfiles vectoriales similares y, por lo tanto, están cerca en el espacio vectorial así definido.

Una implementación estándar de un sistema Word2Vec utiliza alrededor de 300 dimensiones (vector con 300 componentes para cada palabra representada) donde **la distancia euclídea entre dos vectores nos permite representar algún tipo de relación entre las palabras representadas por dichos vectores**. El tamaño del vector (su módulo) nos indica la cantidad de información almacenada en el mismo, aumentando la especificidad aunque manteniendo la dirección del mismo. **Para calcular la similitud semántica entre generalmente se utiliza la similitud del coseno**, es decir, mediante el cálculo del coseno del ángulo entre los dos vectores representando una palabra.

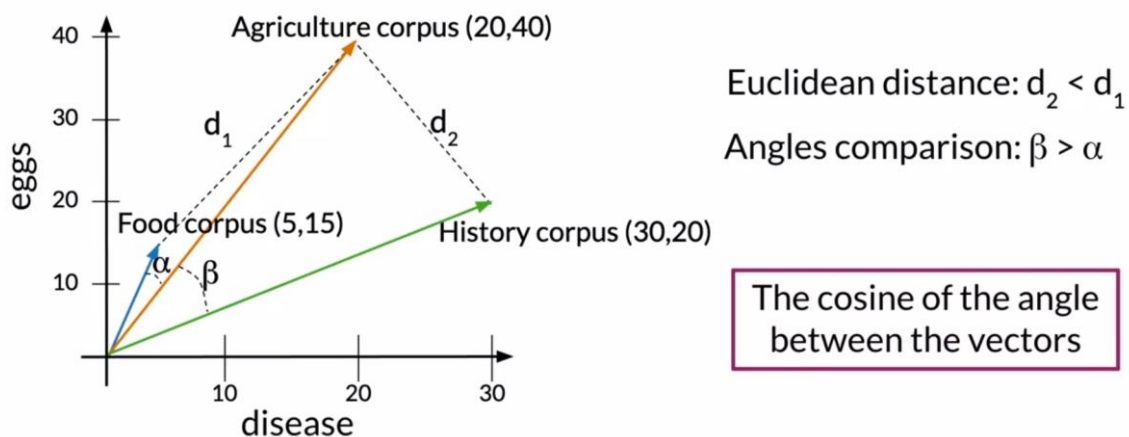


Ilustración 17 Similitud entre palabras mediante uso del coseno (Fuente: (Chadha, 2020))

La medida de similitud por el coseno es un método común para determinar la similitud entre vectores calculando el coseno de su ángulo interno:

- Si el ángulo es pequeño, el coseno tiende a 1.
- A medida que el ángulo se acerca a 90°, el coseno tiende a 0.

- Cuanto mayor sea ángulo entre los dos vectores (más diferentes o separados en el espacio vectorial) menor será la similitud.

En el ejemplo mostrado en la figura anterior (Chadha, 2020), el ángulo α entre la alimentación y la agricultura es más pequeño que el ángulo β entre la agricultura y la historia, lo que indica que la alimentación y la agricultura son más similares que la agricultura y la historia. Por lo tanto, en este caso, el coseno de esos ángulos es una mejor medida de similitud entre estas representaciones vectoriales que su distancia euclidiana.

El modelo word2vec utiliza una arquitectura de red para representar las palabras de entrada y de salida asociadas con mayor probabilidad. Suponiendo que nuestra red neuronal tiene solo una capa oculta, las dos matrices que representan dicha red representan los pesos y sesgos que permiten a la red calcular su representación interna de la función que mapea el vector de entrada al vector de salida. **Los pesos de la red son una representación del mapeo entre la entrada y la salida**, aunque un peso individual no necesariamente representa nada significativo por sí mismo sino que son los pesos de conexión entre todas las unidades en la red neuronal lo que podemos definir como la representación del mapeo de funciones y por eso las redes neuronales a menudo se denominan modelos de “caja negra” ya que es muy difícil interpretar por qué toman decisiones particulares y cómo aprenden.

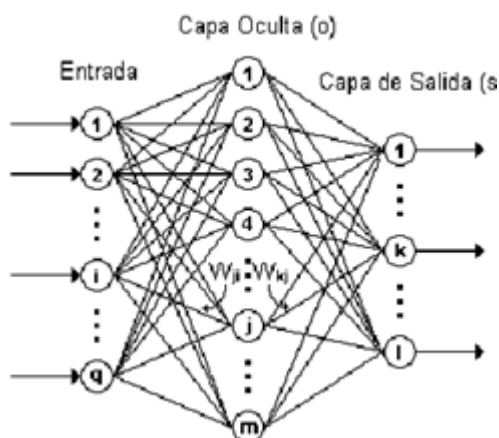


Ilustración 18 Red neuronal con una capa oculta

Los métodos para aprender los “word embeddings” densos siguen siendo un área activa de investigación para comprender la estructura semántica oculta en dichos embeddings y proporcionar información sobre la interpretación de sus dimensiones. Dado que estos embeddings se generan a través de métodos de aprendizaje no supervisado entrenados sobre

grandes corpus no anotados, es posible que algunas relaciones conceptuales que los humanos anticipan pueden pasarse por alto y otras que los humanos no anticipan puedan generarse en el espacio de embeddings. Si quisiéramos proporcionar una interpretación a las dimensiones tendríamos que estudiar cómo, por ejemplo, categorías de conceptos con propiedades semánticas similares se agrupan en una categoría común, preguntándonos cómo está codificado un concepto en particular en una dimensión, o cuánta información se puede extraer de una dimensión con respecto a un concepto en particular.

Word2vec, el más utilizado de estos modelos, utiliza como veremos en el siguiente apartado una red neuronal poco profunda para **construir un espacio vectorial de alta dimensionalidad que no solo refleja las propiedades sintácticas, sino también semánticas del corpus de entrada**. Esto es posible porque, en lugar de usar vectores para representar n-gramas específicos en relación con un vocabulario, las redes neuronales usan vectores (de menor tamaño o dimensión) ponderados más pequeños "ajustados" a las características (features) principales del corpus de entrada mediante la ejecución de varias iteraciones de acuerdo con una función de pérdida (proceso denominado normalmente [backpropagation](#)).

4.2.2 Word2Vec

Word2vec es un algoritmo de aprendizaje **autosupervisado** que utiliza redes neuronales (poco profundas a diferencia de las redes de [deep learning](#) o aprendizaje profundo) para **aprender las asociaciones entre palabras presentes en un corpus de textos**. Las **palabras utilizadas no están etiquetadas, ni categorizadas** y no es necesaria la intervención humana durante el aprendizaje y donde los datos de entrenamiento se etiquetan de forma autónoma. Los datos se etiquetan encontrando y explotando las relaciones entre las palabras en los diferentes textos de entrada al modelo. En lugar de entrenar a la red neuronal para aprender el significado de la palabra objetivo (mediante datos etiquetados por humanos por ejemplo), **Word2Vec enseña a la red a predecir palabras que se encuentran cercanas (en posición dentro del texto) a la palabra objetivo**. Así pues, una vez que el modelo ha sido entrenado podemos detectar palabras sinónimas, con significados cercanos y sugerir palabras que podrían seguir en una frase incompleta.

El concepto básico es que **Word2vec representa cada palabra con un vector** (en nuestro caso un vector no es más que una secuencia de números). Algunos autores podrían considerar esto como un sistema que utiliza etiquetas ya que tenemos etiquetas: las palabras que rodean a la palabra que intentamos predecir (pero al no ser necesarios etiquetadores humanos, Word2Vec es un sistema autosupervisado). Una interpretación geométrica sería que Word2Vec es un modelo que aprende a mapear cada palabra en un espacio vectorial continuo de baja dimensionalidad a partir de las propiedades de la distribución de las palabras observadas en un corpus de texto, y además maximizando la información mutua (Pointwise Mutual Information) entre palabras con contextos “similares”.

Una de las características de las **dimensiones del espacio vectorial** (y por lo tanto de las coordenadas de los “embeddings”) generados en Word2Vec es que, en principio, estas **no tienen un grado de interpretabilidad claro**. Este aspecto de las técnicas vectoriales en NLP es compartido con otras técnicas propias del aprendizaje automático (como por ejemplo las redes neuronales donde existe una discusión acerca de la interpretabilidad relacionada con aspectos éticos del uso de dichas redes).

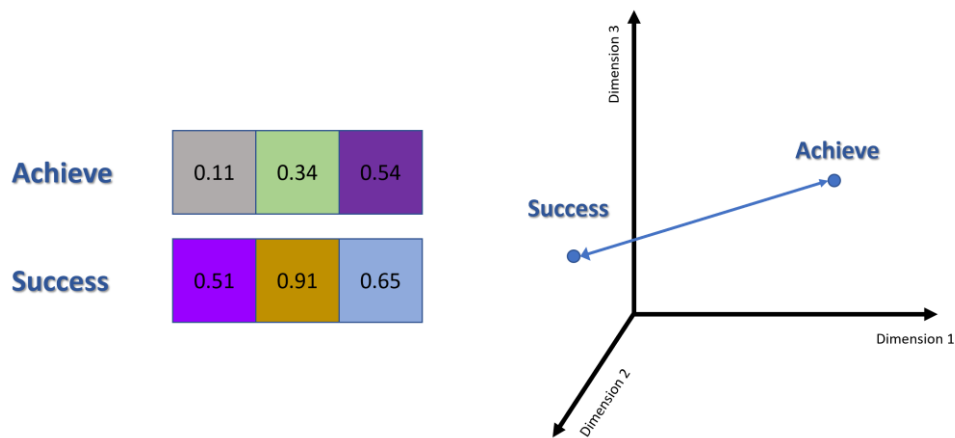


Ilustración 19 Visualización 3D de vectores de palabra (Word vector) después de reducción de dimensiones (Fuente: [Eric Kim](#))

De hecho, Word2vec trata de aprender las representaciones vectoriales para cada palabra que mejor describen el “contexto” de dicha palabra. La asunción principal es que palabras similares tienden a ocurrir o presentarse juntas y por lo tanto tendrán un “contexto” similar. En el ejemplo, “una pera es una fruta” y “una manzana es una fruta”, “pera” y “manzana” comparten un contexto similar (fruta en este caso). La creación de los vectores de cada palabra es un proceso en el que tomando un gran corpus de textos, creamos un vector para cada palabra de forma que ***todas las palabras que comparten contextos comunes en el corpus están ubicadas de forma próxima entre sí en el espacio vectorial***. Word2vec es un algoritmo sencillo que combina dos técnicas: CBOW (“[Continuous bag of words](#)“) y [Skip-gram](#). Ambas son redes neuronales poco profundas que mapean palabras a variables objetivo (que son también palabras) y ambas técnicas también aprenden los pesos (de las etapas de la red neuronal) que son los que actúan como representaciones vectoriales de cada palabra. Entrenar la parte CBOW de la red implica encontrar, utilizando “[propagación hacia atrás de errores](#)” (“[backpropagation](#)”) y “[descenso del gradiente](#)” (“[gradient descent](#)”) los valores de los pesos para los cuales la predicción de la red neuronal es lo más cercana posible al valor etiquetado, dado el contexto de palabras de entrada. En esta parte CBOW por lo tanto, la entrada a la red neuronal es el contexto de una palabra (las palabras que la rodean) y la etiqueta, es decir ***la salida de la red neuronal, es la palabra buscada en sí misma***.

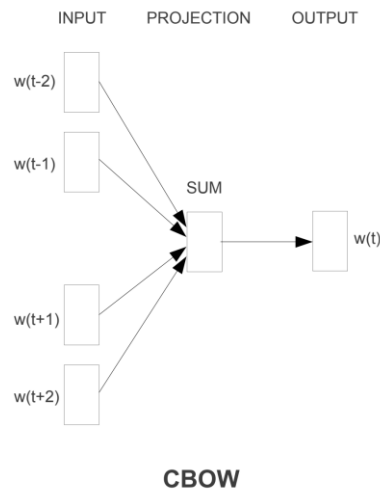


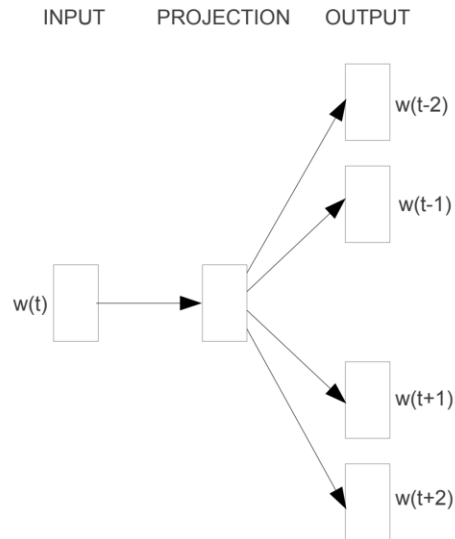
Ilustración 20 Modelo CBOW

La [función objetivo](#) del método CBOW es:

$$J_{\theta} = \frac{1}{T} \sum_{t=1}^T \log p(w_t | w_{t-n}, \dots, w_{t-1}, w_{t+1}, \dots, w_{t+n})$$

Es decir, al entrenar nuestro modelo estamos buscando los pesos de la capa oculta de la red neuronal que minimizan la función de pérdida (“[loss function](#)”). Esto es encontrar, al entrenar la red encontramos los pesos que, dado un contexto (vector de palabras), permiten predecir con la mayor precisión cual es la palabra objetivo (la salida de la red neuronal) que estamos buscando.

En la parte skip-gram, que es donde se computan los word embeddings propiamente, la etiqueta (la salida de la red neuronal) es el contexto (palabras alrededor) de la palabra de entrada. En lugar de utilizar las palabras que rodean a una palabra (tanto anteriores como posteriores) para predecir a dicha palabra (CBOW), **en skip-gram utilizamos como entrada la palabra central (codificada como un número entero) para predecir las palabras que la rodean** (el contexto, es decir, la ventana de palabras que la rodean).



Skip-gram

Ilustración 21 Modelo Skip-gram

La [función objetivo](#) del método skip-gram suma las probabilidades log de las palabras alrededor (anteriores y posteriores, izquierda y derecha en el texto) de la palabra objetivo w_t para producir el siguiente objetivo:

$$J_{\theta} = \frac{1}{T} \sum_{t=1}^T \sum_{-n \leq j \leq n, j \neq 0} \log p(w_{j+1} | w_t)$$

Crear los vectores para cada palabra es un proceso en el que **se toma un corpus de datos extenso y se crean un vector para cada palabra** de forma que las palabras que comparten “contextos” comunes (es decir palabras que aparecen en los mismos “contextos”) se ubican de forma cercana a las otras en el espacio vectorial. Veamos el proceso de generación de los vectores para cada palabra de un corpus en mayor detalle.

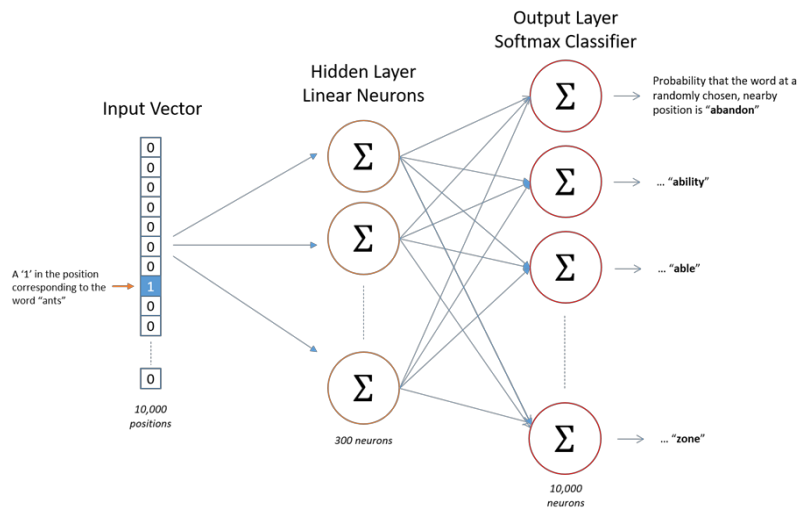


Ilustración 22 Estructura general de la red en Word2Vec. Fuente: [Chris McCormick](#)

Vamos a presentar como ejemplo el entrenamiento de una red neuronal para que dada una palabra específica en el medio de una oración (la palabra de entrada) el algoritmo Word2Vec busque las palabras cercanas y elija una al azar. La red nos dirá la probabilidad de que cada palabra de nuestro vocabulario sea la "palabra cercana" que elegimos. Vamos a representar la palabra de entrada "hormigas" como un [vector one-hot](#) ("[one-hot vector](#)"). Este vector tendrá 10.000 componentes (uno por cada palabra presente en nuestro vocabulario de ejemplo) y tendrá un "1" en la posición correspondiente a la palabra "hormigas" y ceros en todas las demás posiciones. La salida de la red neuronal es un único vector (también con 10.000 componentes) que contiene, para cada palabra de nuestro vocabulario, la probabilidad de que una palabra cercana seleccionada al azar sea esa palabra del vocabulario.

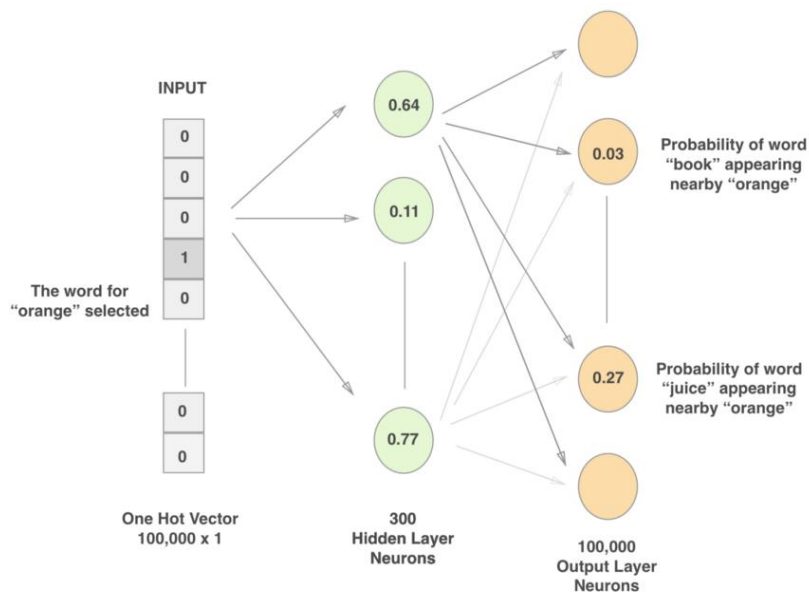


Ilustración 23 Visualización de la red neuronal en Word2Vec. Fuente: [Sanket Gupta](#)

Al final del entrenamiento, la capa oculta contendrá los vectores de palabras entrenados. El tamaño de la capa oculta corresponde al número de dimensiones de nuestro vector (y del espacio). En el ejemplo anterior, cada palabra tendrá un vector de longitud 300 (tamaño del espacio utilizado generalmente en Word2Vec). La red neuronal skip-gram contiene una gran cantidad de pesos, en este ejemplo con 300 dimensiones y un vocabulario de 10.000 palabras, la red tiene 3 millones de pesos en la capa oculta y otros tantos en la capa de salida. Entrenar esta red en un gran conjunto de datos sería prohibitivo y necesitaría una gran capacidad de procesamiento por lo que los autores de word2vec introdujeron varios ajustes para hacer factible la formación.

La calidad de los “embeddings” aumenta con una mayor dimensionalidad, pero a partir de un punto la ganancia marginal disminuye con lo que normalmente la dimensionalidad de los vectores se establece entre 100 y 1.000. Una de las principales limitaciones de los word embeddings no contextuales (y de los modelos de espacio vectorial de palabras en general) es que las palabras con múltiples significados se combinan en una sola representación (un solo vector en el espacio semántico. BERT sin embargo puede generar embeddings contextuales, es decir, que cada palabra (o, en general, cada unidad lingüística) tendrá un embedding diferente dependiendo del contexto en el que aparezca. En otras palabras, *la polisemia y la homonimia no se manejan adecuadamente en el caso de los embeddings no contextuales.* En realidad, la dimensión del vector de la palabra no refleja el tamaño del vocabulario ya que lo que Word2Vec hace es mapear las palabras a su representación en un espacio vectorial de cualquier dimensión que se desee, menor a la dimensión del diccionario inicial generalmente para poder “comprimir” la información y disminuir la capacidad de procesado necesaria en la práctica. *Cada palabra está representada por un punto en este espacio y la dimensión vectorial de palabra son las coordenadas de esta palabra en este espacio.* Otra de las ventajas de este algoritmo es, como ya hemos comentado anteriormente, que las palabras que tienden a aparecer en el mismo contexto aparecen una al lado de la otra en este espacio.

En cuanto a los detalles del algoritmo Word2Vec, como por ejemplo el proceso de “backpropagation”, podemos encontrar una discusión detallada en (Bellei, 2018).

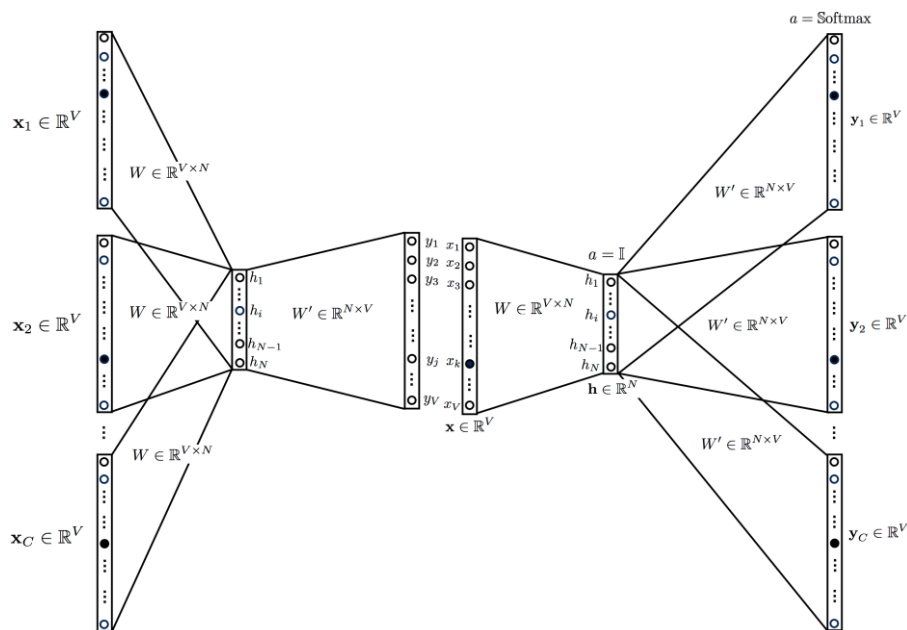


Ilustración 24 Topología de los modelos CBOW y Skip-gram

En lo referente a nuestra investigación, algunos autores como (Kozłowski, Taddy, & Evans, 2019) han mostrado como los “word embedding” son una herramienta útil en el estudio de una sociedad y su cultura:

“Argumentamos que los modelos de word embedding son una herramienta útil para el estudio de la cultura utilizando un análisis histórico de los elementos sociales compartidos. Los word embeddings representan relaciones semánticas entre palabras como relaciones entre vectores en un espacio de alta dimensión, especificando un modelo relacional de significados que es consistente con las teorías contemporáneas de cultura. **Las dimensiones inducidas por las diferencias entre palabras (rico - pobre) en estos espacios se corresponden a dimensiones de significado cultural, y la proyección de las palabras sobre estas dimensiones refleja asociaciones ampliamente compartidas, que validamos con encuestas. Al analizar el texto de millones de libros publicados durante 100 años, mostramos que los marcadores de clase social cambiaron continuamente a través de las transformaciones económicas del siglo XX, sin embargo, las dimensiones culturales básicas de la clase se mantuvieron notablemente estables. La excepción notable es la educación, que se vinculó estrechamente al nivel de riqueza independientemente de su asociación con el gusto cultivado.**”

4.2.3 GloVe

[GloVe](#) (Vectores globales para representación de palabras, “Global Vectors for Word Representation”) es un algoritmo de aprendizaje automático no supervisado desarrollado por [Stanford](#) que permite obtener representaciones vectoriales de palabras. El entrenamiento se realiza sobre la estadística de **coocurrencias global agregada (palabra a palabra) a partir de un corpus** y el resultado permite representar las palabras en forma de vectores en un espacio vectorial. El diseño de GloVe está motivado por la idea de que en lugar de utilizar coocurrencias de palabras locales (como Word2Vec), utilizaremos estadísticas de coocurrencia de palabras agregadas en todo el conjunto de datos. Supongamos que "perro" y "ladrido" ocurren simultáneamente N veces en el conjunto de datos. Hablando en términos generales, el modelo GloVe intenta predecir este número N a partir de los “embeddings” de ambas palabras.

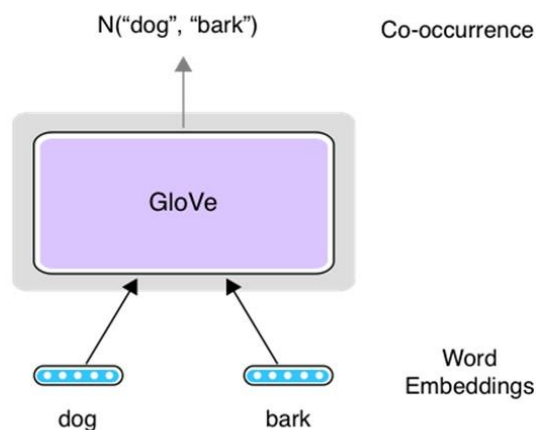


Ilustración 25 Modelo de GloVe

GloVe puede producir matrices equivalentes a la matriz de pesos de entrada y la matriz de pesos de salida de Word2vec, produciendo un modelo con la misma precisión que Word2vec pero en mucho menos tiempo. GloVe además se puede entrenar en corpus más pequeños y aun así converger. Word2vec se basa en la propagación hacia atrás para actualizar los pesos que forman los word embeddings. Sus autores opinan que la propagación hacia atrás (“[backpropagation](#)”) de la red neuronal en Word2Vec es menos eficiente que los algoritmos de optimización más maduros utilizados en SVD para GloVe. Sus mismos autores (Pennington, Socher, & Manning, 2014) presentan GloVe como:

“El resultado es un nuevo modelo de regresión bilineal logarítmica global que combina las ventajas de las dos familias principales de modelos en la literatura: factorización de matriz global y métodos de ventana de contexto local. Nuestro modelo aprovecha de manera eficiente

la información estadística entrenando solo en los elementos distintos de cero en una matriz de coocurrencia palabra-palabra, en lugar de en toda la [matriz dispersa](#) o en ventanas de contexto individuales en un corpus grande. El modelo produce un espacio vectorial con una subestructura con significado, como lo demuestra su rendimiento del 75% en una tarea reciente de analogía de palabras. También supera a otros modelos relacionados en tareas de similitud y [reconocimiento de entidades nombradas \(NER\)](#).”

4.2.4 Autocodificadores y Espacios latentes

El concepto de “espacio latente” (“latent space”) es importante al ser uno de los centrales en aprendizaje automático. Su objetivo sería **“aprender” las características (“features”) de los datos de entrada y simplificar su representación para encontrar patrones en los mismos.**

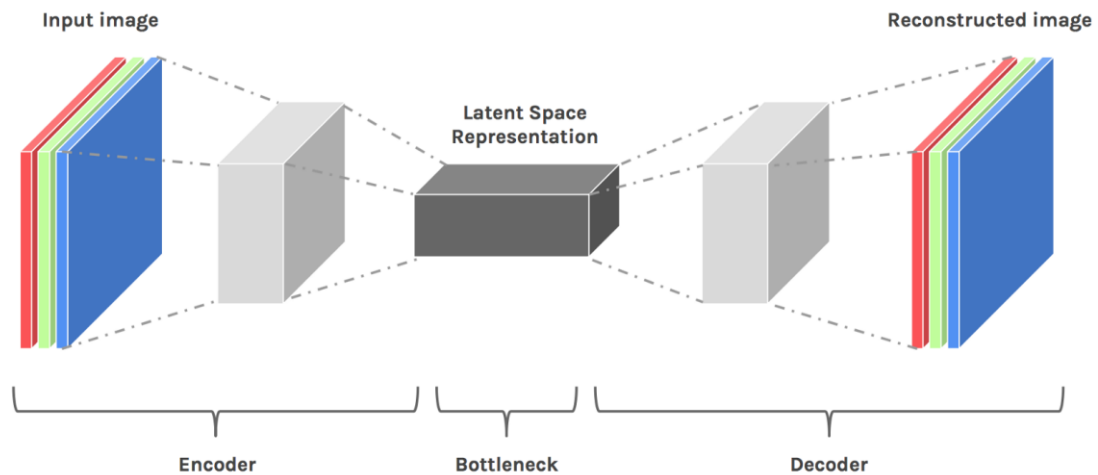


Ilustración 26 Arquitectura de un códec convolucional. Fuente: [Hackernoon](#)

Cada vez que el modelo aprende a través de un punto de datos, la dimensionalidad de la imagen se reduce primero antes de aumentarla finalmente. Esta reducción de la dimensionalidad la podemos considerar como una forma de compresión con pérdida. En este caso de uso, las representaciones del espacio latente se utilizan para transformar formas complejas de datos sin procesar (es decir, imágenes, videos), en representaciones más simples que son "más convenientes de procesar" y analizar.

Un tipo común de modelo de aprendizaje profundo que manipula la "cercanía" de los datos en el espacio latente es el **autocodificador** (“**autoencoder**”), una red neuronal que actúa como una función de identidad, es decir, un codificador automático aprende a generar en su salida los datos de inicio. Cuando forzamos un modelo a convertirse en una función de identidad, lo estamos **forzando a almacenar todas las características relevantes de los datos en una representación comprimida (reducción de dimensiones)** para que haya suficiente información en esa forma comprimida para que el modelo pueda reconstruirlo "con precisión" posteriormente.

Si imaginamos el hipotético corpus de todas las imágenes posibles que podría generar la red, podríamos pensarlo en un espacio de dos dimensiones donde las imágenes más similares estuvieran próximas entre ellas y donde, en consecuencia, podríamos transitar entre dos imágenes cualesquiera transformándolas lentamente una en la otra a partir de pequeñas

diferencias. Una [Red Generativa Antagónica](#) (RGA) o GAN ([Generative adversarial network](#)) entrenada contiene, en cierta manera, un espacio como este, pero no compuesto por dos dimensiones sino por muchas más (del orden de las centenas). El conjunto de estas dimensiones forma un entramado multidimensional donde cada imagen posible se puede entender como un punto en este sistema de coordenadas complejo; este entramado se denomina «espacio latente» (Sancho Caparrini, 2020)

Los autocodificadores variacionales, ([“Variational AutoEncoders”](#), VAE) son modelos de aprendizaje que *mezclan las redes neuronales con distribuciones de probabilidad*. Su principal uso es el de construir modelos generativos que son capaces de producir datos sintéticos que siguen los mismos patrones que los grandes conjuntos de datos de los que se alimentan. Normalmente, se han usado para generar imágenes que asemejan, por ejemplo, características conocidas tales como caras, vehículos, habitáculos, etc. aunque en teoría podrían usarse para la generación de cualquier tipo de dato, siempre y cuando el conjunto de entrenamiento de datos reales sea adecuado (en tamaño y contenido).

Haciendo uso de redes neuronales, un VAE se construye a partir de un autoencoder (autocodificador) formado por dos redes: un encoder (codificador), y un decoder (decodificador), a los que se añade una función de pérdida que mide cuánto se parece un objeto decodificado (la salida de la red) al objeto que se codifica (la entrada de la red). En un autoencoder neuronal, el encoder es una red neuronal que transforma sus entradas en una representación interna (intermedia), normalmente de dimensión muy inferior a la entrada, con el fin de obligarle a aprender una compresión eficiente que extraiga las propiedades principales de los datos de entrada. Posteriormente, esta representación intermedia (salida del encoder) se utiliza como entrada del decodificador (“decoder”) para recuperar, en la medida de lo posible, la entrada original.

Debido a esto, podríamos usar la salida del codificador (“encoder”) como nueva representación del dato para cualquier tarea, no solo para reproducirlo, porque en esa representación se encuentra codificada toda la información relevante del dato de entrada para poder llevar a cabo su reconstrucción. De hecho, es lo que hacen todas las redes neuronales en cualquiera de sus capas intermedias, pero los autoencoder centran su tarea en la capacidad reproductiva de la entrada (o de ligeras variantes de ella).

Normalmente, las dos redes involucradas, codificador (encoder) y decodificador (decoder), se entrenan simultáneamente como una unidad con el fin de acoplar el comportamiento de una al de la otra, y se usa la función de pérdida como mecanismo para dirigir este entrenamiento y

conseguir que el error existente entre el original y la reproducción se vaya reduciendo (desde un punto de vista técnico, la función de pérdida es la función objetivo a minimizar). Debido al objetivo que persiguen los autocodificadores, a esta función de pérdida suele denominarse también *pérdida de reproducción*, y el proceso asegura que se penalicen las configuraciones de la red que crean salidas distintas de la entrada.

Una vez obtenido un autoencoder, podríamos intentar usar el **espacio de representaciones intermedias (también llamado espacio latente)** como un mecanismo independiente para obtener con el decoder salidas similares a los datos de entrada. Para ello, tomaríamos al azar un punto del espacio latente, le aplicaríamos el decoder, y como resultado obtendríamos una salida que debe reflejar un objeto que se mueve en el mismo espacio que las entradas.

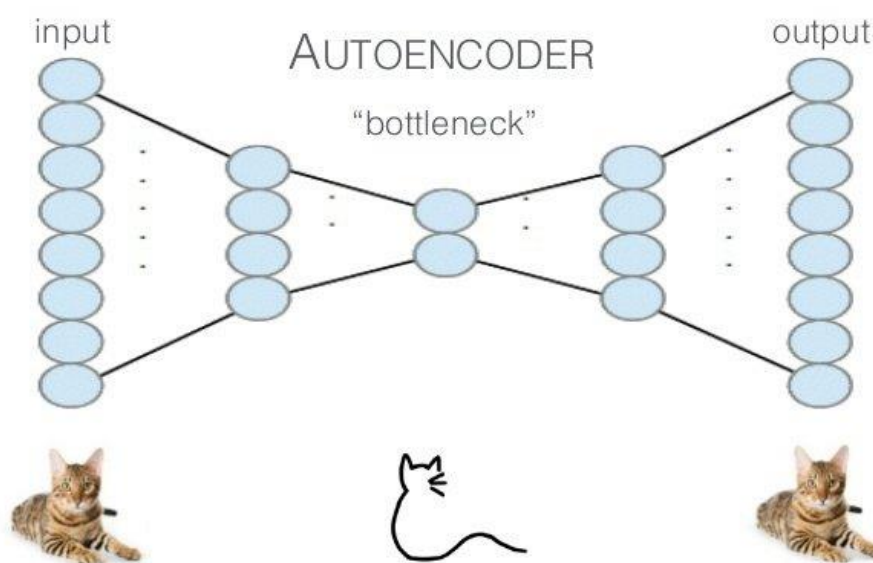


Ilustración 27 El "cuello de botella" de un autocodificador

El principal problema encontrado es que en la mayoría de los casos las representaciones internas que se obtienen (la representación en el espacio latente, salida del encoder, que sirve de entrada al decoder) forman un espacio que no es continuo, sino formado por diversas bolsas aisladas que agrupan en su interior representaciones de datos de entrada similares (clases). Formalmente, el espacio de representación intermedio en general no proporciona buenas propiedades de interpolación y conectividad debido a que su topología es mucho más compleja y no es cóncava. Los VAE precisamente tienen como objetivo eliminar ese problema construyendo explícitamente un espacio latente que ha de ser continuo, permitiendo generar objetos por medio de la interpolación de representaciones latentes de datos de entrada y, en consecuencia, la generación al azar por medio de un muestreo aleatorio del espacio latente. Para conseguir este efecto, un VAE considera que la representación intermedia no viene dada por medio de un vector de tamaño prefijado (la dimensión que se requiera en el espacio latente), sino por medio

de dos vectores del mismo tamaño pero con significados bien distintos: un vector de medias, $\vec{\mu}=(\mu_1,\dots,\mu_n)$, y un vector de desviaciones estándar, $\vec{\sigma}=(\sigma_1,\dots,\sigma_n)$, que conjuntamente forman un vector de variables aleatorias descritas por medio de distribuciones normales $(N(\mu_1,\sigma_1),\dots,N(\mu_n,\sigma_n))$. A partir de este vector de distribuciones normales podemos obtener un vector (que realmente es una muestra al azar siguiendo esas funciones de distribución) que podemos pasar al decoder para que genere una salida. Al ser un muestreo de una variable aleatoria n-dimensional no podemos asegurar que vayamos a obtener siempre la misma salida, sino un vector que, con mayor o menor probabilidad, produce una salida con pequeñas variaciones.

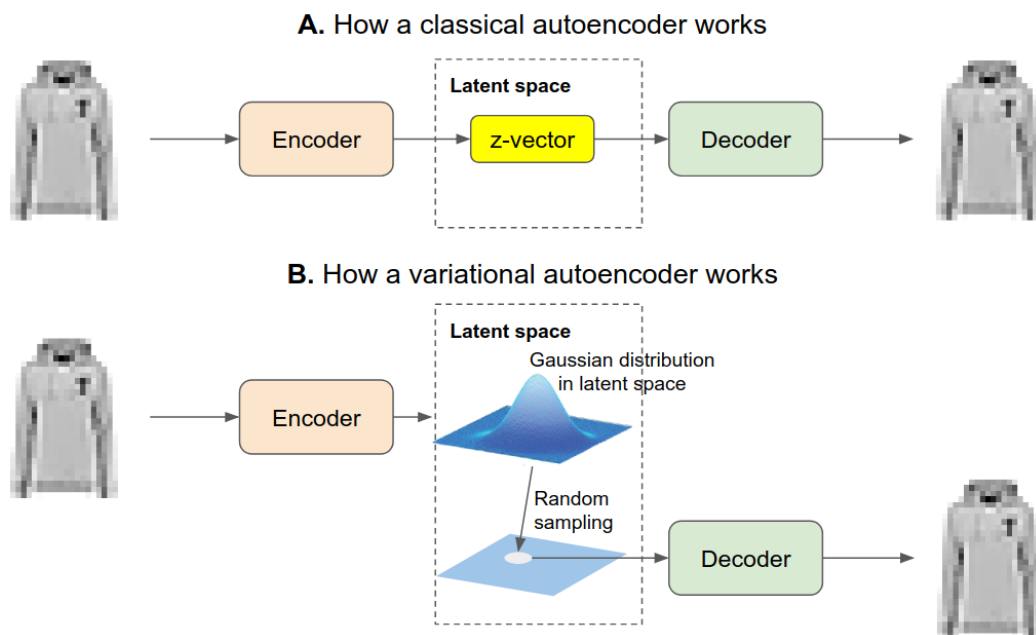


Ilustración 28 Forma en la que trabaja un autocodificador variacional

4.3 Procesamiento de lenguaje natural y análisis político

Presentamos en este apartado diferentes experiencias, métodos y estudios referidos a la utilización de técnicas de NLP en el estudio y análisis del [espectro político](#) que pueden ser de interés para el desarrollo de este trabajo.

4.3.1 La Brújula Política (The Political Compass)

Tal como hemos comentado en el apartado 2, en nuestro estudio no vamos a limitarnos a un análisis semántico de los textos como el realizado por ejemplo por (Falck & Marstaller, 2018). Estos trabajos realizan una extracción de entidades y al análisis de sentimientos de dichos textos (de cada periódico o publicación) respecto a las entidades bajo estudio para poder ubicar partidos políticos y medios de comunicación en un mismo espacio bidimensional como se muestra en la Ilustración 29. En nuestro caso, **nuestro objetivo es un análisis de textos para poder ubicar a los dirigentes políticos de diferentes formaciones políticas** (o a cualquier usuario de Twitter) **en un espacio vectorial multidimensional definido de forma que permita ubicar marcos (frames) culturales**. Este método de utilización de marcos culturales (centrados en el campo político) podríamos considerar que se encuentra dentro del nivel pragmático del análisis del lenguaje.

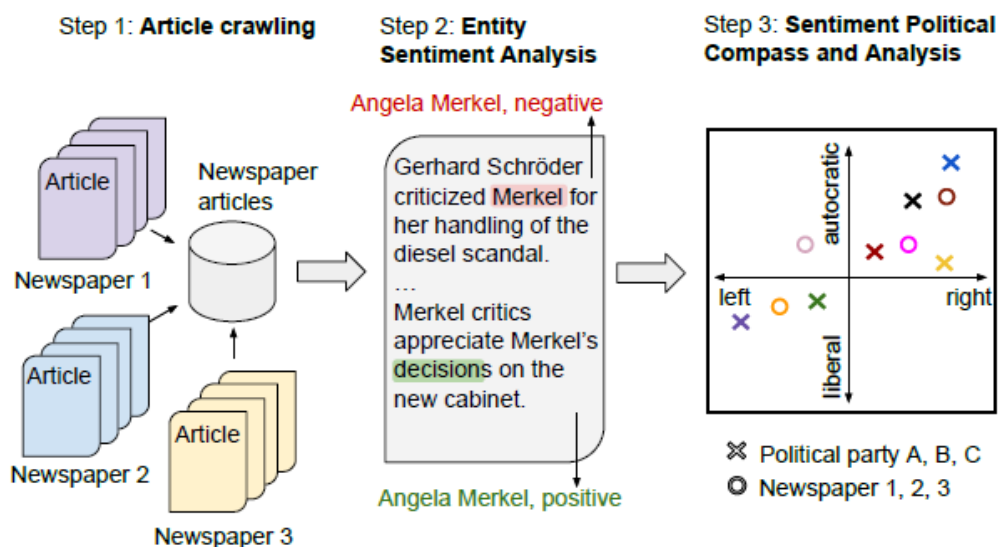


Ilustración 29 Esquema de alto nivel para utilizar la "Brújula de sentimiento político". Fuente: (Falck & Marstaller, 2018)

La “**Brújula Política**” ha sido diseñada y definida por “[The Political Compass](#)”, y en ella se permite *ubicar a todas las formaciones políticas alemanas en un espacio bidimensional*. Esta

representación está basada en “informes, votaciones parlamentarias, manifiestos.... y acciones que hablan más que las palabras. Nos ha implicado un trabajo mayor que simplemente conseguir que los políticos realizaran el test, pero es una evaluación mucho más precisa” (The Political Compass, 2020).

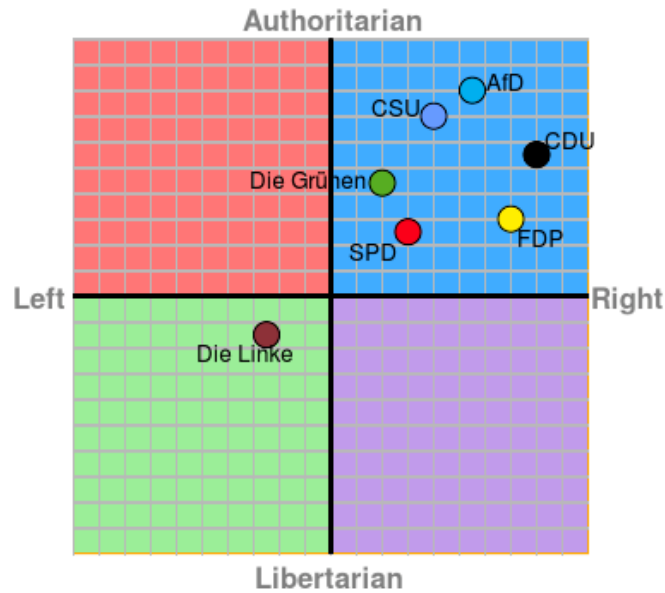


Ilustración 30 Espacio vectorial bidimensional utilizado por la brújula política

Como podemos observar en la Ilustración 30, al eje horizontal estándar derecha-izquierda estos autores añaden un eje vertical que representa la “rigidez ideológica” con un extremo del eje en el extremo ideológico “autoritario” y el otro en el extremo ideológico “libertario”. A partir de este espacio definido únicamente por dos dimensiones, se procede a ubicar a diferentes formaciones políticas, periódicos, televisiones y grupos mediáticos en dicho espacio, e incluso a estudiar cómo se modifica en el tiempo la posición que cada partido mantiene en el espacio vectorial bidimensional.

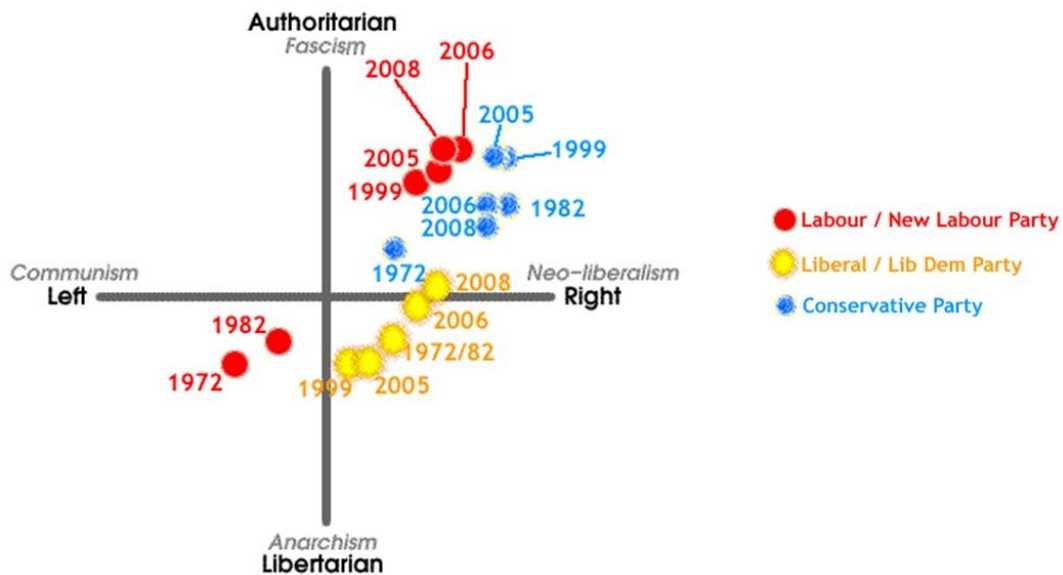


Ilustración 31 Cambio en la posición en el espacio vectorial de partidos en el Reino Unido. Fuente: [Political Compass](#)

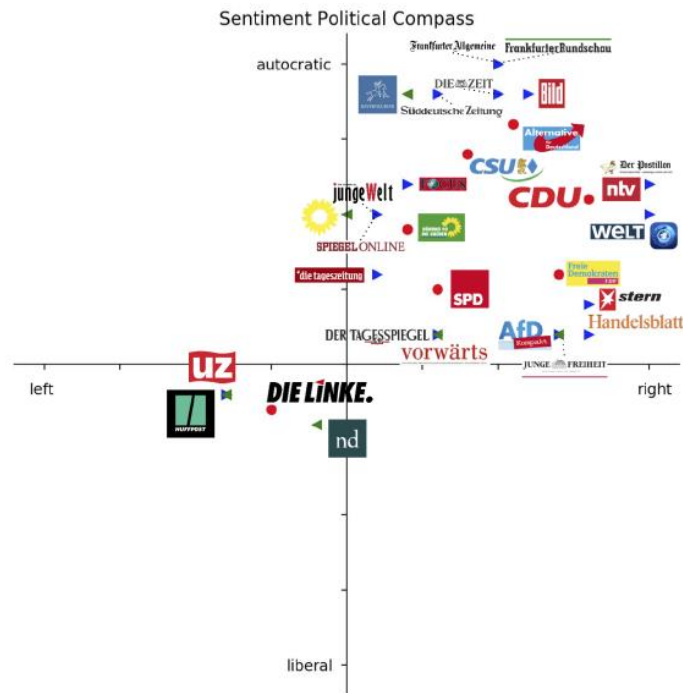


Ilustración 32 Brújula de sentimiento político". Fuente: (Falck & Marstaller, 2018)

Estos autores toman como método de validación y comparación de su técnica los resultados de la utilización de la API del sistema de comprensión del lenguaje natural Watson de IBM para la extracción de sentimientos sobre las entidades bajo estudio.

Un esquema similar es el seguido por [David Nolan](#) cuyos dos ejes están definidos en términos de **grado de libertad personal** (con extremos en un sistema totalitario en un extremo y liberal en el opuesto) y **grado de libertad económica** (con extremos en izquierda y derecha).

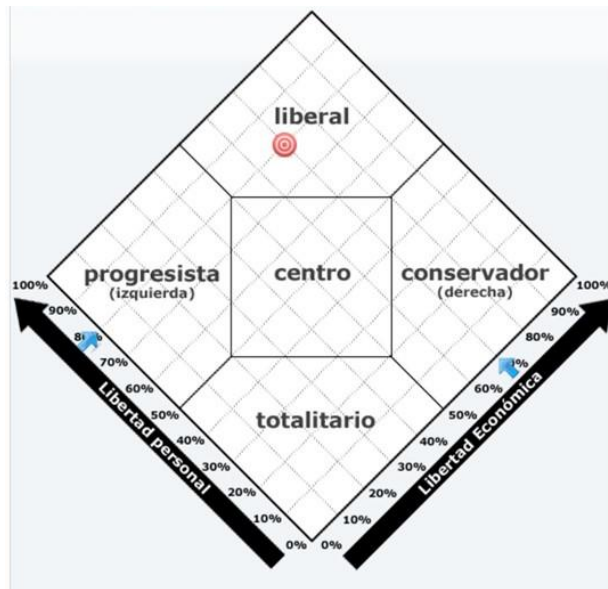


Ilustración 33 Diagrama de Nolan

Otro sistema similar pero con un mayor número de dimensiones (ocho) es el presentado por “[8 Values](#)”.

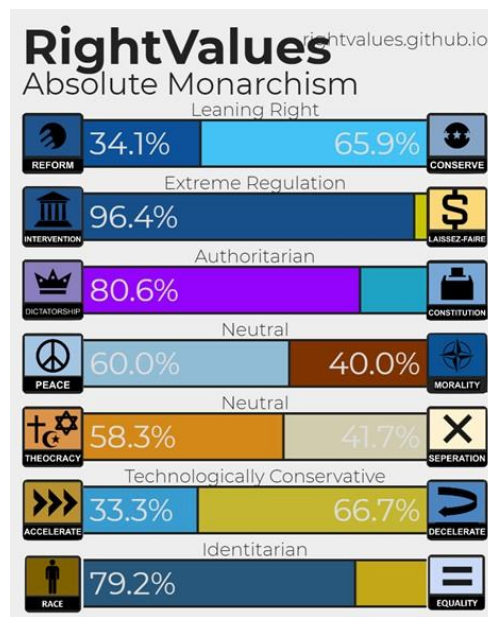


Ilustración 34 Dimensiones políticas utilizadas por “RightValues”

4.3.2 Escalado, Word Embeddings y análisis político

Dos de los enfoques más populares en ciencia política para la extracción de ideología a partir de textos son **WordScores** y **WordFish** que utilizan técnicas denominadas de *escalado de texto* (“**text scaling**”). El escalado (político) de textos tiene como objetivo **ordenar linealmente a partidos y políticos en todas las dimensiones políticas** (por ejemplo, ideología de izquierda a derecha) **en función del contenido textual** (por ejemplo, discursos políticos o programas y manifiestos de partidos) (Glavaš, Nanni, & Ponzetto, 2017). Los modelos existentes generalmente escalan los textos según el uso relativo de palabras.

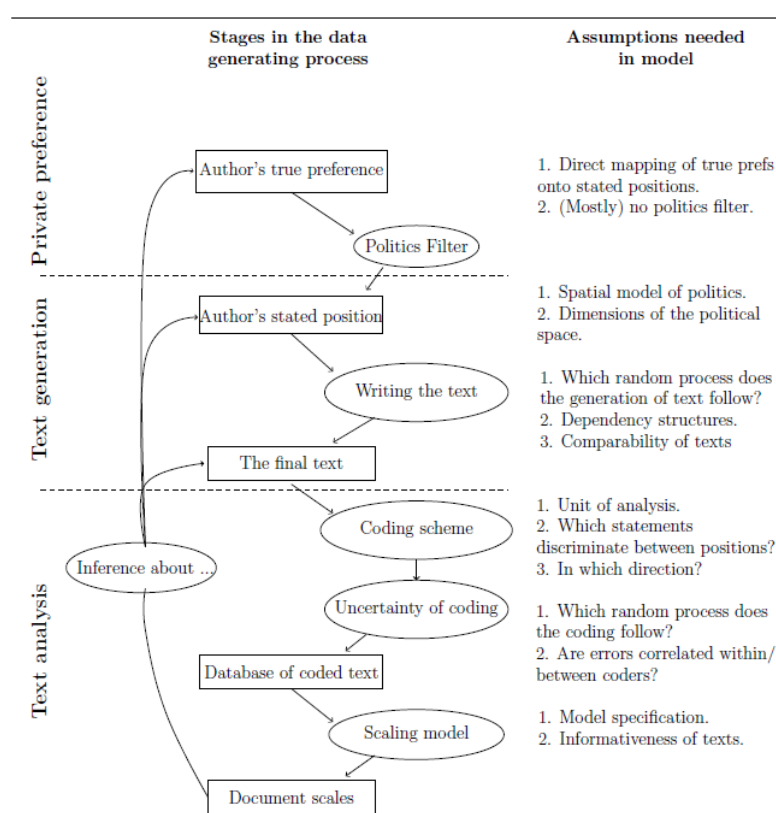


Ilustración 35 Modelo de generación y de análisis de textos políticos (Egerod, B. C. K. & Klemmensen, R., 2020)

WordScores se basa en una muestra de documentos etiquetados, por ejemplo, programas de partidos políticos anotados por expertos. Las probabilidades relativas de ocurrencia de palabras en los documentos etiquetados sirven para producir puntuaciones para cada palabra que pueden verse como indicadores de su carga ideológica.

La metodología de **WordFish** por su parte consiste en ajustar un modelo de regresión donde los *recuentos de las palabras* se proyectan en parámetros para cada partido político, utilizando un algoritmo de maximización de expectativas. Este enfoque permite *evitar la dependencia de*

anotaciones de expertos y equivale a estimar la especificidad del uso de palabras por cada partido en diferentes momentos. *WordFish* sirve para ordenar los documentos en un eje *unidimensional* según cuán similar son entre sí en el uso de ciertas palabras clave estando basado en dos supuestos fundamentales:

1. Las palabras siempre tienen el mismo significado dentro de los textos.
2. Los textos pueden ser ordenados en una dimensión latente X (un eje que articula la diferenciación ideológica de los documentos en un tema dado. La validez de esta suposición se mantiene en la medida en que el corpus de textos utilizados en el análisis es representativo de dicha dimensión.

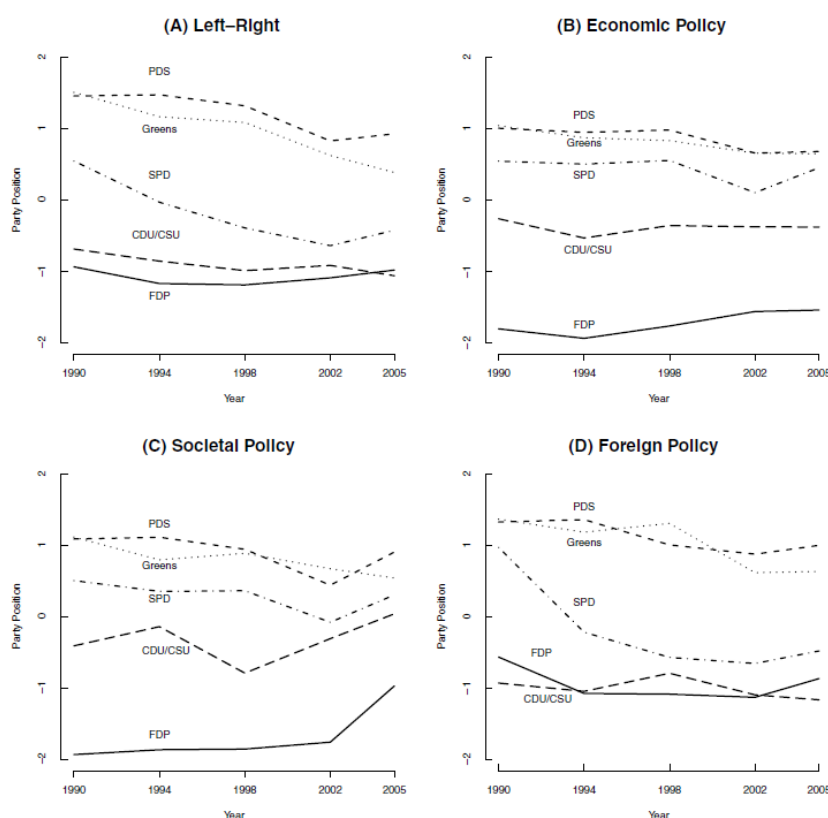


Ilustración 36 Posiciones de partidos alemanes (1999-2005) en varias dimensiones (Slapin & Proksch, 2008, pág. 714)

En lo que a las dimensiones se refiere, [WordScores](#) es una técnica para estimar posiciones en **una dimensión definida ex ante**, que está anclada por textos de referencia, mientras que [WordFish](#) escala el recuento de palabras y **recupera una dimensión que explica la mayor variación en el uso de palabras**. WordScores calcula las puntuaciones de las palabras ("wordscores") utilizando solo información de frecuencia de palabras de documentos con posiciones (políticas) conocidas (documentos de "referencia"). *WordScores funciona a través del supuesto básico de que cada palabra tiene un significado y posición política específica y*

que la posición de un documento se puede encontrar promediando sobre las “puntuaciones” (“wordscores”) de estas palabras.

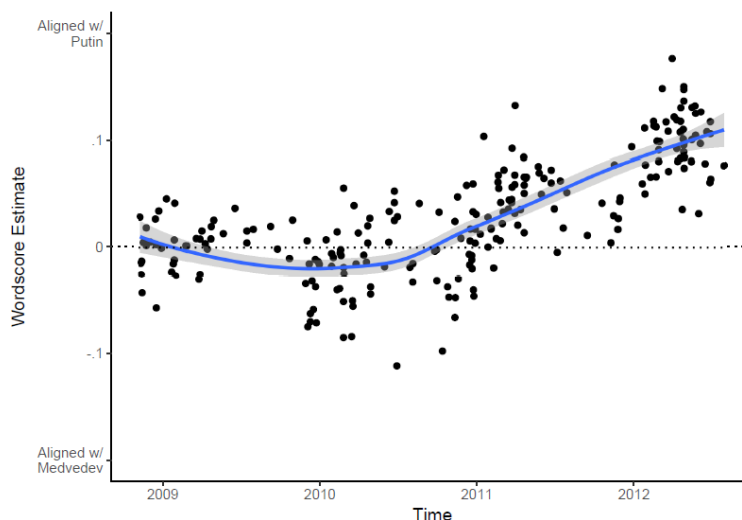


Ilustración 37 Ejemplo de dimensión en espacio político (Egerod, B. C. K. & Klemmensen, R., 2020)

Las aplicaciones de WordScores incluyen el análisis de datos de Twitter para “escalar” (o ubicar en una dimensión dada) a periodistas y medios de comunicación (de liberales a conservadores por ejemplo), la creación de un índice de democracia a partir de artículos periodísticos o el *análisis y medida de ideologías latentes en discursos a nivel de gobiernos, legisladores y países*. Sin embargo, *estos enfoques no consideran realmente el papel de las palabras en el contexto, es decir, realmente estos métodos ignoran la semántica*. Algunas soluciones como [SemScale](#), que se basa en representaciones distribucionales de los documentos bajo estudio, permiten el *escalado (“scaling”) semántico de textos políticos* mediante la *utilización de “word embeddings”*. En lo que respecta al etiquetado POS, lematización y reconocimiento de entidades (NER) esta herramienta utiliza [Spacy 7](#), una biblioteca de Python que ofrece modelos pre-entrenados para diferentes idiomas. Para la vinculación de entidades (“entity linking”) emplean “[DBpedia Spotlight](#)” que ofrece modelos de vinculación de entidades para los idiomas en estudio. Para calcular los “word embeddings” esta solución ha utilizado la herramienta [FastText](#).

Un ejemplo de utilización de *Word embeddings en el análisis político* es el presentado por (Rheault & Cochrane, 2020) donde se puede estimar la ubicación ideológica de los actores bajo estudio a partir de los datos recogidos procedentes de un corpus de documentos del Parlamento y de afiliación política de miembros del parlamento¹. En este trabajo *se mapea actores políticos*

¹ Conjuntos de datos disponibles en <https://github.com/lrheault/partyembed>

y lenguaje en un espacio vectorial común para, por ejemplo, evaluar la deriva ideológica de diversos partidos políticos a lo largo del tiempo. Estos autores, utilizan información adicional (metadatos procedentes del corpus parlamentario) para aumentar (o alimentar diríamos también) la red neuronal que genera los “word embeddings”.

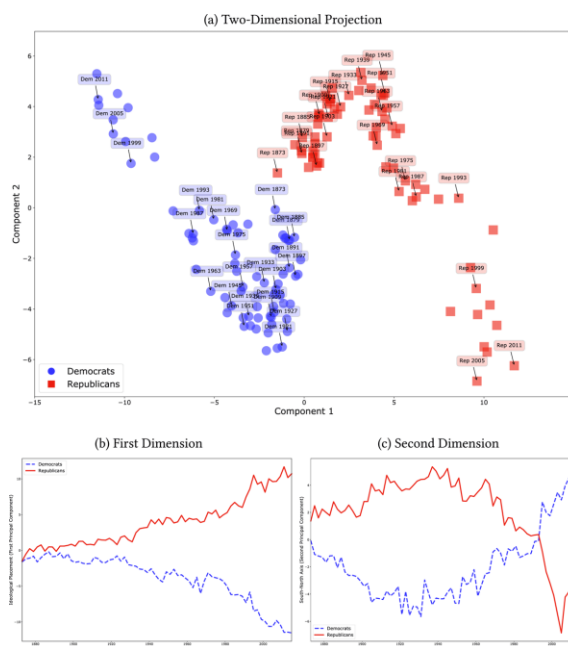


Ilustración 38 Utilización de Word embeddings en el análisis político

Para estos autores, los word embeddings son una herramienta prometedora en lo que a investigación política basada en datos textuales se refiere y su metodología permite la ubicación de actores políticos en un espacio multidimensional. Sin embargo **estos autores no definen los ejes del espacio vectorial a priori** sino que entrenan una red neuronal (de la misma forma que Word2Vec) con un corpus de datos de entrada seleccionado con 200 dimensiones² (capa oculta de word embeddings con lo que estas dimensiones no tienen un “significado” inteligible para el investigador político) para a continuación reducir la dimensionalidad a dos dimensiones mediante el [análisis de componentes principales](#) o [PCA](#) para su visualización (según la Ilustración 38) y **para la búsqueda de “significado” de las dos dimensiones a posteriori**. Estas dos dimensiones son las dos dimensiones que maximizan la varianza a través de los vectores de los word embeddings (Rheault & Cochrane, 2020, pág. 13), restringiendo la segunda dimensión a aquella que presenta una covarianza de cero con la primera componente.

² Tal como se presenta por los autores en [GitHub](#): `model0 = Doc2Vec(vector_size=200, window=20, min_count=50, workers=8, epochs=5)`

4.4 Psicografía

La psicografía (*psychographics*) es una **metodología cualitativa utilizada para encontrar rasgos comunes en grupos de usuarios** que permitan efectuar una segmentación del mercado sobre la base de algunos de sus aspectos psicológicos. La psicografía se ha aplicado al estudio de la personalidad, los valores, las opiniones, las actitudes, los intereses y los estilos de vida y nos permite clasificar a las personas según sus actitudes, aspiraciones y otros criterios psicológicos.

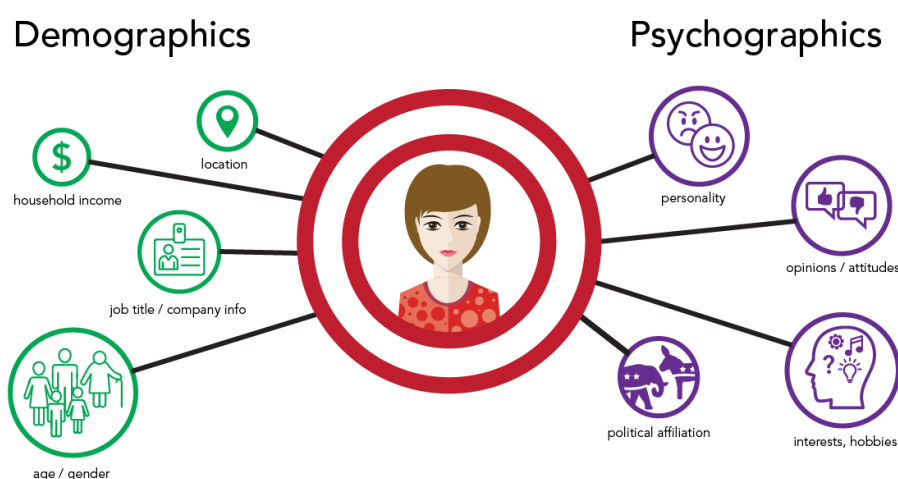


Ilustración 39 Demografía vs Psicografía. Fuente: [Branding Compass](#)

Algunas herramientas como [IBM Watson](#) hacen uso de técnicas NLP de una forma que nos permite deducir ciertos rasgos de personalidad a través de las comunicaciones a través de redes sociales de personas y/o grupos de personas (IBM, 2021), utilizando la analítica lingüística para “deducir los rasgos de personalidad de una persona (como los cinco grandes rasgos, las necesidades y los valores) de las comunicaciones digitales, como correo electrónico, publicaciones de blog, tweets y publicaciones de foro”. El servicio “[Personality Insights](#)” se basa en la psicología del lenguaje, junto con algoritmos de análisis de datos, analizando el contenido que se envía y devolviendo un perfil de personalidad para el autor de la entrada. Este servicio deduce las características de personalidad basándose en tres modelos:

1. Las **cinco grandes características de personalidad** (“[five personality traits](#)”) representan el modelo más utilizado para describir de forma general cómo interactúa una persona con el mundo. El modelo incluye cinco dimensiones primarias y cada una de estas dimensiones de nivel superior tiene seis facetas que caracterizan más a un individuo según la dimensión:

- a. Simpatía es la tendencia de una persona a ser compasiva y cooperadora con los demás.
 - b. Responsabilidad es la tendencia de una persona a actuar de forma organizada o meticulosa.
 - c. Extraversión es la tendencia de una persona a buscar la estimulación en compañía de los demás.
 - d. Rango emocional, también conocido como Neurosis, neuroticismo o reacciones naturales, es la magnitud a la que son sensibles las emociones de una persona a su entorno.
 - e. Apertura es el punto hasta el que una persona está abierta a experimentar distintas actividades.
2. **“Necesidades”** que describe los aspectos de un producto con los que se identifica una persona. El modelo incluye doce valores de necesidades.
 3. **“Valores”** que describe los factores de motivación que influyen en la toma de decisiones de una persona. El modelo incluye cinco valores.

El modelo de los cinco grandes (“Big Five Model”), también conocido como el modelo de los cinco factores (“Five-Factor Model”), es la teoría de la personalidad más aceptada por los psicólogos en la actualidad. La teoría establece que la personalidad se puede reducir a cinco factores centrales, conocidos por el acrónimo CANOE u OCEAN. Estos cinco factores son:

1. **Apertura a la experiencia** (inventivo / curioso vs. consistente / cauteloso) O
2. **Conciencia** (eficiente / organizado vs. extravagante / descuidado) C
3. **Extraversión** (sociable / enérgico vs. solitario / reservado) E
4. **Amabilidad** (amigable / compasivo vs. desafiante / insensible) A
5. **Neuroticismo** (susceptible / nervioso vs. resistente / seguro) N

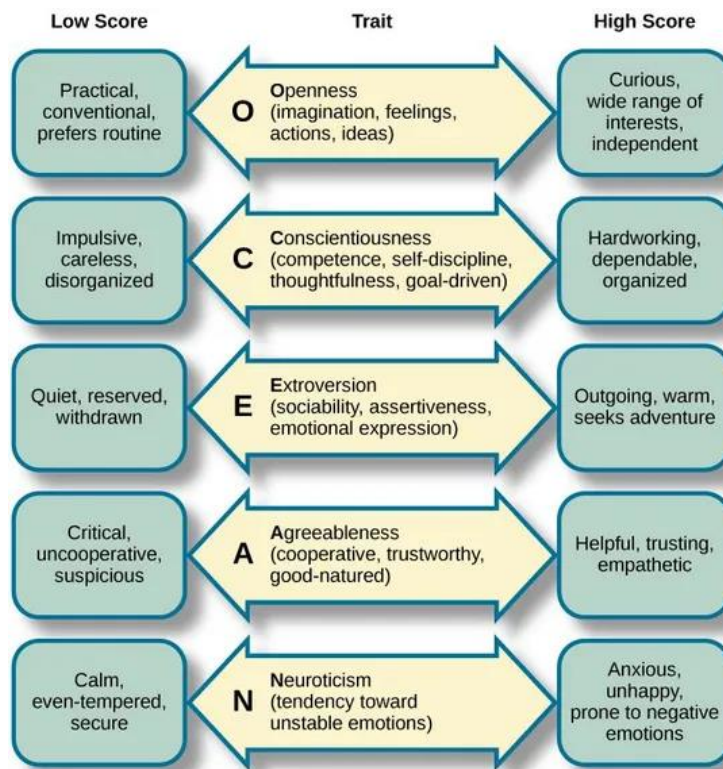


Figure 3 Modelo de los cinco factores de la personalidad humana

Otras organizaciones se han centrado en el estudio de las emociones y su medición, como por ejemplo “[The Junto Institute](#)” han elaborado un “*mapa*” para identificar las principales *emociones humanas*. Creemos que al igual que algunos investigadores son capaces de reconocer, etiquetas con palabras específicas, y clasificar las emociones de las personas a partir de respuestas a cuestionarios, entrevistas personales, o las respuestas emocionales que videos (Cowen & Keltner, 2017) suscitan en las personas, nosotros podemos seguir un método similar utilizando las herramientas que PLN ofrece para clasificar otros aspectos de la personalidad humana (como por ejemplo, las tendencias políticas de una persona, grupo o texto).

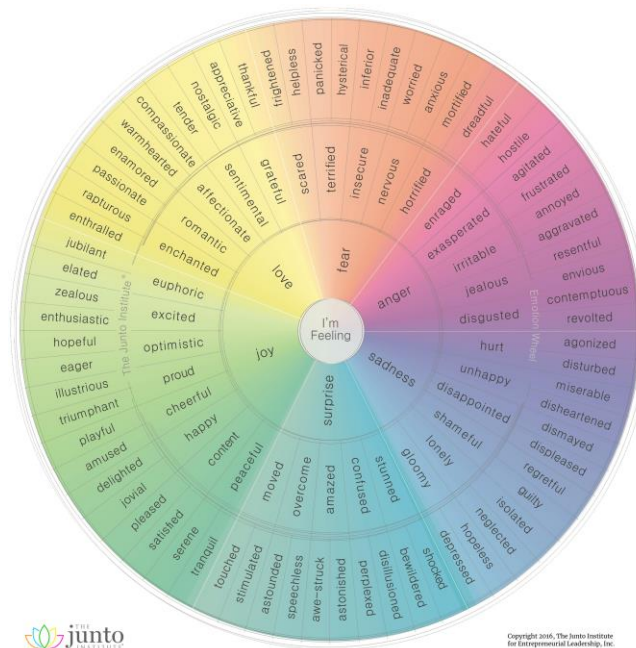


Ilustración 40 Ruleta de las emociones

Otras empresas como [Symanto](#), ofrecen soluciones comerciales de inteligencia artificial que **combinan tecnologías de procesamiento de texto y algoritmos psicolingüísticos** lo que permite predecir los **rasgos de la personalidad** (emocional o racional) y el **estilo de comunicación** de dicha persona (orientado a los hechos, compartidor de la propia opinión, en búsqueda de información, buscando la acción, orientado a la relación).



Figure 4 Rasgos de la personalidad para la empresa Symanto

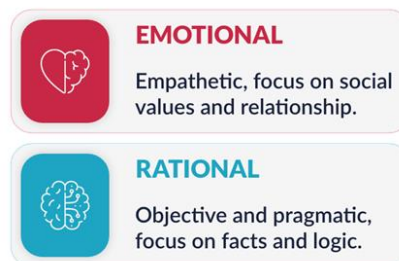


Figure 5 Estilo de comunicación personal para la empresa Symanto

La plataforma de [Symanto](#) (“[Symanto Insights Platform](#)”) es una herramienta de análisis que utiliza IA y que brinda información cualitativa sobre datos textuales. Mediante la carga de los conjuntos de datos disponibles o rastreando información de diferentes fuentes (como redes sociales, sitios de revisión de opiniones en internet o encuestas online), *los sistemas de Symanto analizan los datos y proporcionan información cualitativa sobre el autor*. La plataforma analiza un texto desde dos perspectivas, rasgos de personalidad y estilo de comunicación.

- Predecir los rasgos de personalidad para comprender cómo el autor de un texto escrito toma decisiones, ya sea de una forma *emocional* (orientadas a las relaciones, enfocándose en los valores sociales y la empatía) o de forma *racional* (objetivo y pragmático, centrándose en los hechos y la deducción lógica).
- Detectar el propósito de la comunicación y el estilo del texto. El estilo puede ser: *autorrevelación* (compartir la propia experiencia y opinión), *orientado a hechos* (enfocándose en información objetiva, observaciones o declaraciones objetivas), *búsqueda de información* (planteando preguntas) o de *búsqueda de acción* (con el objetivo de desencadenar la acción de alguien al dar recomendaciones, solicitudes o consejos).

Las principales características del sistema son:

- **Análisis de emociones:** Detecta la emoción expresada en el texto: enfado, disgusto, miedo, alegría, no emoción, tristeza, sorpresa.
- **Análisis de sentimientos:** detecta el sentimiento de un texto o de un fragmento de texto: positivo, negativo.
- **Extracción de tópicos:** detección de temas, categoría de cada tema y análisis de sentimiento hacia cada uno de los temas detectados.
- **Análisis de recomendación de marca:** Detecta si el texto es promotor, detractor o indiferente en diferentes áreas de negocio (automóvil, Banca, Electrónica de Consumo, Empleado, Hotel, farmacéutico, Restaurante, etc.).
- **Detección de idioma:** identifique el idioma del texto de entrada.



Figure 6 Análisis de emociones para la empresa Symanto

La segmentación psicográfica es un enfoque utilizado en la investigación de mercado para agrupar a clientes de acuerdo con las características únicas de su personalidad. Al utilizar la psicografía como base para la segmentación los especialistas en marketing pueden comprender el razonamiento detrás del comportamiento del consumidor y reconocer mejor lo que impulsa inconscientemente a los consumidores a comprar un producto

Symanto utiliza sistemas de inteligencia artificial para **identificar automáticamente los rasgos de personalidad y los estilos de comunicación de los clientes a partir del texto que producen en diferentes canales**, proporcionando una visión profunda de la motivación detrás del comportamiento del consumidor. Un aspecto interesante es que los rasgos de personalidad de una persona se mantienen relativamente estables a lo largo del tiempo, siendo esta una ventaja del uso de la psicografía como base para la segmentación de usuarios.

4.5 Vectores del Pensamiento

Los vectores del pensamiento (“*Thought Vectors*”) son un concepto presentado por [Geoffrey Hinton](#) (investigador en aprendizaje profundo en Google que popularizó la propagación hacia atrás de errores o *backpropagation* en redes neuronales), en su discurso en la Real Sociedad de Londres en 2015. Un “Vector de Pensamiento” es un concepto similar al de “Vector Palabra” (*Word Vector*) utilizado en técnicas como Word2Vec propias de NLP. Así como un “Vector Palabra” es un vector de entre 300 y 500 dimensiones que representa cada palabra de un vocabulario dado, **un “Vector de Pensamiento” representa el significado de una palabra dentro de su contexto (es decir, cuando se relaciona con otras palabras y/o pensamientos) mediante un vector** (una columna de números).

Un “Vector de Pensamiento” es un “pensamiento vectorizado” permitiéndonos relacionar unos pensamientos con otros al igual que en word2Vec se puede relacionar una palabra con otras. Así como las palabras están relacionadas (o quizás están condicionadas) por la gramática de un idioma, también podemos intuir que los pensamientos están relacionados con (o quizás también condicionados por) una cadena de razonamientos (Nicholson, 2020). Y de forma similar a Word2Vec, a partir de un “pensamiento”, (y en un futuro que no es cercano a la década de 2020 en nuestra opinión) una red neuronal (u otro sistema de aprendizaje automático) entrenada con los conjuntos de datos (“pensamientos”) adecuados podría predecir los pensamientos que probablemente aparezcan después. Esto podría definirse en cierto sentido como la construcción artificial del sentido común humano.

Un vector de pensamiento, al ser independiente del idioma utilizado, puede ser utilizado por ejemplo para representar oraciones similares en diferentes idiomas, lo que tendría aplicación en sistemas de traducción automática, por ejemplo. Sin embargo, este campo de conocimiento no ha presentado avances notables en la vectorización de pensamientos desde 2015 que nosotros conozcamos ni hay disponible ninguna implementación en abierto de esta teoría.

4.6 Teoría de Marcos

Las personas adquieren y construyen a lo largo de sus vidas una serie de filtros o marcos mentales emocionales que utilizan para darle sentido al mundo y que están fuertemente influenciados por la cultura y la sociedad en la que esas personas se desarrollan. Los marcos son los esquemas de interpretación en los que se basan los individuos para entender y responder a la información y estímulos que reciben en el día a día.

Para George Lakoff (2004, pág. 4) **los marcos son estructuras mentales que confirman nuestro modo de ver el mundo** y como consecuencia de ello, conforman las metas que nos proponemos, los planes que hacemos y hasta nuestra manera de actuar. En política nuestros marcos conforman las políticas sociales y las instituciones creadas para llevar a cabo dichas políticas. Es interesante la relación que dicho autor encuentra entre los marcos y el lenguaje (y que como posteriormente veremos, será una parte importante en el desarrollo de este trabajo). Para Lakoff, **las palabras activan marcos y conocemos los marcos a través del lenguaje**. Así los marcos formarían parte del “inconsciente cognitivo”, es decir, estructuras en nuestro cerebro a las que no se puede acceder conscientemente pero que conocemos a través de nuestra forma de razonar y se denomina sentido común.

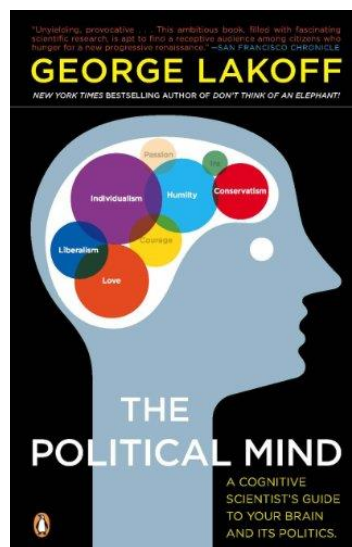


Ilustración 41 La mente política. Fuente: [Amazon](#)

En anexos (Anexo I, la teoría de Marcos (Framing)), se presenta una descripción más detallada del enmarcado, así como de la metodología para encontrar “marcos” en textos utilizada en (Alvarez Prieto, 2018) y otros trabajos, no siendo necesaria su lectura aunque puede servir al lector para adentrarse en este campo.

5. Desarrollo del trabajo

El desarrollo de este trabajo se ha realizado a partir de los conocimientos, herramientas y técnicas presentadas a lo largo del curso 2019-2020 en el “Máster Universitario en Tecnologías del Lenguaje” de la UNED. El trabajo se compone de una parte teórica donde hemos definido las técnicas y herramientas a utilizar, y una segunda parte práctica centrada en el análisis del sentimiento político en España durante el año 2021. Nuestra solución combina las dos grandes líneas mencionadas en el apartado 4.1 (aproximaciones semánticas y aproximaciones basadas en técnicas de aprendizaje automático) y se divide en:

- **Fase I de análisis semántico y adquisición de marcos cognitivos.** A diferencia de los sistemas existentes dentro del área del análisis de sentimiento basados en el uso de diccionarios de términos combinados con una orientación semántica de polaridad de lexicones, los nuestros serán diccionarios de “marcos cognitivos” donde se incorpora para cada marco su orientación de polaridad en cada una de las dimensiones del espacio.
- **Fase II de uso de técnicas de aprendizaje automático.** Esta fase está basada en el uso de *técnicas para la búsqueda automatizada de marcos en textos* y de algoritmos de aprendizaje supervisado basadas en **Transformers (BERT)** para crear un clasificador entrenado con marcos utilizados por una entidad política. En nuestro caso solo nos centraremos en marcos de un ámbito (comunicación política) en un área geográfica (España) y temporal muy restringido.

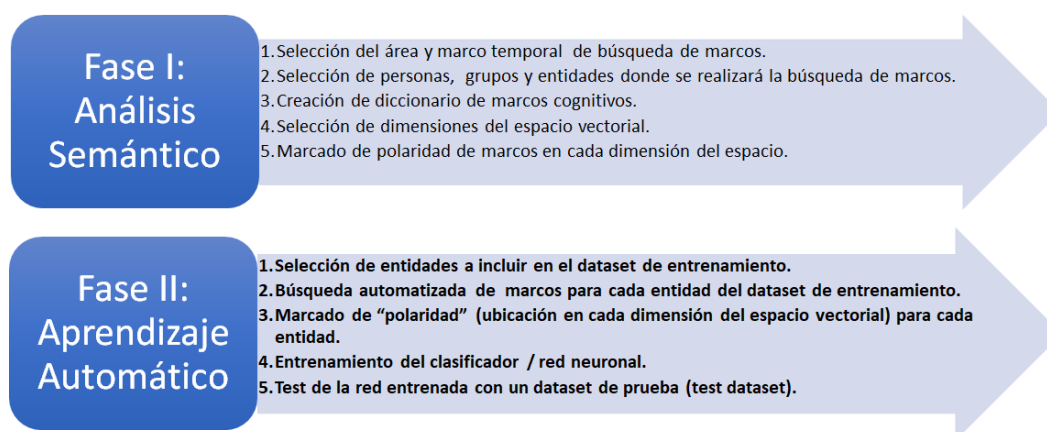


Ilustración 42 Fases de la solución de análisis de sentimiento político (frame embedding)

En una primera fase de este trabajo nos centraremos en la recopilación de recursos de enmarcado (metáforas, eslóganes y narrativas principalmente) utilizados en la comunicación

política en España. Así pues, lo que pretendemos es una ubicación de textos políticos (y a partir de ellos la ubicación de entidades políticas o usuarios a partir de los mensajes compartidos por ellos en las redes sociales) en un espacio vectorial multidimensional definido no por palabras, términos o contenido semántico de los mismos, sino definido por “marcos cognitivos” (a su vez definidos por los elementos que definen dichos marcos como por ejemplo metáforas).



Ilustración 43 Fases I del sistema a desarrollar y ejemplo de lenguaje metafórico

Si tuviéramos que realizar una comparación con un sistema como [Word2Vec](#), nuestro objetivo será tomar como entrada un corpus de textos (como el utilizado por Word2Vec) sobre el que buscaremos los marcos cognitivos relativos al ámbito político en España para generar el **corpus de marcos** de entrada. A continuación **definiremos y generaremos un espacio vectorial multidimensional** (inicialmente en el rango de decenas de dimensiones pero que en un escenario comercial puede ser de varios cientos de dimensiones) *donde cada “marco” único en el corpus tiene asignado un vector en el espacio vectorial*. A diferencia de Word2Vec donde las dimensiones del espacio no tienen ningún significado directo (sino que aparecen como resultado de ejecutar el algoritmo que toma en cuenta las coocurrencias de las palabras dentro de una ventana de texto), *las dimensiones de nuestro espacio son definidas a partir de técnicas como las definidas en los apartados 4.3 (Procesamiento de lenguaje natural y análisis político), 4.4 (técnicas psicográficas) y 4.5 (vectores de pensamiento) y en particular para utilizar técnicas similares a la de “escalado de textos” (text scaling)*.

Los vectores representando las opiniones políticas de una entidad se generan añadiendo la polaridad de cada marco utilizado por esa entidad respecto a cada dimensión del espacio vectorial (por ejemplo si se trata de un marco utilizado por una persona religiosa estaría en el área positiva de este eje o una antirreligiosa y en este caso se ubicaría en el área negativa de este eje) y se posicionan en el espacio vectorial de forma que marcos que comparten un

“contexto” común están ubicados cercanos unos a otros en dicho espacio. *Los marcos se “anotan” por anotadores humanos una única vez al igual que se haría en la creación de un diccionario en un sistema de análisis de sentimientos estándar o en un escalado de textos* como el definido en el apartado 4.3.2. Usuarios que comparten marcos comunes estarán ubicados en áreas cercanas del espacio vectorial (con lo que será posible utilizar de similitud).

Las tareas que desarrollar en las dos fases de nuestra metodología serán las siguientes:



Ilustración 44 Metodología seguida en el proyecto

1. Fase I: Análisis Semántico:

1. **Selección de datos (corpus de entrada).** Nos centraremos en la adquisición de datos presentes en documentos y mensajes de Twitter relativos a comunicación política en España. El resultado de esta fase es el corpus de tuits de entrada al sistema.
2. **Etiquetado de textos de entrada para anotar los marcos** contenidos en dichos textos. El resultado de esta fase será el *corpus de Marcos* (frames) de entrada.
3. **Estudio y diseño (selección de las dimensiones) del espacio vectorial**, así como el diseño de las transformaciones necesarias en las dimensiones del espacio vectorial si fueran necesarias. El resultado de esta fase es el diseño del espacio vectorial propiamente dicho (principalmente el número de dimensiones a utilizar y sus características).
4. **Marcado de polaridad** del corpus de marcos en cada dimensión del espacio. El resultado de esta fase será el *diccionario de Marcos* (frames) donde cada marco se puede ubicar a lo largo de los ejes o dimensiones del espacio definido anteriormente.

2. **Fase II: Operaciones de Aprendizaje Automático (MLOps).** Fase de diseño y entrenamiento del clasificador/red neuronal. Para ello utilizaremos el corpus de entrada anotado como conjunto de datos de entrenamiento y validación, es decir, ubicaremos

los **documentos, partidos políticos o personas en el espacio vectorial** a partir de los marcos utilizados por ellos en Twitter.

1. **Selección de entidades** (personas, partidos políticos, medios de comunicación, etc.) a incluir en el dataset de entrenamiento (pudiendo ser las mismas utilizadas en la Fase I o un número mayor).
2. **Ubicación en el espacio vectorial de cada entidad** mediante la ubicación (en cada dimensión del espacio vectorial) de los marcos utilizados por esa entidad junto con su “polaridad”.
3. **Búsqueda automatizada de marcos** para cada entidad del dataset de entrenamiento en el conjunto de textos de entrada. Esta búsqueda se realizará para un conjunto reducido de marcos y de ejes del espacio, quedando fuera del alcance de este TFM una búsqueda automatizada de potenciales cientos de marcos a lo largo de los 10 ejes del espacio vectorial.
4. **Entrenamiento del [clasificador multiclase](#) / [red neuronal](#)**. Para ello haremos una partición del dataset original para tener un dataset de entrenamiento (training dataset) y otro de validación (validation, development o devset según la literatura).
5. **Evaluación de la red con un dataset** de usuarios con sus marcos correspondientes.
 - i. **Clasificación y ubicación de entidades en el espacio vectorial.** Mediante la utilización de los tuits de una nueva entidad, persona o partido político como entrada a la red neuronal podremos predecir los marcos (estimados como los más cercanos) que esa entidad genera hacia los receptores, y por lo tanto ubicar a la entidad política en el espacio vectorial a partir de las ideas políticas vertidas en Twitter (clasificación de nuevos usuarios de un conjunto de datos de prueba o “test dataset”).
 - ii. **Un objetivo secundario relacionado con el anterior sería a partir de un marco o de los recursos de lenguaje metafórico empleados definir la probabilidad de que existan otros marcos o recursos de lenguaje en un texto.** Esto sería parecido a Word2Vec donde a partir de una palabra podemos estimar las palabras más cercanas a ella.
6. **Evaluación de la solución** (véase apartado 5.3, Resultados).

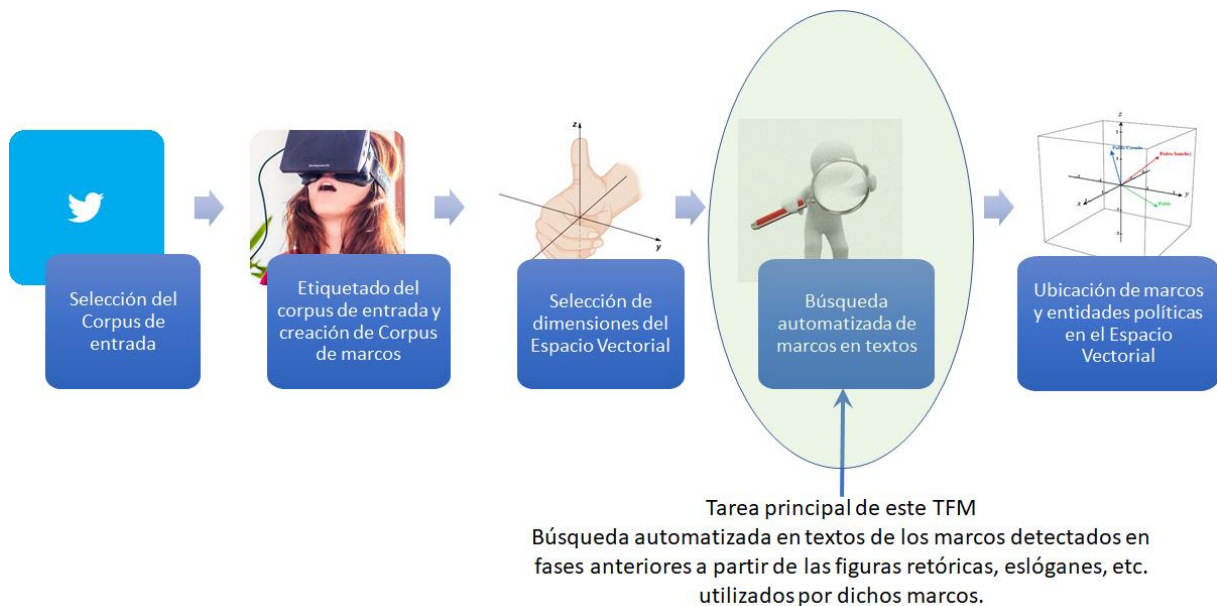


Ilustración 45 Búsqueda automatizada de marcos en textos

Una persona puede estar definida por infinitas dimensiones (culturales, psicológicas, o como se quiera definir una personalidad o mente humana), pero en nuestro caso identificaremos unas pocas dimensiones existentes en la sociedad española en el ámbito político (en el rango de las decenas) y para esas dimensiones (culturales y políticas) identificaremos algunos marcos cognitivos que ubican a una persona en esas dimensiones (y por lo tanto en el espacio vectorial).

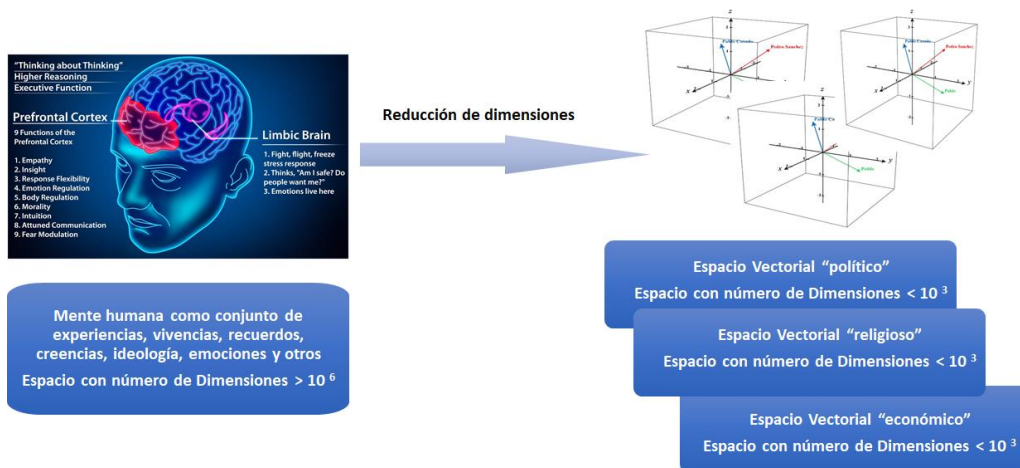
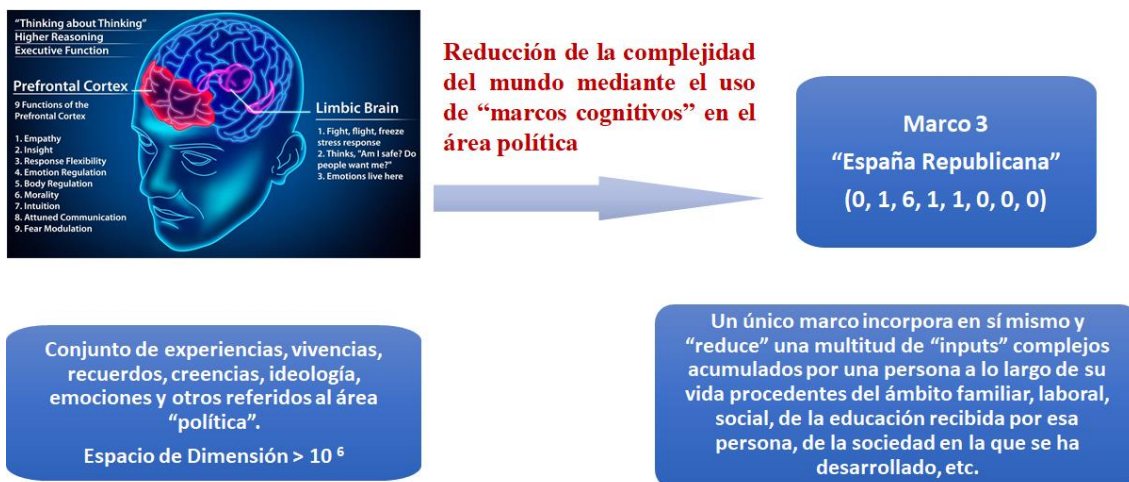


Ilustración 46 De una mente compleja a un espacio vectorial reducido

En cuanto al **modelo del espacio vectorial político**, algunos autores como (Egerod, B. C. K. & Klemmensen, R., 2020) admiten que necesitamos hacer suposiciones sobre cómo un autor (entidad que expresa una opinión política) traduce su posición política en un texto, y sobre cómo se relaciona con los otros autores en el corpus. En particular *el lenguaje utilizado en los textos debe discriminar entre los mensajes políticos de diferentes autores*, es decir, cada autor debe

recibir diferentes niveles de utilidad de su elección de palabras, y esta variación debe estar relacionada con el espacio político que queremos medir (y representar mediante un espacio vectorial). Si autores con diferentes preferencias políticas reciben la misma utilidad de elecciones similares de palabras no podemos utilizar los textos para discriminar entre sus posiciones. Por lo tanto, *los documentos utilizados como corpus de entrada deben informar acerca de las diferencias políticas que buscamos estimar*. En cuanto a la relación entre documentos, un conjunto de documento es “escalable”, si pueden ubicarse sobre el mismo espacio vectorial euclídeo.

Podríamos proponer como teoría que *“la utilización de marcos cognitivos es una transformación que nos permite destacar las diferencias políticas entre entidades y a la vez disminuir las dimensiones del espacio (vectorial) que definen una mente humana”*, al igual que los “Word Embeddings” nos permiten “insertar” un vocabulario completo dentro de un espacio lineal de baja dimensionalidad donde las dimensiones son las principales características o rasgos (“features”) latentes. Es decir, vamos a utilizar en este trabajo la tesis de que el espacio lineal definido son las principales características o rasgos (“features”) de la personalidad humana (algo similar a la teoría que subyace en los estudios que realiza la psicografía) y los “marcos” que una persona utiliza pueden ubicarse a lo largo de dichas dimensiones. Dichos marcos son el resultado de las interacciones personales, sociales, físicas y psicológicas que una persona adquiere a lo largo de su vida y que forman y conforman su mente, y por lo tanto son también una forma de reducir la complejidad de la mente humana para su estudio. Es decir, *los marcos cognitivos “comprimen”, reducen la complejidad y contienen en sí mismos multitud de experiencias y vivencias que una persona ha acumulado a lo largo de su vida referidas a un tema de interés* (tanto para el individuo como para la sociedad y que en el caso de este trabajo es el ámbito político).



Teóricamente podríamos intentar identificar las dimensiones que “definen” una personalidad, o “mente” humana (algo que el campo de la psicografía mencionado en el apartado 4.4 desarrolla a un nivel básico y centrado solo en algunos aspectos de la personalidad humana) y en este caso deberíamos utilizar técnicas para reducir las dimensiones de este espacio. ***Este es un estudio fuera del alcance de este proyecto y más propio de un Doctorado a desarrollar durante un periodo de tiempo más extenso.*** En nuestro caso diseñaremos y utilizaremos un espacio vectorial reducido para ubicar a las organizaciones bajo estudio.

5.1 Fase I: Análisis Semántico

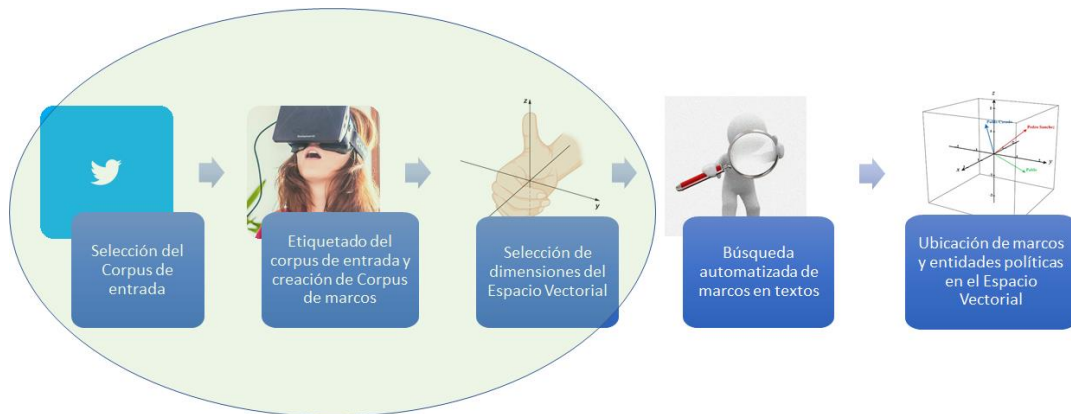


Ilustración 48 Tareas de la Fase I de análisis

5.1.1 Selección de datos y Corpus utilizado

Estando este trabajo centrado en el estudio de la comunicación política a través de las redes sociales, una parte importante de los datos serán recogidos de dichas redes sociales, en particular de **Twitter**. El **corpus de datos de entrada** es un conjunto de textos como programas electorales, entrevistas a dirigentes de formaciones políticas, opiniones vertidas en las redes sociales por los mismos, declaraciones ideológicas, debates parlamentarios (Rheault & Cochrane, 2020), etc., que **se utilizará para generar un diccionario de marcos (con su polaridad política) a ubicar en un espacio vectorial multidimensional** (inicialmente en el rango de decenas de dimensiones pero que en un escenario comercial podrá ser de varios cientos de dimensiones) donde cada “marco” único en el corpus tiene asignado un vector en el espacio vectorial.

Algunos autores en el ámbito de la detección de ideologías políticas en textos han desarrollado *conjuntos de datos (datasets) de ideología política anotados en el nivel de frase* (Iyyer, Enns, Boyd-Graber, & Resnik, 2014, pág. 1114) para mostrar como las redes neuronales recurrentes (RNN) no solo son capaces de etiquetar correctamente oraciones, sino que también mejoran los resultados al incluirse anotaciones adicionales. Estos autores argumentan que *la utilización de modelos basados en la utilización de estadísticas distribucionales a nivel de palabra no es deseable para la detección de ideologías políticas en textos* (Iyyer, Enns, Boyd-Graber, & Resnik, 2014, pág. 1114).

En este proyecto se ha utilizado el *corpus de marcos* definido en (Alvarez Prieto, 2018) con un número muy reducido de marcos políticos utilizados por una formación política para un aspecto

concreto del debate político en España (véase apartado 5.2.2, Búsqueda automatizada de marcos). El **corpus de datos utilizado** en dicha búsqueda de marcos se describe en el apartado 9.7 (Recogida de información y selección de herramientas) de este documento.

5.1.2 Fase de enmarcado

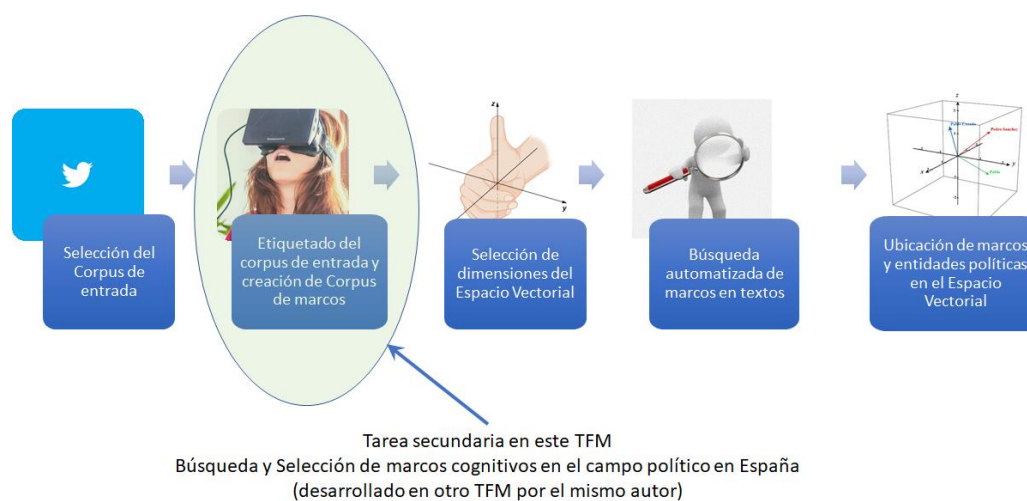


Ilustración 49 Tarea de etiquetado y creación del corpus de marcos cognitivos

En el desarrollo de este trabajo utilizaremos la teoría del enmarcado (Framing) y en particular la metodología de búsqueda de marcos presentada en el Anexo I, la teoría de Marcos (Framing) Dentro de la [teoría del enmarcado \(Framing\)](#) el **lenguaje metafórico (metáforas, similitudes, analogías, eslóganes, etc.)** tiene una especial importancia. Por ejemplo, las metáforas utilizando objetos utilizados por las personas en el día a día **nos ayudan a simplificar y entender aspectos complejos de la realidad, provocando asociaciones con experiencias pasadas o invocando “mitos” sociales y culturales** presentes en un público objetivo. Este público objetivo tiene un determinado contexto cultural que facilita el desarrollo y la transmisión de estas asociaciones, es decir, dichos marcos culturales “resuenan” con mayor fuerza en estas personas. Un ejemplo de metáforas e imágenes (recursos de enmarcado, véase el apartado 9.2) utilizadas por el independentismo (en el marco que podríamos denominar “Independencia”) sería el presentado por (Martínez Fernández, 2016):

Por ejemplo, “desconexión”, “construcción” o “encaje” son algunas de las metáforas a las que, en consonancia con el esquema cognitivo que propone George Lakoff en “No pienses en un elefante”, el independentismo ha dado vida. Siguiendo dicho esquema y partiendo de las premisas expuestas por Lakoff, los denominados “marcos de referencia”, es decir las metáforas que subyacen a nuestros procesos de entendimiento, son los que determinan el hecho de posicionarse a favor o en contra de la independencia de Catalunya.



Ilustración 50 Ejemplo de metáforas utilizadas por el independentismo catalán



Analogía con la [Guerra de Sucesión Española](#) y la batalla y toma de Barcelona del 11 de Septiembre de 1714.



Similitud con la adhesión de nuevos estados a la Unión Europea ("Cataluña, un nuevo estado de Europa").

Un ejemplo de marcos culturales utilizados en la comunicación política en España es el presentado en (Alvarez Prieto, 2018) donde se muestran los *marcos políticos utilizados por la formación política PODEMOS y sus dirigentes y que son los utilizados en este trabajo*:

Marcos utilizados por PODEMOS
Élites culpables de la crisis
El Pueblo como los de "abajo"
Final del Estado del 78
España Plurinacional

Ilustración 51 Ejemplo de marcos utilizados por PODEMOS. Fuente (Alvarez Prieto, 2018)

Por lo tanto, en este proyecto se ha utilizado una versión del *corpus de marcos* definido en (Alvarez Prieto, 2018) con un número muy reducido de marcos políticos utilizados por un par de formaciones políticas (PODEMOS y VOX) y únicamente para un aspecto concreto del debate político en España (véase apartado 5.2.2, Búsqueda automatizada de marcos). Este diccionario, en una fase comercial o de investigación más profunda, requeriría la recopilación y etiquetado de cientos o miles de marcos por parte de codificadores humanos expertos dependiendo del elemento cultural bajo estudio o del alcance perseguido por el sistema. Por ejemplo, *para el estudio de la comunicación política en España estimamos que deberían incluirse en dicho diccionario no menos de 10 marcos cognitivos por cada dimensión del espacio vectorial que se haya diseñado para ubicar "ideas" o marcos políticos* (que en nuestro

caso ha sido diseñado con 10 dimensiones). Lo mismo aplicaría a cualquier otra área de conocimiento (pensamiento económico, psicografía, etc.) que se quisiera estudiar mediante la utilización de marcos cognitivos.

Sería de interés disponer de conjuntos de datos de entrenamiento creados para este fin, como el [dataset Treebank de Stanford para el análisis de sentimientos](#)) con más de 10.000 oraciones procedentes de revisiones de películas donde los sentimientos se clasifican entre 1 y 25, donde uno es el más negativo y 25 es el más positivo), [IMDB](#) u [otros](#). Esta es una tarea que excede el objetivo de este trabajo con lo que como veremos en el apartado 5.1.4 (Marcado de polaridad de marcos), en este trabajo nos centraremos en un diccionario muy reducido resultado del trabajo presentado en (Alvarez Prieto, 2018).

Queremos incidir que en este trabajo no estamos trabajando en la línea de trabajos anteriores en los que el marcado de la polaridad política (a lo largo de un eje principalmente) se realiza basándose en la utilización de n-gramas que se asignan a cada tipo de visión política (en un eje), sino que **buscaremos estructuras de nivel superior (marcos en nuestro caso)**. Otra diferencia adicional es que **no asociamos marcos a una ideología o concepto general** (por ejemplo el independentismo) sino que **los marcos están centrados en un fenómeno con un entorno bajo estudio más concreto y reducido** (por ejemplo el independentismo, en el ámbito geográfico de Cataluña, en las primeras décadas del siglo XXI). Esto nos permite ajustar y realizar un ajuste fino (“fine-tuning”) de los conceptos, frases, y elementos del lenguaje figurado a buscar en los textos para asignarlos a cada marco.

n	Most conservative n-grams	Most liberal n-grams
1	Salt, Mexico, housework, speculated, consensus, lawyer, pharmaceuticals, ruthless, deadly, Clinton, redistribution	rich, antipsychotic, malaria, biodiversity, richest, gene, pesticides, desertification, Net, wealthiest, labor, fertilizer, nuclear, HIV
3	prize individual liberty, original liberal idiots, stock market crash, God gives freedom, federal government interference, federal oppression nullification, respect individual liberty, Tea Party patriots, radical Sunni Islamists, Obama stimulus programs	rich and poor, “corporate greed”, super rich pay, carrying the rich, corporate interest groups, young women workers, the very rich, for the rich, by the rich, soaking the rich, getting rich often, great and rich, the working poor, corporate income tax, the poor migrants
5	spending on popular government programs, bailouts and unfunded government promises, North America from external threats, government regulations place on businesses, strong Church of Christ convictions, radical Islamism and other threats	the rich are really rich, effective forms of worker participation, the pensions of the poor, tax cuts for the rich, the ecological services of biodiversity, poor children and pregnant women, vacation time for overtime pay
7	government intervention helped make the Depression Great, by God in His image and likeness, producing wealth instead of stunting capital creation, the traditional American values of limited government, trillions of dollars to overseas oil producers, its troubled assets to federal sugar daddies, Obama and his party as racialist fanatics	African Americans and other disproportionately poor groups; the growing gap between rich and poor; the Bush tax cuts for the rich; public outrage at corporate and societal greed; sexually transmitted diseases, most notably AIDS; organize unions or fight for better conditions, the biggest hope for health care reform

Ilustración 52 n-gramas utilizados con mayor probabilidad por liberales y conservadores (Iyyer, Enns, Boyd-Graber, & Resnik, 2014, pág. 1121)

5.1.3 Estudio y diseño del espacio vectorial de marcos

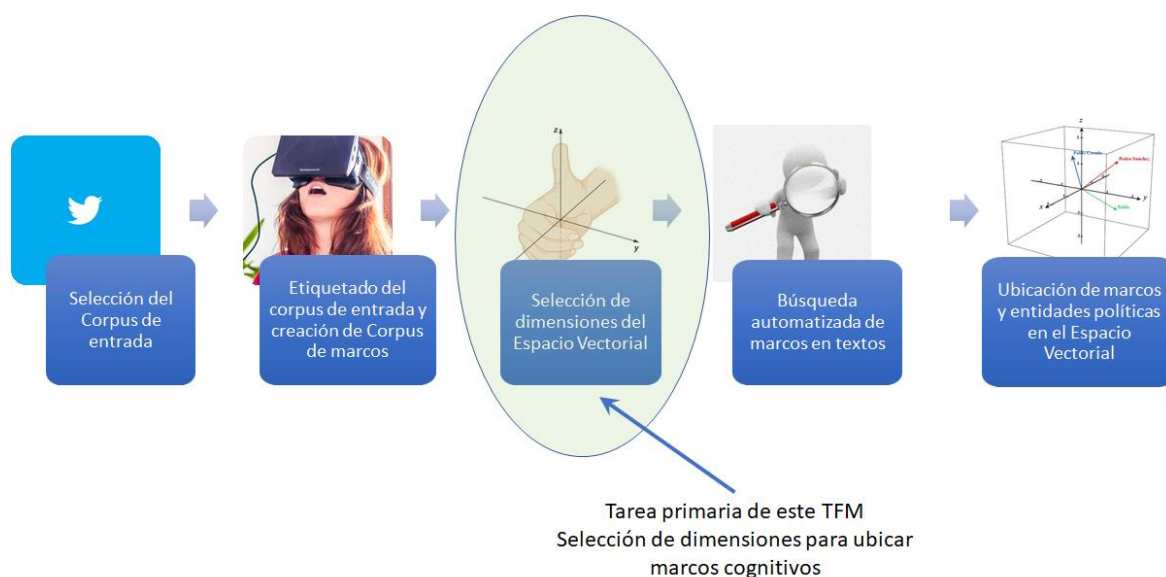


Ilustración 53 Selección de dimensiones del espacio vectorial

Hemos seleccionado en este trabajo varias dimensiones de espacio vectorial (al que denominaremos “**Espacio Vectorial Pragmático**” o EVP) que creemos que son una buena aproximación inicial al estar específicamente diseñadas para ubicar “marcos” (dentro del ámbito político español) y no palabras. Varias de estas dimensiones han sido utilizadas en trabajos anteriores como la “[Brújula Política](#)” (Falck & Marstaller, 2018), el [gráfico de Nolan](#), el [gráfico de Pournelle](#), el [mapa cultural del mundo](#) definido por [Inglehart](#) y Welzel basado en datos recogidos por la “[Encuesta Mundial de Valores](#)”, por Ludovic Rheault y Christopher Cochran (Rheault & Cochran, 2020) y Jonathan B. Slapin y Sven-Oliver Proksch (Slapin & Proksch, 2008) entre otros, como por ejemplo los ejes 1 (Ideas Políticas: izquierda-derecha) y 2 (Libertad Económica: Libertarismo-Estatismo) de la siguiente tabla. Otros autores utilizando [redes neuronales recurrentes \(RNN\)](#) simplemente ubican textos políticos en el eje conservador-liberal (Iyyer, Enns, Boyd-Graber, & Resnik, 2014, pág. 1116) para poder realizar un estudio de posicionamiento político.

Las dimensiones han sido definidas “ex ante” al igual que realiza el algoritmo de WordScores. WordScores calcula las puntuaciones de las palabras (“wordscores”) utilizando solo información de frecuencia de palabras de documentos con posiciones (políticas) conocidas (documentos de “referencia”), sin embargo nosotros (como veremos a continuación) utilizaremos además información acerca del *papel de las palabras en el contexto* mediante el uso de algoritmos de la familia Transformers como BERT.

#	Dimensiones	Ámbito	Extremo (-) del Eje	(-)	Centro (0) del Eje	(+)	Extremo (++) del Eje
1	Ideas políticas (izquierda/derecha)	Político	Izquierda		Centrismo político		Derecha
2	Libertad Económica	Político / Económico	Libertarismo (estado no necesario)		Economía Mixta		Estatismo (estado imprescindible)
3	Organización Política Mundial	Político	Localismo		Estado Unitario		Globalismo
4	Organización Territorial del Estado	Político	Independentismo	Federalismo	Autonomismo		Centralismo
5	Modelo Político del Estado (Forma de Gobierno)	Político	Dictadura	Dictablanda	Monarquía Constitucional	Monarquía Parlamentaria	República
6	Integración supranacional (UE)	Político	Euroescepticismo Fuerte		Euroescepticismo suave		Europeísmo
7	Dimensión sociocultural	Político / Social	Liberalismo cultural		Reformismo		Conservadurismo Social
8	Diversidad cultural	Político / Étnico	Multiculturalismo		Interculturalismo		Nacionalismo étnico
9	Apoyo Progreso Tecnológico	Tecnología	Neoludismo		Tecnorrealismo		Tecnoutopismo
10	Equipo de Fútbol Preferido	Deportivo / Político	FC Barcelona	Lleida	Real Betis y otros	RCD Espanyol	Real Madrid CF

Tabla 2 Dimensiones del Espacio Vectorial Pragmático

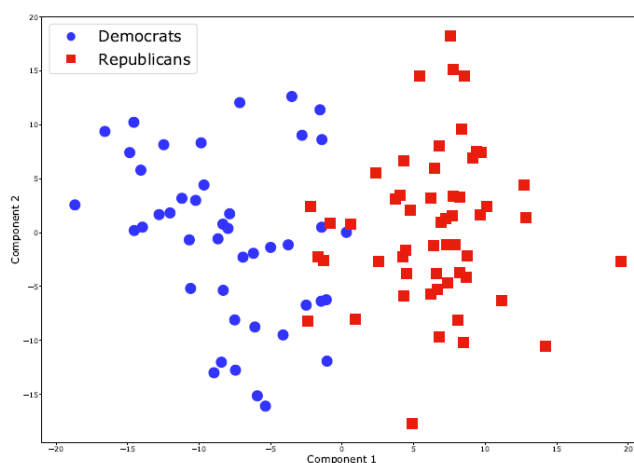


Ilustración 54 Ubicación ideológica de senadores (Congreso 114): Fuente: (Rheault & Cochrane, 2020, pág. 26)

Como presentaremos en apartados posteriores, los dos ejes de este espacio en el que desarrollaremos un sistema de análisis de sentimiento político basado en marcos serán:

- Eje 4 (“Organización Territorial del Estado”).
- Eje 5 (“Modelo Político del Estado” o forma de gobierno).

5.1.4 Marcado de polaridad de marcos

En cuanto a *la ubicación o “polaridad” de los marcos cognitivos a lo largo de los diferentes ejes o dimensiones del espacio vectorial*, vamos a utilizar una metodología similar a la utilizada por algunos autores como (Bruinsma & Gemenis, 2019) para validar el **escalado de textos (text scaling)** técnica mencionada en el apartado 4.3.2. Estos autores utilizan una variedad de fuentes para las puntuaciones de referencia en la escala y para comparar las estimaciones de Wordscores en un ejercicio de validación. Para ello utilizan estimaciones de encuestas de expertos y en lugar de agregar los juicios de múltiples expertos, utilizan a **un solo [codificador](#)** al que se le pide que brinde un juicio informado después de leer un manifiesto, programa de partido político o documento similar. Algunos autores utilizan varios codificadores (dependiendo del volumen de información a etiquetar) restringiendo dichos codificadores a un país o región dada (Iyyer, Enns, Boyd-Graber, & Resnik, 2014, pág. 1117). Estos autores restringen los codificadores a aquellos con una dirección IP válidas dentro de los Estados Unidos de América para etiquetar 3.412 oraciones procedentes del dataset de textos políticos [Ideological Books Corpus](#) (IBC).

La validación en métodos supervisados como WordScores debe implicar demostrar que el método asistido por computadora puede reproducir los resultados en un conjunto de documentos para los que se conocen las puntuaciones reales del documento de interés. ***Dado que no existen puntuaciones verdaderas de las posiciones ideológicas de los partidos (y podríamos decir que como se toma en este trabajo, estas posiciones son relativas)***, estos siguen dos enfoques de validación diferentes. En primer lugar, comparan las estimaciones de Wordscores con los juicios humanos de las posiciones de los partidos y, en segundo lugar, usan las estimaciones de WordScores para predecir otras cantidades conocidas de interés.

En nuestro caso, **hemos realizado el marcado inicial de polaridad para los dos marcos bajo estudio para la formación política “PODEMOS” como se muestra en la siguiente tabla:**

Marco 1	“República”	Eje 5	“Modelo político del Estado”
Marco 2	“Estado Plurinacional”	Eje 4	“Organización Territorial del Estado”

#	Dimensiones	Ámbito	Extremo (--) del Eje	(-)	Centro (0) del Eje	(+)	Extremo (++) del Eje
1	Ideas políticas (izquierda/derecha)	Político	Izquierda		Centrismo político		Derecha
2	Libertad Económica	Político / Económico	Libertarismo (estado no necesario)		Economía Mixta		Estatismo (estado imprescindible)
3	Organización Política Mundial	Político	Localismo		Estado Unitario		Globalismo
4	Organización Territorial del Estado	Político	Independentismo	Federalismo	Autonomismo		Centralismo
5	Modelo Político del Estado (Forma de Gobierno)	Político	Dictadura	Dictablanda	Monarquía Constitucional	Monarquía Parlamentaria	República
6	Integración supranacional (UE)	Político	Euroescepticismo Fuerte		Euroescepticismo suave		Europeísmo
7	Dimensión sociocultural	Político / Social	Liberalismo cultural		Reformismo		Conservadurismo Social
8	Diversidad cultural	Político / Étnico	Multiculturalismo		Interculturalismo		Nacionalismo Étnico
9	Apoyo Progreso Tecnológico	Tecnología	Neoludismo		Tecnorrealismo		Tecnoutopismo
10	Equipo de Fútbol Preferido	Deportivo / Político	FC Barcelona	Lleida	Real Betis y otros	RCD Espanyol	Real Madrid CF
Marco 7	"España Republicana"	Eje 5	"Modelo político del Estado"				
Marco 2	"Estado Plurinacional"	Eje 4	"Organización Territorial del Estado"				

Ilustración 55 Mercado de polaridad en marcos para PODEMOS

Es decir, para el eje de “organización política del estado”, estimamos (a priori y en base a la opinión experta del codificador en base a las guías de codificación acordadas en cada proyecto) que PODEMOS está ubicado en el área negativa (-) del eje 4 (Organización Territorial del Estado), alrededor del concepto “Federalismo” (Coll, 2016), mientras que para el eje 5 (“Modelo político de Estado” o Forma de Gobierno) PODEMOS se sitúa en el extremo más positivo (++) de dicho eje alrededor del concepto “República”. En el caso de VOX, esta formación se ha ubicado en el eje 4 en el área positiva (++) del eje (alrededor del concepto “Centralismo” de un estado uninacional), mientras que en el eje 5 se ha ubicado en el área alrededor del centro del eje (“Monarquía Constitucional”).

#	Dimensiones	Ámbito	Extremo (--) del Eje	(-)	Centro (0) del Eje	(+)	Extremo (++) del Eje
1	Ideas políticas (izquierda/derecha)	Político	Izquierda		Centrismo político		Derecha
2	Libertad Económica	Político / Económico	Libertarismo (estado no necesario)		Economía Mixta		Estatismo (estado imprescindible)
3	Organización Política Mundial	Político	Localismo		Estado Unitario		Globalismo
4	Organización Territorial del Estado	Político	Independentismo	Federalismo	Autonomismo		Centralismo
5	Modelo Político del Estado (Forma de Gobierno)	Político	Dictadura	Dictablanda	Monarquía Constitucional	Monarquía Parlamentaria	República
6	Integración supranacional (UE)	Político	Euroescepticismo Fuerte		Euroescepticismo suave		Europeísmo
7	Dimensión sociocultural	Político / Social	Liberalismo cultural		Reformismo		Conservadurismo Social
8	Diversidad cultural	Político / Étnico	Multiculturalismo		Interculturalismo		Nacionalismo Étnico
9	Apoyo Progreso Tecnológico	Tecnología	Neoludismo		Tecnorrealismo		Tecnoutopismo
10	Equipo de Fútbol Preferido	Deportivo / Político	FC Barcelona	Lleida	Real Betis y otros	RCD Espanyol	Real Madrid CF
Marco 5	"España Monárquica"	Eje 5	"Modelo político del Estado"				
Marco 6	"Estado Uninacional"	Eje 4	"Organización Territorial del Estado"				

Ilustración 56 Mercado de polaridad en marcos para VOX

No obstante, esta “polaridad” o ubicación a lo largo de dichos ejes tiene importancia principalmente para *medir de forma relativa a lo largo de dichos ejes la ubicación de diferentes fuerzas y entidades políticas en cada eje en las fases posteriores de procesamiento automático de textos mediante el uso de BERT*. Esta forma de ubicar a fuerzas similares en áreas cercanas del espacio vectorial es lo que nos permitirá comparar a dichas entidades políticas una vez encontrados los marcos que dichas entidades “*emiten*” en sus textos.

Se ofrece una descripción del marcado de polaridad por parte de los etiquetadores humanos en este proyecto en el Anexo III de este trabajo.

5.2 Fase II: Operaciones de Aprendizaje Automático (MLOps)

En esta fase es en la que propiamente realizaremos las tareas que nos permita ubicar entidades políticas dentro del espacio vectorial definido en las fases anteriores. Nuestro dataset de entrenamiento incluirá diferentes entidades (personas, partidos políticos y documentos). **Dichos datos de entrenamiento etiquetados incluirán para cada entidad los marcos cognitivos detectados para cada una de ellas** (así como los elementos que los definen), **buscando una solución para que dicha detección de marcos se haga de forma no supervisada** (a partir de los términos y metáforas propias de cada marco detectadas en un texto, por ejemplo). El diccionario incluirá también la **“polaridad”** de cada marco (y por lo tanto de cada entidad política) respecto a cada una de las dimensiones del espacio vectorial definido.

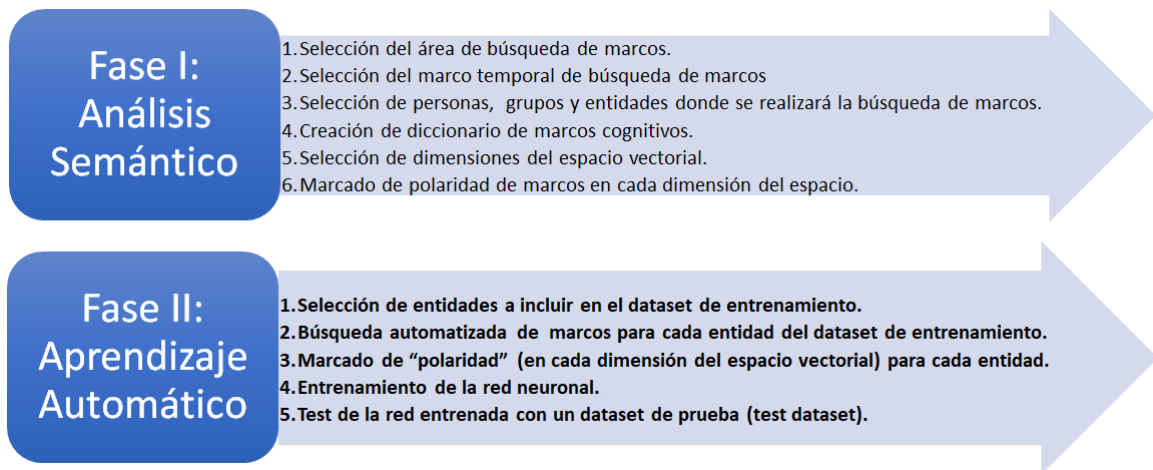


Ilustración 57 Fase II de aprendizaje automático

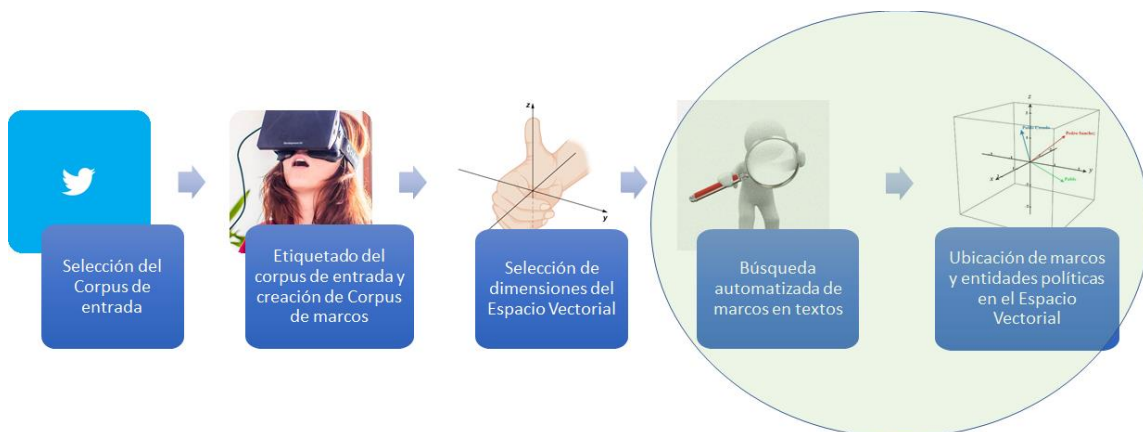


Ilustración 58 Tareas a realizar en la Fase II

Esta es también la fase de diseño y entrenamiento del clasificador/red neuronal. Para ello utilizaremos un corpus de entrada anotado como conjunto de datos de entrenamiento y validación, es decir, ubicaremos los **documentos, partidos políticos o personas en el espacio vectorial** a partir de los marcos utilizados por ellos.

5.2.1 Dataset de Entrenamiento

En esta fase se realiza la selección de entidades (personas, partidos políticos, medios de comunicación, etc.) a incluir en el dataset de entrenamiento de un clasificador basado en BERT, pudiendo ser dichas entidades las mismas utilizadas en la Fase I o un número mayor.

En nuestro caso hemos seleccionado las formaciones políticas PODEMOS y VOX para ubicarlas en áreas diferentes de los ejes del espacio vectorial. Utilizaremos como **dataset de entrada el conjunto de textos etiquetados como pertenecientes a dos marcos por cada uno de los dos ejes del espacio vectorial político bajo estudio**. Para ello disponemos de los textos seleccionados como componentes de los marcos bajo estudio que tomaremos como textos etiquetados para entrenar a la red. A diferencia del caso mostrado en la siguiente figura, **en nuestro caso simplemente tendremos textos etiquetados según el marco al que hagan referencia a lo largo de un eje**.

Para el eje 4, Organización Territorial del Estado:

- 0 para el marco “Estado Plurinacional”.
- 1 para “no marco” (neutral, imparcial o textos que no tratan sobre la “Organización Territorial del Estado”, eje 4).
- 2 para el marco correspondiente a “Estado Uninacional”.

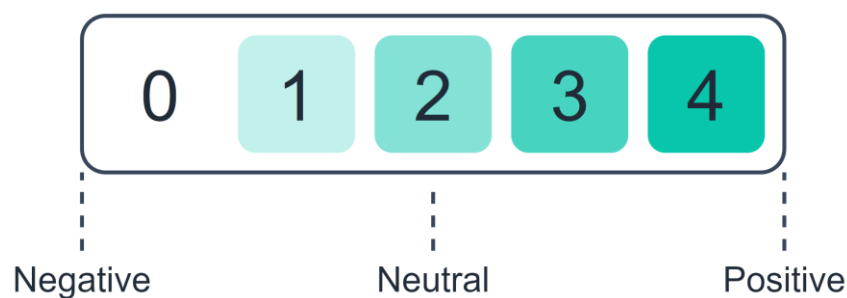


Ilustración 59 Dataset "Rotten Tomatoes" con su escala de sentimiento

Para el eje 5, “Modelo Político del Estado” (Forma de Gobierno):

- 1 para el marco “España Monárquica”.
- 0 para “no marco” (neutral, imparcial o textos que no tratan sobre la “Organización Territorial del Estado”, eje 4).
- 2 para el marco correspondiente a “España Republicana”.

La mayor parte de los datos de entrenamiento marcados como “**No Marco**” proceden del dataset público “[MAS Corpus](#)” (Corpus for Marketing Analysis in Spanish) que consta de 3.763 tuits aproximadamente. Este corpus contiene un conjunto de tweets etiquetados manualmente en idioma español de interés con fines de marketing y que hemos seleccionado como primera aproximación para filtrar tuits no relevantes al elemento político bajo estudio.

El dataset de entrada consta para el primer clasificador (véase 5.2.5, Test y evaluación de la solución) consta de los siguientes casos etiquetados:

- Casos de Entrenamiento: 3088
 - No Marco 2954
 - **Estado Uninacional 85**
 - **Estado Plurinacional 49**
- Casos de Validación: 772
 - No Marco 739
 - **Estado Uninacional 21**
 - **Estado Plurinacional 12**

Para el segundo clasificador el conjunto de datos de entrada tiene las siguientes características:

- Casos de Entrenamiento: 3011
 - No Marco 2954
 - **España Republicana 30**
 - **España Monárquica 27**
- Casos de Validación: 753
 - No Marco 739
 - **España Republicana 8**
 - **España Monárquica 6**

Como podemos observar, en este trabajo y dada la falta de datasets públicos e incluso privados desarrollados para este tipo de investigación política, el autor ha tenido que etiquetar del orden de 167 tuits en un eje y 71 tuits en el otro. Este es un dataset de pequeño tamaño pero que como posteriormente veremos, nos puede ofrecer resultados interesantes mediante el uso de transformadores BERT. Una investigación más profunda requeriría que este proceso de anotación fuera realizado por varios anotadores y éstos tuvieran un grado mínimo de acuerdo.

5.2.2 Búsqueda automatizada de marcos

En este apartado describimos el diseño de la solución de búsqueda automatizada de marcos para cada entidad del dataset de entrenamiento en el conjunto de textos de entrada. Esta búsqueda se realiza para un **conjunto reducido de marcos y de ejes del espacio (dos en este TFM)**, quedando fuera del alcance de este TFM una búsqueda automatizada de los potenciales cientos de marcos que podrían representar la dimensión política de una persona a lo largo de los 10 (o más) ejes del espacio vectorial definido.

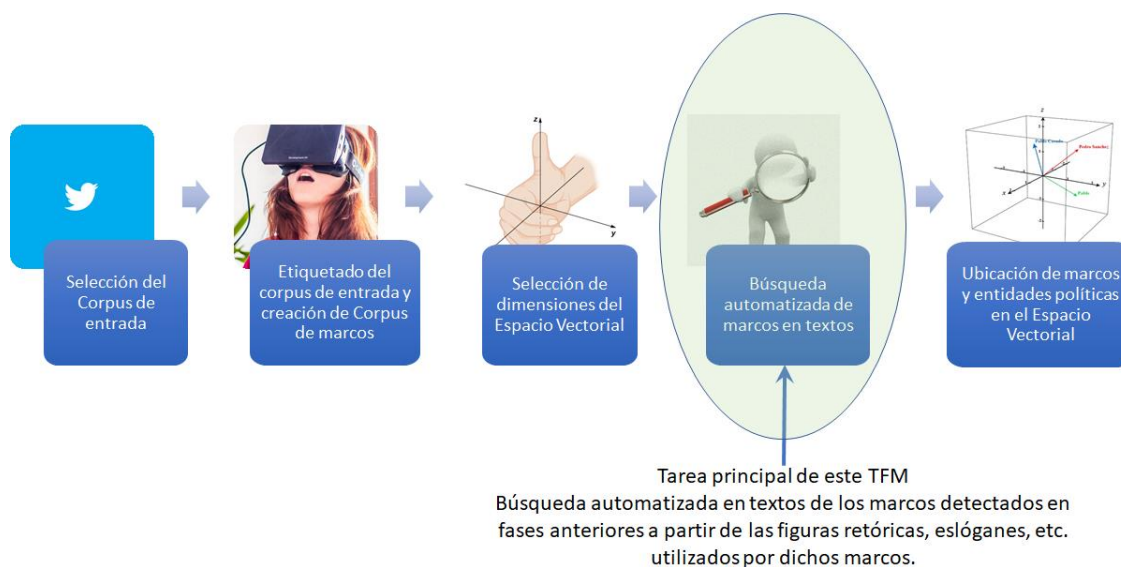


Ilustración 60 Tarea de búsqueda automatizada de marcos en textos

En este proyecto hemos seleccionado **dos de los cuatro marcos** utilizados por la formación política **PODEMOS** definidos en (Alvarez Prieto, 2018), el marco **“República (El Final del Estado del 78)”** como potencial marco a ubicar principalmente a lo largo del **eje 5 (“Modelo político del Estado / Forma de Gobierno”)** definido en el apartado 5.1.3 y el marco **“La España Plurinacional”** como potencial marco a ubicar principalmente a lo largo del **eje 4 (“Organización Territorial del Estado”)**.

Los cuatro marcos identificados en dicho trabajo (véase el apartado 9.8) fueron:

1. El Pueblo como los de "abajo".
2. Las Élités como culpables de la crisis.
3. El Final del Estado del 78.
4. La España Plurinacional.

Marcos utilizados por PODEMOS		España Plurinacional
Recursos de Razonamiento	Definición del Problema	Marco en el que PODEMOS define a España como un estado plurinacional (no reconocido todavía) como solución a los problemas políticos y de todo tipo existentes.
	Culpable	Élites centrales. Monarquía.
	Juicio Moral	España plurinacional y respetuosa que ya existe en las ciudades y en los pueblos y que aún no ha entrado en los partidos y en las instituciones. Hará saltar por los aires uno de los pactos cruciales de la Transición (Generalitat de Catalunya). Es necesario defender España asumiendo el derecho del pueblo catalán a decidir su futuro en un referéndum.
	Solución	Referéndum legal y pactado. Nuevo encaje constitucional que reconociera a Catalunya como nación y profundizara su autogobierno. Fórmulas, federales o confederales, para afrontar la plurinacionalidad de España. Nueva España social, republicana y plurinacional. Iniciativas constituyentes que afirman la soberanía de los pueblos que la habitan. Nueva Constitución que recoja las relaciones entre las diferentes regiones, autonomías y naciones que cohabitan en el Estado.
Recursos de Enmarcado	Figuras retóricas (Metáforas, Sinédoques, hipérbolos, preguntas, etc)	Administración colonial de Catalunya. El virrey Rajoy. Dar forma al sujeto constituyente aragonés. Tejer soluciones de fraternidad. ¿Existe la nación española dentro de Catalunya?
	Eslóganes	España será plurinacional o no será. "España es plurinacionalidad y esto no es discutible" Referéndum pactado es la única que puede resolver el conflicto
	Relatos	Las imágenes de gente con banderas españolas e independentista unidas. El pueblo por delante de las instituciones. El virrey Rajoy querrá administrar Catalunya. La reinstauración de una institución republicana como la Generalitat, reconocida por la Constitución de 1978, fue la base del amplio apoyo social al texto constitucional en Catalunya.

Marcos utilizados por		Final del Estado del 78
Recursos de Razonamiento	Definición del Problema	Marco en el que PODEMOS define la necesidad de finalizar el estatus político resultante del consenso del 78. A diferencia de ERC, para PODEMOS este final está centrado en un proceso constituyente que cambie las estructuras del Estado.
	Culpable	La estructura de poder de nuestro país. Élites
	Juicio Moral	Nuevo proyecto de país asociado a la justicia social y a la soberanía popular.
	Solución	Nuevo encaje constitucional que reconociera a Catalunya como nación. Configuración estatal que acepte de una vez la realidad plurinacional de nuestra patria. Cambiar el marco constitucional. Proceso constituyente.
Recursos de Enmarcado	Figuras retóricas (Metáforas, Sinédoques, hipérbolos, preguntas, etc)	Saltar por los aires uno de los pactos cruciales de la Transición. "¿Para qué sirve la monarquía?"
	Eslóganes	España será plurinacional o no será.
	Relatos	-

Ilustración 61 Marcos a encontrar en la búsqueda automatizada de marcos en textos

Para la definición de dichos marcos se siguió la metodología mencionada en el Anexo I, la teoría de Marcos (Framing). Se muestra a continuación como referencia la tabla resumen de fuentes (421 documentos) y referencias etiquetadas (1.446 en total) en la fase de codificación y búsqueda de los elementos componentes de los marcos (recursos de enmarcado y recursos de razonamiento).

Nombre	Fuentes	Referencias
Recursos de Enmarcado	16	116
Eslóganes	11	22
Eslóganes ERC	3	5
Eslóganes FN	0	0
Eslóganes PODEMOS	8	18
El pueblo como los de abajo (PODEMOS)	5	9
Élites culpables de la crisis (PODEMOS)	4	7
España Plurinacional (PODEMOS)	3	3
Final del Estado del 78 (PODEMOS)	1	1
Figuras retóricas	15	54
Figuras Retóricas ERC	4	9
Figuras Retóricas FN	1	1
Figuras Retóricas PODEMOS	10	45
El pueblo como los de abajo (PODEMOS)	6	13
Élites culpables de la crisis (PODEMOS)	7	15
España Plurinacional (PODEMOS)	5	8
Final del Estado del 78 (PODEMOS)	2	2
Relatos	11	40
Relatos ERC	3	19
Relatos FN	1	1
Relatos PODEMOS	8	21
El pueblo como los de abajo (PODEMOS)	4	8
Élites culpables de la crisis (PODEMOS)	6	9
España Plurinacional (PODEMOS)	2	3
Final del Estado del 78 (PODEMOS)	0	0
Recursos de Razonamiento	22	249
Culpables	19	64
Culpables ERC	3	25
Culpables FN	4	9
Culpables PODEMOS	12	30
El pueblo como los de abajo (PODEMOS)	2	3
Élites culpables de la crisis (PODEMOS)	11	24
España Plurinacional (PODEMOS)	2	2
Final del Estado del 78 (PODEMOS)	1	1
Definición del Problema	17	66
Definición del Problema ERC	3	24
Definición del Problema FN	3	7
Definición del Problema PODEMOS	11	34
El pueblo como los de abajo (PODEMOS)	6	11
Élites culpables de la crisis (PODEMOS)	7	8
España Plurinacional (PODEMOS)	3	4
Final del Estado del 78 (PODEMOS)	5	5
Juicio Moral	7	24
Juicio Moral ERC	3	13
Juicio Moral FN	0	0

Tabla 3 Recursos de razonamiento y de enmarcado

Un ejemplo de referencia para el marco “La España plurinacional” utilizado por PODEMOS sería la siguiente:

*<INTERNALS\ESPAÑA - PODEMOS\PODEMOS Y LOS 'COMUNES' SE LANZAN A EXPLICAR LA PLURINACIONALIDAD A ESPAÑA> - § 3 REFERENCES CODED [1,61% COVERAGE]
REFERENCE 1 - 0,75% COVERAGE
"ESPAÑA ES PLURINACIONALIDAD Y ESTO NO ES DISCUTIBLE", HA SENTENCIADO IGLESIAS.*

Como se puede observar, en la fase de búsqueda de marcos se ha generado un listado de textos (frases seleccionadas a partir de documentos) que se asignan a dichos marcos.

Para realizar esta tarea desarrollaremos una herramienta **en Python basada en BERT preentrenado para encontrar los marcos presentes en un texto dado a partir de los elementos descriptivos de cada marco** (palabras, eslóganes, metáforas, etc.) encontrados en Fase I. Este diseño es similar al presentado por (Briggs, 2021) para el **análisis de sentimientos y clasificación multiclase mediante el uso de Transformers (BERT)**. Otro ejemplo es el presentado por el [CSMaP](#), Centro para las redes sociales y política de la Universidad de Nueva York (Politics, 2020) que utiliza [SMaBERTa](#) basado en **RoBERTa** para clasificar textos (basado a su vez en las librerías de Hugging Face de Transformers simples). El [dataset de entrada](#) en este ejemplo consiste en **texto libre acompañado de sus etiquetas** para la tarea en particular que en su caso es **a partir de titulares de noticias realizar la clasificación en temas** (25 etiquetas o temas posibles cada una representada por un número separado).

1. Macroeconomics
2. Civil Rights, Minority Issues, and Civil Liberties
3. Health
4. Agriculture
5. Labor, Employment, and Immigration
6. Education
7. Environment
8. Energy
10. Transportation
12. Law, Crime, and Family Issues
13. Social Welfare
14. Community Development and Housing Issues
15. Banking, Finance, and Domestic Commerce
16. Defense
17. Space, Science, Technology and Communications
18. Foreign Trade
19. International Affairs and Foreign Aid
20. Government Operations
21. Public Lands and Water Management
24. State and Local Government Administration
26. Weather and Natural Disasters
27. Fires
28. Arts and Entertainment
29. Sports and Recreation
30. Death Notices
31. Churches and Religion
99. Other, Miscellaneous, and Human Interest

Ilustración 62 Conjunto de etiquetas utilizados en dataset de entrada (Fuente: [Amber Ellen Boydston](#))

El estudio realizado por (Terechshenko, 2020) sugiere que los modelos Transformer, y *RoBERTa* en particular, puede ayudar a los científicos políticos a clasificar textos con niveles altos de precisión en relación con otros métodos más tradicionales de aprendizaje automático, y sobre todo cuando disponemos de un conjunto de datos pequeño y/o hay relativamente pocos datos etiquetados disponibles como es nuestro caso.

5.2.3 Ubicación de entidades políticas en espacio vectorial

Ubicación en el espacio vectorial de cada entidad (en cada dimensión del espacio vectorial) de los marcos utilizados por esa entidad junto con su “polaridad”. A partir de la representación de documentos e información extraída de la red social Twitter (mediante la detección y ubicación de los marcos presentes en ellos), extraeremos información política que permita, por ejemplo:

1. Ubicar a una formación política dentro del espacio vectorial y compararla con otras formaciones políticas.
2. Ubicar a una personalidad política o legislador (Rheault & Cochrane, 2020, pág. 25), o mejor dicho, las opiniones vertidas por dicha persona en la red social Twitter, dentro del espacio vectorial y ubicarla dentro de un clúster de personas con ideas políticas similares.

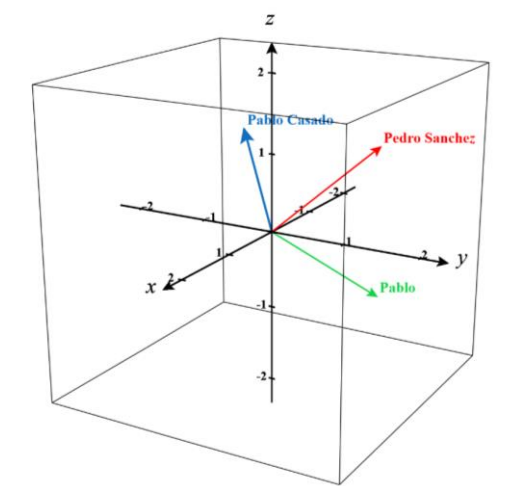


Ilustración 63 Ubicación de tres líderes políticos en el espacio vectorial pragmático

En nuestro caso, en el que hemos ubicado *dos marcos utilizados por la formación política PODEMOS y otros dos marcos utilizados por VOX a lo largo de las dos dimensiones del espacio vectorial* bajo estudio. Para las nuevas entidades políticas que se quieran estudiar (por ejemplo el PP o el PSOE), hemos identificado (mediante el clasificador basado en BERT en el apartado anterior) si utilizan (o no) dichos marcos, y por lo tanto **ubicaremos dichas entidades a lo largo de dichos ejes siempre y cuando utilicen los mismos marcos utilizados por PODEMOS y VOX** (en caso contrario no se ubicarán hasta que para dichas entidades se detecten otros marcos fuera del alcance de este trabajo).

5.2.4 Entrenamiento del clasificador

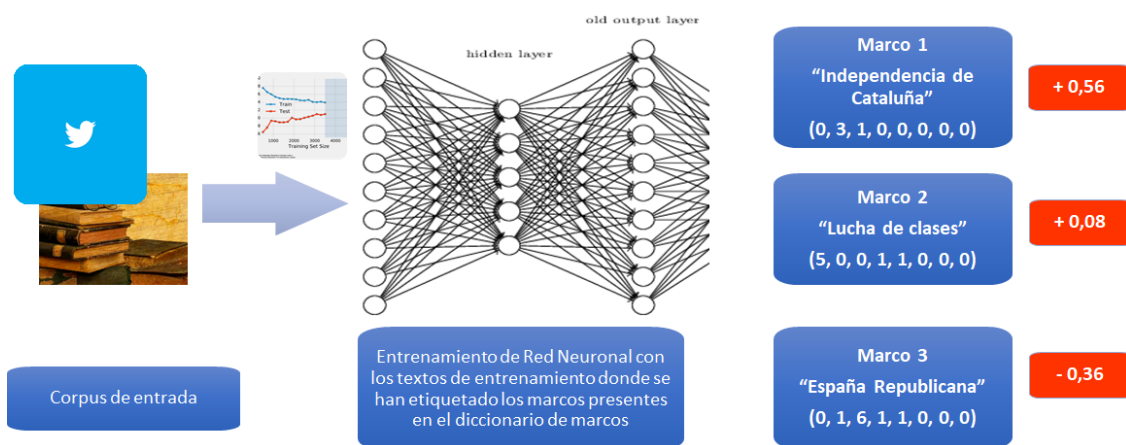


Ilustración 64 Entrenamiento del clasificador para la detección automatizada de marcos

Como hemos comentado anteriormente, se ha realizado el diseño en Python de un clasificador multiclase basado en BERT (SMaBERTa) entrenado con los textos etiquetados indicando los marcos contenidos en dichos textos.

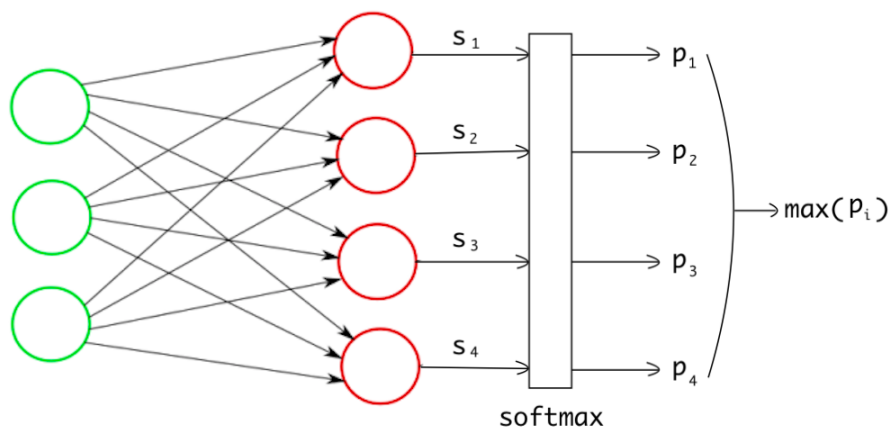


Ilustración 65 Clasificación multiclase ("Multiclass classification")

En la **clasificación multiclase** la etapa de salida tiene el mismo número de nodos de salida que el número de clases (en nuestro caso el número de marcos sobre el eje de estudio, es decir, dos más un conjunto de datos para el caso de "no marco". Cada nodo de salida pertenece a una clase y genera una puntuación para esa clase. Las puntuaciones de la última etapa se pasan a través de una función de activación (sigmoide, softmax en la figura anterior, ReLU u otros tipos) que convierte la puntuación en valores de probabilidad (y en general los datos se clasifican en la clase correspondiente al valor de probabilidad más alto).

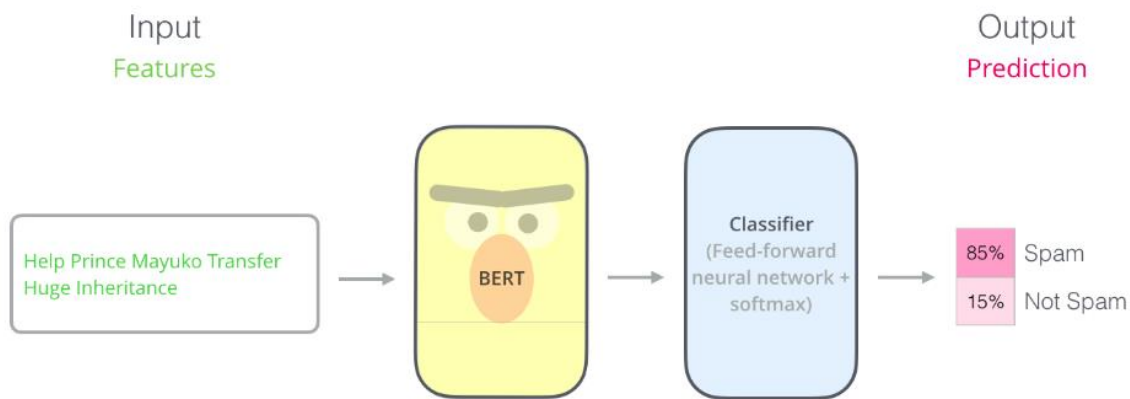


Ilustración 66 BERT en clasificación de textos. Fuente: [Swatimeena](#)

En nuestro caso el modelo de lenguaje (LM) es el transformer **RoBERTa** con una etapa de clasificación (etapa lineal simple con dropout) en la etapa final. Debajo de esta solución se encuentran las [librerías para Transformers pre-entrenados de Huggingface](#). En nuestro caso utilizaremos el modelo “[bert-base-multilingual-cased](#)” que incluye entre otros idiomas el [español](#) y que está compuesto por 12 capas o etapas, **768 dimensiones en la capa oculta (es decir, tamaño de los embeddings)**, **12 cabezas (heads) de transformer** y **110 millones de parámetros**. Este modelo ha sido **entrenado sobre todos los textos en mayúsculas de Wikipedia** para los 104 idiomas principales de Wikipedia.

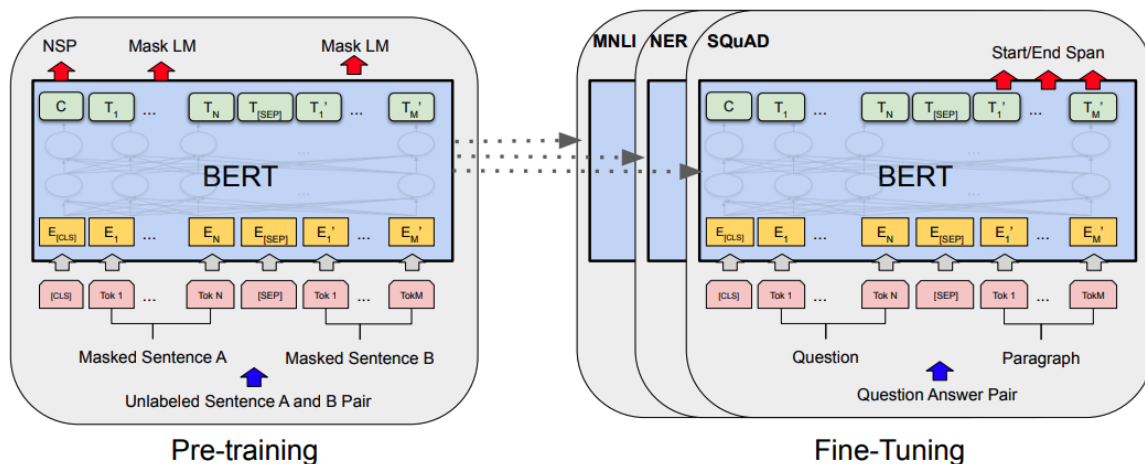


Ilustración 67 Preentrenamiento y ajuste fino en BERT. Fuente: [BERT paper](#)

De manera similar a un clasificador tradicional, **en la fase de entrenamiento se utilizan las secuencias de texto con las etiquetas como argumentos**. El Transformer primero realiza la codificación de las oraciones en función de su tokenizador, seguido de la etapa hacia adelante

(forward pass) a través de la red neuronal y **finalmente se ejecuta la etapa de optimización mediante la utilización de la función de pérdida de entropía cruzada** (“cross entropy loss”). En aplicaciones de clasificación de textos, **la representación BERT del token de clasificación especial “<cls>” codifica la información de toda la secuencia de texto de entrada** (un tuit o una frase en nuestro caso). La representación del texto de entrada (salida del transformer BERT), se utiliza como entrada para alimentar una red neuronal (colocada a la salida del transformer BERT) que consta de varias capas completamente conectadas (densas) para generar la distribución de todos los valores discretos de etiqueta y predecir así la etiqueta correspondiente al texto de entrada.

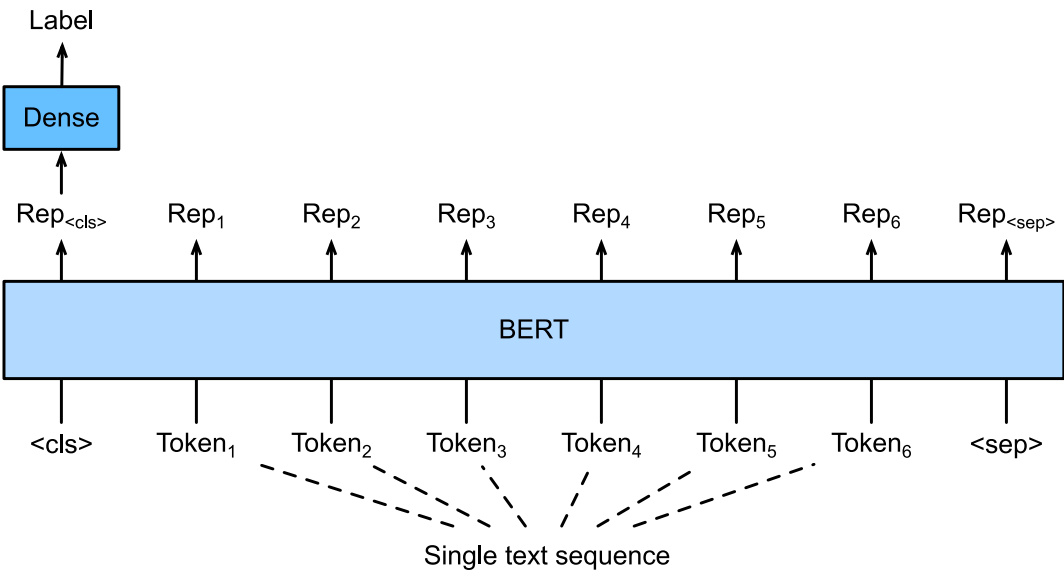


Ilustración 68 Ajuste fino de un transformer BERT para clasificación de textos

Al utilizar BERT en PyTorch, podemos utilizar los Transformers proporcionados por **Hugging Face** (modelos de preentrenamiento de BERT), incluidos algunos modelos para tareas “aguas abajo” (“downstream”) específicas, como por ejemplo el modelo denominado “*bertForSequenceClassification*” que es el utilizado en nuestro caso. El código para dicha clase se encuentra disponible en la web de HuggingFace (Hugging Face , 2020).

bertForSequenceClassification

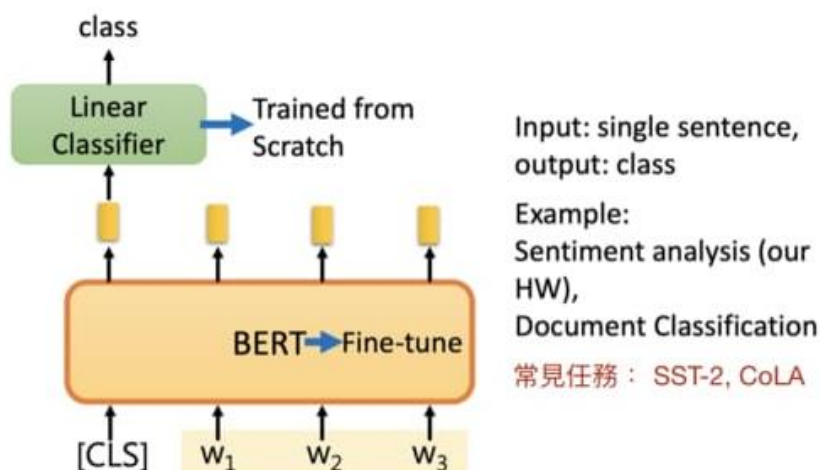


Ilustración 69 BERT for Sequence Classification

Como podemos observar, **añadimos una etapa lineal a la ya existente en el transformer BERT** (hay una capa oculta entre el estado oculto final y la salida agrupada). Esta solución ha sido comparada con otras opciones a la salida de la etapa BERT (Ma, 2019) con resultados comparables a las otras opciones (como por ejemplo redes convolucionales o con más etapas a la salida). En general, para tareas de clasificación dos capas ocultas son suficientes. Este “fine tuning” permite a la red seleccionar para cada tarea (en nuestro caso la clasificación) las componentes del embedding generado por BERT más importantes para dicha tarea. De este modo, por ejemplo, para una tarea de tipo NER la información que se usaría durante el fine-tuning de BERT sería diferente.

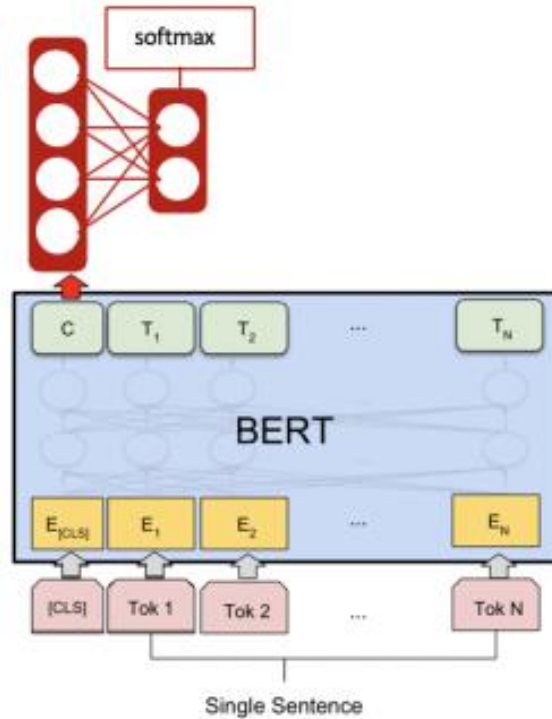


Ilustración 70 Etapa final de clasificación de frases

Entre las ventajas de este ajuste fino de un sistema BERT está que **los pesos del modelo BERT previamente entrenado ya codifica mucha información sobre nuestro idioma** y por lo tanto lleva mucho menos tiempo entrenar nuestro modelo con textos de entrenamiento. Podríamos decir que es como si ya hubiéramos entrenado las capas inferiores de nuestra red de manera intensiva (con millones de textos utilizados para entrenar el modelo BERT inicial) y solo necesitaríamos ajustarla de forma mínima para después utilizar su salida como características (features) de nuestra etapa final de clasificación.

En la fase de producción, si por ejemplo seleccionásemos una entidad (documento, persona, partido político) con ideología similar a la de la formación política PODEMOS, a partir de los textos generados por dicha entidad (datos de test, tuits, por ejemplo), al alimentar a la red neuronal con los textos utilizados por dicha entidad, la salida del clasificador debería generar la probabilidad de que dicha entidad esté utilizando (y comparta con PODEMOS) los marcos bajo estudio.

5.2.5 Test y evaluación de la solución (BERT + Clasificador)

El objetivo de esta fase es, después de haber realizado el entrenamiento del clasificador en la fase anterior, realizar la **evaluación del clasificador con el conjunto de datos de prueba** (test/holdout set en la literatura), es decir, con un conjunto de textos etiquetados con sus marcos correspondientes.

En cuanto a la evaluación del clasificador basado en BERT, hemos evaluado nuestro sistema mediante el **coeficiente de correlación de Mathews (MCC)** (Chicco & Jurman, 2020). La mayor parte de los sistemas utilicen métricas como la “**Accuracy**” (exactitud) y el valor **F1** (media armónica que combina los valores de la precisión, precision, o y de la exhaustividad o recall) para valorar la efectividad de sistemas de análisis de sentimiento. MCC por su parte al igual que el valor F1 también tiene en cuenta los cuatro valores en la matriz de confusión. Un valor alto (cercano a 1) significa que todas las clases se predicen bien, incluso si una clase está desproporcionadamente subrepresentada (o sobrerrepresentada). Creemos que es preferible la utilización del coeficiente de correlación de Mathews por ser más informativo que el valor F1 y la precisión en la evaluación de problemas de clasificación **al tomar en cuenta las relaciones de equilibrio de las cuatro categorías de la matriz de confusión** (verdaderos positivos, verdaderos negativos, falsos positivos, falsos negativos) evitando los problemas en casos de datos con “class imbalance” y asimetría de la “accuracy” y F1 (Chicco & Jurman, 2020).

$$\begin{aligned}
 \text{MCC} &= \frac{\text{cov}(X, Y)}{\sqrt{\text{cov}(X, X) \cdot \text{cov}(Y, Y)}} \\
 &= \frac{\sum_{k, l, m=1}^N C_{kk} C_{ml} - C_{lk} C_{km}}{\sqrt{\sum_{k=1}^N \left[\left(\sum_{l=1}^N C_{lk} \right) \left(\sum_{f, g=1, f \neq k}^N C_{gf} \right) \right]} \sqrt{\sum_{k=1}^N \left[\left(\sum_{l=1}^N C_{kl} \right) \left(\sum_{f, g=1, f \neq k}^N C_{fg} \right) \right]}}
 \end{aligned}$$

Ilustración 71 Generalización del coeficiente de correlación de Mathews para un caso multiclase (Jurman, Giuseppe et al. “, 2021)

El entrenamiento del clasificador basado en BERT se ha realizado con los siguientes parámetros:

- Learning Rate: 1e-05

- Train Epochs: 5
- Number Labels: 3

En este primer prototipo se ha decidido crear **dos clasificadores diferentes, uno por cada eje del espacio vectorial** definido en este proyecto (eje4 “Organización Territorial del Estado” y eje 5 “Modelo Político del Estado”). En un sistema comercial podría evaluarse utilizar un único clasificador que tome todos los datos de entrada para generar la ubicación de una entidad sobre diferentes ejes sobre un único clasificador al mismo tiempo.

Si posteriormente dispusiéramos de datos etiquetados de marcos ubicados en los otros 8 ejes del espacio vectorial definido (no utilizados en este proyecto) podríamos crear clasificadores similares a lo largo del resto de ejes y ubicar a cada entidad política bajo estudio (individuos, partidos políticos, textos políticos) en la totalidad de las 10 dimensiones del espacio vectorial.

En cuanto al conjunto de datos de entrada, éste tiene las siguientes características:

- Casos de Entrenamiento: 3088
 - No Marco 2954
 - **Estado Uninacional 85**
 - **Estado Plurinacional 49**
- Casos de Validación: 772
 - No Marco 739
 - **Estado Uninacional 21**
 - **Estado Plurinacional 12**

Los resultados de la fase de validación (disponibles en el siguiente enlace al [cuaderno de Google Colab](#)) han sido los siguientes:

{'mcc': 0.861134666847093}

Number of predictions: 772

Total Accuracy: 98.83%

También presentamos a continuación el valor de precisión para cada una de las tres etiquetas, así como el detalle de los textos clasificados erróneamente en otra etiqueta y dicha etiqueta.

[0 0 0 0 0 **2 0 2** 0 0 0 0]

[0 0 0 0 0 **0 0 0** 0 0 0 0]

Clase Inicial: 2, Estado **Plurinacional**

Accuracy: 10/12, = **83.33%**

Clase Inicial: 0, **No Marco**

Accuracy: 736/739, = **99.59%**

[2 2 2 2 2 2 2 2 2 **1 1** 2 2 **0 0** 2 2 2 2 2]

[2 2 2 2 2 2 2 2 2 **2 2** 2 2 **2 2** 2 2 2 2 2]

Clase Inicial: 5, **Estado Uninacional**

Accuracy: 17/21, = **80.95%**

5.2.5.2 Clasificador BERT para el eje 5 “Modelo Político del Estado”

El conjunto de datos de entrada para éste segundo clasificador tiene las siguientes características:

- Casos de Entrenamiento: 3011
 - No Marco 2954
 - **España Republicana 30**
 - **España Monárquica 27**
- Casos de Validación: 753
 - No Marco 739
 - **España Republicana 8**
 - **España Monárquica 6**

Los resultados de la fase de validación (disponibles en el siguiente enlace al [cuaderno de Google Colab](#)) han sido los siguientes:

{'mcc': 0.9276059465250862}

Number of predictions: 753

Total Accuracy: 99.73%

En cuanto al valor de precisión para cada una de las tres etiquetas, así como el detalle de los textos clasificados erróneamente en otra etiqueta y dicha etiqueta, tenemos:

[1 1 1 1 1 1]

[1 1 1 1 1 1]

Clase Inicial: 6, **España Monárquica**

Accuracy: 6/6, = **100.00%**

Clase Inicial: 0, **No Marco**

Accuracy: 738/739, = **99.86%**

[2 2 2 **0** 2 2 2 2]

[2 2 2 2 2 2 2 2]

Clase Inicial: 7, **España Republicana**

Accuracy: 7/8, = **87.50%**

5.2.6 Solución tecnológica

En este apartado describiremos la solución utilizada, que básicamente está construida alrededor de una **etapa inicial** basada en un **modelo de lenguaje preentrenado BERT**, seguida de una etapa utilizado para clasificar los textos de entrada en los “marcos” a los que dichos textos hacen referencia (clasificador). Dicha solución está desarrollada en Python haciendo uso (allá donde ha sido posible) de librerías existentes como Pandas, HuggingFace y Tweepy. En la siguiente figura representamos los componentes principales de alto nivel de esta solución: entrada de datos desde Twitter, utilización de Python, análisis de marcos, ubicación de textos en un espacio definido por marcos.

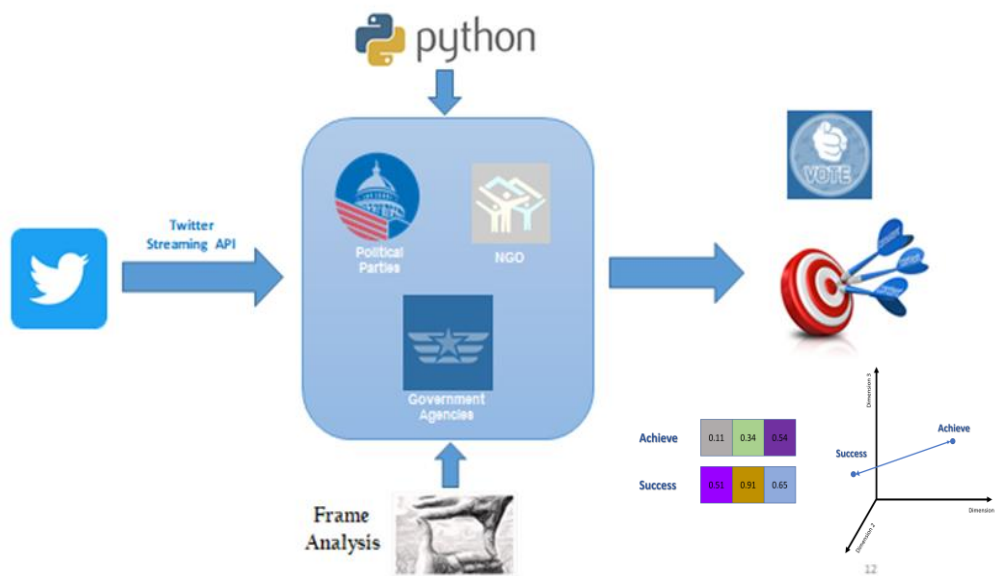
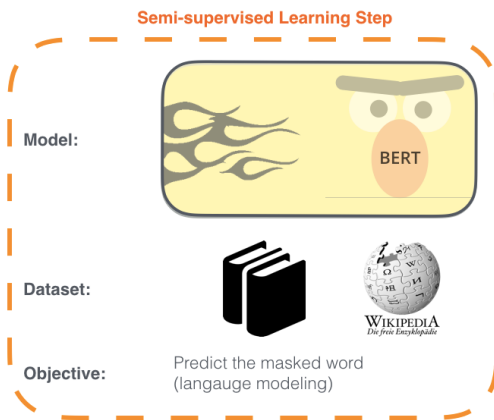


Ilustración 72 Visión global del sistema a desarrollar

Nuestra solución está basada en un **modelo de lenguaje preentrenado BERT**, seguida de una etapa de clasificación tal como se muestra en la solución número 2 de la siguiente figura (entrenamiento supervisado en una tare específica mediante el uso de un conjunto de datos etiquetados):

1 - **Semi-supervised** training on large amounts of text (books, wikipedia..etc).

The model is trained on a certain task that enables it to grasp patterns in language. By the end of the training process, BERT has language-processing abilities capable of empowering many models we later need to build and train in a supervised way.



2 - **Supervised** training on a specific task with a labeled dataset.

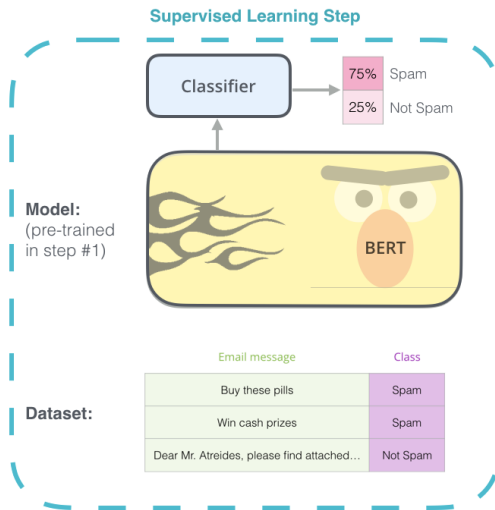


Ilustración 73 Utilización de BERT para clasificación de textos

5.2.6.1 Plataforma MLOps (Machine Learning Operations)

En la realización de proyectos de aprendizaje automático e inteligencia artificial, una de las [mejores prácticas](#) actuales es la utilización de una plataforma de operaciones de aprendizaje automático (MLOps). Entre otras podemos mencionar a Dataiku, H2O.ai, Databricks, DataRobot, Alteryx, etc.

MLOps (una palabra compuesta de "aprendizaje automático" y "operaciones") es una práctica de *colaboración y comunicación entre científicos de datos y profesionales de operaciones* para ayudar a **administrar el ciclo de vida de ML** (o DL) incluyendo la fase de producción. De manera similar a los enfoques DevOps o DataOps, MLOps busca aumentar la automatización y mejorar la calidad del aprendizaje automático en producción, mientras que también toma en cuenta y se enfoca en los requisitos comerciales y regulatorios. MLOps está evolucionando hacia un enfoque independiente para la gestión del ciclo de vida completo de ML, desde la integración con la generación de modelos (ciclo de vida de desarrollo de software, integración continua / entrega continua), la orquestación y la implementación, hasta el diagnóstico, la gobernanza y las métricas de negocio.

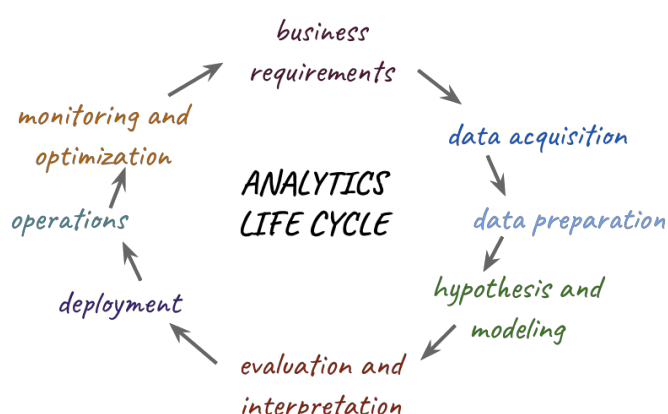


Ilustración 74 Fases en operaciones de aprendizaje automático (MLOps)

MLOps es un conjunto de mejores prácticas que gira en torno a hacer que el aprendizaje automático en producción sea más sencillo y fluido. El propósito es cerrar la brecha entre la experimentación y la producción para hacer que el aprendizaje automático sea reproducible, colaborativo y continuo.

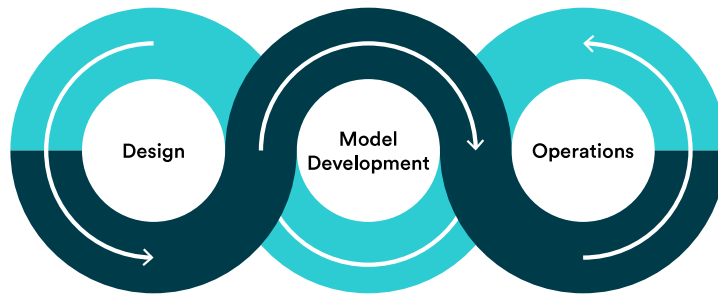


Ilustración 75 Fases de diseño, desarrollo del modelo y operaciones en aprendizaje automático

Dado el alcance limitado de este proyecto donde solo existe un desarrollador y codificador de textos con un dataset reducido, no hemos utilizado ninguna solución comercial de MLOps. El desarrollo software se ha realizado en el lenguaje **Python**, en el entorno de desarrollo integrado de “**Google Colab**” y en local con **Spyder 3.3.6**. La fase de entrenamiento y evaluación se ha ejecutado completamente sobre la siguiente plataforma de aceleración GPU [Tesla TPU-100 de NVIDIA](#) con una RAM disponible de 27,3 Gbytes:

```

Mon Jun 14 09:17:13 2021
+-----+
| NVIDIA-SMI 465.27          Driver Version: 460.32.03    CUDA Version: 11.2     |
+-----+-----+-----+-----+-----+-----+
| GPU  Name                Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp   Perf   Pwr:Usage/Cap|      Memory-Usage | GPU-Util  Compute M. |
|                                           MIG M.         |
+-----+-----+-----+-----+-----+-----+
|   0   Tesla P100-PCIE...    Off          | 00000000:00:04:0  Off |                    0 |
| N/A   41C    P0     28W / 250W |      0MiB / 16280MiB |      0%      Default |
|                                           |                      | N/A   |
+-----+-----+-----+-----+-----+-----+

+-----+
| Processes:                                                       GPU Memory |
|  GPU   GI    CI          PID    Type   Process name                               Usage      |
|-----+-----+-----+-----+-----+-----+
| No running processes found
+-----+

```

Ilustración 76 Plataforma de ejecución GPU utilizada

Para la realización de este trabajo, en la parte de recopilación de datos de redes sociales nos centraremos en la red social Twitter a través de la **API Tweepy** para recoger datos de Twitter. En este caso se ha desarrollado el código necesario (en Python) para la recogida de tuits relevantes para este trabajo.

- Twitter: **Tweepy** es una librería Open Source escrita en Python y alojada inicialmente en GitHub que permite la comunicación con Twitter por medio de su API. Provee de acceso a todos los métodos de la API Rest de Twitter.
- La búsqueda se ha limitado a los **3.000 tuits más recientes para cada entidad** bajo estudio en Twitter (limitación de la API de Twitter libre y sin coste) y el código desarrollado se encuentra disponible en el [siguiente enlace al cuaderno de Google Colab](#).
- El objetivo es la **recogida automatizada de tuits del timeline de usuarios de Twitter**. Esta tarea podría realizarse a través de la búsqueda únicamente de “recursos de enmarcado” y “recursos de razonamiento” componentes de los marcos mencionados en 9.5.1 (Metodología de búsqueda de marcos en la fase Inductiva) en lugar de la descarga de todos los tuits (queda fuera del alcance de este TFM, pero podría ser implementada como continuación de este trabajo).

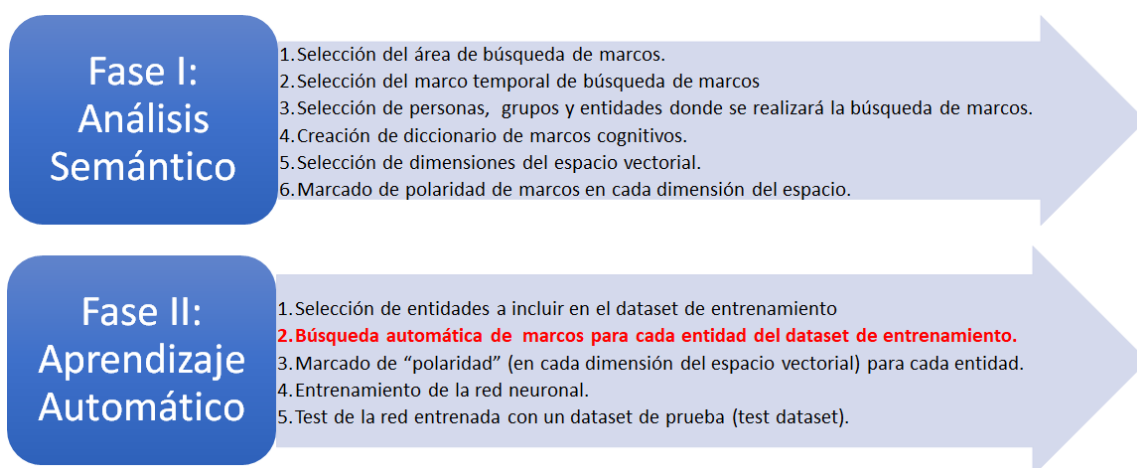


Ilustración 77 Búsqueda automatizada de marcos en Twitter

En este trabajo realizaremos una codificación manual de diferentes marcos encontrados en el corpus de documentos y tuits de entrada.

Utilizaremos **MongoDB**, un sistema de bases de datos de los conocidos como “No SQL”, es decir, que no utiliza un sistema tradicional como en los sistemas de almacenamiento de base de datos relacional. En su lugar, MongoDB permite guardar estructuras de datos dinámicas, para lo cual utiliza BSON (<https://www.mongodb.com/json-and-bson>), una variante basada en JSON y que significa Binary JSON. Los objetos BSON son listas ordenadas de elementos. La versión utilizada ha sido la **MongoDB Atlas 4.2.3**. MongoDB Atlas es el **servicio global de bases de datos en la nube** utilizado en aplicaciones modernas.

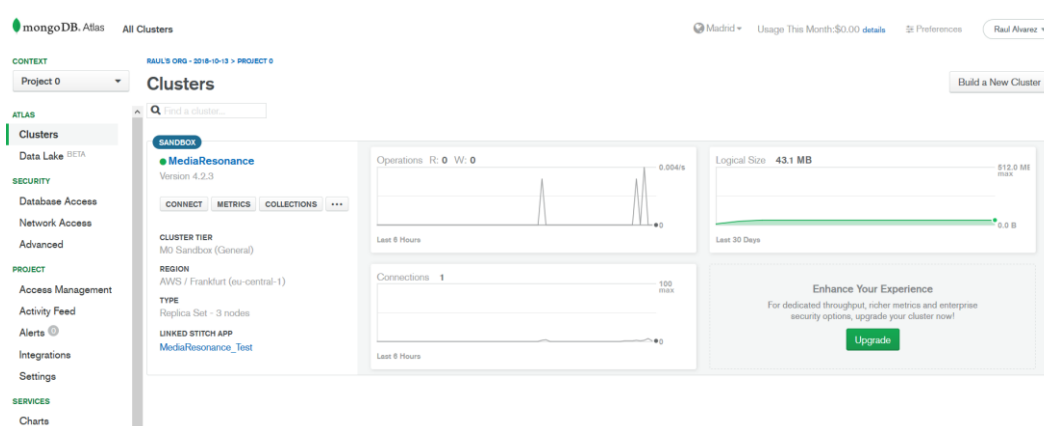


Ilustración 78 MongoDB Atlas 4.2.3

MongoDB guarda la estructura de los datos (en nuestro caso los tuits) en documentos BSON con un esquema dinámico, lo que implica que no existe un esquema predefinido. Los elementos de los datos se denominan documentos y se guardan en colecciones (que pueden tener un número indeterminado de documentos). Las colecciones son como tablas y los documentos son registros en la tabla en una base de datos relacional. La diferencia es que en una base de datos relacional cada registro en una tabla tiene la misma cantidad de campos, mientras que en MongoDB cada documento en una colección puede tener diferentes campos.³

5.2.6.4 Solución de Procesamiento Lingüístico

En este apartado describiremos las diferentes herramientas que utilizaremos en los niveles de procesamiento lingüístico:

1. Análisis morfológico y léxico.
2. Análisis sintáctico.
3. Análisis semántico.
4. Análisis pragmático.

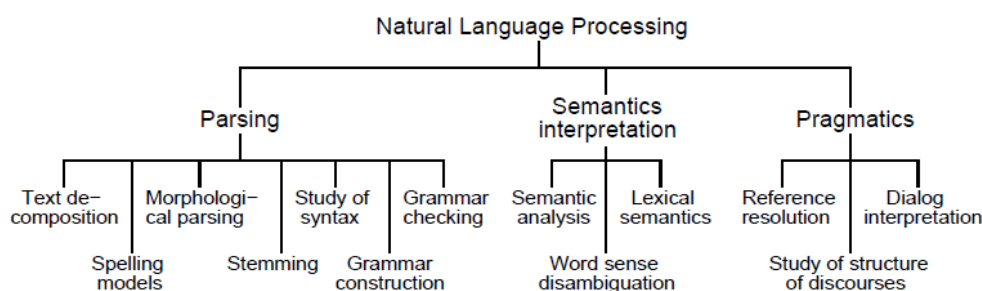


Ilustración 79 Tareas en los diferentes niveles de procesado lingüístico. Fuente: (Lensu, 2002)

Para las capas inferiores de procesado lingüístico (morfológico, sintáctico, semántico) hemos seleccionado en nuestro proyecto la librería [HuggingFace](#), una biblioteca “open source” escrita en Python para el Procesamiento del Lenguaje Natural. Proporciona un conjunto de bibliotecas de procesamiento de texto para la clasificación, tokenización, stemming, etiquetado, análisis y de razonamiento semántico, contenedores para las bibliotecas de NLP de potencia industrial, sobre todo para Transformers. En nuestro caso como hemos comentado trabajaremos sobre [RoBERTa](#) cuyo **tokenizador**, derivado del tokenizador GPT-2, utiliza **codificación de pares de bytes a nivel de bytes** (byte-level Byte-Pair-Encoding). Este tokenizador ha sido entrenado para **tratar los espacios como partes de los tokens** por lo que una palabra se codificará de manera diferente si está al principio de la oración (sin espacio) o en otra posición.

La etapa de análisis en niveles superiores (nivel pragmático) como se ha mostrado anteriormente, está basada en el análisis mediante marcos cognitivos a través de una metodología propia basada en la detección y creación de un diccionario de marcos en el área política para su ubicación en un espacio vectorial diseñado para ubicar los marcos utilizados.

5.3 Resultados

En este paso se alimenta al clasificador con **datos no vistos hasta ahora, que son textos generados por otras entidades políticas o partidos y recogidos de Twitter**, y se comprueba que la salida de la red neuronal clasifica adecuadamente los marcos utilizados por dicho individuo. *La salida de dicho clasificador es un vector que estima la probabilidad de aparición de cada marco del diccionario en función del texto de entrada a la red neuronal.*

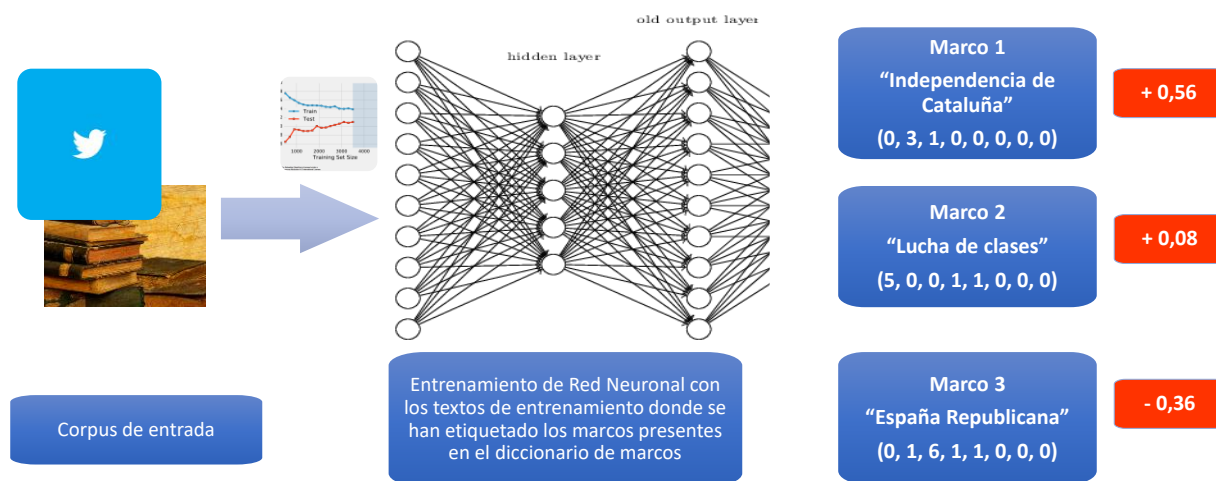


Ilustración 80 Entrenamiento del clasificador para la detección automatizada de marcos

Nuestro sistema basado en BERT proporciona una valoración relativa al hecho de que una entidad política esté utilizando un marco (por ejemplo, el marco “Estado plurinacional” o “Estado Uninacional”) *a lo largo de un eje del espacio vectorial* (“organización territorial del estado en este ejemplo”). *Este valor lo podremos comparar con el valor medido para otra entidad política* lo que nos permite por ejemplo comparar formaciones políticas entre sí.

Para determinar las puntuaciones asignadas por nuestro sistema a cada entidad política hemos realizado la *clasificación y ubicación de nuevas entidades políticas y documentos en el espacio vectorial* mediante la utilización de los tuits de cada entidad política como entrada a la red neuronal basada en el transformer BERT. Así, **hemos estimado, a lo largo de un eje, la cercanía a los dos marcos utilizados como referencia (estado plurinacional y estado uninacional)** de los textos que esa entidad genera hacia la sociedad y los medios. Esto nos permite ubicar a dicha entidad en el espacio vectorial a partir de las ideas políticas vertidas en Twitter (clasificación de nuevas entidades de un conjunto de datos de prueba o “test dataset”).

Nuestras mediciones, calculadas automáticamente a partir de los tuits y textos recogidos, han sido principalmente tres:

1. **Polaridad Bruta (PB):** Número de mensajes haciendo referencia a un marco menos el número de mensajes haciendo referencia al marco contrapuesto ($PB = M1 - M2$). El número de mensajes neutros (“No Marco”) no se toman en cuenta.
2. **Polaridad Relativa (PR):** Número de mensajes haciendo referencia a un marco dividido entre el número de mensajes haciendo referencia al marco contrapuesto ($PR = M1 / M2$). El número de mensajes neutros (“No Marco”) no se toman en cuenta.
3. **Polaridad Neta (PN):** Polaridad Bruta dividido entre todos los mensajes (tuits o frases) procesados, $PN = PB / (M1 + M2 + M0)$. En este caso el número de mensajes neutros (“No Marco”) si se toman en cuenta. El Valor de la polaridad neta está entre +1 (100%) y -1 (-100%).

Los resultados proporcionados por nuestro sistema BERT para las más de 30 entidades políticas estudiadas se presentan en las siguientes tablas:

Entidad Política	Eje 4 (Plurinacional-Uninacional)			Eje 5 (Monarquía-Republica)			Tipo de Entidad	Correlación PEARSON (PN1, PN2)
	PB1	PR1	PN1	PB2	PR2	PN2		
vox_es_tweets	247	163,17%	7,61%	679	1128,79%	20,92%	Partido Político	-68,58%
psoe_tweets	-114	80,81%	-3,51%	822	2517,65%	25,30%	Partido Político	
populares_tweets	-83	87,09%	-2,55%	993	1282,14%	30,55%	Partido Político	
podemos_tweets	208	149,76%	6,42%	606	1831,43%	18,69%	Partido Político	
CiudadanosCs_tweets	-19	97,31%	-0,58%	934	2624,32%	28,74%	Partido Político	
IzquierdaUnida_tweets	344	179,81%	10,59%	738	1704,35%	22,72%	Partido Político	
IzRepublicana_tweets	79	128,52%	2,49%	603	1216,67%	18,97%	Partido Político	
fedelajsons_tweets	209	140,74%	6,60%	682	554,67%	21,53%	Partido Político	-74,58%
webpcpe_tweets	-182	66,67%	-5,60%	859	7909,09%	26,44%	Partido Político	
ume_umes_tweets	402	192,63%	12,53%	-1149	21,57%	-35,82%	Organización	
rielcano_tweets	350	278,57%	10,78%	194	670,59%	5,98%	Organización	85,86%
CasaReal_tweets	1012	495,31%	31,14%	-1412	11,53%	-43,45%	Organización	
CCivicaCatalana_tweets	262	142,46%	8,11%	959	1255,42%	29,70%	Organización	
publico_es_tweets	332	228,19%	10,22%	618	1607,32%	19,02%	Medios	-78,82%
marca_tweets	9	212,50%	0,28%	11		0,34%	Medios	
elpais_espana_tweets	367	238,49%	11,29%	1040	1353,01%	32,00%	Medios	
elnacionalcat_e_tweets	182	180,53%	5,60%	512	1144,90%	15,75%	Medios	
eldiarios_tweets	343	263,33%	10,55%	674	2908,33%	20,74%	Medios	
abcespana_tweets	329	189,40%	10,12%	1190	1467,82%	36,62%	Medios	
RubenSanchezTW_tweets	192	268,42%	5,92%	249	1878,57%	7,68%	Individuo	
raul_alvarez_pr_tweets	73	661,54%	2,26%	28	333,33%	0,87%	Individuo	-74,73%
ramoncotarelo_tweets	-139	73,32%	-4,31%	538	3264,71%	16,69%	Individuo	
MonederoJC_tweets	-141	75,44%	-4,35%	695	1978,38%	21,45%	Individuo	
mikelayestaran_tweets	86	151,50%	2,65%	336	1868,42%	10,34%	Individuo	-9,32%
Pablo Iglesias y el proyecto de República	-6	64,71%	-15,38%	20	766,67%	51,28%	Documento	
Pablo Iglesias - La sociedad española no a	6	146,15%	17,14%	9	212,50%	25,71%	Documento	
La Monarquía, cinco razones por la que e	10	171,43%	21,74%	-11	52,17%	-23,91%	Documento	
52-medidas-autonomias VOX	61	2133,33%	78,21%	5	155,56%	6,41%	Documento	
2017_10_23_carta_consideraciones_Cat	-38	25,49%	-58,46%	45	662,50%	69,23%	Documento	
santi_abascal_tweets	379	176,88%	11,67%	823	891,35%	25,33%	Dirigente Político	-9,32%
sanchezcastejon_tweets	396	164,60%	12,18%	667	625,20%	20,52%	Dirigente Político	
PabloIglesias_tweets	432	199,54%	13,58%	717	1478,85%	22,55%	Dirigente Político	
PabloEchenique_tweets	13	208,33%	9,15%	16	633,33%	11,27%	Dirigente Político	
pablocasado_tweets	404	155,34%	12,43%	1161	979,55%	35,73%	Dirigente Político	
InesArrimadas_tweets	224	131,37%	6,92%	1089	2278,00%	33,64%	Dirigente Político	
FulgencioColl_tweets	118	257,33%	11,61%	145	637,04%	14,27%	Dirigente Político	

Ilustración 81 Resultados del clasificador BERT para dos dimensiones

Hemos tomado como base para el análisis de resultados de nuestro estudio las medidas de “Polaridad Neta (PN)” al ser las que mejor nos permiten independizar los resultados del número de textos procesados por entidad (miles de tuits recogidos por usuario frente a decenas o centenas de oraciones por documentos), así como tomar en cuenta los textos no relativos a las posiciones políticas bajo estudio (textos sin marco). La siguiente tabla muestra la lista ordenada de polaridades netas (desde el extremo negativo correspondiente a la Plurinacionalidad hasta el extremo positivo correspondiente a la Uninacionalidad):

Eje 4 (Plurinacional-Uninacional)	
Entidad Política	PN1
2017_10_23_carta_consideraciones_Cat	-58,46%
Pablo Iglesias y el proyecto de República	-15,38%
webpcpe_tweets	-5,60%
MonederoJC_tweets	-4,35%
ramoncotarelo_tweets	-4,31%
psoe_tweets	-3,51%
populares_tweets	-2,55%
CiudadanosCs_tweets	-0,58%
marca_tweets	0,28%
raul_alvarez_pr_tweets	2,26%
IzRepublicana_tweets	2,49%
mikelayestaran_tweets	2,65%
elnacionalcat_e_tweets	5,60%
RubenSanchezTW_tweets	5,92%
podemos_tweets	6,42%
fedelasjons_tweets	6,60%
InesArrimadas_tweets	6,92%
vox_es_tweets	7,61%
CCivicaCatalana_tweets	8,11%
PabloEchenique_tweets	9,15%
abcespana_tweets	10,12%
publico_es_tweets	10,22%
eldiarioes_tweets	10,55%
IzquierdaUnida_tweets	10,59%
rielcano_tweets	10,78%
elpais_espana_tweets	11,29%
FulgencioColl_tweets	11,61%
santi_abascal_tweets	11,67%
sanchezcastejon_tweets	12,18%
pablocasado_tweets	12,43%
ume_umes_tweets	12,53%
PabloIglesias_tweets	13,58%
Pablo Iglesias - La sociedad española no a	17,14%
La Monarquía, cinco razones por la que e	21,74%
CasaReal_tweets	31,14%
52-medidas-autonomias VOX	78,21%

Ilustración 82 Evaluación mediante clasificador BERT de diferentes entidades políticas en el eje 4 (Organización Territorial del Estado)

La siguiente tabla muestra la lista ordenada de polaridades netas (desde el extremo negativo correspondiente a la Monarquía hasta el extremo positivo correspondiente a la República):

Eje 5 (Monarquía-República)	
Entidad Política	PN2
CasaReal_tweets	-43,45%
ume_umes_tweets	-35,82%
La Monarquía, cinco razones por la que es	-23,91%
marca_tweets	0,34%
raul_alvarez_pr_tweets	0,87%
rielcano_tweets	5,98%
52-medidas-autonomias VOX	6,41%
RubenSanchezTW_tweets	7,68%
mikelayestaran_tweets	10,34%
PabloEchenique_tweets	11,27%
FulgencioColl_tweets	14,27%
elnacionalcat_e_tweets	15,75%
ramoncotarelo_tweets	16,69%
podemos_tweets	18,69%
IzRepublicana_tweets	18,97%
publico_es_tweets	19,02%
sanchezcastejon_tweets	20,52%
eldiarios_tweets	20,74%
vox_es_tweets	20,92%
MonederoJC_tweets	21,45%
fedelasjons_tweets	21,53%
PabloIglesias_tweets	22,55%
IzquierdaUnida_tweets	22,72%
psoe_tweets	25,30%
santi_abascal_tweets	25,33%
Pablo Iglesias - La sociedad española no a	25,71%
webpcpe_tweets	26,44%
CiudadanosCs_tweets	28,74%
CCivicaCatalana_tweets	29,70%
populares_tweets	30,55%
elpais_espana_tweets	32,00%
InesArrimadas_tweets	33,64%
pablocasado_tweets	35,73%
abcespana_tweets	36,62%
Pablo Iglesias y el proyecto de República d	51,28%
2017_10_23_carta_consideraciones_Cat	69,23%

Ilustración 83 Evaluación mediante clasificador BERT de diferentes entidades políticas en el eje 5 (Modelo Político del Estado)

5.4 Evaluación de resultados

Evaluaremos la calidad del sistema final desde los puntos de vista correspondientes a las siguientes ciencias:

1. Ciencia Informática (Procesamiento de Lenguaje Natural).
2. Ciencia Política.

Desde el punto de vista de la Ciencia Informática la evaluación debería estar basada en comparar los resultados de nuestra solución de procesamiento de lenguaje natural con los resultados de otras arquitecturas a nivel tecnológico tomando una de ellas como “**baseline**”.

Desde el punto de vista de la Ciencia Política la evaluación debe estar basada en comparar los resultados de nuestro sistema con un “**Gold standard**” o conjunto de resultados que sirve como criterio externo aprobado. En nuestro caso, este gold standard está definido a partir de datos acerca de *cómo las formaciones políticas se ubican ellas mismas en los dos ejes bajo estudio*. Las entidades bajo estudio han sido los principales partidos y líderes políticos de España (PP, PSOE, VOX, PODEMOS, Ciudadanos, IU, etc.) así como diferentes personalidades e instituciones de interés en cada eje como por ejemplo la Casa Real en el eje 4 (Modelo político de estado).

5.4.1 Análisis de resultados desde la Ciencia Informática

Para este análisis, debemos recordar que la evaluación del clasificador basado en BERT ya fue presentada en el apartado 5.2.5 (Test y evaluación de la solución mediante el coeficiente de correlación de Mathews). En este apartado vamos a realizar una comparativa de resultados para diferentes arquitecturas de la solución de procesamiento de lenguaje natural como forma de **conocer la mejora que una solución BERT nos ofrece respecto a otras soluciones** utilizadas en este campo. En nuestro caso hemos decidido presentar los datos presentados en un caso similar llevado a cabo por (Ma, 2019) que utiliza soluciones LSTM o redes neuronales profundas en dicha comparativa:

1. La arquitectura tomada como “**baseline**” por este autor ha sido una red neuronal **LSTM bidireccional** de capa simple que utiliza embeddings GloVe (200 dimensiones) preentrenados a partir de datos de Twitter.
2. **BERT + etapa lineal (BERT para clasificación de secuencias)**, utilizado como **modelo BERT por defecto**. La salida de BERT (estado oculto final CLS de la primera palabra) está conectada a una etapa completamente conectada (solución lineal).
3. **BERT + etapa no lineal**. La salida de BERT (estado oculto final CLS de la primera palabra) está conectada a una red neuronal de 3 capas utilizando una función de activación “leaky RELU” que introduce la no-linealidad.
4. **BERT + LSTM**. Se añade una etapa bidireccional LSTM a la salida de BERT.
5. **BERT + CNN**. Se introduce una red neuronal convolucional que toma los estados ocultos de todas las capas BERT.

Como se presenta en dicho trabajo, **todos los modelos BERT presentan mejoras respecto al sistema base LSTM** en términos de exactitud, coeficiente de Matthews y F1. **Los tres sistemas BERT evaluados (etapa adicional no lineal, CNN y LSTM) no proporcionan mejoras evidentes en términos de exactitud sobre el sistema BERT por defecto (BERT+ etapa lineal)** y suelen empeorar otras métricas de evaluación (F1, recall) como se presenta en la siguiente tabla:

Table 4: Evaluation metrics

Model	Accuracy	Matthews coef	Macro precision	Macro recall	Macro F-1
baseline	0.64	0.56	58.00	68.43	60.71
default BERT	0.67	0.59	60.43	71.14	64.00
BERT+NL	0.67	0.59	60.57	68.00	63.14
BERT+LSTM	0.67	0.60	61.29	69.86	64.00
BERT+CNN	0.67	0.59	60.86	69.29	63.43

Ilustración 84 Métricas de evaluación utilizados por (Ma, 2019).

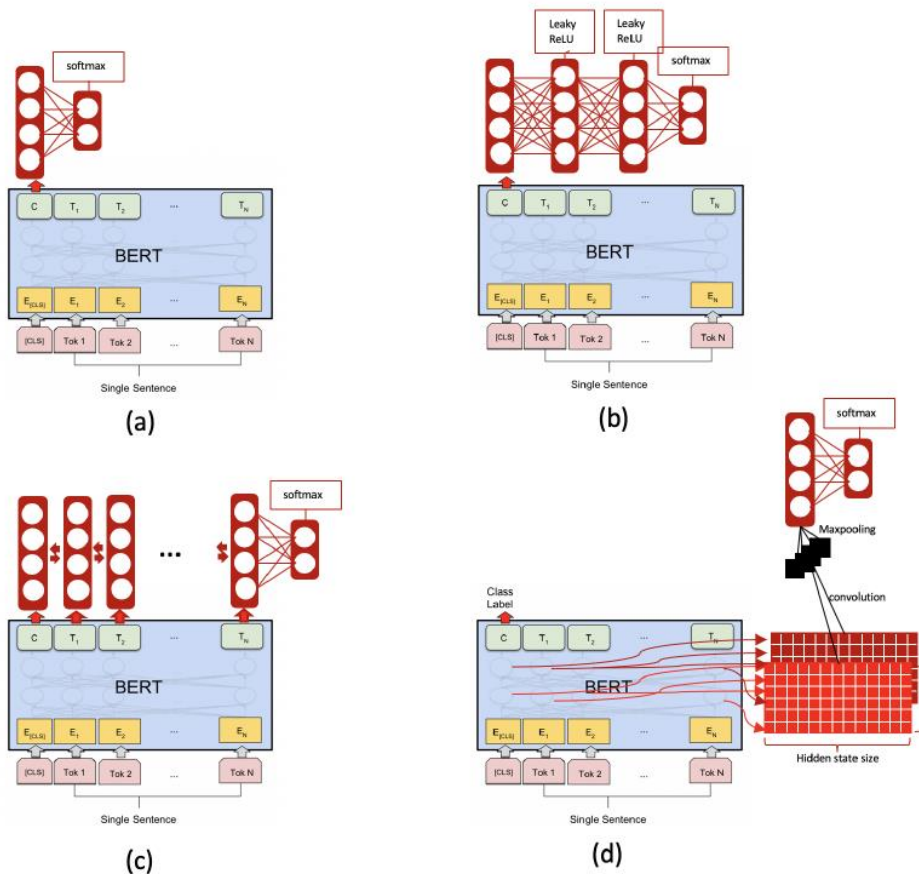


Ilustración 85 Modelos BERT utilizados por (Ma, 2019).

Este autor disponía de un conjunto de datos etiquetados representativos del estudio a realizar (tuits para la gestión de crisis) como CrisisLexT26 y CrisisNLP que ofrecen información sobre la situación local durante un desastre. Este autor ha compilado los conjuntos de datos etiquetados en uno único con un **total de 74.346 tuits etiquetados**, utilizando **64.346 de ellos en la fase de entrenamiento**, 5.000 tuits etiquetados como conjunto de datos de validación y finalmente otros **5.000 tuits etiquetados como datos de test** (test dataset). Desgraciadamente en nuestro caso **no disponemos de datasets públicos** para la realización de nuestro estudio comparativo al ser un **campo de estudio** (comparativa de dos dimensiones políticas muy concretas en España) **no investigado hasta ahora según nuestro conocimiento**.

No hemos creído interesante utilizar como baseline la red simple LSTM (Ma, 2019) debido a la necesidad de creación de todos los embeddings GloVe y la dificultad que creemos que hay en la comparativa respecto a la solución BERT. Hemos preferido utilizar como “baseline” una solución SVM para comparar con nuestra solución BERT para clasificación (BERT + etapa lineal) utilizado por la Universidad de Nueva York (Politics, 2020). En nuestro caso, hemos

decidido tomar como baseline un modelo SVM. Esto nos permitirá añadir resultados adicionales al estudio anteriormente mencionado.

- Arquitectura tomada como “**baseline**”: Máquinas de vectores de soporte (Support Vector Machine, SVM) e TF-IDF para representar los tuits.
- Arquitectura de nuestro trabajo: **BERT + etapa lineal**.

Modelo	Accuracy	Matthews (MCC)
SVM (baseline)	97,49%	0,3061
BERT + Etapa Lineal	99,47%	0,8869

Es importante notar que no solo el nivel de precisión (accuracy) en el caso **BERT mejora en casi un 2% a la solución basada en SVM**, si no que **el valor del coeficiente de Matthews es notablemente mejor para BERT**. Ello es debido a que la solución SVM no ha sido capaz de distinguir correctamente los marcos “Monarquía” y “Republica”, tal como se presenta en la siguiente tabla:

Modelo	precision	recall	f1-score
España Monárquica	1.00	0.22	0.36
España Republicana	0.00	0.00	0.00
No Marco	0.97	1.00	0.99

Es en este tipo de problemas en los que una clase está desproporcionadamente subrepresentada (o sobrerrepresentada como es el caso de “No Marco”) en los que la utilización de MCC nos ofrece una visión de que tipo de solución es la más precisa y adecuada. Recordemos que un valor de MCC valor alto (cercano a 1) significa que **todas las clases** se predicen bien como es el caso de la solución BERT utilizada en nuestro trabajo.

5.4.2 Análisis de resultados desde la Ciencia Política

Vamos a realizar en este apartado una aproximación a los resultados desde una visión propia de la Ciencia Política. *Creemos que es interesante aportar este punto de vista de forma que el lector pueda valorar el conocimiento y capacidades añadidas que un análisis basado en el procesamiento del lenguaje natural (PLN) puede aportar a las Ciencias Sociales y en particular a las Ciencias Políticas.* Pretendemos que los politólogos puedan disponer de nuevas herramientas para investigar la realidad social y política que nos rodea de una forma novedosa. Es decir, vamos a intentar relacionar lo “sociológico” y “político” con la ciencia informática y el procesamiento del lenguaje natural allá donde sea posible. Para ello en los apartados 5.4.2.1 y 5.4.2.25.4.2.2 **presentaremos información acerca de cómo las formaciones políticas se ubican ellos mismos en los dos ejes bajo estudio.** Buscando una similitud con NLP, **dichos políticos se “autoanotan” o “autoetiquetan”** para que así podamos comparar con los resultados de nuestro sistema BERT. Así **podremos apoyar en resultados, como por ejemplo la potencial correlación entre partidos políticos en ambos ejes estudiados, la utilidad que las técnicas de NLP tienen para las Ciencias Sociales en general y para la Ciencia Política en particular.** Los textos y cuestionarios de referencia donde los partidos políticos se “auto etiquetan” se adjuntan en el Anexo III. Un estudio más detallado y complejo de los datos a incluir en este gold standard podrían ser parte del trabajo a realizar en un Doctorado posterior ya que un estudio detallado del interés e importancia que los partidos otorgan a cada marco (lo que probablemente este íntimamente ligada a su ideología) va más allá de los objetivos de este trabajo por lo que aquí solo hemos realizado una somera exploración con datos proporcionados por un único anotador/evaluador humano. Presentamos a continuación los resultados detallados de polaridades para los partidos políticos bajo estudio:

Entidad Política	Eje 4 (Plurinacional-Uninacional)			Eje 5 (Monarquía-Republica)		
	PB1	PR1	PN1	PB2	PR2	PN2
VOX	247	163,17%	7,61%	679	1128,79%	20,92%
PSOE	-114	80,81%	-3,51%	822	2517,65%	25,30%
PP	-83	87,09%	-2,55%	993	1282,14%	30,55%
PODEMOS	208	149,76%	6,42%	606	1831,43%	18,69%
Cs	-19	97,31%	-0,58%	934	2624,32%	28,74%
IU	344	179,81%	10,59%	738	1704,35%	22,72%

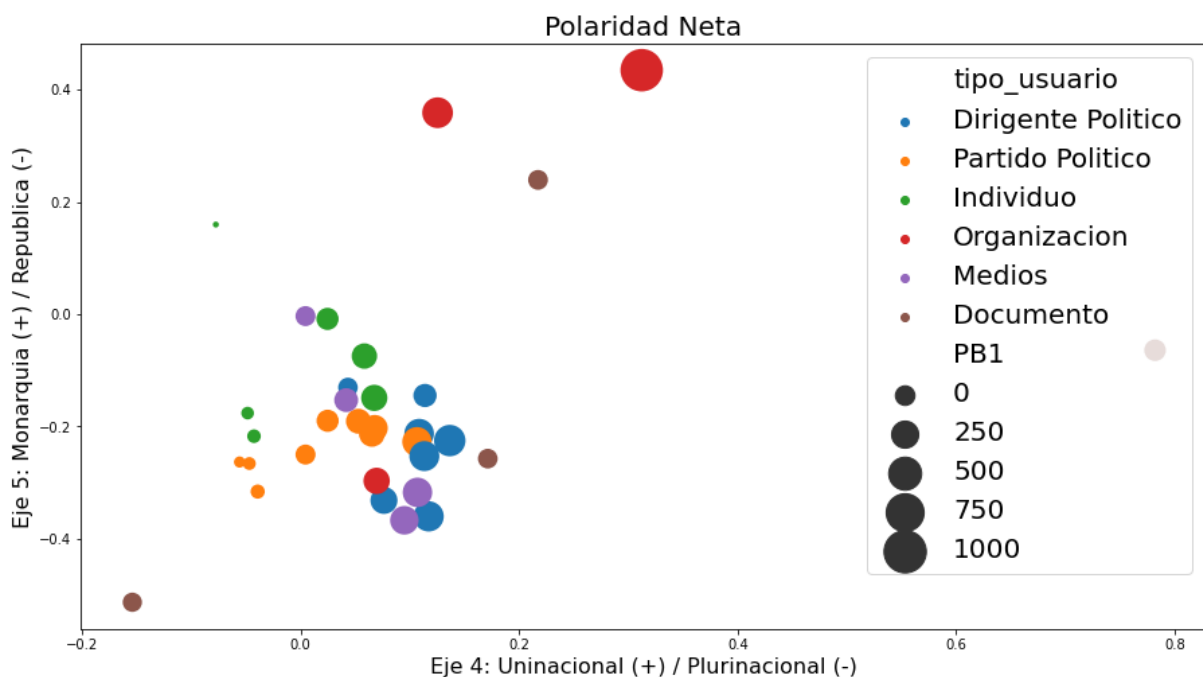


Ilustración 86 Ubicación de usuarios según Polaridad Neta (PN) en ejes 4 y 5 por tipo de usuario (tamaño del círculo = Polaridad Bruta, ejes invertidos)

Si medimos el **grado de correlación** de los usuarios bajo estudio (según su tipología) en lo que a las mediciones para los dos ejes se refiere utilizando para ello el coeficiente de correlación de Pearson:

- **Perfecto:** si el valor está cerca de ± 1 se dice que es una correlación perfecta: a medida que una variable aumenta, la otra también tiende a aumentar (si es positiva) o disminuir (si es negativa).
- **Grado alto:** si el valor del coeficiente se encuentra entre $\pm 0,50$ y ± 1 , se dice que hay una correlación fuerte.
- **Grado moderado:** si el valor se encuentra entre $\pm 0,30$ y $\pm 0,49$ se dice que es una correlación media.
- **Grado bajo:** cuando el valor está por debajo de $\pm 0,29$, se dice que es una correlación pequeña.

Como hemos comentado anteriormente, hemos tomado como base para este estudio las **medidas de Polaridad Neta (rango -1 a +1 en el binomio de marcos bajo estudio)**. Los dos **usuarios de control** que hemos tomado (diario Marca y el usuario de Twitter raul_alvarez_pr que tuitea mayoritariamente en lengua inglesa), están ubicados en el área central de los ejes.

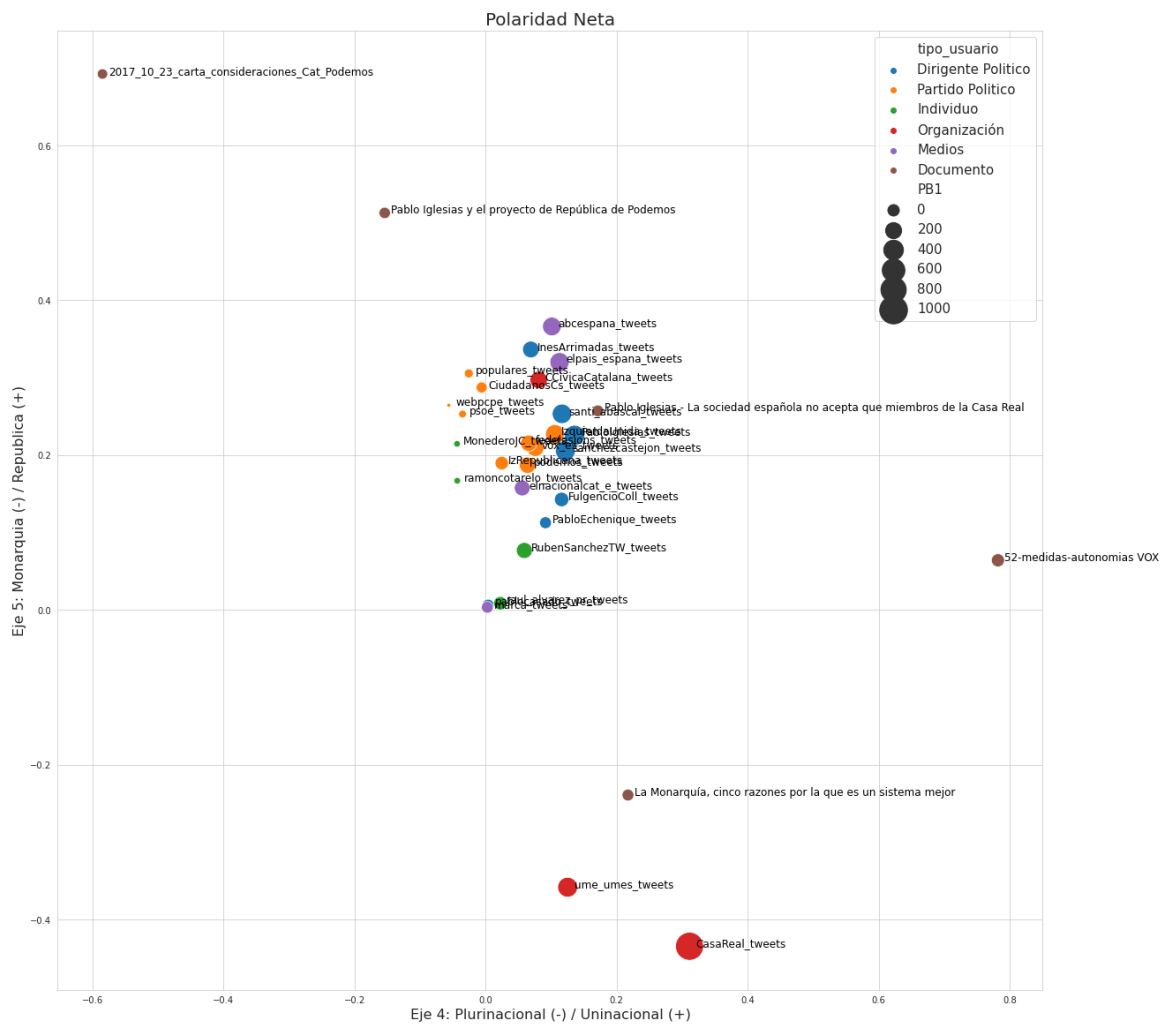


Ilustración 87 Ubicación de entidades políticas en espacio vectorial de marcos

Si realizamos un sencillo estudio de correlación entre los resultados en ambos ejes tenemos:

Tipo		PN1	PN2
Dirigente Politico	PN1	1	-0.093238
	PN2	-0.093238	1
Documento	PN1	1	-0.747290
	PN2	-0.747290	1
Individuo	PN1	1	-0.788157
	PN2	-0.788157	1
Medios	PN1	1	0.858599
	PN2	0.858599	1
Organización	PN1	1	-0.745778
	PN2	-0.745778	1
Partido Politico	PN1	1	-0.685805
	PN2	-0.685805	1

Tabla 4 Correlación de resultados entre ambos ejes

Observamos que **la correlación entre los resultados para ambos ejes es alta (0,68) para el conjunto de todos los partidos políticos** evaluados (para un subconjunto formado por PSOE, PP, PODEMOS, Vox, IU y Ciudadanos esta correlación sube al 0,76). Esto podría interpretarse (con todas las salvedades) como que **los partidos políticos presentan una correlación (lineal) en sus opiniones sobre la organización territorial del estado y el modelo político del estado**. En nuestro caso, *cuantas más menciones a la “república” haga un partido, más menciones a un estado plurinacional hace también ese partido, y cuantas más menciones a la monarquía, más menciones a un estado uninacional por parte de los partidos políticos.*

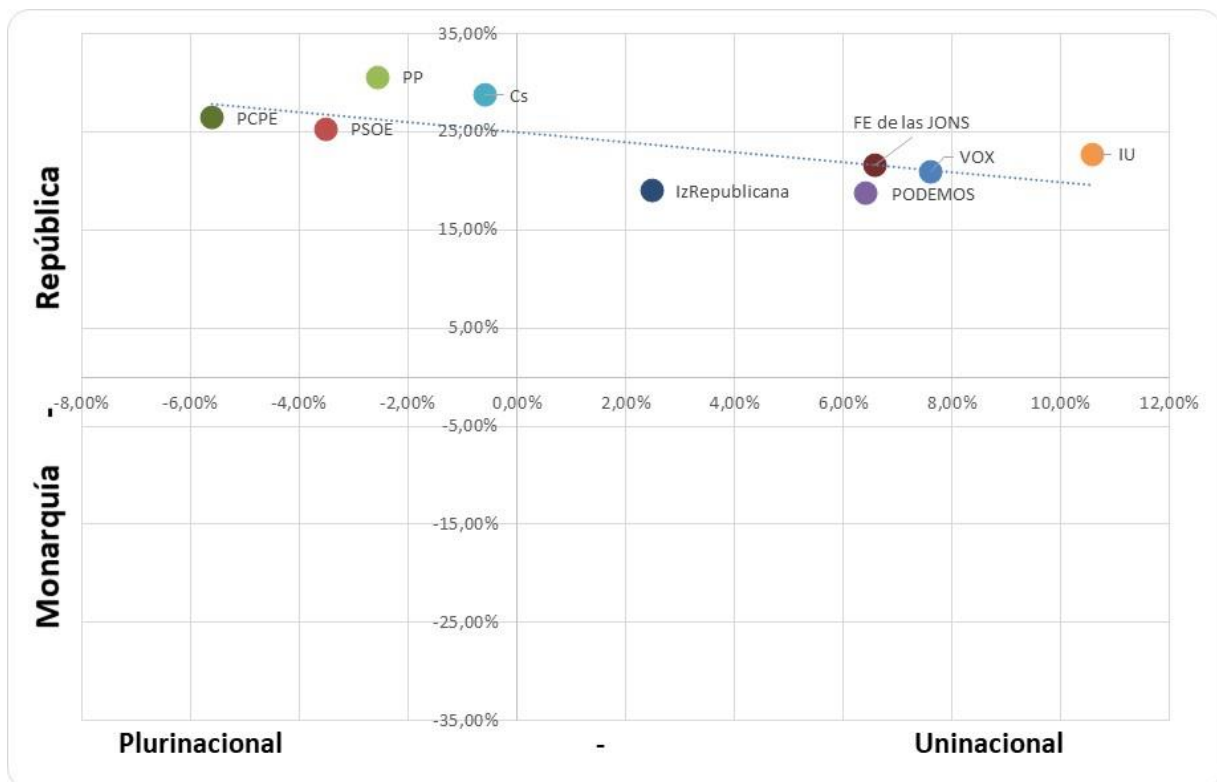


Ilustración 88 Ubicación de partidos políticos en el cuadrante definido por los dos ejes bajo estudio

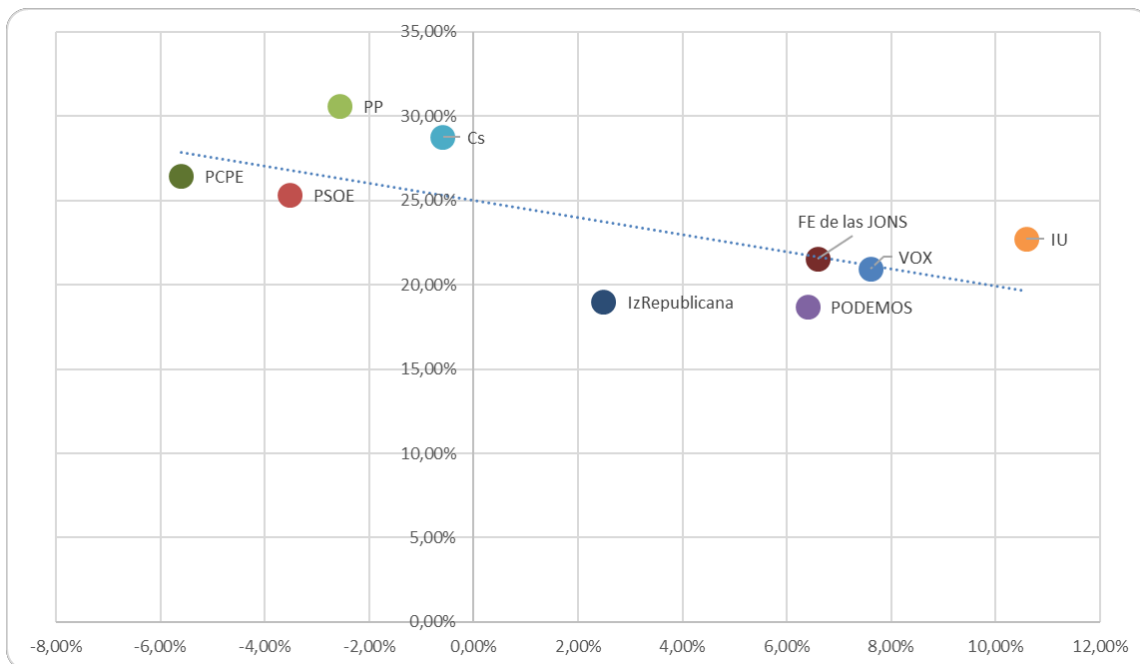


Ilustración 89 Detalle de la ubicación de partidos políticos en ejes

Se muestran a continuación los valores de polaridad neta para estos partidos políticos en ambos ejes.

Eje 4 (Plurinacional-Uninacional)	
Entidad Política	PN1
PCPE	-5,60%
PSOE	-3,51%
PP	-2,55%
Cs	-0,58%
IzRepublicana	2,49%
PODEMOS	6,42%
FE de las JONS	6,60%
VOX	7,61%
IU	10,59%

Ilustración 90 Polaridad Neta en el Eje 1 para partidos políticos

Eje 5 (Monarquía-República)	
Entidad Política	PN2
PODEMOS	18,69%
IzRepublicana	18,97%
VOX	20,92%
FE de las JONS	21,53%
IU	22,72%
PSOE	25,30%
PCPE	26,44%
Cs	28,74%
PP	30,55%

Ilustración 91 Polaridad Neta en el Eje 2 para partidos políticos

Podemos imaginar como este tipo de herramientas pueden utilizarse para la detección de necesidades, la *planificación de objetivos y estrategias de partidos políticos y administraciones públicas*, así como herramienta de valoración de la opinión pública ante un argumento o propuesta política. Vamos a realizar un breve análisis de los resultados de este proyecto para los marcos y ejes políticos bajo estudio. Como hemos comentado anteriormente, hemos definido en este proyecto un dataset básico centrado en los marcos (República, Monarquía, Estado plurinacional, Estado uninacional) y ejes bajo estudio (organización territorial y política de España) para la evaluación de los resultados proporcionados por el sistema BERT. Vamos ahora a analizar algunos de los aspectos que creemos más relevantes de los resultados presentados hasta ahora.

En cuanto a la comparativa entre los resultados de nuestro sistema con datos propios de la Ciencia Política (Gold Standard), al ser difícil encontrar un conjunto de datos estándar para el campo de estudio, es decir en lo que a ideología y similitud política se refiere centrada en los marcos bajo estudio (organización territorial y política de España), hemos definido uno para este proyecto. Tómese en cuenta que dado el alcance de este proyecto, no se ha podido realizar ningún tipo de encuesta ni estudio cualitativo referente a los objetos de estudio. Este “dataset” en un caso ideal debería incluir el **conjunto de entidades políticas** (personas, partidos políticos, medios, etc.) con al menos 10 juicios de personas/expertos políticos acerca de su ubicación en los ejes bajo estudio.

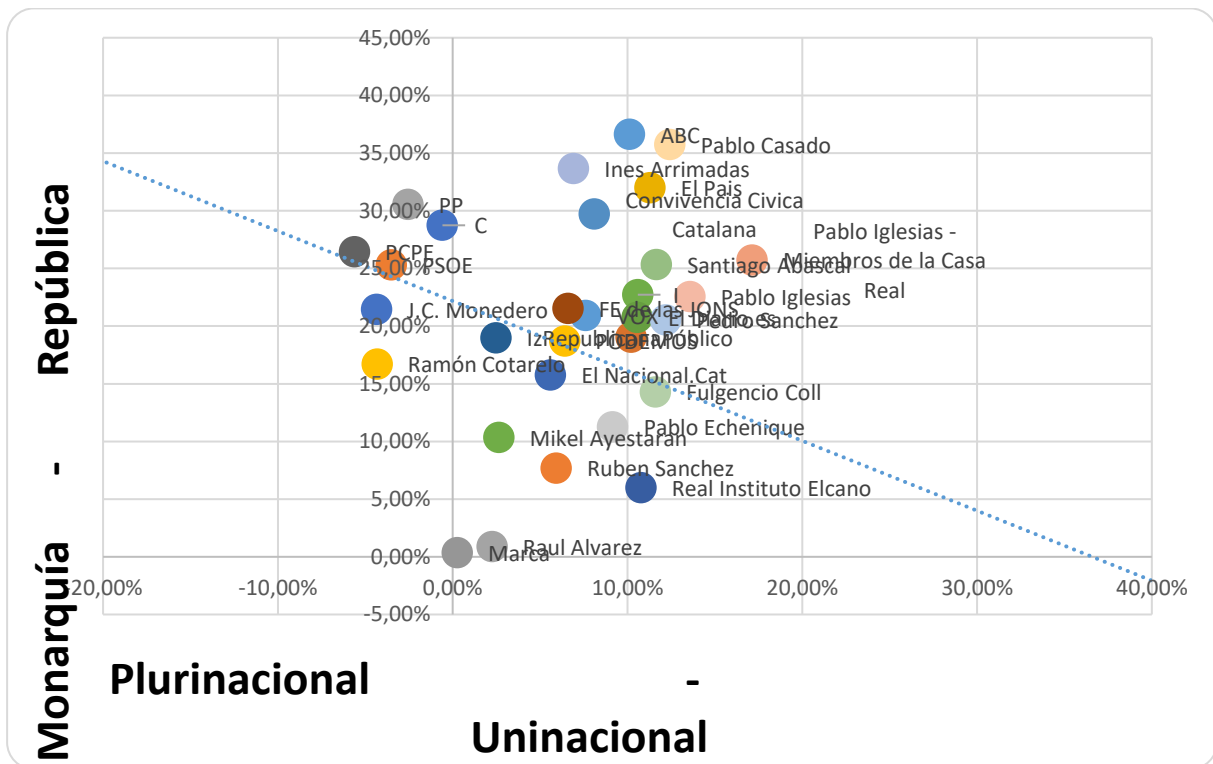


Ilustración 92 Ubicación de entidades en el cuadrante Republicano-Uninacional

Para evaluar la calidad de un sistema de evaluación política como el nuestro una opción sería *calcular la correlación entre las puntuaciones encontradas a partir de nuestro sistema y aquellas encontradas por los juicios humanos en referencia a las entidades políticas y marcos bajo estudio*. El evaluador humano evalúa los textos y proporciona la ubicación en los ejes “organización política del estado” y “organización territorial del estado” a partir de los mismos textos evaluados por el clasificador BERT o de otros de interés como por ejemplo programas electorales o de posición acerca de los aspectos bajo estudio. En el caso de disponer de un equipo multidisciplinar que incluya politólogos, sociólogos, etc. podríamos utilizar sus valoraciones para realizar dicha comparación. Otra opción, común en tareas simples como el etiquetado de imágenes y búsqueda de objetos en las mismas) será utilizar servicios como “[Amazon Mechanical Turk](#)”, plataformas de crowdsourcing que nos permiten realizar tareas repetitivas y en grandes volúmenes mediante la contratación de etiquetadores humanos.

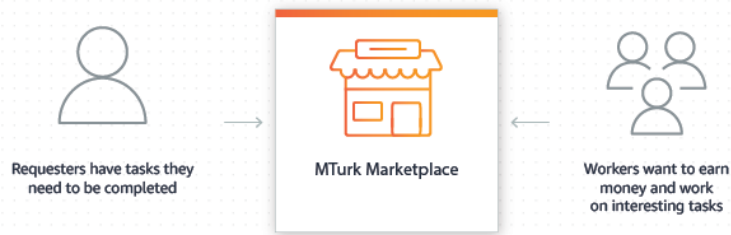


Ilustración 93 Servicio Amazon Mechanical Turk". Fuente: Amazon

En un proyecto como el que se ha desarrollado en este TFM, la forma que creemos óptima sería la **utilización de un equipo multidisciplinar de etiquetadores y especialistas** en lugar de etiquetadores no especializados como los de “Amazon Mechanical Turk”. No obstante y dado que ambas opciones deben procesar miles/millones de tuits y textos políticos, cada proyecto específico debe plantearse la opción más conveniente y realizable en la práctica tomando en cuenta la calidad, coste y plazo requerido.

En este trabajo hemos utilizado **7.624 textos etiquetados** (tuits, frases) en total, de ellos un **80% (6.099 textos) para el entrenamiento del clasificador y el 20% restante (1.525 textos) para la validación del sistema**. La mayoría de estos textos (97% aproximadamente) son textos etiquetados como “No marco” de forma que podamos tener unas condiciones en las que las referencias a un marco dado sean escasas respecto a todos los mensajes que una persona o entidad emite a lo largo de un periodo de tiempo, caso que estimamos más cercano a la realidad en lo que a comunicación política se refiere.

Posteriormente **hemos utilizado alrededor de 3.000 tuits por entidad política en la fase de test** (datos no etiquetados que queremos evaluar en un sistema en producción) donde se ubican las entidades políticas a lo largo de los dos ejes. Para la valoración del “Gold Standard” en este proyecto estimamos que **el equipo de etiquetadores debería etiquetar del orden de 36.000 textos** (3.000 x 6 partidos políticos x 2 ejes bajo estudio). De esta forma podríamos comparar los resultados de nuestro sistema con los procedentes de un criterio externo aprobado.

Dado que este trabajo también excede los objetivos y recursos a destinar, en lugar de realizar una etiquetación de dichos 18.000 tuits (cada uno a etiquetas en los dos ejes con lo que el trabajo de etiquetado es el doble) hemos decidido evaluar los resultados finales (ubicación en ambos ejes) en sí mismos. **Hemos utilizado para dicha evaluación únicamente las encuestas y textos**

presentados en el Anexo III para ubicar a los 6 partidos políticos en los dos ejes y desafortunadamente sólo a través de las opiniones de un único evaluador experto (el autor de este TFM). Entendiendo que esta es una aproximación burda, podemos utilizarla para comparar los resultados procesados finales de nuestro sistema (la ubicación de los 5 partidos políticos sobre los dos ejes) con los resultados de una evaluación humana.

5.4.2.1 Análisis de resultados para el eje 4 “Organización Territorial del Estado”

Las puntuaciones del evaluador humano (Gold Standard en la siguiente tabla) están basadas en el rango [0,10] definido por:

- Estado Plurinacional: 0
- Estado Uninacional: 10

Es decir, el evaluador humano a ubicado a cada partido político de acuerdo a las opiniones presentadas en el Anexo II y a su conocimiento de la ideología política de dichos partidos.

Presentamos a continuación la tabla donde se presenta una comparativa de resultados para los principales partidos políticos de España. En ella se muestran los resultados de Polaridad Neta de nuestro sistema BERT, los resultados de la evaluación humana (Gold Standard) así como la correlación entre ambos.

Eje 4: Organización Territorial del Estado (Plurinacional-Uninacional)			
Entidad Política	Polaridad Neta 1	Gold Standard	Correlación
PODEMOS	6,42%	2	-27,08%
PP	-2,55%	7	
PSOE	-3,51%	4	
IU	10,59%	2	
VOX	7,61%	9	
Cs	-0,58%	6	

De estos resultados observamos una **correlación moderada** o media. Es interesante hacer notar que **nuestro sistema ha evaluado a algunas formaciones como Podemos e IU como más cercanas al extremo “Uninacional” de esta dimensión junto a VOX**, y a formaciones como PP, PSOE y Ciudadanos más escoradas hacia al extremo Plurinacional. Una posible explicación por la que algunos partidos como **PP, PSOE y Cs emitan (moderadamente) el marco “Estado Plurinacional”**, sería la de proyectar una visión de “neutralidad” o “moderación” para alejarse de la visión que muchos votantes puedan tener acerca de su partido como un partido de corte “centralista” o “uninacional” (tal como nuestro evaluador humano ha ubicado

a estos partidos), sobre todo a la hora de captar votos en regiones más proclives a una visión plurinacional de España como puedan ser Cataluña y el País Vasco.

5.4.2.2 Análisis de resultados para el eje 5 “Modelo Político del Estado”

Las puntuaciones del evaluador humano (Gold Standard en la siguiente tabla) están basadas en el rango [0,10] definido por:

- Monarquía: 0
- República: 10

Presentamos a continuación la tabla donde se presenta una comparativa de resultados para los principales partidos políticos de España. En ella se muestran los resultados de Polaridad Neta de nuestro sistema BERT, los resultados de la evaluación humana (Gold Standard) así como la correlación entre ambos.

Eje 5: Modelo Político del Estado (Monarquía-República)			
Entidad Política	Polaridad Neta 2	Gold Standard	Correlación
PODEMOS	18,69%	8	-74,55%
PP	30,55%	2	
PSOE	25,30%	5	
IU	22,72%	9	
VOX	20,92%	4	
Cs	28,74%	3	

De estos resultados observamos una **correlación alta** o fuerte. Es interesante hacer notar que *aunque nuestro sistema ha ubicado a todas las formaciones en cuadrantes “Republicanos”, ha evaluado también a VOX como más cercana a Podemos e IU que al PP, PSOE y Cs en la ubicación a lo largo de este eje*. Estas tres formaciones son las que en menciones netas hacen al marco “Monarquía” (a pesar de encontrarse todas las formaciones en cuadrantes “Republicanos” como se ha comentado anteriormente).

6. Discusión y Conclusiones

Este apartado resume el resultado de la ejecución del proyecto de fin de Máster. Como hemos presentado en los anteriores apartados, **hemos desarrollado una teoría y las herramientas necesarias para da respuesta a las preguntas de investigación** presentadas al inicio de este trabajo:

1. *¿Es posible avanzar en la teoría de marcos a partir del uso de técnicas de procesamiento de lenguaje natural?*
2. *¿Son de utilidad las técnicas de procesamiento del lenguaje natural basadas en Transformers para una tarea de detección de marcos en una red de microblogging?*

Siendo las respuestas a las anteriores preguntas las siguientes:

1. **El uso de técnicas de NLP nos ha permitido avanzar en la teoría de marcos, llevando a la práctica una herramienta automatizada que identifica marcos y evalúa textos en función de un conjunto de marcos prediseñados.** Creemos que en particular nos puede permitir avanzar (aunque sea mínimamente) hacia la Inteligencia Artificial Fuerte mediante la utilización de conceptos propios del nivel pragmático de la lingüística (marcos) sobre la información que soluciones como BERT son capaces de captar en los niveles inferiores (niveles sintáctico y semántico).
2. **La utilización de soluciones NLP basadas en Transformers, nos ha permitido ofrecer soluciones innovadoras a los investigadores en el campo de la Ciencia Política.** A partir de los resultados de nuestro trabajo, hemos podido cuantificar el grado de mejora que una solución BERT ofrece respecto a otras soluciones utilizadas como SVM.

Un aspecto interesante es que en este proyecto nos hemos centrado en **encontrar los marcos que una entidad política “proyecta” en sus textos hacia las audiencias objetivo**, es decir, si se centran en proyectar el marco “monarquía” o en el marco “republica”, **pero sin entrar a valorar si las menciones a dicho marco se realizan de una forma positiva o negativa**. Esto podría ser llevado a cabo añadiendo una segunda fase de clasificación mediante Transformers una vez detectados los marcos (fuera del alcance de este proyecto). Por ejemplo, hay partidos que probablemente proyectan el marco “Republica” como una forma de “motivar” a sus

seguidores a los que mayoritariamente no agrada dicho concepto. Lo mismo podría decirse de dirigentes de partidos que emiten mayoritariamente el marco “Estado Plurinacional” como una forma de proyectar una visión de “neutralidad” o “moderación” para alejarse de la visión que muchos votantes puedan tener acerca de su partido como un partido de corte “centralista” (véase en la siguiente figura PP y PSOE) o “uninacional” con el fin de captar votos en regiones más proclives a una visión plurinacional de España como puedan ser Cataluña y el País Vasco.

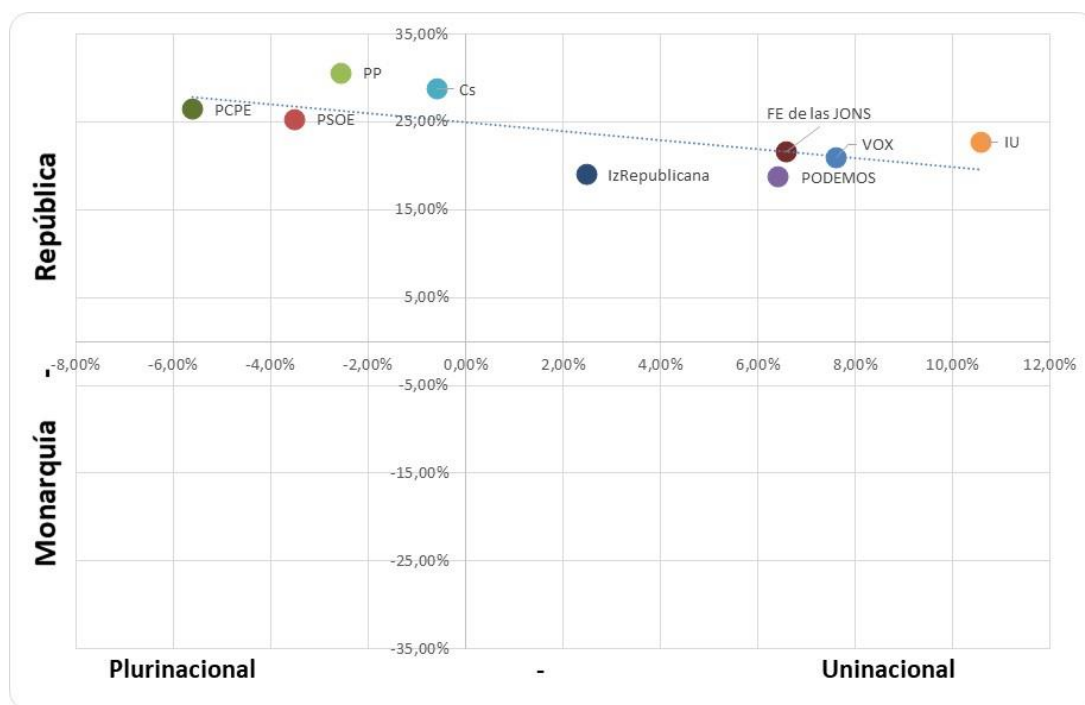


Ilustración 94 Ubicación de partidos políticos en el cuadrante definido por los dos ejes bajo estudio

7. Conocimientos adquiridos, limitaciones del trabajo y futuras líneas de investigación

A partir de la representación de documentos e información extraída de la red social Twitter, hemos extraído información política para:

- Ubicar a una formación política dentro del espacio vectorial multidimensional (y no bidimensional) y compararla con otras formaciones políticas.
- Ubicar a una entidad política (mejor dicho, las opiniones vertidas por esa entidad en la red social Twitter) dentro del espacio vectorial y ubicarla dentro de un clúster de personas con ideas políticas similares.

En cuanto a las limitaciones de este trabajo podemos mencionar:

1. Es **difícil definir qué dimensiones identifican a una persona** (o personalidad, o mente humana) dentro de todas las potenciales dimensiones (que podrían estar en el orden de millones) que pueden intentar definir una “maquinaria” tan compleja. En este trabajo nos hemos limitado a diseñar un espacio dentro de la dimensión política con un número reducido de dimensiones, para un marco geográfico reducido (España), así como para un marco temporal de duración definida (2021).
2. *Es difícil estudiar y encontrar los marcos que definen un campo de estudio en una cultura concreta* de una forma totalmente aceptada por la comunidad científica (subjetividad), aunque la metodología definida en el apartado 9.5 puede ser una solución inicial a este problema.
3. **Para la creación de un diccionario de marcos ajustados a un área de investigación (política en este caso) y geográfica (España), es necesario recopilar una gran cantidad de datos anotados** para disponer del conjunto de datos de marcos y lenguaje figurado necesarios para cualquier investigación avanzada.
 - a. Este trabajo de generación de un diccionario de marcos puede ser gradual e incremental, de forma que podamos crear un diccionario de marcos para un ámbito (pensamiento político en este caso) en España con un número de dimensiones reducido e ir ampliando gradualmente (a más dimensiones, a

otros entornos geográficos/culturales y para otros campos (como el campo religioso, el deportivo, el de marketing de productos, etc.).

- b. En el caso de este TFM se ha etiquetado un conjunto reducido de datos para crear dicho diccionario de marcos, todos por un único codificador humano. En el caso de una solución comercial masiva sería deseable realizar esta tarea mediante equipos multidisciplinares trabajando el diseño y ajuste de dicho diccionario de marcos siguiendo unas guías (“guidelines”) de codificación definidas con anterioridad.
4. Es recomendable disponer de un conjunto de datos que puedan ser utilizados como **“Gold Standard”** o conjunto de resultados que sirvan como criterio externo aprobado a la hora de comparar diferentes sistemas y arquitecturas NLP. Este “dataset” en un caso ideal debería incluir el conjunto de entidades políticas (personas, partidos políticos, medios, etc.) con al menos 10 juicios de personas/expertos políticos acerca de su ubicación en los ejes bajo estudio.
5. Sería interesante profundizar en la forma de **sobreponderar ciertos elementos de marcos** (lemas o figuras retóricas, por ejemplo) que por su importancia en una cultura lleven una carga cognitiva o emocional adicional que “active” ciertos marcos en la población.
6. **Sería necesario disponer de sistemas preentrenados (como BERT) con grandes datasets de textos en español**, y a ser posible sobre datasets centrados en el área bajo estudio (la política en España en el siglo XXI) como forma de mejorar la detección de marcos y la clasificación de opiniones políticas. Como ejemplo tenemos BERTO (Cañete, Chaperon, Fuentes, & Pérez, 2002). Esto nos permitiría adaptar el modelo preentrenado de BERT a nuestro dominio (técnicas de transfer learning).
7. Es imprescindible incluir en el estudio todas las **consideraciones éticas y legales** pertinentes para este tipo de tecnologías, sobre todo las destinadas a **anonimizar** los datos utilizados y los resultados de este tipo de soluciones, así como para **reducir o eliminar todos los sesgos que pudieran existir en los datos etiquetados de entrada**.

8. Referencias

- Masuch, L. (7 de July de 2015). *Trend detection and analysis on Twitter*. Obtenido de SlideShare: https://www.slideshare.net/LuMa921/trend-detection-and-analysis-on-twitter?from_action=save
- Řehůřek, R. (2019). *Distance metrics*. Obtenido de Gensim: https://radimrehurek.com/gensim/auto_examples/tutorials/run_distance_metrics.html#distance-metrics-for-topic-distributions
- Alvarez Prieto, R. (2018). *Estudio comparativo del mensaje político de PODEMOS mediante la utilización del análisis de marcos*. Madrid.
- Alvarez-Melis, D., & Saveski, M. (2016). Topic Modeling in Twitter: Aggregating Tweets by Conversations. *Proceedings of the Tenth International AAAI Conference on Web and Social Media (ICWSM 2016)*, (págs. 519-522).
- Aydogan, E., & Akcayol, M. (2016). A comprehensive survey for sentiment analysis tasks using machine learning techniques. *International Symposium on INnovations in Intelligent SysTems and Applications (INISTA), 1-7*.
- Bayón , E. (2017). *Cerrando el círculo: Podemos, la articulación de un partido nacional-popular en la España en crisis*.
- Bellei, C. (06 de 01 de 2018). *The backpropagation algorithm for Word2vec*. Obtenido de Marginalia: <http://www.claudiobellei.com/2018/01/06/backprop-word2vec/>
- Benedicto, J. (2015). *Y en esto llegó la crisis». Transformaciones y quiebra de la matriz cultural de la democracia española*.
- Bhadury, A. (8 de February de 2017). *Clustering Similar Stories Using LDA*. Obtenido de <https://engineering.flipboard.com/2017/02/storyclustering>
- Bian, J., Gao, B., & Liu, T.-Y. (2014). Knowledge-Powered Deep Learning for Word Embedding. En Springer (Ed.), *ECML PKDD 2014: Machine Learning and Knowledge Discovery in Databases. Lecture Notes in Computer Science, vol 8724*, (págs. 132-148). doi:https://doi.org/10.1007/978-3-662-44848-9_9
- Blanco, J. (14 de 04 de 2019). *La monarquía asoma en campaña aunque los partidos la eluden en sus programas*. Recuperado el 29 de July de 2021, de La Vanguardia: <https://www.lavanguardia.com/politica/20190414/461639951587/la-monarquia-asoma-en-campana-aunque-los-partidos-la-eluden-en-sus-programas.html>
- Bogage, J. (23 de June de 2016). *Whom are you voting for? This guy can read your mind*. Obtenido de The Washington Post: [133](https://www.washingtonpost.com/news/the-</p></div><div data-bbox=)

switch/wp/2016/06/23/whom-are-you-voting-for-this-guy-can-read-your-mind/?utm_term=.63055bdfc465

- Briggs, J. (21 de March de 2021). *Multi-Class Classification With Transformers*. (T. D. Science, Ed.) Recuperado el 03 de June de 2021, de Preprocess, train, and predict with BERT: <https://towardsdatascience.com/multi-class-classification-with-transformers-6cf7b59a033a>
- Bruinsma, B., & Gemenis, K. (2019). Validating Wordscores: The Promises and Pitfalls of Computational Text Scaling. *Communication Methods and Measures*, *Volume 13*, 2019 - Issue 3, págs. 212-227.
- Cañete, J., Chaperon, G., Fuentes, R., & Pérez, J. (2020). Spanish pre-trained BERT model and evaluation data. *PML4DC, ICLR 2020*. Obtenido de https://pml4dc.github.io/iclr2020/program/pml4dc_10.html
- Casqueiro, J. (06 de Diciembre de 2020). *Los partidos políticos analizan la Constitución y sus posibles reformas*. Recuperado el 29 de July de 2021, de El País: <https://elpais.com/espana/2020-12-05/los-partidos-politicos-analizan-la-constitucion-y-sus-posibles-reformas.html>
- Chadha, A. (2020). Distilled Notes for the Natural Language Processing Specialization on Coursera (offered by deeplearning.ai). Recuperado el 21 de 07 de 2021, de <https://www.aman.ai>
- Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, *21*(6). doi:<https://doi.org/10.1186/s12864-019-6413-7>
- Coll, J. (18 de 06 de 2016). *España federal o plurinacional*. Obtenido de El Periodico: <https://www.elperiodico.com/es/opinion/20160618/espana-federal-o-plurinacional-5215559>
- Cowen, A., & Keltner, D. (Sep de 2017). Varieties of reported emotional experience. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1702247114
- D'Angelo, P., & Kuypers, J. (2010). *Doing News Framing Analysis: Empirical and Theoretical Perspectives*.
- Desagulier, G. (05 de 04 de 2018). *Word embeddings: the (very) basics*. Obtenido de Around the world: A corpus linguist's research notebook: <https://corpling.hypotheses.org/495>
- DL4J. (09 de November de 2020). *Thought Vectors, Deep Learning & the Future of AI*. Obtenido de DeepLearning4J:

- <https://web.archive.org/web/20170211043631/https://deeplearning4j.org/thoughtvectors>
- Egerod, B. C. K., & Klemmensen, R. (2020). Scaling Political Positions from Text: Assumptions, Methods and Pitfalls. *The SAGE Handbook of Research Methods in Political Science and International Relations. Volume 1* (págs. 498-521). SAGE Publications. doi:<https://doi.org/10.4135/9781526486387.n30>
- Falck, F., & Marstaller, J. (2018). Sentiment Political Compass: A Data-driven Analysis of Online Newspapers regarding Political Orientation. *IPP 2018*. Obtenido de <http://blogs.oii.ox.ac.uk/policy/ipp-conference/papers/>
- Fernández de Mosteyrín, L., & Morán, M. (2014). *Encontrar la cultura: estrategias de indagación para el análisis sociopolítico*.
- Glavaš, G., Nanni, F., & Ponzetto, S. (2017). Texts, Unsupervised Cross-Lingual Scaling of Political. *15th Conference of the European Chapter of the Association for Computational Linguistics. Volume 2, Short Papers*, págs. 688–693. Valencia: Association for Computational Linguistics. doi:<https://www.aclweb.org/anthology/E17-2109.pdf>
- Hugging Face . (2020). *Source code for transformers.models.bert.modeling_bert*. Recuperado el 29 de July de 2021, de Hugging Face: https://huggingface.co/transformers/_modules/transformers/models/bert/modeling_bert.html#BertForSequenceClassification
- IBM. (2021). *Watson Personality Insights*. Recuperado el 2021, de IBM: <https://www.ibm.com/es-es/cloud/watson-personality-insights>
- Institute, F. (19 de Enero de 2018). *Overview*. Obtenido de <http://www.frameworksinstitute.org/overview.html>
- Iyyer, M., Enns, P., Boyd-Graber, J., & Resnik, P. (2014). Political Ideology Detection Using Recursive Neural Networks. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. (Volume 1: Long Papers)*, págs. 1113–1122. Association for Computational Linguistics. doi:10.3115/v1/P14-1105
- Jurman, Giuseppe et al. “. (2021). A comparison of MCC and CEN error measures in multi-class prediction. *PloS one*, vol. 7,8 (2012): e41882. doi:<https://doi.org/10.1371/journal.pone.0041882>
- Knott, D., Muers, S., & Aldridge, S. (2008). *Achieving Cultural Change: A Policy Framework*. Cabinet Office.

- Kozlowski, A., Taddy, M., & Evans, J. (2019). The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings. *American Sociological Review*, 84(5), 905-949. doi:<https://doi.org/10.1177/0003122419877135>
- Krul, A., & et al. (2006). Corpus design based on the Kullback-Leibler divergence for Text-To-Speech synthesis application. *INTERSPEECH 2006*, (págs. 2023-2034). ittsburgh, Pennsylvania. Obtenido de <https://www.semanticscholar.org/paper/Corpus-design-based-on-the-kullback-leibler-for-Krul-Damnati/dbd513007fea0368ce0c4f9e1e7e919a2eab9bc0>
- Lakoff, G. (2004). *No pienses en un elefante*. Editorial Complutense.
- Lensu, A. (2002). Computationally intelligent methods for qualitative data analysis.
- Liu, Y., Bi, J.-W., & Fan, Z.-P. (2017). Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic fuzzy set theory. *Information Fusion*, 36, págs. 149-161. doi:<https://doi.org/10.1016/j>.
- Ma, G. (2019). Tweets Classification with BERT in the Field of Disaster Management.
- Martínez Fernández, A.-C. (31 de Diciembre de 2016). *Lakoff, la comunicación y el independentismo*. Obtenido de <https://debate21.es/2016/12/31/lakoff-la-comunicacion-independentismo/>
- Matthes, J. (2012). Framing Politics : An Integrative Approach. *American Behavioral Scientist* 56(3) , 247-259.
- Meijer, P. (7 de January de 2014). *Red dragon or white knight? Analysing Dutch news coverage on Analysing Dutch news coverage on*. University of Leiden, Dutch Language and Culture - Journalism and New Media. Obtenido de <https://openaccess.leidenuniv.nl/bitstream/handle/1887/35003/ScriptiePetraMeijer.pdf?sequence=1>
- Minkov, M. (2011). *Cultural differences in a globalizing world*. Emerald.
- Nicholson, C. (29 de November de 2020). *Pathmind*. Obtenido de Thought Vectors: <https://wiki.pathmind.com/thought-vectors>
- Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (págs. 1532–1543). Doha, Qatar: Association for Computational Linguistics. doi:10.3115/v1/D14-1162
- Politics, T. N. (8 de December de 2020). *Text Classification Using a Transformer-Based Model*. Recuperado el 3 de June de 2021, de Medium: <https://medium.com/the-center-for->

social-media-and-politics/text-classification-using-a-transformer-based-model-1bed832f22fe

- Reagan, A., Danforth, C., Tivnan, B., & et al. (2017). Sentiment analysis methods for understanding large-scale texts: a case for using continuum-scored words and word shift graphs. *EPJ Data Science*, 6(28). doi:<https://doi.org/10.1140/epjds/s13688-017-0121-9>
- Rheault, L., & Cochrane, C. (2020). Word Embeddings for the Analysis of Ideological Placement in Parliamentary Corpora. *Political Analysis*, 28(1), 112-133. doi:<https://doi.org/10.1017/pan.2019.26>
- Rivera, C. (25 de June de 2016). *Could neuroscience explain what Trump voters are thinking?* Obtenido de <https://galindes.wordpress.com/2016/06/25/political-neuroscience/>
- Rivera, C. (17 de January de 2017). *Politics, Emotions, and the Brain*. Recuperado el 17 de Marzo de 2018, de POLITICAL PSYCHOLOGY: <https://galindes.wordpress.com/2017/01/12/politics-emotions-and-the-brain/>
- Rivera, C. (4 de March de 2018). *Neural collates of heightened brain reactivity in some conservatives*. Obtenido de <https://galindes.wordpress.com/2018/03/04/neural-collates-of-heightened-brain-reactivity-in-some-conservatives/>
- Rosset, C. (13 de February de 2020). *Turing-NLG: A 17-billion-parameter language model by Microsoft*. Obtenido de Microsoft: <https://www.microsoft.com/en-us/research/blog/turing-nlg-a-17-billion-parameter-language-model-by-microsoft/>
- Saifee, M. (30 de MAy de 2020). *GPT-3: The New Mighty Language Model from OpenAI*. Obtenido de Towards Data Science: <https://towardsdatascience.com/gpt-3-the-new-mighty-language-model-from-openai-a74ff35346fc>
- Saldaña, J. (2009). *The Coding Manual for Qualitative Researchers*. SAGE.
- Sampson, G. (2000). *The SUSANNE Corpus: Documentation*. Obtenido de <https://www.grsampson.net/SueDoc.html>
- Sancho Caparrini, F. (2 de Marzo de 2020). *Variational AutoEncoder*. Obtenido de Fernando Sancho Caparrini: <http://www.cs.us.es/~fsancho/?e=232>
- Sarkar, D. (2018). *Robust Word2Vec Models with Gensim*. Obtenido de KDnuggets: <https://www.kdnuggets.com/2018/04/robust-word2vec-models-gensim.html>
- Seele, N. (14 de April de 2021). <https://www.linkedin.com/pulse/isaleph-alpha-last-company-why-we-invested-nils-seele/>. Obtenido de LinkedIn: <https://www.linkedin.com/pulse/isaleph-alpha-last-company-why-we-invested-nils-seele/>

- Slapin, J., & Proksch, S.-O. (9 de July de 2008). A scaling model for estimating time-series party positions from texts. (B. P. Inc, Ed.) *American Journal of Political Science*, 52(3), 705-722. doi:<https://doi.org/10.1111/j.1540-5907.2008.00338.x>
- Soricut, R., & Lan, Z. (20 de December de 2019). *ALBERT: A Lite BERT for Self-Supervised Learning of Language Representations*. Obtenido de Google AI: <https://ai.googleblog.com/2019/12/albert-lite-bert-for-self-supervised.html>
- Suau Gomila, G. (2015). *El Fenómeno de Podemos: Origen y mediatización de la formación política*. Tesis - Universidad de las Islas Baleares (CESAG).
- Terechshenko, Z. a. (5 de November de 2020). A Comparison of Methods in Political Science Text Classification: Transfer Learning Language Models for Politics. 25. doi:<https://dx.doi.org/10.2139/ssrn.3724644>
- The decision Lab. (2021). Why do our decisions depend on how options are presented to us? Framing Effect, explained. Recuperado el 28 de 07 de 2021, de <https://thedecisionlab.com/biases/framing-effect/>
- The Political Compass*. (07 de November de 2020). Obtenido de <https://politicalcompass.org/>
- van den Oord, A., Li, Y., & Vinyals, O. (22 de Jan de 2019). Representation Learning with Contrastive Predictive Coding.
- Van Gorp, B. (2005). *Where is the Frame? Victims and Intruders in the Belgian Press Coverage of the Asylum Issue*. European Journal of Communication Copyright - SAGE Publications.
- Van Gorp, B. (2007). The Constructionist Approach to Framing: Bringing Culture Back In. *Journal of Communication* 57 ISSN 0021-9916, 60-78.
- Zhou, L., & Hripcsak, G. (April 2007). Temporal reasoning with medical data—A review with emphasis on medical natural language processing. *Journal of Biomedical Informatics* Volume 40, Issue 2, 183-202.

9. Anexo I, la teoría de Marcos (Framing)

Este anexo está basado principalmente en el Trabajo de Fin de Master (Alvarez Prieto, 2018) titulado “Estudio comparativo del mensaje político de PODEMOS mediante la utilización del análisis de marcos”, un trabajo centrado en la utilización de “marcos cognitivos” dentro de la comunicación política en España donde se diseñó una metodología de búsqueda de marcos en textos políticos y se mostraron los marcos cognitivos utilizados por diversas formaciones políticas en España referidos a dos temas: la **globalización** (y sus perdedores) y el “**Pueblo**” (como sujeto de la soberanía).

9.1 Los marcos y el lenguaje

Es muy importante para nosotros definir claramente desde un punto de vista tanto teórico como práctico *lo que entendemos por marcos (frames) y su relación cercana con el lenguaje*. Gran parte del trabajo desarrollado estará basado en la premisa de que **mediante el estudio del lenguaje** (ya sea tanto en el utilizado en la documentación oficial de las formaciones políticas, como del lenguaje empleado por sus principales dirigentes en las redes sociales), **seremos capaces de llegar** (mediante un proceso inicialmente inductivo y posteriormente deductivo como se verá en apartados posteriores) **a descubrir los principales marcos cognitivos utilizados en la comunicación política por parte de** diferentes formaciones políticas o personas. Para George Lakoff (2004, pág. 4) **los marcos son estructuras mentales que confirman nuestro modo de ver el mundo** y como consecuencia de ello, conforman las metas que nos proponemos, los planes que hacemos y hasta nuestra manera de actuar. En política nuestros marcos conforman nuestras políticas sociales y las instituciones que creamos para llevar a cabo dichas políticas. Cambiar nuestros marcos es cambiar todo esto, **el cambio de marco es cambio social**.

Es interesante la relación que dicho autor encuentra entre los marcos y el lenguaje (y que como posteriormente veremos, será una parte importante en el desarrollo de este trabajo). Para Lakoff, **las palabras activan marcos y conocemos los marcos a través del lenguaje**. Así los marcos formarían parte del “inconsciente cognitivo”, es decir, estructuras en nuestro cerebro a las que no se puede acceder conscientemente pero que conocemos a través de nuestra forma de razonar y se denomina como sentido común.

*También conocemos los marcos a través del lenguaje. Todas las palabras se definen en relación a marcos conceptuales. Cuando se oye una palabra, se activa en el cerebro su marco (o su colección de marcos). **Cambiar de marco es cambiar el modo que tiene la gente de ver el mundo.** Es cambiar lo que se entiende por sentido común. Puesto que el lenguaje activa los marcos, los nuevos marcos requieren un nuevo lenguaje. Pensar de modo diferente requiere hablar de modo diferente.*

Podríamos decir que **las palabras son estructuras que aparecen en la superficie pero que intentan representar estructuras más profundas.** Las palabras sirven para enmarcar nuestras experiencias realzando algunos aspectos de las mismas y dejando otros aspectos en el fondo. Nuestros modelos mentales están contruidos desde el lenguaje y a su vez las palabras nos permiten expresar emociones complejas. Creemos que es por lo tanto posible utilizar palabras tanto para reflejar como para moldear las experiencias mentales de una persona, grupo y sociedad.

Si como indican D'Angelo y Kuypers (2010, pág. 89) por una parte los marcos son parte de una cultura y no puramente individuales, y por el otro, los individuos son necesarios como agentes para realizar la conexión entre un texto y el conjunto de los marcos culturales existentes en una sociedad, entonces ***el conjunto de los marcos culturales no está por encima de la gente, sino “entre” la gente.*** Así cuando una persona (o grupo) “habla” mediante marcos, lo que está haciendo es integrar dichos marcos con sus experiencias y conexiones mentales personales (nótese que no todas estas experiencias personales deben estar necesariamente conectadas y ser consistentes con las manifestaciones externas de los marcos culturales presentes en la sociedad). Es más, se podría argumentar que **los marcos son el vocabulario común que permiten a las elites (políticas) y a los ciudadanos compartir significados y formar parte de una misma conversación.** Los marcos son para Van Gorp (2007, pág. 65) una forma de metacomunicación que enlazan una serie de elementos relacionados en un texto y, lo que es mar importante, que **definen cómo un asunto va a ser definido y entendido por una audiencia.**

9.2 Marcos y Cultura

Para Van Gorp (2007, pág. 61), ***el repertorio compartido de marcos dentro de una cultura nos proporciona el nexo entre la producción de noticias y su consumo.*** La cultura para él es la base primaria que constituye el conocimiento y significado del mundo exterior y por ello es utilísimo **desarrollar la utilidad de los marcos como puente entre la cognición y la cultura, esforzándonos en encontrar cómo los marcos (como parte de la cultura) se introducen e**

imbrican en los contenidos mediáticos, cómo trabajan y cómo interactúan con los esquemas mentales tanto de los periodistas como de las audiencias.

Para ello vamos a iniciar este apartado con una primera introducción de varios autores a lo que definimos como cultura en general, y cultura política en particular. La **definición general de cultura** y que creemos que es importante a la hora de relacionar marcos y cultura aquí utilizada es la presentada por Fernández de Mosteyrín y Morán (2014) que define la cultura como el conjunto de códigos o repertorios públicos comunes que intervienen en la capacidad de las personas de pensar y comunicar ideas. **Las culturas, por tanto, son sistemas simbólicos compartidos por ciertos grupos o incluso por sociedades**, pero, además, son también prácticas semióticas.

Como **definición extendida de cultura política** utilizaremos la presentada por Benedicto (2015) que creemos que es fácil relacionar con el estudio de los marcos en este TFM:

*“Cultura política como el contexto de **significados y de representaciones compartidas en el que se mueve la vida política** de una comunidad determinada y que se expresa/plasma en diferentes formas o estructuras de contenido simbólico (**discursos, mitos, lenguajes, rituales, emociones, etc.**), las cuales proporcionan a los individuos los recursos necesarios e imprescindibles para configurar sus universos políticos y crear sus identidades colectivas.”*

Estos **códigos comunes** manejados por los diferentes actores son los que permiten comprenderse entre sí, incluso (o sobre todo podríamos decir) cuando las representaciones comunes no se comparten o se sienten como propias por alguno de los actores (imaginemos una minoría étnica, lingüística o religiosa dentro de un estado pluricultural donde se supone una cultura hegemónica propia de la mayoría). Para (Benedicto, 2015) los elementos culturales *“ocupan un lugar específico, en tanto en cuanto **establecen la gramática de la acción política**, lo que implica –siguiendo con la metáfora lingüística- establecer las reglas de pertinencia, los límites de lo posible y las cadenas de significación”*.

Las ideologías de los movimientos sociales y formaciones políticas están inmersas y son influenciadas por los **componentes culturales preexistentes**, y a su vez sustentan las estrategias de acción que posibilitan que en los periodos de transformación cultural estas nuevas ideologías triunfen o fracasen. Aquí es donde creemos que el análisis de marcos tiene una función más allá del análisis de las noticias generadas por un actor político, y es donde creemos que el análisis de marcos nos permite entroncar con los estudios culturales más allá del estudio de la propia comunicación política.

Es por lo tanto importante identificar los marcos dominantes dentro del conjunto de marcos culturales que son parte de una cultura, entender cómo persisten y cómo se distinguen de otros

marcos con los que compiten para llegar al receptor, entender cómo los receptores son capaces de conectar los recursos utilizados por un marco con su propia experiencia cultural con la que ya están familiarizados y finalmente, entender cómo aparecen y desaparecen los marcos culturales a través de la interacción social.

Para Van Gorp (D'Angelo & Kuypers, 2010, pág. 91) cada marco se presenta como una estructura integrada de recursos de enmarcado (framing devices) y una cadena lógica de recursos de razonamiento (reasoning devices) utilizados para demostrar cómo funciona el marco a la hora de representar una cierta realidad. Los elementos en un texto que funcionan como indicadores de la existencia de un marco son los denominados “recursos de enmarcado”, y entre ellos los más representativos están:

- las metáforas,
- los casos históricos de los que derivar lecciones,
- los eslóganes o consignas,
- los relatos y
- las imágenes visuales (incluyendo gráficos e infografías).

Por otra parte, los “recursos de razonamiento” (reasoning devices) incorporan la función definitoria de los marcos, evocando un cierto razonamiento causal (justificación, causa y consecuencia) cuando un cierto elemento/evento/asunto es asociado con un marco. Estos recursos de razonamiento no necesitan estar presentes ni formar parte explícita del mensaje mediado. *Es en el momento en el que el receptor interpreta el mensaje cuando dicho receptor realiza la conexión mental entre el texto (o la palabra, o la imagen) y el marco cognitivo para a continuación establecer una relación causal en línea con el recurso de razonamiento (implícito) asociado con dicho marco cognitivo evocado* (D'Angelo & Kuypers, 2010, pág. 91).

Una de las funciones que definen los marcos es por lo tanto indicar cuál es la causa de un problema, que hay que hacer para resolverlo, quien es el responsable de las causas y soluciones, y además incorporar al mismo tiempo un juicio moral. Como vemos, aquí es donde encontramos implícita o explícitamente los “recursos de razonamiento” propios de los marcos. La incorporación de unos “recursos de enmarcado” explícitos (por ejemplo, una metáfora) en una noticia puede promover implícitamente un cierto razonamiento causal en un sentido (o en otro total o parcialmente diferente). Algunos recursos de enmarcado pueden ser tan poderosos que su simple referencia es capaz de activar un esquema mental, y a su vez una cadena causal de “recursos de razonamiento” asociados a un marco. La forma de presentar una noticia acerca del cambio climático puede hacer que los lectores sean más proclives a entender que los culpables de dicho problema/asunto sean los países, las industrias, o incluso los individuos, y sobre todo,

puede hacer que los lectores entiendan que la responsabilidad en la resolución del mismo problema recae en uno u otro actor. En este asunto el recurso de razonamiento es a quien culpar por el cambio climático.

En resumen, para Van Gorp (D'Angelo & Kuypers, pág. 92) el elemento central del análisis de marcos es **encontrar los “recursos de enmarcado” (metáforas, eslóganes, relatos, etc.) y los “recursos de razonamiento” (afirmaciones explícitas o implícitas que tratan sobre las causas, justificaciones y consecuencias en un orden temporal) de forma que podamos condensarlos en un “símbolo” que forma parte de la cultura compartida.** Este símbolo es lo que en el presente trabajo consideramos como un “marco”, elemento compartido que tiene sentido dentro de una cultura, cultura de la que los individuos han “bebido” y cultura que a su vez ha sido generada por el conjunto de los individuos a lo largo de diversas generaciones y etapas en la sociedad. El mismo Van Gorp reconoce (2007, pág. 69) que **cuando el tema cultural constituye la idea central del enmarcado, habrá muy probablemente una base más fuerte para la resonancia entre un texto mediático y los esquemas mentales de los receptores.**

9.3 Marcos cognitivos utilizados en la comunicación política

En lo que a la comunicación (y en particular la comunicación política) se refiere, el **enmarcado o encuadre** (framing) se puede definir como el proceso por el cual **los medios de masas, movimientos u organizaciones políticas o sociales influyen en la percepción de los individuos acerca de los significados atribuidos a ideas o conceptos**, alentando ciertas interpretaciones y desalentando otras. Un encuadre sobre un problema puede cambiar la percepción del público acerca del asunto. Enmarcar es pues, seleccionar algunos aspectos de la realidad percibida para resaltarlos de una forma que se promueva:

1. una definición de un problema,
2. una interpretación causal,
3. una evaluación moral y
4. una recomendación de solución para dicho problema.

No solo se trata de enmarcar correctamente un aspecto o tema de discusión, sino que como Van Gorp (2007, pág. 68) indica existe una parte muy importante de convencer estratégicamente a los medios para que acepten dicho marco. En situaciones específicas, como conferencias de prensa o comunicados del gobierno, ciertos patrocinadores de un marco particular (grupos de interés, spin doctors, anunciantes, y otros), intentan convencer estratégicamente a los medios

para que cubran una situación de acuerdo con “su” marco. Y para ello toman la **decisión estratégica respecto a la manera en la que anunciarán su punto de vista**. Los patrocinadores de un marco están preocupados en dirigir la percepción y la selección de marcos de los periodistas al informar sobre un evento.

En cuanto a las estrategias de comunicación de los partidos políticos y su utilización de frases que activan marcos, Lakoff (2004, pág. 41), para el caso particular de Estados Unidos, nos indica que el ala derecha ha utilizado la **estrategia de repetir continuamente frases que evocan sus marcos y que definen las cuestiones importantes a su manera**. Estas repeticiones llegan a conseguir que su lenguaje parezca normal, que sus marcos parezcan normales y lleguen a ser modos cotidianos de pensar acerca de las cuestiones importantes.

Vemos pues que el lenguaje es una forma de influir y facilitar ciertas interpretaciones de la realidad por parte del público objetivo. Es tal la influencia ejercida, que si nos referimos a la persistencia de los marcos cuando estos son contradichos por la realidad, Lakoff (2004, pág. 59) indica que uno de los descubrimientos fundamentales de la ciencia cognitiva es aquel según el cual **la gente piensa en términos de marcos y de metáforas**, es decir, estructuras conceptuales. Los marcos están en las sinapsis de nuestro cerebro, presentes físicamente bajo la forma de circuitos neuronales. **Cuando los hechos no encajan en los marcos, los marcos se mantienen y los hechos se ignoran**.

Si nos centramos en la forma de imponer unos marcos, Lakoff (2004, pág. 69) apunta al uso de una **“guerra civil cultural”** por parte de algunos partidos políticos. No es extraño ni impensable que en el caso europeo, y más específicamente en el caso español, algunos partidos políticos hayan seguido estrategias similares. El método utilizado en los Estados Unidos para conseguirlo ha sido la guerra civil cultural, enfrentando a los americanos de la **moral del padre estricto** (llamados conservadores) con los americanos de **la moral familiar protectora** (los odiados liberales), a los que se retrata como una amenaza para el estilo de vida americano y para las identidades culturales, religiosas y personales de los conservadores.

Las campañas de comunicación externa de dichos partidos, la **utilización de “think-tanks”, de agencias publicitarias**, de especialistas en el lenguaje y en medios, son elementos a los que se hace especial atención y que todos los partidos políticos han utilizado y utilizan de forma más o menos sistemática. Gracias a estos think-tanks, agencias publicitarias y especialistas en medios, Lakoff indica que los conservadores han puesto en marcha una **revolución en el pensamiento y en el lenguaje** durante las últimas décadas:

“Utilizando el lenguaje, han tildado a los liberales (cuyas políticas son populistas) de elitistas decadentes, de despilfarradores no patrióticos, de liberales de limusina, de liberales frívolos,

de liberales de muchos impuestos y mucho gasto, de liberales hollywoodenses, de liberales de la Costa Este, de élite liberal, de liberales inconsistentes, etc.”

Otro de los puntos importantes a destacar es la relación que para Lakoff existe entre los programas políticos, los valores y los principios de una formación política. Claramente para este autor **los valores y los principios prevalecen en el discurso político sobre los programas** (2004, pág. 71). Los progresistas tienden a hablar de programas, pero la mayoría de los americanos no quieren saber nada de programas. La mayoría de los americanos quieren saber lo que defiendes, si tus valores son los mismos que los de ellos, por qué principios te riges y en qué dirección quieres llevar al país. **En el discurso público, los valores prevalecen sobre los programas, los principios prevalecen sobre los programas y la dirección política del país prevalece sobre los programas.**

Es clave desde el punto de vista de un investigador el conocer hasta qué punto podemos ser capaces de utilizar el discurso público de los partidos políticos (de sus principales dirigentes al menos) como forma de analizar y “descubrir” sus principales valores, su ideología, y los marcos utilizados para comunicar dichos principios políticos. Si bien este es un objetivo secundario en esta investigación, el mismo Lakoff presenta su opinión acerca de la correlación directa entre valores, principios y direcciones políticas de las formaciones políticas conservadoras estadounidenses, y las palabras (evocando marcos cognitivos) utilizadas en la comunicación política por dichas formaciones:

*“hace más de treinta años que los conservadores han **definido sus valores, sus principios y sus direcciones políticas**, exponiéndolos a la opinión pública con tanta eficacia que pueden evocarlos todos con una **filosofía expresada en diez palabras**: Defensa Fuerte, Mercados Libres, Impuestos Bajos, Menos Gobierno, Valores Familiares.”*

Progresistas (Demócratas)	Conservadores (Republicanos)
América (más) Fuerte	Defensa Fuerte
Amplia Prosperidad	Mercados Libres
Futuro Mejor	Impuestos Bajos
Gobierno Eficaz	Menos Gobierno
Responsabilidad Compartida	Valores Familiares

Tabla 5 Marcos utilizados en USA por Demócratas y Republicanos

Pero ¿cómo se transmiten estos marcos generados por un gobierno o formación política hacia el público final? Una de las teorías que nos parecen más atractivas debido a la realimentación implícita en la misma es la **teoría de la cascada de marcos**.

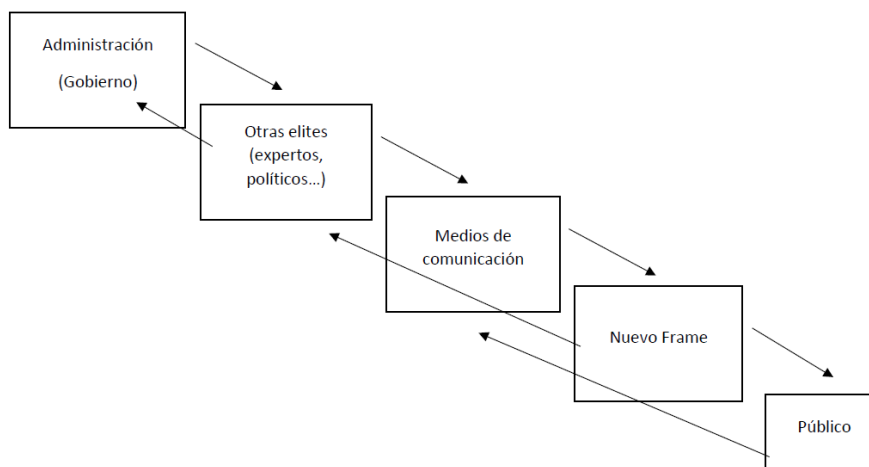


Ilustración 95 Modelo cascada de marcos (Suau Gomila, 2015, pág. 83)

Para Suau Gomila (2015, pág. 83) según el modelo de cascada nos encontramos en un proceso en el que los medios elaboran un marco (en cuya elaboración influye el Gobierno y las élites) que posteriormente llega al público quien a su vez influye en los medios. Los marcos influyen sobre las élites quienes a su vez ejercen presión y condicional al propio Gobierno. Este proceso dista de ser un proceso lineal y sencillo ya que existe una interacción entre todos los agentes que a su vez se realimentan entre sí. Pero ciertamente los medios gozan de cierta centralidad al ser los agentes que el resto de los actores necesitan para poder ejercer su influencia sobre los demás.

Es decir, partiendo del hecho de que las formaciones políticas generan una serie de mensajes hacia el exterior, la cadena de transmisión hacia el público final no es unidireccional sino bidireccional; los efectos generados sobre los medios y sobre el público en general se transmiten de vuelta a las formación políticas de forma que los mensajes que son capaces de “resonar” (aquellos que despiertan la atención del público y/o para los cuales hay marcos que resuenan en el público objetivo) son refinados, mejorados, adaptados y finalmente amplificadas a lo largo de la cadena mediática y social. Suau lo explica (2015, pág. 84) de forma similar ya que para él, el éxito del encuadre viene determinado parcialmente por el grado de verosimilitud que presente con la propia realidad y con la coherencia con la que la audiencia recibe el encuadre. Otro esquema de transmisión de los marcos es el presentado por Matthes (2012, pág. 255) y que podemos observar gráficamente en la Ilustración 96. Para él, existen tres tipos de actores involucrados en los procesos de comunicación política, los actores políticos, los medios, y el público ciudadano. Los procesos y decisiones que tienen lugar en cada uno de estos actores y

los tres conjuntos de procesos recíprocos uniendo cada par de actores son los que resumen la **generación y propagación de marcos entre los actores**.

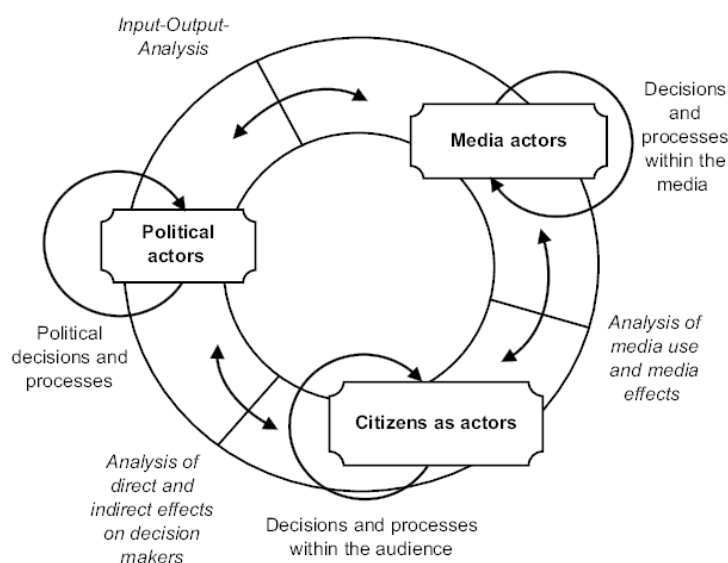


Ilustración 96 Actores en los procesos de comunicación política (Matthes, 2012)

9.4 Identificando Marcos en los medios de comunicación

Si bien es cierto que muchos autores han expresado su preocupación acerca de la validez de la investigación en el campo del framing realizada a partir del análisis de contenidos, esta preocupación está generalmente basada en la dificultad para demostrar la relación entre un marco (abstracto) y el patrón de elementos dentro de un texto que permiten que un receptor evoque dicho marco. **Partiendo de la posición de que existe una estructura de significado de base (marcos) que puede ser observada independientemente del sujeto que lo interpreta (en este caso el investigador), es posible utilizar el análisis de contenidos para descubrir esta estructura oculta, es decir, los marcos presentes en un texto** (D'Angelo & Kuypers, 2010, pág. 90). Este es un punto clave: el marco sólo existe empíricamente si los receptores responden a él y es en esta respuesta donde encontramos la relevancia de estructura oculta.

La **utilización de técnicas sistemáticas para llevar a cabo el análisis de contenidos** puede ayudar a limitar la subjetividad de la(s) persona(s) que realiza(n) el análisis de marcos. Para estos autores, en la investigación de marcos generalmente se utilizan **técnicas deductivas**, es decir, **se utiliza un conjunto predeterminado y restringido de marcos y el objeto de estudio empírico se limita a decidir en qué medida dichos marcos son utilizados en los medios de comunicación**, en las campañas, en la comunicación política, **en internet y las redes sociales**,

etc. Pero los mismos D'Angelo y Kuypers se preguntan de dónde vienen esos marcos y a qué nivel y mediante qué técnicas se pueden observar estos marcos (2010, pág. 90). Su respuesta es clara y está basada en la **utilización de un análisis de marcos inductivo mediante el cual se identifica el conjunto de marcos posibles y que son relevantes** para el tema bajo estudio.

Para D'Angelo & Kuypers es pues necesario un paso previo donde el investigador utilice técnicas inductivas de análisis para identificar el rango de marcos utilizados en un tópico de estudio. La intención de dicho análisis inductivo es reconstruir los marcos que es necesario conocer y que son útiles a la hora de definir un asunto. Un análisis sistemático con el objeto de encontrar los recursos de razonamiento y los recursos de marco relacionados con un marco permite a los codificadores identificar la presencia de dicho marco en una segunda fase deductiva, limitando e incluso eliminando la subjetividad en el análisis de marcos (2010, pág. 92). De hecho, una vez identificado un marco, y dado que este no está únicamente relacionado con un tópico en particular para Van Gorp (2007, pág. 66) debería ser posible identificar marcos que definan una situación idéntica de una forma diferente. Es decir, un tema puede ser enmarcado de diferentes formas, y un marco puede ser aplicado a varios temas. En cuanto a la necesidad de la presencia y notoriedad de los recursos de razonamiento y de enmarcado, para Van Gorp (2007, pág. 67) la función principal de los recursos de enmarcado es activar o disparar los esquemas mentales y por lo tanto la notoriedad es crucial, mientras que los recursos de razonamiento y los fenómenos culturales que soportan el verdadero peso de los marcos y de los procesos de enmarcado no necesitan estar presentes dentro de los contenidos mediáticos.

En un estudio en el campo del enmarcado Gorp (2007, pág. 72) incide en que es necesaria la **utilización tanto de técnicas cualitativas como cuantitativas. Una fase inicial inductiva para a continuación pasar a una segunda fase deductiva y todo ello centrado en la generación de una matriz de marcos** ya que la naturaleza abstracta de los marcos implica la combinación de métodos de investigación cuantitativa con las perspectivas interpretativas de los métodos cualitativos. Se podría comenzar con un borrador inductivo de los marcos encontrados en una primera revisión de la literatura existente, del discurso y de los contenidos publicados en los medios para identificar los recursos de enmarcado más sobresalientes de los marcos. Seguidamente mediante el uso de técnicas deductivas se mide la aparición de dichos recursos en el conjunto de datos manejado.

Para ello, en la fase inductiva se reconstruyen los marcos mediante la representación de paquetes de marco en una matriz donde se representen en las columnas los recursos de enmarcado y los recursos de razonamiento encontrados. Para construir esta matriz de marcos comenzaríamos con el análisis de un conjunto de textos seleccionados estratégicamente en los

medios para determinar qué elementos y proposiciones pueden funcionar probablemente como recursos de marco y recursos de razonamiento. A partir de aquí, podemos **identificar cadenas lógicas de recursos de razonamiento y recursos de enmarcado a través de diferentes textos**. Los recursos más representativos pueden ser identificados y finalmente integrados en los paquetes de marco que se presentan en la matriz de marcos. **Esta es una forma práctica de demostrar como el tema central de enmarcado se mantiene a través de una estructura coherente de recursos**.

9.4.1 Fase I. Estudio Inductivo

Como se ha mencionado en el apartado anterior, el resultado final de la fase inductiva es una matriz de marcos en la que para cada marco encontraremos en las columnas el conjunto de recursos de enmarcado y de razonamiento mediante los cuales dicho marco se manifiesta en un texto. Los marcos toman forma gradualmente durante el proceso de recopilación, codificación y análisis de los textos. Es interesante remarcar que, en nuestra opinión y dado que queremos reconstruir los marcos que subyacen y están imbuidos culturalmente en un texto, los investigadores preferentemente deben formar parte de la cultura estudiada.

Las diferentes fases que para Van Gorp (D'Angelo & Kuypers, 2010, pág. 93) deben seguirse son las siguientes:

- **Recopilación de fuentes originales:** ya que los marcos son parte de la cultura, dichos marcos gozan de una circulación más amplia que simplemente las noticias. Es por ello *recomendable recopilar los marcos de fuentes o patrocinadores de marcos* que son aquellas fuentes que utilizan los marcos para conseguir sus objetivos estratégicos (por ejemplo, ONG, grupos de presión y partidos políticos). En este tipo de patrocinadores de marcos, las notas de prensa, folletos y panfletos suelen incluir un marco dominante fácilmente identificable.
- **Codificación abierta de los textos:** en los textos seleccionados se busca un conjunto de indicadores empíricos que puedan ayudar al lector a entender los textos, *centrándonos en cómo se cuenta la historia* ya que esta forma de contar la historia es el resultado de la selección de varias opciones al contarla.
- **Ordenación de los códigos alrededor de ejes de significado:** *búsqueda de patrones en los recursos (de enmarcado y de razonamiento) mediante su unión a ideas o temas preponderantes*. Se revisan las diferencias, similitudes y contrastes entre los diferentes

recursos para reducirlos a una o varias dimensiones, incrementándose el nivel de abstracción y la separación respecto al texto original.

- **Codificación selectiva:** Ordenación de los códigos mediante el relleno de las celdas de la matriz de marcos. Al finalizar el proceso cada fila representaría un *paquete de marco (frame package, conjunto de recursos organizados lógicamente que funcionan como kit identificativo de un marco)* y las columnas los recursos de enmarcado y de razonamiento correspondientes.

Es decir, el primer paso es completar la matriz de marcos para a continuación buscar combinaciones lógicas de las diferentes columnas (definición, responsabilidad causal, soluciones, responsabilidades, etc.) **hasta conseguir un conjunto de marcos mutuamente excluyentes y que en sí mismos sean coherentes. Darle un nombre al marco implica asociar dicho marco a un motivo cultural que pueda funcionar como idea central fusionando los diferentes recursos en una unidad coherente** (D'Angelo & Kuypers, 2010, pág. 96). Una vez que el repertorio de marcos ha sido definido debemos preguntarnos por su validez, por ejemplo, comprobando que los marcos son suficientemente abstractos y pueden ser aplicados sin problema a otras problemáticas o asuntos.

Table 4.1 Framing and Reasoning Devices in a Newspaper Article About Poverty

Source text	Framing devices	Reasoning devices
[...] The Christians and the crack addicts meet for breakfast every Thursday in a downtown park. [...]	Contrast of actors: Christians vs. crack addicts; description of visual scene: having breakfast as an everyday activity	
It is not a happy scene. Sadly, it is too familiar. [...]	Emotional appeal: sadness, compassion	'Familiar' refers to omnipresence of poverty
But here in B[ritish] C[olumbia]'s bible belt, feeding the poor and the afflicted is cause for concern. [...]	Context: Bible Belt refers to the ascetic Protestants. Lexical choice: afflicted	Feeding the poor as iconic and ultimate act of charity
"Cheerios build relationships." According to others, Cheerios cause trouble. [...]	Symbol / synecdoche: A brand of cereals is used as 'pars pro toto', a symbol for charity	Seemingly illogical causal relationship between harmless charity and problems
The place is overrun with pushers and drug users. [...]	Lexical choices: overrun, pushers, and drug users	Drug users cause problems just by being there (no actual examples of them causing trouble are given); do drug users belong to the poor people?
Even with its stunning scenery, the mountains and the rich farmland that surround it, it is like any other community in Canada: Threatened by drug user use and uncomfortable with the rising tide of homelessness.	Description of visual scene with contrast: idyllic scenery vs. misery; metaphor 'rising tide' that refers to an unstoppable overwhelming force	Drug users but also homeless people are perceived as threatening

Note. From "Please don't feed the homeless," by B. Hutchinson, May 24, 2008, National Post, pp. A7, A8.

Ilustración 97 Recursos de enmarcado y de razonamiento en un artículo de periódico acerca de la pobreza (D'Angelo & Kuypers, 2010)

9.4.2 Fase II. Estudio Deductivo

En la fase deductiva de un análisis de marco los recursos de enmarcado y los recursos de razonamiento son cruciales porque son reconocibles, demostrables, cuantificables y porque pueden ser organizados y manipulados como mencionan D'Angelo y Kuypers (2010, pág. 99). La fase de análisis deductivo está basada en técnicas cuantitativas de análisis de contenidos que nos permiten medir en qué medida los paquetes de marco reconstruidos inductivamente son realmente aplicados y utilizados en un conjunto representativo de textos.

Es decir, a partir de los paquetes de marco identificados en la fase 1 inductiva, los autores se centran en medir en qué medida los marcos identificados se encuentran presentes en el corpus de datos utilizado en la Fase 2 (Deductiva). Los diferentes pasos en esta fase se dividen en:

1. Creación de un “**libro de códigos**” (**codebook**) a partir de la matriz de marcos.
2. Búsqueda de agrupaciones (clúster) en los recursos de enmarcado y de razonamiento codificados.
3. Evaluación del peso de los paquetes de marco.

9.4.2.1 Creación del libro de códigos

Para asegurar la fiabilidad de los resultados, debemos por una parte **limitar la fase deductiva a los marcos que sean claramente excluyentes entre sí**, y por otra debemos limitar el número de codificadores (personas involucradas en la codificación de los textos) utilizados para interpretar los materiales de estudio como también mencionan D'Angelo y Kuypers (2010, pág. 99). Estos autores proponen que se desarrollen sesiones de entrenamiento donde dichos codificadores puedan absorber en sus mentes los esquemas cognitivos que estén en línea con el contenido proposicional de los marcos estudiados. En este entrenamiento los codificadores comparten sus experiencias de codificación y acuerdan los protocolos de codificación a utilizar en su trabajo posterior. **Para estos autores los mejores resultados se consiguen cuando únicamente dos codificadores realizan todas las tareas en una investigación o estudio** (nótese que en el presente trabajo de fin de Master una única persona realizará toda la codificación caso de estudio). En los casos en que los recursos de enmarcado (**metáforas**, por ejemplo) no sean utilizados frecuentemente, no podemos esperar un alto grado de acuerdo entre los diferentes codificadores por lo que compartimos la opinión de dichos autores y **hemos mantenido en la realización de este trabajo un único codificador para todos los textos**

utilizados. La codificación en nuestro trabajo seguirá en general las líneas presentadas por Saldaña en su obra “The Coding Manual for Qualitative Researchers” (2009).

9.4.2.2 Búsqueda de agrupaciones en recursos de enmarcado y de razonamiento

En lugar de codificar cada recurso de enmarcado o recurso de razonamiento independientemente, D’Angelo y Kuypers (2010, pág. 100) proponen que estos se codifiquen simultáneamente en un mismo paquete de marco en vez de individualmente. Indiscutiblemente es necesario asegurar que los elementos identificados dentro de un mismo paquete de marco constituyen realmente una agrupación o clúster asociados alrededor de un marco latente. Una forma sería representar gráficamente los recursos alrededor de dos ejes o dimensiones para comprobar si realmente se presentan simultáneamente en cualquiera de los textos utilizados en el estudio y así comprobar la fuerza de dicha relación y la coherencia de su agrupación (análisis de homogeneidad). Estas representaciones gráficas pueden ser de gran utilidad a la hora de evaluar la coherencia de la matriz de marcos.

9.4.2.3 Evaluación del peso de los paquetes de marco

En este paso simplemente se determina en que cantidad los paquetes de marco identificados se encuentran en el conjunto de datos y textos bajo estudio. Para ello simplemente contaremos el número de recursos de enmarcado y de razonamiento que pertenecen a cada paquete de marco en particular. Cuanto mayor sea este número, mayor será la probabilidad de que los diferentes recursos que aparecen en un texto evoquen un esquema en la mente del lector que este en línea con el marco al que dicho recurso se refiere. Así como mencionan D’Angelo y Kuypers (2010, pág. 101) podemos encontrar si un marco domina en un texto o si varios marcos están presentes al mismo tiempo en otro texto, realizar comparaciones temporales, encontrar cuando un marco desaparece, etc. En resumen, se trataría de contar el número de veces que un recurso, parte de un paquete de marco, aparece en un texto. Por otra parte, es razonable tomar también en cuenta la prominencia de un recurso en particular presentado dentro de una historia ya que existen recursos cuya simple mención una única vez es capaz de evocar en la mente del lector y activar un marco determinado. Por ejemplo, un titular tiene mayor prominencia que una nota al pie de una página. Para resolver este aspecto, se puede usar una escala de coeficientes para ayudar en los cálculos para medir la prominencia de dichos paquetes de marco.

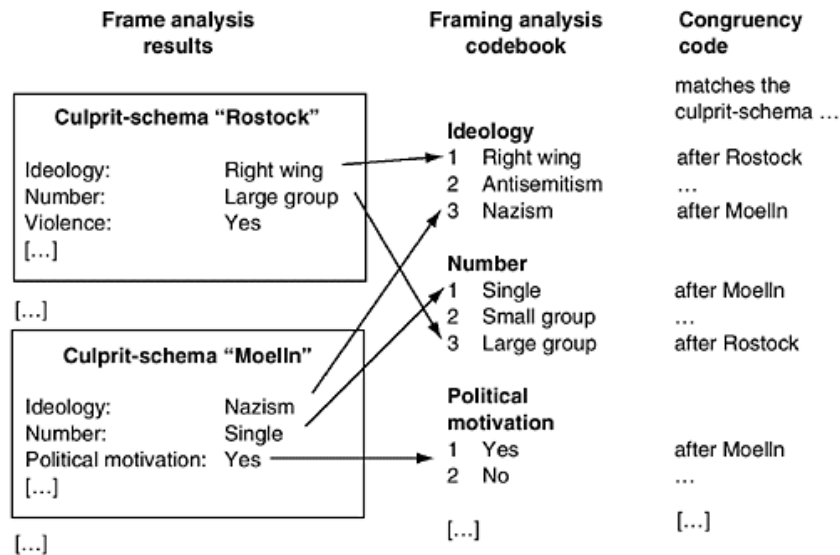


Ilustración 98 Transformar esquemas en categorías del libro de códigos (D'Angelo & Kuypers, 2010, pág. 127)

Table 1 Frame matrix for the asylum issue with their representative reasoning and framing devices

	Type of asylum-seeker	Role of asylum-seeker	Problem definition	Problem source	Responsibility	Policy solution	Moral and emotional basis	Metaphor/ stereotype	Lexical choices	Visuals
<i>Victim-frame</i>	Refugees according to a broad interpretation of the definition of the UNHCR (1)	Passive / people in need of help (2) and people who frequently fall into the hands of human traffickers (3)	How to receive refugees who are victims of persecution and other kinds of harm (4) and who are forced to leave their country? (1)	Violence, persecution, poverty (5) and the gap between rich and poor countries (6)	Borne by democratic countries (on the basis of e.g. the Geneva Convention) (6)	A humane (7) and a flexibly, cautiously and effectively applied (8) asylum policy	The moral duty to help people in affliction (9). Emotion: compassion (10)	Metaphors: shelter (11), hunting (12), fortress Europe (13), gate (14). Stereotype: helpless and anxious victim (15)	"The enemy is waiting in the homeland"; 'returning home = death' (16). Terminology: refugee (17), documentation (18)	Families with children (15); pictures that represent distress, fear and misery (10)
<i>Intruder-frame</i>	The would-be asylum-seeker who abuses the right to seek asylum (A)	Active / adventurer or criminal who resorts to illicit practices (B) and makes a deal with human smugglers (C)	How to stop people who form a threat to 'our' culture, achievements and social services (D)?	A lax asylum and deportation policy (E)	Borne by policy-makers; and the asylum-seekers themselves (F)	Deportation of asylum-seekers and a discourse oriented towards discouragement (G)	Protect the interests of the native people (H). Emotion: xenophobia; aversion to (I) and distrust of (J) strangers	Metaphors: flood (K), garbage (L). Stereotype: the dangerous stranger (M)	"The degeneration of the neighbourhood" (N); 'Belgium can't be the world's Public Centre for Social Welfare' (O). Terminology: [would-be] asylum-seeker (P)	A batch of asylum-seekers (K), especially single men

Ilustración 99 Matriz de marcos con sus recursos de enmarcado y razonamiento (Van Gorp, 2005, pág. 491)

9.5 Método de búsqueda de marcos en los medios y redes sociales

En primer lugar, analizaremos un conjunto estratégico de datos (documentación programática, entrevistas a dirigentes de formaciones políticas, etc.) para abstraer los marcos presentes en los medios relativos a varios tópicos de estudio (fase inductiva que nos permitirá evitar en la medida de lo posible encontrar marcos impuestos por el investigador y en la que buscaremos los marcos ofrecidos a la sociedad por varios “promotores de marcos”).

Estos marcos identificados en la fase inductiva permiten conformar la “matriz de marcos” o “corpus de marcos” y la creación de un libro de códigos utilizados para poder codificar cuantitativamente el conjunto de los materiales bajo estudio en una **segunda fase (fase deductiva) en la que utilizaremos principalmente datos recogidos de Twitter**. Para ello nos hemos basado en la metodología de búsqueda de marcos propuesta por Baldwin Van Gorp en el apartado 9.4 (Identificando Marcos en los medios de comunicación) que ha sido modificada de acuerdo a los siguientes apartados.

9.5.1 Metodología de búsqueda de marcos en la fase Inductiva

En la fase inductiva utilizaremos un **conjunto de documentos y datos generados por los denominados “promotores de marcos”**. En este estudio hemos tomado como promotores a fundaciones afines a los partidos políticos, los mismos partidos políticos a través de sus programas electorales y a los principales dirigentes de cada partido. **El total de documentos en esta fase será inferior al centenar** y una cantidad similar de tuits seleccionados. Entre otros documentos utilizaremos:

1. Programas electorales.
2. **Entrevistas a dirigentes** de formaciones políticas, así como las opiniones vertidas en las redes sociales por los mismos.
3. Declaraciones ideológicas.

El resultado de esta fase será la matriz de marcos (a la que nos referiremos como **matriz de marcos cualitativa**) que será utilizada posteriormente en la fase deductiva. Un ejemplo de dicha matriz de marcos para el caso de estudio de “inmigración y derecho de asilo” en Bélgica presentado por Van Gorp (2005, pág. 491) donde se ha llegado a una matriz de marcos con dos elementos sería la presentada a continuación:

Table 1 Frame matrix for the asylum issue with their representative reasoning and framing devices

	Type of asylum-seeker	Role of asylum-seeker	Problem definition	Problem source	Responsibility	Policy solution	Moral and emotional basis	Metaphor/stereotype	Lexical choices	Visuals
<i>Victim-frame</i>	Refugees according to a broad interpretation of the definition of the UNHCR (1)	Passive / people in need of help (2) and people who frequently fall into the hands of human traffickers (3)	How to receive refugees who are victims of persecution and other kinds of harm (4) and who are forced to leave their country? (1)	Violence, persecution, poverty (5) and the gap between rich and poor countries (6)	Borne by democratic countries (on the basis of e.g. the Geneva Convention) (6)	A humane (7) and a flexibly, cautiously and effectively applied (8) asylum policy	The moral duty to help people in affliction (9). Emotion: compassion (10)	Metaphors: shelter (11), hunting (12), fortress Europe (13), gate (14). Stereotype: helpless and anxious victim (15)	'The enemy is waiting in the homeland'; 'returning home = death' (16). Terminology: refugee (17), documentation (18)	Families with children (15); pictures that represent distress, fear and misery (10)
<i>Intruder-frame</i>	The would-be asylum-seeker who abuses the right to seek asylum (A)	Active / adventurer or criminal who resorts to illicit practices (B) and makes a deal with human smugglers (C)	How to stop people who form a threat to 'our' culture, achievements and social services (D)?	A lax asylum and deportation policy (E)	Borne by policy-makers; and the asylum-seekers themselves (F)	Deportation of asylum-seekers and a discourse oriented towards discouragement (G)	Protect the interests of the native people (H). Emotion: xenophobia; aversion to (I) and distrust of (J) strangers	Metaphors: flood (K), garbage (L). Stereotype: the dangerous stranger (M)	'The degeneration of the neighbourhood' (N); 'Belgium can't be the world's Public Centre for Social Welfare' (O). Terminology: [would-be] asylum-seeker (P)	A batch of asylum-seekers (K), especially single men

Ilustración 100 Matriz de marcos con recursos de enmarcado y de razonamiento representativos

En este trabajo utilizaremos una tabla similar a Van Gorp y otros para la búsqueda de marcos. Hemos sin embargo preferido delimitar claramente en dicha tabla las dos partes que nos permiten definir un marco, por un lado, los “**recursos de enmarcado**” encontrados (**metáforas, eslóganes, narraciones**) y por el otro los “**recursos de razonamiento**” que también nos definen un marco concreto (lo que incluye la definición de un problema, sus culpables, un juicio o evaluación moral, y una propuesta de solución). Un ejemplo de dicha tabla es el siguiente:

Recursos de Enmarcado			
Paquete de Marco	Metáforas	Eslóganes	Relatos
Elites culpables de la crisis		Casta Los poderosos	Pueblo oprimido por los poderosos
Final del Estado del 78			
España Plurinacional		Nación como comunidad imaginaria	
El Pueblo como los de "abajo"	Los perdedores	Justicia social	Pueblo oprimido por los de "arriba"
...			

Ilustración 101 Matriz de Marcos con sus "Recursos de Enmarcado"

Recursos de Razonamiento				
Paquete de Marco	Definición del Problema	Culpable	Juicio Moral	Solución
Elites culpables de la crisis	Crisis económica	Elite económica y política del 78	Casta corrupta / injusticia Obligación de ayudar al Pueblo	PODEMOS
Final del Estado del 78	Agotamiento político y fin de elites del 78	Elites, Monarquía		REPUBLICA
España Plurinacional	Estado del 78 agotado	Elites	Obligación de ayudar a los "Pueblos" de España	PLURINACIONALIDAD Cambio en la Constitución
El Pueblo como los de "abajo"	El Pueblo empobrecido por la pérdida de derechos	Elites	Obligación de recuperar derechos perdidos	Programa de Podemos
...				

Ilustración 102 Matriz de Marcos con sus "Recursos de Razonamiento"

9.5.2 Metodología de búsqueda de marcos en la fase Deductiva

En la fase deductiva estudiaremos principalmente los tuits de los principales responsables políticos de las formaciones y de los perfiles en Twitter de dichas formaciones políticas: el conjunto de datos utilizados en esta fase estará en el orden de varios cientos de miles de tuits. Tal y como mencionamos en el apartado 9.4 (Identificando Marcos en los medios de comunicación) en la investigación de marcos generalmente se utilizan técnicas deductivas en las cuales **se utiliza un conjunto predeterminado y restringido de marcos y el objeto de estudio empírico se limita a decidir en qué medida dichos marcos son utilizados en los medios de comunicación, en internet y las redes sociales**, etc. Y este es el objetivo de esta fase en nuestro trabajo, es decir, verificar en qué medida el conjunto de marcos encontrado en la fase 1 (Inductiva) se encuentra presente en las redes sociales utilizadas por los partidos políticos bajo estudio. El resultado de esta fase será una matriz de marcos revisada (a la que nosotros nos referiremos como matriz de marcos cuantitativa) donde los diferentes marcos encontrados en la fase inductiva han sido revisados, recodificados y testados sobre el conjunto de datos bajo estudio seleccionado en Twitter. **Este será nuestro diccionario de marcos que utilizaremos a la hora de ubicarlos en nuestro espacio vectorial (véase 5.2.6.4 Solución de Procesamiento Lingüístico).**

9.5.3 Metodología de Codificación de datos Cualitativos

Utilizaremos, tanto en la fase inductiva como en la fase deductiva la metodología de codificación de datos cualitativos:

1. **Codificación de Primera Fase:** método de **codificación Descriptiva**. Buscaremos resumir en una palabra o frase corta el tópico básico (el que define la sustancia del mensaje) de un pasaje de datos (ya sea un documento o un conjunto de tuits).
2. **Codificación de Segunda Fase:** método de **codificación por Patrones**. Códigos que identifican una respuesta emergente reuniendo en una unidad de análisis una gran cantidad de material del corpus de tuits y documentos investigados. Muchos de estos códigos de patrones se capturarán cuando sea posible en la forma de **metáforas** que sinteticen grandes bloques de datos.

9.5.4 Esquema gráfico de la Metodología de búsqueda de marcos

En la siguiente figura resumimos la metodología utilizada: Método Abductivo, selección de fuerzas políticas, codificación de primer y según nivel que junto a la utilización de herramientas automatizadas nos permite llegar a la **matriz de marcos generada a través de dos fases** también, una primera inductiva utilizando documentación generada por los promotores de marcos, y una segunda matriz final deductiva refinada a través de los miles de tuits de los principales dirigentes políticos de las tres formaciones.

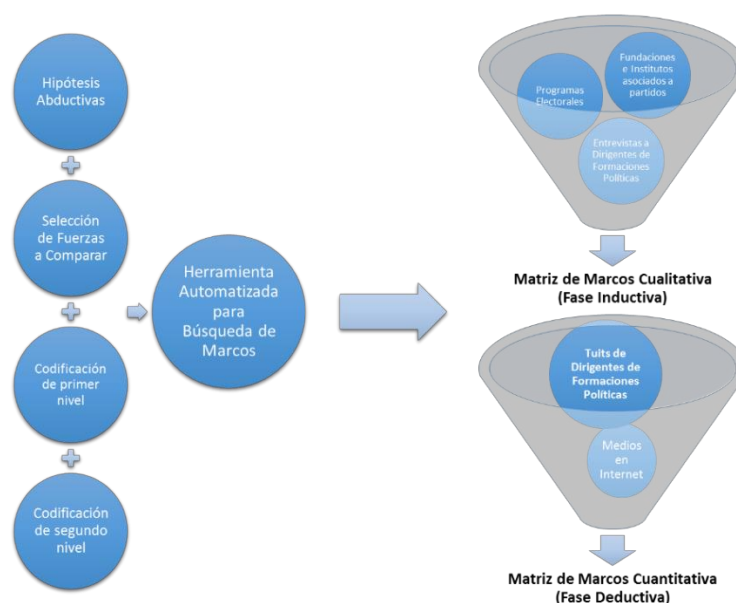


Ilustración 103 Metodología de Búsqueda de Marcos

En la siguiente figura ampliamos la información acerca de las herramientas, fuentes y metodologías y referencias concretas utilizadas.

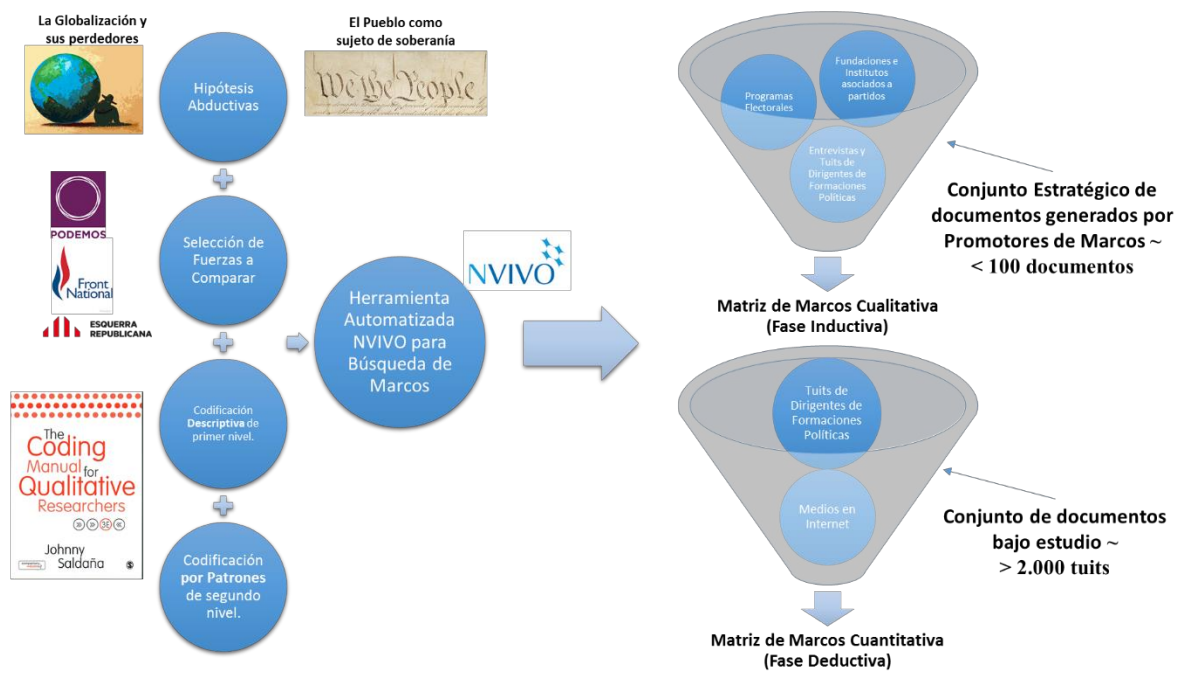


Ilustración 104 Metodología y herramientas utilizadas en la investigación

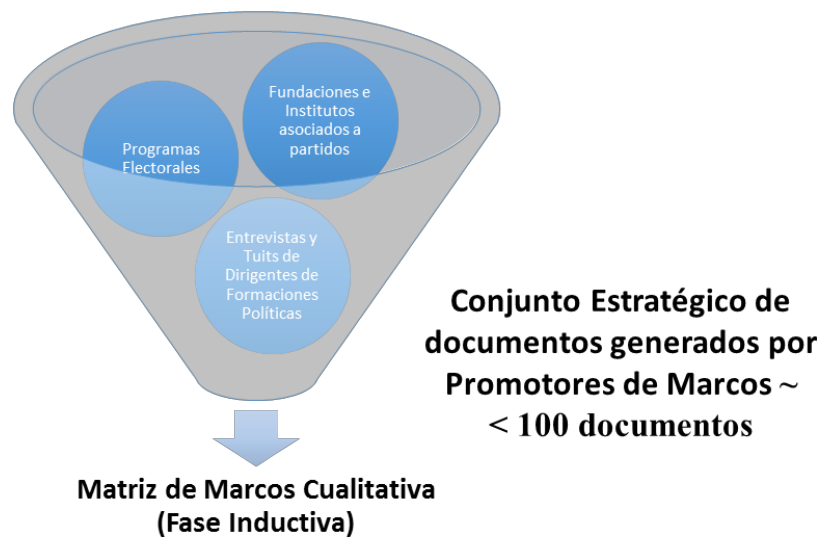


Ilustración 105 Metodología en la Fase Cualitativa – Inductiva

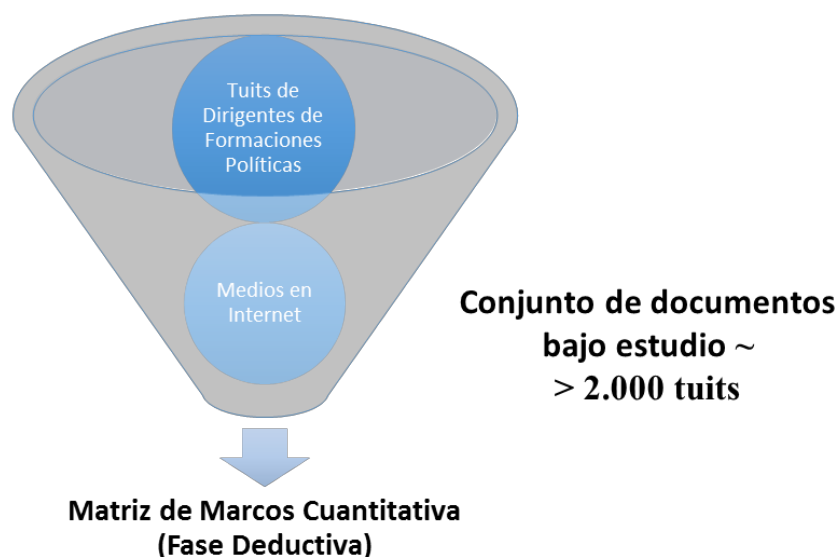


Ilustración 106 Metodología en la Fase Cuantitativa – Deductiva

No quisiéramos terminar este apartado sin incluir algún **ejemplo de trabajos y métodos utilizados por otros investigadores en investigaciones similares**. Presentamos aquí los principales resultados de la investigación desarrollada por Meijer sobre China (2014, pág. 2), donde el corpus primario de documentación utilizado en la **fase inductiva** fueron los **1.905 artículos sobre China publicados en dos periódicos** (cada uno en un extremo del espectro ideológico y político) **en un periodo de un año** (del 1 de agosto de 2011 al 1 de agosto de 2012) y donde la **fase deductiva simplemente implicó un análisis en el que una muestra aleatoria de dicho corpus (48 artículos en total) fue recodificada y codificada por un codificador diferente** para calcular la fiabilidad inter códigos. Los artículos fueron seleccionados mediante búsquedas online en las webs de dichos periódicos utilizando palabras clave de búsqueda (China, chino). Sobre los 4.176 artículos encontrados se eliminaron aquellos que fueron tomados como no relevantes (China aparecía simplemente como elemento geográfico o ningún recurso de enmarcado o de razonamiento pudo ser detectado en el texto) resultando en 1.905 artículos relevantes para su estudio (2014, pág. 31). La fase de codificación fue realizada con la ayuda de la **herramienta QCA** y después de **una docena de rondas de codificación** los pasajes relevantes fueron conceptualizados, **dichos conceptos fueron trasladados a categorías que a su vez se trasladaron y transformaron finalmente en marcos**. Durante la **fase de codificación abierta** se encontraron diferentes categorías que después de varias rondas se redujeron a **15 categorías por Meijer** (2014, pág. 40). Al mismo tiempo se realizó la **codificación axial** de dichas categorías a través del software Prezi para reordenar dichas categorías.

1. Economic threat
2. Economic opportunity
3. White Knight (economic savior and financial help)
4. Miracle about to end
5. Cross-grained in international community
6. Moral frame: human rights / Tibet / Taiwan
7. No such thing as a free lunch / opportunistic
8. Communist / authoritarian / dictatorial / censorship
9. Military threat
10. Filthy rich
11. Panda Diplomacy / soft power
12. Changing world order
13. Environmental damage
14. Imitation goods
15. Producer of bad quality, cheap junk

Ilustración 107 Categorías encontradas (Figura 12 (Meijer, 2014, pág. 41))



Ilustración 108 Codificación Axial (Meijer, 2014, pág. 42)

Después de varias fases de codificación, las categorías se redujeron (mediante técnicas de combinación y abstracción) a las 6 que finalmente definieron los marcos dominantes encontrados:

Pro	Contra
Booming business	Red dragon
White Knight	Enfant terrible
	Moral frame
Miracle about to end	Miracle about to end

Ilustración 109 Marcos dominantes encontrados (Meijer, 2014, pág. 43)

Para cada marco, esta autora presenta seis apartados para describirlos:

1. Descripción.
2. **Preguntas:** para ayudarnos a detectar si el marco está presente en un texto.
3. **Ejemplos:** párrafos de ejemplo en textos donde se ha detectado el marco.
4. **Reglas de decisión:** para determinar que el marco está presente en un texto.
5. Recursos de enmarcado.
6. Recursos de Razonamiento.

9.6 Selección de fuerzas políticas utilizadas en el estudio comparativo

El trabajo desarrollado por (Alvarez Prieto, 2018) está basado principalmente en la búsqueda y comparación de los **principales marcos utilizados por PODEMOS, ERC (España) y el Frente Nacional (Francia)**. Una de las razones principales por las que se han seleccionado estas tres formaciones es que las tres fuerzas comparten de inicio ciertas características comunes:

1. Son **fuerzas con importante soporte electoral** pero que no gozan de una mayoría para gobernar.
2. Las tres **piensan que están fuera de los juegos ideológicos comunes** (fuera del eje derecha-izquierda clásico, fuerzas transversales, etc.).
3. Las tres **apelan a un sujeto político** (la gente o el pueblo) **utilizado en diversos grados por las fuerzas populistas** (“som un poble”, el pueblo como búsqueda de la homogeneización por encima de las fracturas izquierda-derecha, etc.).

9.7 Recogida de información y selección de herramientas

Para la recogida de información a utilizar como datos de entrada en varias fases del estudio realizado por (Alvarez Prieto, 2018), este autor se ha centrado en:

- **Selección de documentación estratégica** para identificar los marcos principales en la **fase de estudio inductiva**. Estos datos son lo que generan los “**promotores de marcos**” que para nosotros serán un conjunto de fundaciones afines a los partidos políticos y los mismos partidos políticos a través de sus programas electorales.
- Selección de documentación en la fase deductiva para encontrar los marcos compartidos entre las tres formaciones políticas. Utilizaremos los *tuits de los principales responsables políticos de las tres formaciones políticas, así como artículos recuperados de periódicos y medios en internet* relacionados con los tópicos de estudio.

9.7.1 Recogida de documentación ideológica y programática de las formaciones políticas

Tal y como menciona el Frameworks Institute en (Institute, 2018), los **promotores de marcos** son un aspecto muy importante en la comunicación y el enmarcado de un tema de interés. Los **marcos necesitan promotores (sponsors)**, comunicadores que puedan elevar y redefinir un asunto, atraer a nuevos electores y construir una voluntad política y publica. El Frameworks Institute se ha aliado con filántropos, ONG, coaliciones, científicos y expertos para apoyarles para ser más efectivos y mejores “enmarcadores” de forma que **organizaciones enteras y sectores pueden alinear sus comunicaciones y amplificar un nuevo marco**.

En nuestro caso hemos tomado como promotores de los marcos a las mismas formaciones políticas a través de sus publicaciones online (ya sea en sus propias páginas web o en las redes sociales) y a aquellos organismos, asociaciones o fundaciones afines a los mismos y que suelen ser la avanzadilla, liderar, ayudar y “promover” los principales argumentarios políticos de dichos partidos. Esta información será la que se utilice primariamente en la fase inductiva de la búsqueda de marcos. Nos centraremos aquí en:

1. Recogida de información publicada por las propias formaciones políticas a través de sus páginas web.
 - a. Podemos: <https://podemos.info/>
 - b. ERC: <https://www.esquerra.cat/ca/inici>

- c. FN: <http://www.frontnational.com/>
- 2. Recogida de información programática e ideológica publicada por fundaciones y organismos relacionados con las formaciones políticas.
 - a. Podemos:
 - i. “[Instituto 25 de Mayo para la Democracia](#)”, fundación vinculada a Podemos “cuyo fin es el análisis, la formación y la creación política y cultural”.
 - ii. Programas electorales: <https://lasonrisadeunpais.es/programa/>
 - b. ERC:
 - i. Programas Electorales: <https://www.esquerra.cat/ca/programes-electorals>
 - ii. Documentación presente en la web de ERC acerca de varias comisiones y grupos como por ejemplo la **Comisión sectorial de la UE y de Política Internacional** (“Comissió sectorial d'UE i Política Internacional”).
 - c. FN:
 - i. Programas Electorales:
 - <http://www.frontnational.com/?s=mondialisation>
 - <http://www.frontnational.com/?s=peuple>
 - ii. Información ideológica y sobre el proyecto:
 - <http://www.frontnational.com/le-projet-de-marine-le-pen/>

9.7.2 Recogida de información en redes sociales

A través de la herramienta [NCapture](#) de la plataforma NVIVO hemos recopilado información procedente de medios en internet y de redes sociales (Twitter) que se han integrado principalmente en la fase deductiva en la que hemos investigado sobre los marcos efectivamente presentes en las redes sociales (y entre ellas en Twitter) para las tres formaciones políticas bajo estudio. Nos hemos centrado por lo tanto en la información publicada por los partidos y por los principales dirigentes de las formaciones políticas en Twitter y **esta información será la que se utilice masivamente en la fase deductiva de la búsqueda de marcos:**

1. **Podemos:** búsqueda de los términos “pueblo” y “globalización”.
 1. Podemos: <https://twitter.com/ahorapodemos>
 2. Instituto 25M: <https://twitter.com/Instituto25M>
 3. Pablo Iglesias: https://twitter.com/pablo_iglesias

4. Iñigo Errejón: <https://twitter.com/ierrejon>
 5. Irene Montero: https://twitter.com/irene_montero
 6. Pablo Echenique: <https://twitter.com/pnique>
 7. Teresa Rodriguez: <https://twitter.com/teresarodr>
2. **ERC**: búsqueda de los términos “**poble**” y “**globalització**”.
 1. Esquerra Republicana de Catalunya: [@Esquerra ERC](https://twitter.com/Esquerra)
 2. ERC Internacional: [@ERC_INTL](https://twitter.com/ERC_INTL)
 3. ERC Economía: [@ERC_economia](https://twitter.com/ERC_economia)
 4. Oriol Junqueras: [@junqueras](https://twitter.com/junqueras)
 5. Raul Romeva: [@raulromeva](https://twitter.com/raulromeva)
 6. Marta Rovira: [@martarovira](https://twitter.com/martarovira)
 7. Roger Torrent: [@robertorrent](https://twitter.com/robertorrent)
 3. **FN**: búsqueda de los términos “**people**” y “**mondialisation**”.
 1. Front Nationale: [@FN_officiel](https://twitter.com/FN_officiel)
 2. Marine Le Pen: [@MLP_officiel](https://twitter.com/MLP_officiel)
 3. FN Jeunesse: [@FNJ_officiel](https://twitter.com/FNJ_officiel)



Ilustración 110 Recogida de datos en Twitter mediante la herramienta Capture-NVivo

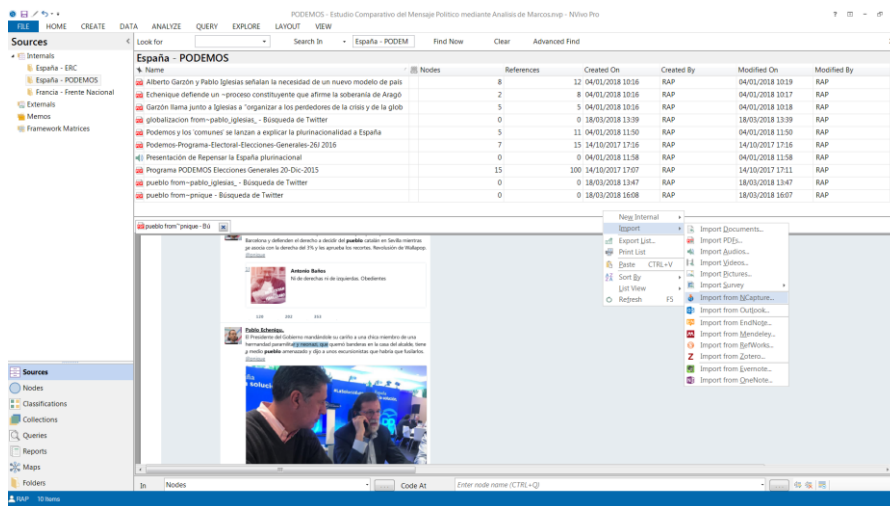


Ilustración 111 Importación de datos procedentes de Twitter en la herramienta NVivo

9.7.3 Herramientas de gestión de información cualitativa utilizadas

En la realización del presente trabajo hemos utilizado la plataforma de gestión de información **NVIVO** para la gestión de toda la documentación y tuits utilizados tanto en la fase inductiva como en la fase deductiva. NVivo es un software enfocado a la investigación con métodos cualitativos y mixtos diseñado para organizar, analizar y encontrar perspectivas en datos no estructurados o cualitativos como entrevistas, respuestas de encuestas con preguntas abiertas, artículos, contenido de las redes sociales y la web. La investigación ayudada por este software facilita la tarea de descubrir de conexiones entre datos. Además NVivo proporciona un lugar donde organizar y gestionar nuestros materiales permitiendo hacer preguntas sobre los datos de modo más eficiente.

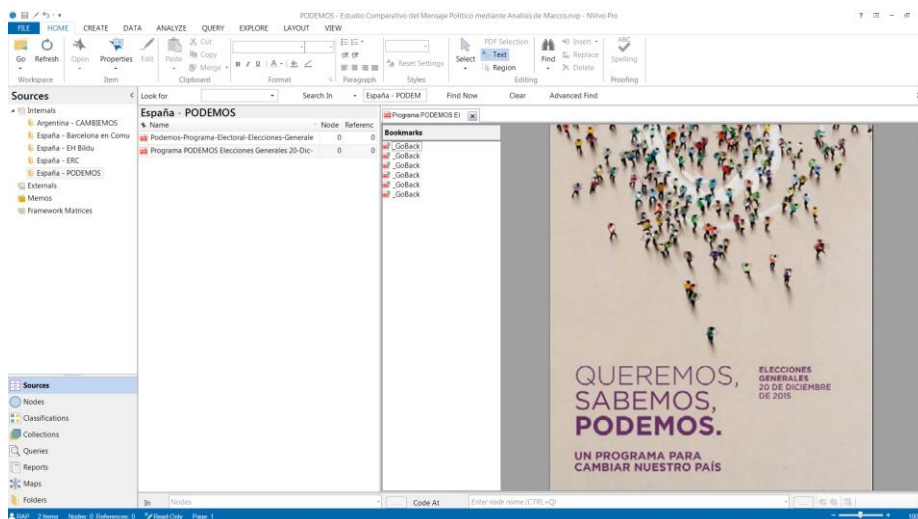


Ilustración 112 Pantalla de inicio del proyecto PODEMOS en la plataforma NVIVO

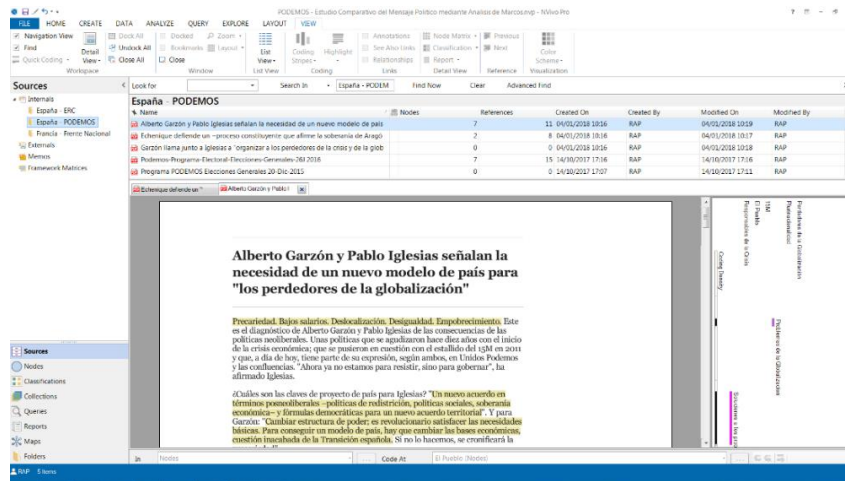


Ilustración 113 Ejemplo de Codificación de documento recogido en internet relativo a Podemos y los Perdedores de la "Globalización"

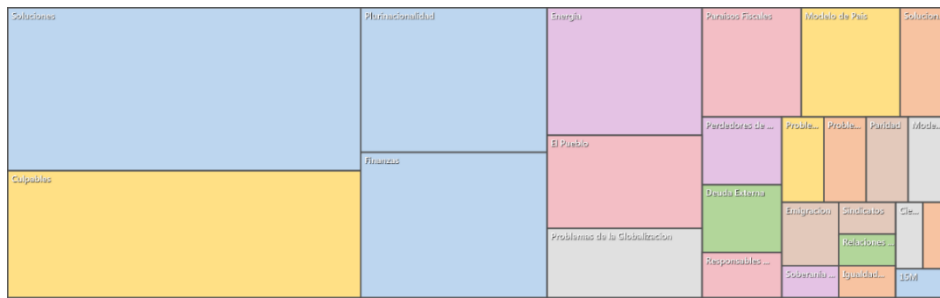


Ilustración 114 Nodos comparados por el número de referencias en la codificación

9.8 Marcos Cognitivos utilizados por PODEMOS

Presentamos en este apartado el conjunto de marcos resultado de la investigación realizada por (Alvarez Prieto, 2018) siguiendo la metodología presentada en los apartados anteriores de este anexo. **Estos marcos son el resultado de una primera fase inductiva donde se ha tomado como fuente principal la documentación presentada por los promotores de marcos en cada formación política** (programas electorales, fundaciones asociadas a partidos, principales dirigentes) y de una **segunda fase deductiva donde se refina la matriz de marcos a través de la información publicada en Twitter por dichas formaciones políticas y sus principales dirigentes**. En este trabajo se restringió la búsqueda de marcos políticos a aquellos relevantes para el estudio del caso presentado en (Alvarez Prieto, 2018), es decir, la discusión política acerca de dos temas: la **globalización** (y sus perdedores) y el **“Pueblo”** (como sujeto de la soberanía).

9.8.1 Matriz de Marcos Cualitativa (fase inductiva)

En el trabajo, se llegó a una matriz de marcos cualitativa conformada por cuatro marcos principales:

1. Las élites como culpables de la crisis.
2. El Pueblo como los de “abajo”.
3. España Plurinacional.
4. El final del Estado del 78.

Marcos utilizados por PODEMOS
Élites culpables de la crisis
El Pueblo como los de "abajo"
Final del Estado del 78
España Plurinacional

Ilustración 115 Marcos utilizados por PODEMOS (Fase Inductiva)

Presentamos a continuación el estudio detallado de cada uno de los cuatro marcos identificados en esta investigación incluyendo los recursos de razonamiento (definición del problema, culpables, juicio moral y solución) y recursos de enmarcado (figuras retóricas, eslóganes y relatos) asociados a dichos marcos.

Marcos utilizados por PODEMOS		Élites culpables de la crisis
Recursos de Razonamiento	Definición del Problema	Los culpables de la crisis económica y pérdida de derechos son las élites políticas y financieras.
	Culpable	Elite económica, mediática y política del 78. Bancos y oligarquías. Gobiernos cobardes. Monarquía. Grandes fortunas.
	Juicio Moral	La nueva extrema derecha que representa Ciudadanos. Los monárquicos es que jamás entendieron España. Ellos son los antisistema.
	Solución	Plan Nacional contra la Corrupción y para la Transparencia Institucional. No tener políticos en los consejos de administración. Construcción de un sistema mundial de gobernanza económica y financiera más justo. Eliminación del secreto bancario privilegiado de los paraísos fiscales.
Recursos de Enmarcado	Figuras retóricas (Metáforas, Sinécdoques, hipérboles, preguntas, etc)	TC hiciera saltar por los aires el Estatut. Basta ya de secuestrar la soberanía. Bloque monárquico" (PP, PSOE y Ciudadanos). Ley de Paredes de Cristal. Puertas giratorias. LOS INTOCABLES de la crisis financiera
	Eslóganes	Corrupción es que un 1% de ricos posea lo mismo que el 70% de la población. Hoy soñamos con una Europa de los ciudadanos. No de los mercaderes y los bancos. Frente a las mentiras de los corruptos y los que los mantienen en el Gobierno, nuestro pueblo.
	Relatos	Los ladrones más peligrosos usan gomina y corbata. La conjura monárquica. Mil veces nos contaron el fin de los asesinos de Viriato y mil veces la historia se repite. Los enemigos del pueblo. Ya modificaron la Constitución para arrebatarle la soberanía a nuestro pueblo y dársela a las élites económicas.

Ilustración 116 Marco "Élites culpables de la crisis" utilizado por PODEMOS

Marcos utilizados por PODEMOS		El Pueblo como los de "abajo"
Recursos de Razonamiento	Definición del Problema	El Pueblo empobrecido por la perdida de derechos en contraposición (principalmente económica) a los de arriba (la élite).
	Culpable	Élites económicas. Corruptos y los que los mantienen en el Gobierno.
	Juicio Moral	Juventud a la que le habían dicho que iban a vivir mejor que nuestros padres.
	Solución	Locos dignos que se enfrenten a los poderosos. Proceso constituyente que active al soberano de nuestra Constitución, el pueblo. Organizar a los perdedores de la crisis. Conseguir un modelo de país que pueda satisfacer los derechos fundamentales. Que el desbordamiento popular tome las riendas del gobierno. Que el poder constituyente, el pueblo soberano, de origen al poder constituido: el legislativo, ejecutivo y el judicial.
Recursos de Enmarcado	Figuras retóricas (Metáforas, Sinédoques, hipérbolos, preguntas, etc)	El andamiaje de un movimiento popular. Esa gente valiente y humilde está en nuestro ADN. Yo estoy dispuesto a dar la batalla por la gente. Sentarnos en la mesa de la democracia. Somos hijos del 15M.
	Eslóganes	Fraternidad, compañerismo, "Si se puede". Bandera de los perdedores. Ver a un pueblo sonreír. Podemos nació de la caradura de unos pocos y del dolor de unos muchos. Solo el pueblo ayuda al pueblo. Un país lo hace su pueblo y no sus autopistas o su bandera.
	Relatos	Jóvenes estudiantes y trabajadores jugándose todo por la dignidad de nuestro país. Si vivimos con miedo somos súbditos. Podemos nació en código de construcción de mayoría, nació de un desbordamiento ciudadano y popular gestado a base de cierres, recortes, despidos y desahucios.

Ilustración 117 Marco "El Pueblo como los de Abajo" utilizado por PODEMOS

Marcos utilizados por		Final del Estado del 78
Recursos de Razonamiento	Definición del Problema	Marco en el que PODEMOS define la necesidad de finalizar el estatus político resultante del consenso del 78. A diferencia de ERC, para PODEMOS este final está centrado en un proceso constituyente que cambie las estructuras del Estado.
	Culpable	La estructura de poder de nuestro país. Élites
	Juicio Moral	Nuevo proyecto de país asociado a la justicia social y a la soberanía popular.
	Solución	Nuevo encaje constitucional que reconociera a Catalunya como nación. Configuración estatal que acepte de una vez la realidad plurinacional de nuestra patria. Cambiar el marco constitucional. Proceso constituyente.
Recursos de Enmarcado	Figuras retóricas (Metáforas, Sinédoques, hipérboles, preguntas, etc)	Saltar por los aires uno de los pactos cruciales de la Transición. "¿Para qué sirve la monarquía?"
	Eslóganes	España será plurinacional o no será.
	Relatos	-

Ilustración 118 Marco "Final del Estado del 78" utilizado por PODEMOS

Marcos utilizados por PODEMOS		España Plurinacional
Recursos de Razonamiento	Definición del Problema	Marco en el que PODEMOS define a España como un estado plurinacional (no reconocido todavía) como solución a los problemas políticos y de todo tipo existentes.
	Culpable	Élites centrales. Monarquía.
	Juicio Moral	España plurinacional y respetuosa que ya existe en las ciudades y en los pueblos y que aún no ha entrado en los partidos y en las instituciones. Hará saltar por los aires uno de los pactos cruciales de la Transición (Generalitat de Catalunya). Es necesario defender España asumiendo el derecho del pueblo catalán a decidir su futuro en un referéndum.
	Solución	Referéndum legal y pactado. Nuevo encaje constitucional que reconociera a Catalunya como nación y profundizara su autogobierno. Fórmulas, federales o confederales, para afrontar la plurinacionalidad de España. Nueva España social, republicana y plurinacional. Iniciativas constituyentes que afirman la soberanía de los pueblos que la habitan. Nueva Constitución que recoja las relaciones entre las diferentes regiones, autonomías y naciones que cohabitan en el Estado.
Recursos de Enmarcado	Figuras retóricas (Metáforas, Sinédoques, hipérboles, preguntas, etc)	Administración colonial de Catalunya. El virrey Rajoy. Dar forma al sujeto constituyente aragonés. Tejer soluciones de fraternidad. ¿Existe la nación española dentro de Catalunya?
	Eslóganes	España será plurinacional o no será. "España es plurinacionalidad y esto no es discutible" Referéndum pactado es el único que puede resolver el conflicto
	Relatos	Las imágenes de gente con banderas españolas e independentista unidas. El pueblo por delante de las instituciones. El virrey Rajoy querrá administrar Catalunya. La reinstauración de una institución republicana como la Generalitat, reconocida por la Constitución de 1978, fue la base del amplio apoyo social al texto constitucional en Catalunya.

Ilustración 119 Marco "España Plurinacional" utilizado por PODEMOS

9.8.2 Matriz de Marcos Cuantitativa (fase deductiva)

Los documentos utilizados en la fase deductiva han estado concentrados principalmente en información publicada a través de Twitter por las cuentas de las formaciones políticas así como de sus principales dirigentes. En esta segunda fase deductiva **se ha utilizado un conjunto predeterminado y restringido de marcos (los identificados en la fase inductiva anterior) y el objeto de estudio empírico se limita a decidir en qué medida dichos marcos son utilizados en los medios de comunicación, en internet y las redes sociales**, etc. y si es posible, revisar la matriz de marcos de forma que lleguemos a una matriz de marcos que denominaremos **matriz de marcos deductiva o cuantitativa**. Y este es principalmente el objetivo de esta fase, es decir, **verificar en qué medida el conjunto de marcos encontrado en la fase 1 (Inductiva) se encuentra presente en las redes sociales utilizadas por los tres partidos políticos bajo estudio**.

Así pues, el resultado de esta fase ha sido una **matriz de marcos revisada** (a la que nosotros nos referiremos como **matriz de marcos cuantitativa**) donde los diferentes marcos encontrados en la fase inductiva han sido revisados, recodificados y testados sobre el conjunto de datos bajo estudio seleccionado en Twitter. El resultado inicial que podríamos decir que es más notable al evaluar los datos recogidos en Twitter es el **escaso interés que la globalización despierta en los partidos como PODEMOS y ERC** que contrasta con el relativo interés que esta despierta en el FN (con un total del 15% de los tuits seleccionados). Lo que es **común en las tres formaciones son las referencias en Twitter al “Pueblo”**. Este dato es relevante y está alineado con los datos que ya presentamos en la fase inductiva de este estudio.

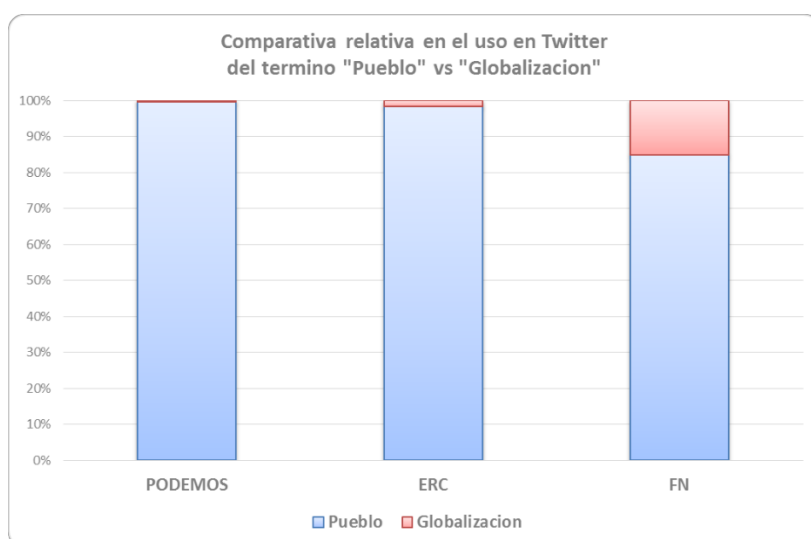


Ilustración 120 Comparativa entre fuerzas políticas en Twitter del uso de "Pueblo" y "Globalización"

Si representamos gráficamente los resultados de las búsquedas en Twitter para todas las fueras y dirigente políticos incluidos en este estudio tenemos:

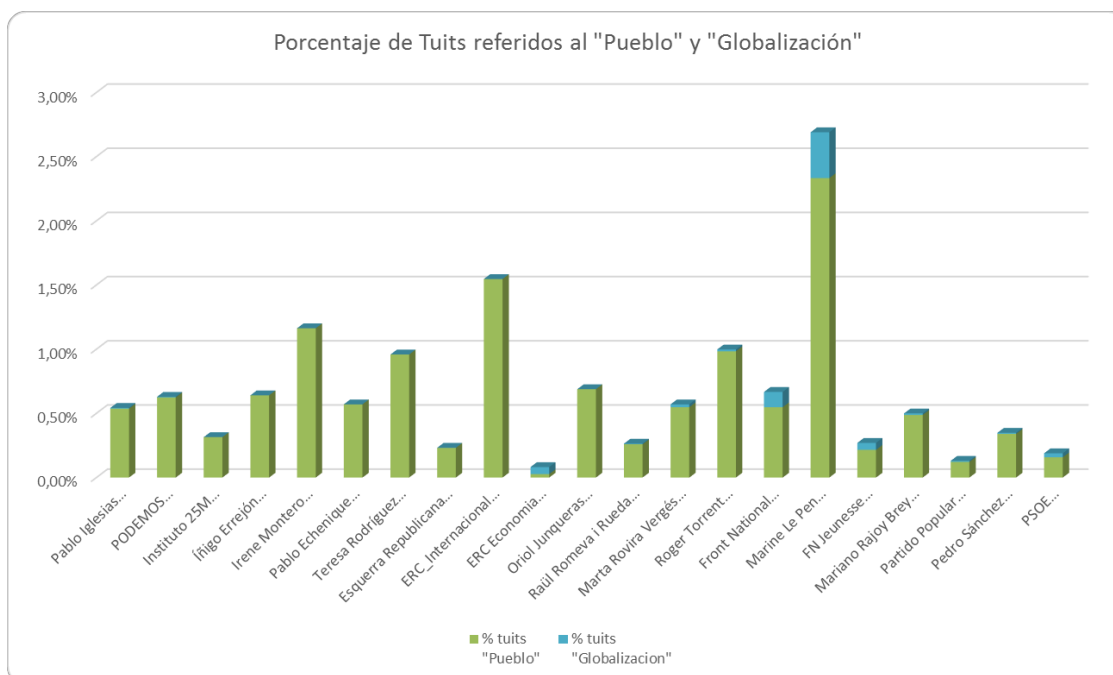


Ilustración 121 Porcentaje de tuits referidos a "Pueblo" y "Globalización" para diversas formaciones y dirigentes políticos

	Tuits Totales	Tuits "Pueblo"	% tuits "Pueblo"	Tuits "Globalizacion"	% tuits "Globalizacion"
Pablo Iglesias @Pablo_Iglesias_	18.600	100	0,54%	1	0,01%
PODEMOS @ahorapodemos	91.100	569	0,62%	4	0,00%
Instituto 25M @Instituto25M	3.489	11	0,32%	0	0,00%
Íñigo Errejón @ierrejon	18.100	116	0,64%	0	0,00%
Irene Montero @Irene_Montero_	9.712	113	1,16%	0	0,00%
Pablo Echenique @pnique	5.434	31	0,57%	0	0,00%
Teresa Rodríguez @TeresaRodr_	14.700	141	0,96%	0	0,00%
TOTAL	161.135	1.081	0,67%	5	0,00%

Ilustración 122 Utilización de los términos "Pueblo" y "Globalización" en Twitter por PODEMOS

	Tuits Totales	Tuits "Pueblo"	%	Tuits "Globalizacion"	%
Esquerra Republicana @Esquerra_ERC	87.800	203	0,23%	1	0,00%
ERC Internacional @ERC_INTL	3.297	51	1,55%	0	0,00%
ERC Economia @ERC_economia	7.317	2	0,03%	4	0,05%
Oriol Junqueras @junqueras	13.800	95	0,69%	0	0,00%
Raül Romeva i Rueda @raulromeva	20.800	54	0,26%	1	0,00%
Marta Rovira Vergés @martarovira	8.939	49	0,55%	2	0,02%
Roger Torrent @rogertorrent	7.013	69	0,98%	1	0,01%
TOTAL	148.966	523	0,35%	9	0,01%

Ilustración 123 Utilización de los términos "Pueblo" y "Globalización" en Twitter por ERC

Presentamos a continuación una parte de la matriz de marcos cuantitativa resultado de refinar la matriz de marcos con los resultados recogidos en Twitter para ERC. Es de resaltar que a pesar de la **gran similitud en los marcos proyectados**, hay algunos aspectos novedosos como la **utilización de hashtags (que quizás podrían verse como equivalentes a eslóganes o figuras retóricas en algunos casos)** y que hemos incluido como recursos de enmarcado en los marcos correspondientes (#dandoEjemplo, #laSonrisadeunPais, etc.). En cuanto a los resultados para PODEMOS después de esta fase deductiva, solamente remarcar la profusión de eslóganes y relatos utilizados en la comunicación en Twitter, y que estos están totalmente alineados con los marcos que hemos recopilado en la fase inductiva.

Marcos utilizados por ERC		Final del Estado del 78
Recursos de Razonamiento	Definición del Problema	Marco en el que ERC define la necesidad de finalizar el estatus político resultante del consenso del 78. A diferencia de PODEMOS, para ERC este final está centrado en la Independencia de Cataluña mas que en un proceso constituyente que cambie las estructuras del Estado.
	Culpable	-
	Juicio Moral	Solo la independencia por parte de todas las naciones que lo deseen puede garantizar el pleno ejercicio de los derechos democráticos nacionales.
	Solución	La forma de estado republicana es la mejor garantía para el ejercicio de la democracia. Consecución de la independencia.
Recursos de Enmarcado	Figuras retóricas (Metáforas, Sinédoques, hipérbolos, preguntas, etc)	EL pueblo ha llegado a la mayoría de edad.
	Eslóganes	Régimen post-franquista. #DeclaracionDeSoberania
	Relatos	La forma de estado republicana es la mejor garantía para el ejercicio de la democracia. Respetaremos el mandato del pueblo que lo que va a decir es que quería hacer Republica. Exiliados.

Ilustración 124 Marco fase cuantitativa "Final del Estado del 78" utilizado por ERC

9.8.3 Discusión de resultados

En los apartados anteriores se han presentado los marcos utilizados por Podemos encontrados en (Alvarez Prieto, 2018), y para cada uno de ellos:

- Los recursos de razonamiento utilizados: culpable, solución, acciones. una definición de un problema, una interpretación causal, una evaluación moral y una recomendación de solución para dicho problema
- Todos los recursos de marco utilizados: las metáforas, los casos históricos de los que derivar lecciones, los eslóganes o consignas, los relatos, etc.

Los cuatro marcos identificados han sido:

1. El Pueblo como los de "abajo".
2. Las Élite como culpables de la crisis.
3. El Final del Estado del 78.
4. La España Plurinacional.

Estos resultados están alineados con lo que otros autores han detectado en el discurso de PODEMOS, como por ejemplo (Bayón , 2017, pág. Pos. 427) que también ha mostrado la contraposición que hace PODEMOS entre los de “arriba” (las élites, “la casta”) y la **“gente” (aquellos que están “abajo”)**.

10. Anexo II

Presentamos en este anexo los cuestionarios y referencias utilizadas en la evaluación del “Gold Standard”.

10.1.1.1 Referencias utilizadas en el eje 4 “Organización Territorial del Estado”

Aunque sea complejo determinar el punto de vista respecto al binomio “Estado Plurinacional-Estado Uninacional” que unas organizaciones políticas plurales o un dirigente político en particular puedan tener, creemos interesante utilizar como aproximación para el análisis el artículo del diario “El País” (Casqueiro, 2020) en el que se presentan los resultados de un cuestionario sobre el futuro de la Ley Fundamental planteado a las ocho fuerzas representativas del arco parlamentario (PSOE, PP, Vox, Unidas-Podemos, Ciudadanos, JxC, PNV y Mas País).

No debemos olvidar que en nuestro trabajo *estamos más interesados en conocer la importancia que dichos partidos otorgan a cada marco en sus estrategias de comunicación política*, es decir, si utilizan mayoritariamente el marco de “Estado Uninacional” o el de “Estado Plurinacional”. De esta forma podremos realizar una comparativa con los datos procedentes de nuestro sistema automatizado. Un estudio detallado del interés e importancia que dichos partidos otorgan a cada marco, y que probablemente este íntimamente ligada a su ideología, va más allá de los objetivos de este trabajo por lo que aquí solo hemos realizado una somera exploración por un único anotador de este interés. Este cuestionario está centrado en los siguientes nueve apartados de nuestra Constitución:

1. Forma de Estado.
2. Soberanía y unidad de España
3. Lenguas
4. Sanidad pública
5. Corona
6. Investidura
7. Estado de alarma
8. Control del gasto
9. Organización territorial

Referente a la pregunta: ¿Es partidario de modificar alguno de estos dos artículos para cambiar el sujeto de soberanía o eliminar la referencia a la unidad de España?

- VOX:
 - No se deben modificar.
 - No creemos que se deban modificar estos artículos con esas finalidades. Sí comprendemos a los que señalan el uso perverso que se ha hecho del término “nacionalidades”, cuya introducción en la Constitución ya fue duramente criticada en 1978 por muchos; entre ellos, por los fundadores del actual PP. Y de aquellos polvos, estos lodos.
 - Suspensión inmediata de la autonomía de Cataluña.
 - Ilegalizar partidos, asociaciones u ONGs que persigan la destrucción de la unidad territorial.
- PODEMOS:
 - La soberanía democrática debería pensarse también en plural.
 - El artículo 2 es uno de los más tortuosos de la Constitución, ya que su versión final fue impuesta por sectores del Ejército ligados al franquismo. Un proyecto republicano y plurinacional en común, como el que defendemos, exigiría pensar la soberanía democrática también en plural. Y asegurar que esta pueda compartirse, sin imposiciones, por la libre decisión de los diferentes pueblos y gentes que conviven en el Estado.
 - Apostar por un referéndum pactado y defender un nuevo encaje para Cataluña en España.
- PSOE:
 - La soberanía debe residir en el conjunto del pueblo.
 - No. El artículo 2 de la Constitución Española es la clave sobre la que descansa la estructura territorial del poder en una España plural al proclamar dos principios: la unidad de la Nación española y el reconocimiento y garantía del derecho al autogobierno de sus nacionalidades y regiones. Ello abrió paso a la transformación de un Estado centralista en un Estado que se sitúa entre los más descentralizados del planeta.
 - La soberanía debe residir en el conjunto del pueblo español, no en partes diferenciadas del mismo. Por cierto, es tan contrario a la Constitución ir en

contra de la unidad de España como ir en contra de la autonomía política de las nacionalidades y regiones que la integran, pues la Constitución, en el art. 2 CE, las sitúa al mismo nivel.

- Impulsar el diálogo con Cataluña dentro de los límites constitucionales y del Estatuto de Autonomía.
- PP:
 - Garantiza la unidad de la nación y nuestra pluralidad.
 - El artículo 2 representa muy bien la grandeza del sistema político construido en la Transición: garantiza la unidad de la nación española y reconoce y garantiza nuestra pluralidad. España no solo se encuentra entre las democracias más avanzadas del mundo, sino también entre las más descentralizadas.
 - Introducir en el Código Penal el cumplimiento íntegro de las penas para condenados por rebelión o sedición.
 - Recuperar los delitos de referéndum ilegal y de rebelión impropia.
- Ciudadanos:
 - El poder corresponde al pueblo en su conjunto.
 - En absoluto. Con los partidos separatistas tan crecidos tras la bocanada de oxígeno que les ha brindado el Gobierno del Partido Socialista y Podemos, abrir en canal algo tan fundamental como la cuestión de la soberanía nacional no solo es innecesario y un disparate, sino un auténtico peligro. Pero es que, además, la norma es la que contienen prácticamente todas las constituciones del mundo. Y es de pura lógica que el poder corresponda al pueblo en su conjunto y no esté troceada.
 - Abogar por la aplicación del artículo 155 para asegurar el restablecimiento de la Constitución en Cataluña.

10.1.1.2 Referencias utilizadas en el eje 5 “Modelo Político del Estado”

Aunque sea complejo determinar el punto de vista respecto al binomio “Monarquía-República” que unas organizaciones políticas plurales o un dirigente político en particular puedan tener, creemos interesante utilizar como aproximación para el análisis el artículo presentado en La Vanguardia referente a este “dilema” político presente en España (Blanco, 2019). También utilizaremos los datos presentados en el artículo de “El País” en el que se presentan los

resultados de un cuestionario sobre el futuro de la Ley Fundamental planteado a las ocho fuerzas representativas del arco parlamentario como en el apartado anterior.

No debemos olvidar que en nuestro trabajo *estamos más interesados en conocer la importancia que dichos partidos otorgan a cada marco en sus estrategias de comunicación política*, es decir, si utilizan mayoritariamente el marco de “Monarquía” o el de “República”. De esta forma podremos realizar una comparativa con los datos procedentes de nuestro sistema automatizado. Un estudio detallado del interés e importancia que dichos partidos otorgan a cada marco, y que probablemente este íntimamente ligada a su ideología, va más allá de los objetivos de este trabajo por lo que aquí solo hemos realizado una somera exploración por un único anotador de este interés.

- VOX:
 - Y Abascal garantizó el apoyo al rey siempre que siga actuando como hasta ahora.
 - "Si la monarquía sigue siendo una institución al servicio de la unidad y la permanencia de España, con toda lealtad con la monarquía, *pero si alguien convence a la monarquía ahora o en el futuro para dejarse llevar por los cantos de sirena de un estado confederal con varias soberanías -precisó-, sé que habrá muchos españoles que prefieran una república nacional unitaria*".
 - Por si quedaban dudas, le ratificó su apoyo: "A día de hoy, por suerte, podemos decir que Felipe VI ha estado en la vanguardia y al frente de la defensa de la unidad nacional. Por tanto, toda nuestra lealtad".
- PODEMOS:
 - No hay tampoco mención explícita alguna a la república entre los compromisos de Podemos, que sí citan en sólo una ocasión al rey y a la monarquía.
 - En concreto lo hacen al prometer "*suprimir los aforamientos y acabar con la irresponsabilidad del rey*", en alusión al artículo constitucional que determina que el monarca "es inviolable y no está sujeto a responsabilidad".
 - "Trabajaremos -señala Podemos- en una reforma constitucional para eliminar el privilegio judicial más grave, que es la ausencia de responsabilidad del rey".

- A renglón seguido promete también "una comisión de investigación sobre los negocios de la monarquía con el objetivo de determinar si, por ser irresponsable, es una institución idónea para nuestra democracia".
- PSOE:
 - El último Congreso Federal del PSOE, celebrado en junio de 2017 y en el que Pedro Sánchez fue ratificado como secretario general, debatió una propuesta del líder de sus juventudes, Nino Torre, para "avanzar e implantar la república". Se rechazó esa pretensión y lo que se aprobó fue ***incorporar al ideario del partido el fortalecimiento de "los valores republicanos" sin cuestionar a la monarquía.***
 - Sánchez alude al rey en varias ocasiones en su libro "Manual de Resistencia", en el que respalda el discurso del 3 de octubre de 2017 pronunciado por Felipe VI ante el desafío independentista catalán y criticado duramente por los soberanistas. Además, asegura que estableció con el monarca "una relación muy franca" ante las conversaciones para formar gobierno tras las elecciones de 2015 y ambos se reconocieron "mutuamente -escribe en el libro- como las personas que íbamos a sacar al país del riesgo de bloqueo".
 - Eso no es óbice para que ***el presidente del Gobierno haya defendido públicamente que se elimine la inviolabilidad del rey*** y haya asegurado que Felipe VI lo vería con buenos ojos. Sin embargo, no plantea propuesta alguna para ello.
- PP:
 - ***El PP ha defendido en todo momento la institución monárquica***, y Pablo Casado, además de su apoyo de esta semana, ha hecho bandera de ello.
 - "Monarquía, sin duda. Tenemos una monarquía que es un lujo. ***Tenemos un rey que es un profesional y un patriota (...), que defiende el papel constitucional que le otorgan los españoles***", resaltó Casado.
- Ciudadanos:
 - Rivera también elogió a Felipe VI: "Yo soy de esos que ***me considero más del rey que de la monarquía.*** Creo que tenemos un gran rey, que es un gran activo de España"
 - "Felipe VI -añadió- no es el problema, es parte de la solución (...) Es un gran jefe de Estado".

11. Anexo III, Etiquetado de datos

Se presenta en este anexo un extracto de los datos etiquetados utilizados en el entrenamiento del sistema BERT.

11.1 España Plurinacional

Texto	Etiqueta	Marco
El 1 de octubre no se produjo solo una exhibición del poder de movilización del soberanismo, sino también la expresión de una voluntad mayoritaria de la sociedad catalana de decidir su futuro en las urnas y un ejemplo de movilización pacífica frente a la represión ordenada por el Gobierno	2	España Plurinacional
Cuando el PP forzó que el Tribunal Constitucional (TC) hiciera saltar por los aires el Estatut	2	España Plurinacional
como consecuencia de la decisión del PP y sus magistrados afines en el TC, hoy debemos afirmar que la celebración de un referéndum en Catalunya es imprescindible.	2	España Plurinacional
me indigna como español que la estrategia negacionista hacia el problema catalán por parte de las élites centrales pretenda impedir que Catalunya ayude a la formación de una nueva España.	2	España Plurinacional
configuración estatal que acepte de una vez la realidad plurinacional de nuestra patria y construya un proyecto de país asociado a la justicia social y a la soberanía popular.	2	España Plurinacional
La patria es esa comunidad que asegura que se protege a todos los ciudadanos, que respeta la diversidad nacionalidad, que asegura que todos los niños, sea cual sea el color de su piel, van limpios y calzados a una escuela pública, la patria es esa comunidad que asegura que a los enfermos se les atiende en los mejores hospitales con los mejores medicamentos	2	España Plurinacional
están en marcha en varios territorios del Estado iniciativas constituyentes que afirman	2	España Plurinacional

la soberanía de los pueblos que la habitan" y pide "dar forma al sujeto constituyente aragonés y este que le dé a Aragón la capacidad de actuar en todos estos procesos en marcha y de promover un proceso propio"		
pueblo aragonés	2	España Plurinacional
Podemos quiere crear un "foro constituyente de Aragón" para "dar forma al sujeto constituyente aragonés	2	España Plurinacional
el pueblo valenciano se ha unido para reivindicar un #FinançamentJust. La plurinacionalidad es la única salida para construir una España justa.	2	España Plurinacional
Hay un Madrid que trabaja por un Estado donde el derecho a decidir lo administre la política y el pueblo"	2	España Plurinacional
Antes que poner en riesgo sus privilegios y su impunidad, las élites decidieron arriesgar la integridad territorial de España.	2	España Plurinacional
La victoria de las bases del PSOE contra el aparato de su partido y contra los principales poderes mediáticos del país se basó en tres pilares: plurinacionalidad, mayor cercanía a Podemos y una oposición real al PP en la que no se descartaba la moción de censura.	2	España Plurinacional
me indigna como español que la estrategia negacionista hacia el problema catalán por parte de las élites centrales pretenda impedir que Catalunya ayude a la formación de una nueva España.	2	España Plurinacional
Hacienda federal	2	España Plurinacional
Ayer las mujeres luchaban por el voto, hoy lo hace el pueblo catalán, pero es la misma lucha por toda la ciudadanía	2	España Plurinacional
PP, el PSOE y la nueva extrema derecha que representa Ciudadanos implica de facto la ruptura del pacto territorial	2	España Plurinacional
155 como política de vulneración de derechos y libertades democráticas no tiene por qué quedarse en Catalunya	2	España Plurinacional

situación de administración colonial de Catalunya.	2	España Plurinacional
El virrey Rajoy querrá administrar Catalunya y se encontrará resistencias que solo podrá afrontar con represión y más encarcelamientos.	2	España Plurinacional
El problema de una declaración de independencia no es tanto su ilegalidad (o su unilateralidad) como su ilegitimidad.	2	España Plurinacional
El 1 de octubre no se produjo solo una exhibición del poder de movilización del soberanismo, sino también la expresión de una voluntad mayoritaria de la sociedad catalana de decidir su futuro en las urnas y un ejemplo de movilización pacífica frente a la represión ordenada por el Gobierno	2	España Plurinacional
Un referéndum legal y pactado, además de ser una solución democrática, es la única solución que puede asegurar que Catalunya siga formando parte de España	2	España Plurinacional
hizo saltar por los aires buena parte de las bases del pacto territorial que había hecho viable España como un Estado que integraba una territorialidad plurinacional compleja.	2	España Plurinacional
En cualquier caso, el referéndum no debería limitarse a dos opciones. La opción más deseada en Catalunya (según las encuestas) sería un nuevo encaje constitucional que reconociera a Catalunya como nación y profundizara su autogobierno.	2	España Plurinacional

11.2 España Uninacional

Texto	Etiqueta	Marco
España debe tener un único parlamento, que legisle lo mismo para todos los españoles, con igualdad de derechos y de oportunidades	5	España Uninacional
Fin de las Autonomías.	5	España Uninacional
VOX plantea que las autonomías deberían dejar de ser una fórmula de reparto competencial territorial, y por tanto, no	5	España Uninacional

deberán existir 17 parlamentos, 17 gobiernos, que generan a su vez un estado de duplicidades que sobredimensionan el gasto público, y que no mejoran la vida a los ciudadanos, pero sí la encarecen vía aumentos en los impuestos.		
Devolución de competencias.	5	España Uninacional
modelo territorial sin 17 parlamentos que dividan a los españoles y generen duplicidades de gasto, y más cargas en la que tener políticos	5	España Uninacional
devolviendo competencias al Estado	5	España Uninacional
Todos los españoles tributarán lo mismo con independencia de donde residan.	5	España Uninacional
las consecuencias sobre las tasas e impuestos han generado una gran diferenciación entre territorios.	5	España Uninacional
Armonizar la unidad de mercado. Impulsaremos medidas que faciliten la unidad de mercado	5	España Uninacional
En tanto en cuanto el Estado no recupere la competencia, propondremos normas armonizadoras que garanticen la igualdad entre españoles tanto de servicios prestados como de costes de financiación.	5	España Uninacional
Cobertura sanitaria de ámbito nacional y no autonómica: Una misma cartilla debe ser válida para todos los españoles	5	España Uninacional
Simplificar la legislación y las reglamentaciones burocráticas a nivel de las comunidades autónomas y recuperar ciertas competencias para centralizarlas	5	España Uninacional
Muchas de las cuestiones son de competencia estatal, pero algunas dependen de las comunidades autónomas. Es por ello, que hasta la devolución de competencias VOX considera	5	España Uninacional
se suprimirán los actuales Tribunales Superiores de Justicia de las CCAA	5	España Uninacional
Política de protección del español	5	España Uninacional

se promoverá a través del Instituto Cervantes y la R.A.E. el conocimiento, conservación y difusión de la lengua española y de las demás lenguas regionales por todo el mundo.	5	España Uninacional
Se derogarán cualquiera de las leyes llamadas de inmersión lingüística que impongan como obligatorio el uso de cualquier lengua, o que impidan la libre elección del español en la enseñanza, la Administración, el comercio o la Justicia.	5	España Uninacional
VOX unificará los textos normativos actuales que hacen que los requisitos sean diferentes entre territorios, llegando incluso a limitar hasta la prohibición la fiesta nacional	5	España Uninacional
La España de las autonomías ha generado la desconfianza y mala rivalidad entre españoles, estableciendo barreras mentales en las que estimamos de forma inexacta que los territorios somos competidores entre nosotros, desplazando la idea de que todos somos compatriotas y que unidos podremos lograr una mayor prosperidad.	5	España Uninacional
VOX apuesta por cultivar la identidad española, y fomentar los valores de la solidaridad y el afecto entre territorios.	5	España Uninacional
Santiago Abascal, ha acusado a PSOE, PP y Ciudadanos, de ser "cómplices" con el independentismo en Cataluña, mientras que ha presentado a su partido como el único con la "voluntad firme" de defender la unidad de España.	5	España Uninacional
España ni se negocia, ni se rinde, ni se dialoga, sino que se defiende con todas las consecuencias	5	España Uninacional

11.3 España Monárquica

Texto	Etiqueta	Marco
En España han fallado los políticos y han funcionado los tribunales y la corona	6	España Monárquica

El Rey entrega los Despachos de secretario de Embajada a la LXXII promoción de la Carrera Diplomática	6	España Monárquica
El Rey conversa por videoconferencia con distintas unidades de las Fuerzas Armadas, en el Cuartel General del Estado Mayor.	6	España Monárquica
Saludos previos a la Cena de Estado. Visita de Estado de la República de Corea a España.	6	España Monárquica
Viva El Rey De España (VERDE)	6	España Monárquica
V.E.R.D.E	6	España Monárquica
defenderá "hasta la última gota de sangre" la Monarquía aplicando "ley y orden"	6	España Monárquica
los españoles "apoyan a la Corona" y "la nación no va a permitir" que "una cuadrilla de políticos sin escrúpulos acabe con una de las instituciones más ligadas a la unidad nacional".	6	España Monárquica
"el pueblo español apoya a la Corona" y quieren salvaguardar su figura como "garante de la unidad y permanencia de España".	6	España Monárquica
la Corona "goza de muy buena salud", incluso "mucho mejor" que otras instituciones como los partidos políticos, la justicia o los medios de comunicación	6	España Monárquica
a sociedad española valora a la Corona" y ve a Felipe VI como "la cabeza de la nación".	6	España Monárquica
Felipe VI "reina pero no gobierna"	6	España Monárquica
El Rey ha dejado claro desde el minuto uno que la Jefatura del Estado, como no puede ser de otra manera, que es garante de la unidad nacional	6	España Monárquica
entre sus funciones como Rey está la de sancionar las leyes que el Gobierno de Pedro Sánchez apruebe en Consejo de Ministros.	6	España Monárquica
que la Corporación cumpla con su obligación legal de defender y difundir la actividad relativa a la Corona e incluso le sugiere que le dedique más horas para no "ningunear" al Jefe del Estado.	6	España Monárquica

"justo cuando el país más necesitaba los valores de unidad y estabilidad que representa la Monarquía".	6	España Monárquica
guardar y proteger el orden constitucional y la Unidad de España, manteniendo siempre la máxima lealtad hacia la Corona y la Carta Magna.	6	España Monárquica
Audiencia del Rey a un grupo de coroneles y capitanes de Navío	6	España Monárquica
El Rey, en Valencia de Alcántara para visitar el Puesto de la Guardia Civil perteneciente a la Comandancia de Cáceres	6	España Monárquica
Visita del Rey al Consejo de Estado con motivo de la presentación de la Memoria del año 2020.	6	España Monárquica
El Rey, en el Congreso de los Diputados. 40º Aniversario del 23 de febrero de 1981.	6	España Monárquica
La Corona es el mejor símbolo de España	6	España Monárquica
Defensa de los valores que encarnan el Rey y la Cultura Monárquica	6	España Monárquica
la cultura monárquica como vehículos de convivencia social, progreso económico y bienestar para todos los españoles	6	España Monárquica

11.4 España Republicana

Texto	Etiqueta	Marco
La República es un hito democrático en nuestro país. Hoy tenemos que extraer la lección política: construir pueblo, ganar el país #14abril	7	España Republicana
"Monarquía para mí no es democracia". Irene es una de los muchos jóvenes que este domingo han conmemorado la II República en Madrid y han pedido un referéndum para poder decidir el modelo de Estado en el que quiere vivir.	7	España Republicana
no hay dos sin tres, República otra vez	7	España Republicana
Felipe, acelera, que llega la tercera (republica)	7	España Republicana
España mañana será republicana	7	España Republicana
Estamos aquí como cada 14 de abril reivindicando una república como único modelo de Estado verdaderamente democrático	7	España Republicana
Qué se proponían los legisladores republicanos? Algo tan simple como: -- Implantar y consolidar en España una República democrática y burguesa.	7	España Republicana
¡Viva la República!	7	España Republicana
No pasaran	7	España Republicana
Ni todos los conventos de Madrid valen la vida de un republicano	7	España Republicana
El espíritu constituyente del 15M debe impulsar la nueva España a la que aspiramos; social, republicana y plurinacional.	7	España Republicana
reserva democrática de valor incalculable, un espíritu republicano que debe dejar de ser una nostalgia asociada a símbolos del siglo xx para acompañar el impulso constituyente que inició el 15M.	7	España Republicana
La república es la única esperanza que tenemos	7	España Republicana

Pensamos que nuestro país, en lo económico y en términos de respeto de las identidades será mejor cuando haya una República en España y vamos a trabajar para que haya una República en España, aunque esto moleste a sectores del Estado profundo que tienen muchas dificultades para aceptar los resultados de la democracia cuando no ganan los suyos	7	España Republicana
convoquen “urnas” para que “la gente pueda elegir la jefatura del Estado, para que sea elegida democráticamente y para que responda de sus actos”, al entender que la Corona está “a punto de implosionar” por una “crisis de materiales”.	7	España Republicana
Mayoral ha recordado que su formación es republicana	7	España Republicana
en otras ocasiones históricas se han producido cambios de régimen hacia la república después de unas elecciones municipales o “tras la salida de los componentes de la jefatura del Estado, como Isabel II”	7	España Republicana

12. Anexo IV. Tablas y gráficos

Entidad Política	Eje 4 (Plurinacional-Uninacional)			Eje 5 (Monarquía-Republica)			Tipo de Entidad	Correlación PEARSON (PN1, PN2)
	PB1	PR1	PN1	PB2	PR2	PN2		
VOX	247	163,17%	7,61%	679	1128,79%	20,92%	Partido Politico	-68,58%
PSOE	-114	80,81%	-3,51%	822	2517,65%	25,30%	Partido Politico	
PP	-83	87,09%	-2,55%	993	1282,14%	30,55%	Partido Politico	
PODEMOS	208	149,76%	6,42%	606	1831,43%	18,69%	Partido Politico	
Cs	-19	97,31%	-0,58%	934	2624,32%	28,74%	Partido Politico	
IU	344	179,81%	10,59%	738	1704,35%	22,72%	Partido Politico	
IzRepublicana	79	128,52%	2,49%	603	1216,67%	18,97%	Partido Politico	
FE de las JONS	209	140,74%	6,60%	682	554,67%	21,53%	Partido Politico	
PCPE	-182	66,67%	-5,60%	859	7909,09%	26,44%	Partido Politico	
Unión Monárquica de España	402	192,63%	12,53%	-1149	21,57%	-35,82%	Organización	-74,58%
Real Instituto Elcano	350	278,57%	10,78%	194	670,59%	5,98%	Organización	
Casa Real	1012	495,31%	31,14%	-1412	11,53%	-43,45%	Organización	
Convivencia Civica Catalana	262	142,46%	8,11%	959	1255,42%	29,70%	Organización	
Público	332	228,19%	10,22%	618	1607,32%	19,02%	Medios	85,86%
Marca	9	212,50%	0,28%	11		0,34%	Medios	
El Pais	367	238,49%	11,29%	1040	1353,01%	32,00%	Medios	
El Nacional.Cat	182	180,53%	5,60%	512	1144,90%	15,75%	Medios	
El Diario.es	343	263,33%	10,55%	674	2908,33%	20,74%	Medios	
ABC	329	189,40%	10,12%	1190	1467,82%	36,62%	Medios	
Ruben Sanchez	192	268,42%	5,92%	249	1878,57%	7,68%	Individuo	-78,82%
Raul Alvarez	73	661,54%	2,26%	28	333,33%	0,87%	Individuo	
Ramón Cotarelo	-139	73,32%	-4,31%	538	3264,71%	16,69%	Individuo	
J.C. Monedero	-141	75,44%	-4,35%	695	1978,38%	21,45%	Individuo	
Mikel Ayestaran	86	151,50%	2,65%	336	1868,42%	10,34%	Individuo	
Pablo Iglesias - El proyecto de	-6	64,71%	-15,38%	20	766,67%	51,28%	Documento	-74,73%
Pablo Iglesias - Miembros de la	6	146,15%	17,14%	9	212,50%	25,71%	Documento	
La Monarquía, cinco razones p	10	171,43%	21,74%	-11	52,17%	-23,91%	Documento	
VOX 52-medidas-autonomias	61	2133,33%	78,21%	5	155,56%	6,41%	Documento	
PODEMOS - Consideraciones C	-38	25,49%	-58,46%	45	662,50%	69,23%	Documento	
Santiago Abascal	379	176,88%	11,67%	823	891,35%	25,33%	Dirigente Politico	-9,32%
Pedro Sanchez	396	164,60%	12,18%	667	625,20%	20,52%	Dirigente Politico	
Pablo Iglesias	432	199,54%	13,58%	717	1478,85%	22,55%	Dirigente Politico	
Pablo Echenique	13	208,33%	9,15%	16	633,33%	11,27%	Dirigente Politico	
Pablo Casado	404	155,34%	12,43%	1161	979,55%	35,73%	Dirigente Politico	
Ines Arrimadas	224	131,37%	6,92%	1089	2278,00%	33,64%	Dirigente Politico	
Fulgencio Coll	118	257,33%	11,61%	145	637,04%	14,27%	Dirigente Politico	

Ilustración 125 Resultados detallados en dos ejes bajo estudio

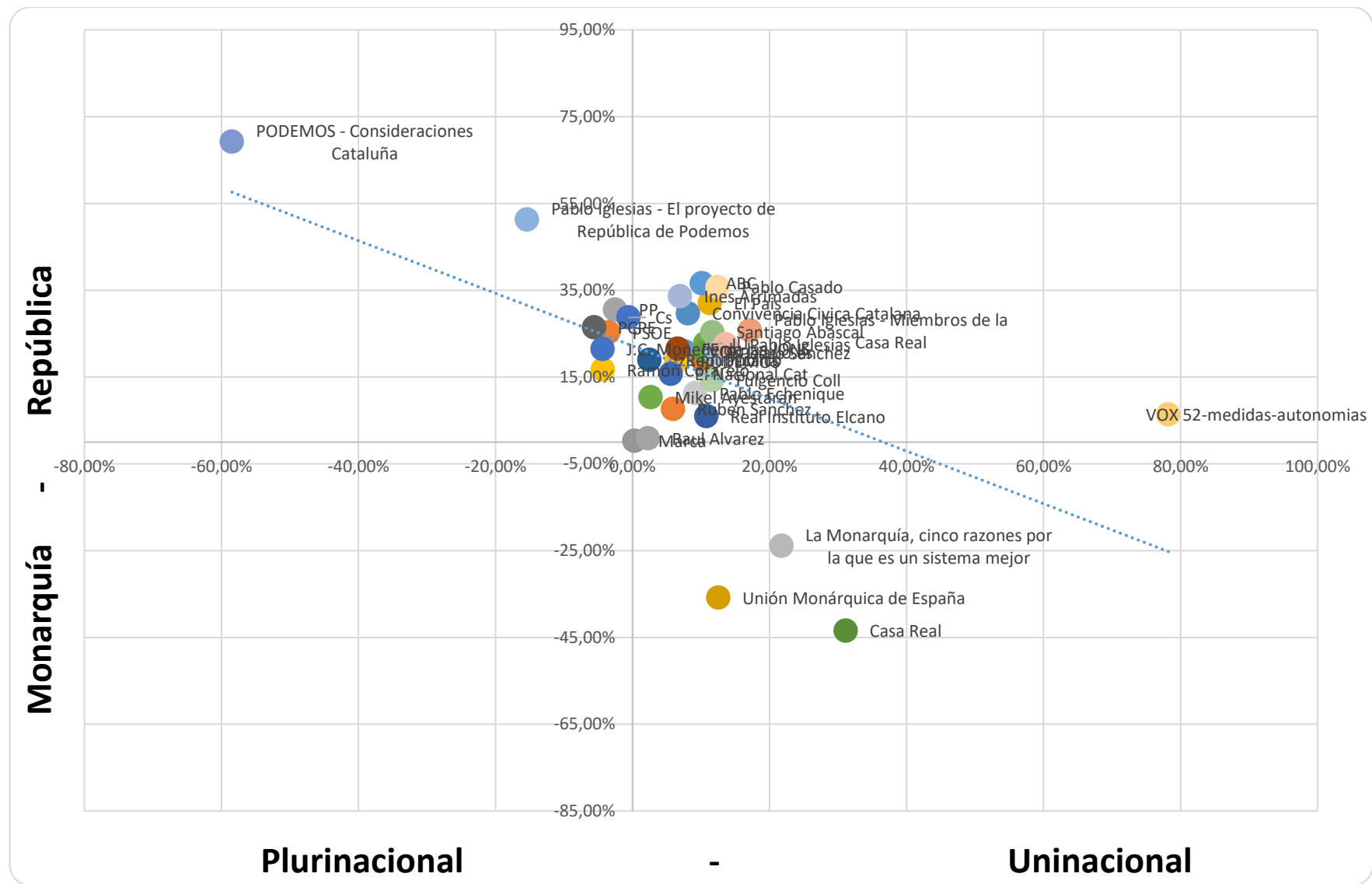


Ilustración 126 Ubicación de entidades políticas en los cuatro cuadrantes

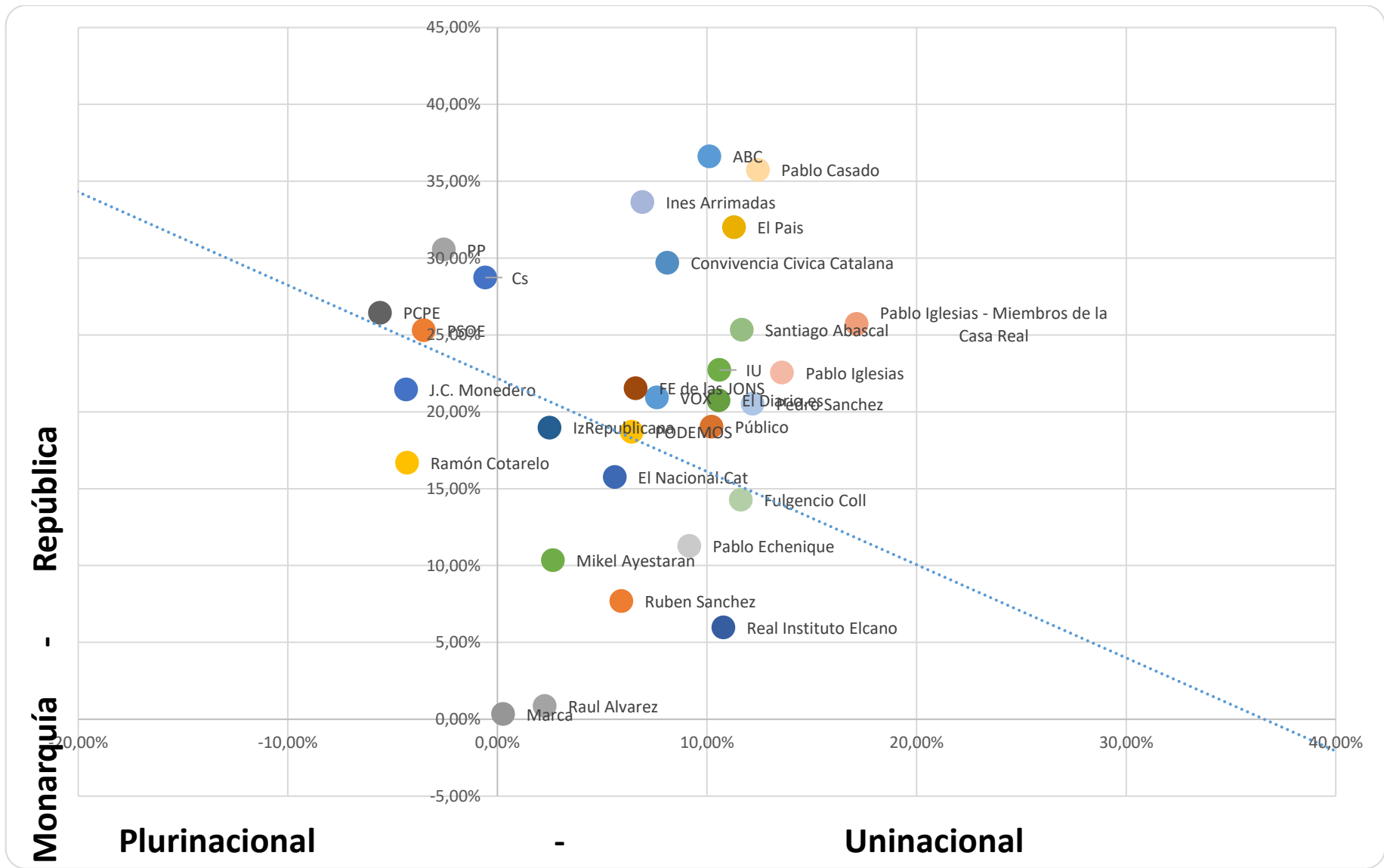


Ilustración 127 Ubicación de entidades políticas en el cuadrante República-Uninacional

Ubicación en eje: (-) Monarquía / (+) República	Textos o Tuits evaluados	Marco 2 (República) ++	Marco 0 (No marco)	Marco 1 (Monarquía) -	Polaridad Bruta (Marco 2- Marco 1)	Polaridad Relativa (Marco 1) / (Marco 2)	Polaridad Neta (PB / Textos evaluados)
CasaReal	3250	559	1110	1581	-1022	35%	-31,4%
ume_umes (Unión Monar	3203	764	1387	1052	-288	73%	-9,0%
marca	3250	12	3238	0	12	N/A	0,4%
raul_alvarez_pr	3232	63	3164	5	58	1260%	1,8%
RubenSanchezTW	3248	289	2957	2	287	14450%	8,8%
FulgencioColl	1009	193	798	18	175	1072%	17,3%
elnacionalcat_e	3250	574	2669	7	567	8200%	17,4%
ramoncotarelo	3226	572	2652	2	570	28600%	17,7%
IzRepublicana	3177	668	2501	8	660	8350%	20,8%
podemos	3243	706	2527	10	696	7060%	21,5%
psoe	3250	753	2485	12	741	6275%	22,8%
vox_es	3241	766	2448	27	739	2837%	22,8%
IzquierdaUnida	3245	852	2378	15	837	5680%	25,8%
MonederoJC	3246	885	2342	19	866	4658%	26,7%
CiudadanosCs	3242	891	2342	9	882	9900%	27,2%
sanchezcastejon	3250	990	2183	77	913	1286%	28,1%
PabloIglesias	3169	914	2234	21	893	4352%	28,2%
fedelasjons	3165	966	2129	70	896	1380%	28,3%
santi_abascal	3246	1002	2213	31	971	3232%	29,9%
CCivicaCatalana	3228	1113	2091	24	1089	4638%	33,7%
populares	3246	1151	2065	30	1121	3837%	34,5%
InesArrimadas	3246	1258	1962	26	1232	4838%	38,0%
elpais_espana	3260	1289	1938	33	1256	3906%	38,5%
abcespana	3249	1350	1872	27	1323	5000%	40,7%
pablocasado	3248	1540	1625	83	1457	1855%	44,9%

Ilustración 128 Resultados eje "Monarquía-República"

Ubicación en eje: (-) Plurinacionalidad / Nación única (+)	Textos o Tuits evaluados	Marco 0 (Estado Plurinacional) --	Marco 1 (No marco)	Marco 2 (Estado Uninacional) ++	Polaridad Bruta (Marco 2- Marco 0)	Polaridad Relativa (Marco 2) / (Marco 0)	Polaridad Neta (PB / Textos evaluados)
ramoncotarelo	3226	542	2298	386	-156	71%	-4,8%
psoe	3250	583	2236	431	-152	74%	-4,7%
MonederoJC	3220	569	2218	433	-136	76%	-4,2%
populares	3244	655	2061	528	-127	81%	-3,9%
marca	3248	4	3225	19	15	475%	0,5%
CiudadanosCs	3240	616	1993	631	15	102%	0,5%
raul_alvarez_pr	3224	11	3122	91	80	827%	2,5%
elnacionalcat_e	3209	226	2631	352	126	156%	3,9%
CCivicaCatalana	2137	454	1124	559	105	123%	4,9%
podemos	3219	411	2225	583	172	142%	5,3%
RubenSanchezTW	3139	105	2751	283	178	270%	5,7%
vox_es	2910	356	1998	556	200	156%	6,9%
InesArrimadas	3246	698	1602	946	248	136%	7,6%
abcespana	3249	366	2208	675	309	184%	9,5%
santi_abascal	2856	429	1696	731	302	170%	10,6%
elpais_espana	3250	272	2358	620	348	228%	10,7%
IzquierdaUnida	3240	416	2060	764	348	184%	10,7%
sanchezcastejon	3249	632	1633	984	352	156%	10,8%
pablocasado	3245	740	1385	1120	380	151%	11,7%
PabloIglesias	3167	432	1871	864	432	200%	13,6%

Ilustración 129 Resultados eje "Plurinacionalidad-Nación única"

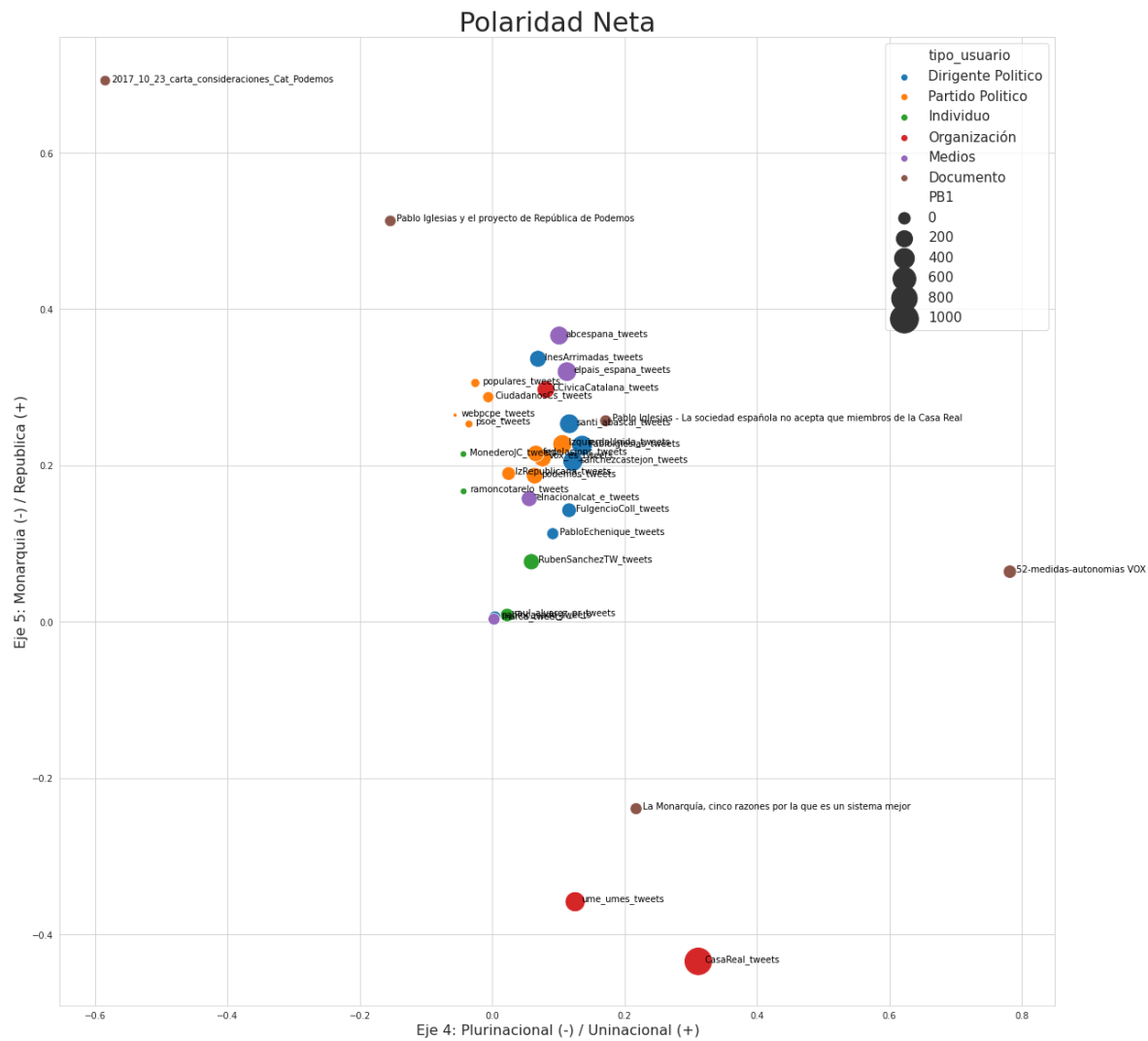


Ilustración 130 Gráfico "Polaridad Neta"

Espacio Vectorial Pragmático

#	Dimensiones	Ámbito	Extremo (--) del Eje	(-)	Centro (0) del Eje	(+)	Extremo (++) del Eje
1	Ideas políticas (izquierda/derecha)	Político	Izquierda		Centrismo político		Derecha
2	Libertad Económica	Político / Económico	Libertarismo (estado no necesario)		Economía Mixta		Estatismo (estado imprescindible)
3	Organización Política Mundial	Político	Localismo		Estado Unitario		Globalismo
4	Organización Territorial del Estado	Político	Independentismo	Federalismo	Autonomismo		Centralismo
5	Modelo Político del Estado (Forma de Gobierno)	Político	Dictadura	Dictablanda	Monarquía Constitucional	Monarquía Parlamentaria	República
6	Integración supranacional (UE)	Político	Euroescepticismo Fuerte	-	Euroescepticismo suave		Europeísmo
7	Dimensión sociocultural	Político / Social	Liberalismo cultural		Reformismo		Conservadurismo Social
8	Diversidad cultural	Político / Étnico	Multiculturalismo		Interculturalismo		Nacionalismo Étnico
9	Apoyo Progreso Tecnológico	Tecnología	Neoludismo		Tecnorrealismo		Tecnoutopismo
10	Equipo de Fútbol Preferido	Deportivo / Político	FC Barcelona	Lleida	Real Betis y otros	RCD Espanyol	Real Madrid CF

Ilustración 132 Tabla definición dimensiones del espacio vectorial pragmático

#	Dimensiones	Ámbito	Extremo (--) del Eje	(-)	Centro (0) del Eje	(+)	Extremo (++) del Eje
1	Ideas políticas (izquierda/derecha)	Político	Izquierda		Centrismo político		Derecha
2	Libertad Económica	Político / Económico	Libertarismo (estado no necesario)		Economía Mixta		Estatismo (estado imprescindible)
3	Organización Política Mundial	Político	Localismo		Estado Unitario		Globalismo
4	Organización Territorial del Estado	Político	Independentismo	Federalismo	Autonomismo		Centralismo
5	Modelo Político del Estado (Forma de Gobierno)	Político	Dictadura	Dictablanda	Monarquía Constitucional	Monarquía Parlamentaria	República
6	Integración supranacional (UE)	Político	Euroescepticismo Fuerte	-	Euroescepticismo suave		Europeísmo
7	Dimensión sociocultural	Político / Social	Liberalismo cultural		Reformismo		Conservadurismo Social
8	Diversidad cultural	Político / Étnico	Multiculturalismo		Interculturalismo		Nacionalismo Étnico
9	Apoyo Progreso Tecnológico	Tecnología	Neoludismo		Tecnorealismo		Tecnoutopismo
10	Equipo de Fútbol Preferido	Deportivo / Político	FC Barcelona	Lleida	Real Betis y otros	RCD Espanyol	Real Madrid CF

Marco 7	"España Republicana"	Eje 5	"Modelo político del Estado"	PODEMOS
Marco 2	"Estado Plurinacional"	Eje 4	"Organización Territorial del Estado"	

Ilustración 133 Ubicación de los marcos utilizados por PODEMOS en el espacio vectorial pragmático

#	Dimensiones	Ámbito	Extremo (--) del Eje	(-)	Centro (0) del Eje	(+)	Extremo (++) del Eje
1	Ideas políticas (izquierda/derecha)	Político	Izquierda		Centrismo político		Derecha
2	Libertad Económica	Político / Económico	Libertarismo (estado no necesario)		Economía Mixta		Estatismo (estado imprescindible)
3	Organización Política Mundial	Político	Localismo		Estado Unitario		Globalismo
4	Organización Territorial del Estado	Político	Independentismo	Federalismo	Autonomismo		Centralismo
5	Modelo Político del Estado (Forma de Gobierno)	Político	Dictadura	Dictablanda	Monarquía Constitucional	Monarquía Parlamentaria	República
6	Integración supranacional (UE)	Político	Euroescepticismo Fuerte	-	Euroescepticismo suave		Europeísmo
7	Dimensión sociocultural	Político / Social	Liberalismo cultural		Reformismo		Conservadurismo Social
8	Diversidad cultural	Político / Étnico	Multiculturalismo		Interculturalismo		Nacionalismo Étnico
9	Apoyo Progreso Tecnológico	Tecnología	Neoludismo		Tecnorrealismo		Tecnoutopismo
10	Equipo de Fútbol Preferido	Deportivo / Político	FC Barcelona	Lleida	Real Betis y otros	RCD Espanyol	Real Madrid CF

Marco 5	"España Monárquica"	Eje 5	"Modelo político del Estado"	VOX
Marco 6	"Estado Uninacional"	Eje 4	"Organización Territorial del Estado"	

Ilustración 134 Ubicación de los marcos utilizados por VOX en el espacio vectorial pragmático

© 2021 por Raúl Álvarez Prieto

Todos los contenidos de este documento están sujetos a derechos de propiedad por las leyes de Derechos de Autor y demás Leyes relativas Internacionales a Raúl Álvarez Prieto y de terceros titulares de los mismos que han autorizado debidamente su inclusión. En ningún caso se entenderá que se concede licencia alguna o se efectúa renuncia, transmisión, cesión total o parcial de dichos derechos ni se confiere ningún derecho, y en especial, de alteración, explotación, reproducción, distribución o comunicación pública sobre dichos contenidos sin la previa autorización expresa de Raúl Álvarez Prieto o de los titulares correspondientes. El uso de imágenes, fragmentos de vídeos y demás material que sean objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos, y cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor. Queda prohibido copiar, reproducir, distribuir, publicar, transmitir, difundir, o en cualquier modo explotar cualquier parte de este documento sin la autorización previa por escrito de Raúl Álvarez Prieto o de los titulares correspondientes.