

Psicología, lenguaje y comunicación: análisis con la herramienta LIWC de los discursos y tweets de los candidatos a las elecciones gallegas de 2012

Mercedes Fernández-Cabana
Universidad de Santiago de Compostela

José Rúas-Araújo
Universidad de Vigo

M. Teresa Alves-Pérez

Grupo NECOM (Neurocomunicación, publicidad y política)

Se analizaron los mítines de las jornadas de apertura y cierre de las elecciones así como los tweets publicados, durante los 15 días de duración oficial de la campaña, por los candidatos.

Se aplicó la prueba Kruskal-Wallis para la determinación de diferencias en el porcentaje de uso de categorías de palabras entre los textos de los siete candidatos, resultando significativas ($p \leq 0,01$) en 16 de las categorías analizadas.

Se discuten las posibles interpretaciones de estos resultados, resaltando la utilidad del programa LIWC como estrategia complementaria para el análisis de textos en política.

Palabras clave: análisis de textos, LIWC, comunicación, estilo de lenguaje.

Psychology, language and communication: Analysis with the tool LIWC of the speeches and tweets from the candidates to 2012 Galician elections

Linguistic styles of the 7 major candidates for election to the Parliament of Galicia in 2012 were analyzed using the Spanish version of the Linguistic Inquiry and Word Count (LIWC) program.

We analyzed the speeches of the opening and closing sessions of the campaign and the tweets posted during the 15-day official campaign period by the candidates.

The Kruskal-Wallis test was used to determine differences in the percentage of use of categories of words in the texts of the seven candidates, resulting significant ($p \leq 0.01$) in 16 of the categories analyzed.

Possible interpretations of these results are discussed, highlighting the usefulness of LIWC as a complementary strategy for the analysis of political texts.

Keywords: Text analysis, LIWC, communication, language style.

Introducción

En el análisis de textos y discursos políticos en los medios digitales, se clasifican tres tipos básicos de corrientes teóricas: las que buscan monitorizar los discursos, las que analizan sentimientos y reacciones frente a los comportamientos de los políticos y, finalmente, las que intentan predecir resultados electorales a partir de lo publicado en las redes sociales.

La aplicación de las denominadas tecnologías de minería para valorar el caudal de los tweets generados por los usuarios, dio lugar a una corriente de investigación favorable sobre su capacidad de predicción de resultados electorales (e.g. Tumasjan, Sprenger, Sandner y Welpe, 2010) y otra crítica (e.g. Gayo-Avello, Metaxas y Mustafaraj, 2011). En España Congosto y Aragón (2012), trataron de medir la participación ciudadana en Twitter en las elecciones generales del 2011, y analizar el nivel de emotividad de los políticos en los mensajes; mientras que Borondo, Morales, Losada y Benito (2012) observaron que el 40% de dichos mensajes eran re-tweets, lo que les llevó a concluir la falta de discusión y debate sobre dicha campaña en Twitter, además de sugerir que la atención colectiva estaba dirigida por un porcentaje muy pequeño de usuarios.

En cualquier caso, parece existir cierto consenso a la hora de considerar la contribución de las redes sociales en la activación del denominado capital social (Gil de Zúñiga, Jung y Valenzuela, 2011) y la modificación de las conductas de las personas que las utilizan.

Una herramienta de análisis lingüístico que se ha aplicado también al lenguaje en política, es el programa *Linguistic Inquiry and Word Count* (LIWC) de Pennebaker, Francis y Booth (2001), desarrollado para la evaluación cognitiva y emocional de textos a través de una serie de categorías psicológicas y estructurales. Este programa analiza textos, palabra por palabra, en una clasificación de 72 variables lingüísticas, que incluyen categorías de lenguaje estándar (artículos, preposiciones, pronombres, etc.), procesos psicológicos (categorías de emociones positivas y negativas, variables cognitivas, etc.), palabras relacionadas con la relatividad (tiempo y espacio) y dimensiones tradicionales de contenido (trabajo, dinero, logro, etc.) organizadas de forma jerárquica, y ofrece sus resultados en forma de porcentaje de uso de cada una de dichas categorías. El LIWC 2001 ha sido validado en español (Ramírez-Esparza, Pennebaker, García y Suriá, 2007), lo cual ha supuesto un aumento del número de categorías analizadas así como del número de palabras y raíces

de palabras incluidas en su diccionario (más de 7000). Las estrategias de análisis cuantitativo como el LIWC, asumen que las palabras que elegimos al hablar dan información independientemente de su sentido literal o del contexto semántico en que se usen; y que dicha elección ofrece información sobre quiénes somos, a quién hablamos y en qué contexto lo hacemos, así como sobre nuestro género, edad, estatus y motivaciones (Pennebaker, Mehl y Nierderhoffer, 2003).

El LIWC se ha utilizado para analizar comparativamente el lenguaje de hombres y mujeres (Newman, Groom, Handelman y Pennebaker, 2008), evaluar sus modificaciones a lo largo del ciclo vital (Pennebaker y Stone, 2003) o comparar historias falsas y reales redactadas por sujetos (Newman, Pennebaker, Berry y Richards, 2003). El lector puede consultar la excelente revisión de Tausczik y Pennebaker (2010) para conocer otros estudios sobre el LIWC.

En el ámbito de la política, se ha aplicado al análisis de tweets durante la campaña electoral alemana del 2008 (Tumasjan et al., 2010) o al análisis de las transcripciones de los comunicados del grupo terrorista Al-Qaeda en comparación con otros grupos extremistas (Pennebaker y Chung, 2008). Pennebaker y Lay (2002) analizaron el lenguaje de R. Giuliani durante su mandato como alcalde de la ciudad de Nueva York, y hallaron que éste fue variando a medida que se enfrentaba a crisis tanto personales como de su comunidad, y que estas variaciones en su lenguaje suponían también cambios en el modo en que era percibido por sus conciudadanos. Slatcher, Chung, Pennebaker y Stone (2007) analizaron el estado psicológico y la personalidad de los candidatos a la presidencia y a la vicepresidencia de EEUU a través de su estilo lingüístico, en este caso buscando indicadores de características como honestidad (con un mayor porcentaje de uso de auto referencias, referencias a otros y palabras que indican exclusión) o complejidad cognitiva (aumento de palabras que indican exclusión, palabras largas, tentativas, negaciones y discrepancias). Estos autores hallaron que el lenguaje de Dick Cheney era el que mostraba una mayor complejidad cognitiva, que los republicanos mostraban un estilo más optimista y que el lenguaje de los aspirantes a la vicepresidencia (Dick Cheney y John Edwards) mostraba más indicadores de honestidad que el lenguaje de los aspirantes a la presidencia (George W. Bush y John Kerry). Todos estos estudios se aplicaron a muestras de texto en inglés.

En nuestro país, Carrera-Fernández, Guardia-Olmos y Peró-Cebollero (2013) aplicaron la versión española del LIWC al análisis del estilo lingüístico de los candidatos del PP y el PSOE a la presidencia del Gobierno de España, comparando, entre otros aspectos, sus intervenciones en los debates televisados celebrados en las campañas de las elecciones generales de 2008 y 2011. También se ha aplicado el LIWC en español al análisis de los mítines de los candidatos a las elecciones al Parlamento de Galicia, celebradas en año 2012 (Rúas, Fernández-Cabana y Puentes, 2013); aunque hasta el momento es muy reducida la aplicación del LIWC al análisis de textos políticos en español.

En este trabajo combinamos el análisis lingüístico de los discursos de apertura y cierre de campaña de los candidatos a las elecciones al Parlamento de Galicia del

2012, con el análisis de los tweets publicados en sus cuentas de Twitter oficiales durante la campaña, ya que entendemos que las redes sociales han de ser cada vez más tenidas en cuenta como posibles medios de influencia. Nuestro objetivo ha sido analizar si existen diferencias en el lenguaje utilizado por cada uno de ellos, partiendo de la idea de que las palabras que utilizan producen impresiones en los votantes y que éstas influyen, al menos en parte, la elección del voto (Slatcher *et al.*, 2007).

Material y Métodos

Analizamos con el LIWC los tweets publicados por los siete principales candidatos a las elecciones gallegas de Octubre de 2012, así como sus discursos de apertura y cierre de campaña. Los candidatos estudiados fueron Xoan Bascuas (candidato por el partido Compromiso por Galicia, CxG), Xosé Manuel Beiras (candidato por Alternativa Galega de Esquerda, AGE), Francisco Jorquera (candidato del Bloque Nacionalista Galego, BNG), Pachi Vázquez (candidato del Partido Socialista de Galicia, PSdG), José Canedo (candidato por Unión Progreso y Democracia, UPyD), Alberto Núñez Feijoo (candidato del Partido Popular, PP), y Mario Conde (candidato por Sociedad Civil y Democracia, SCD).

Recopilamos de forma retrospectiva todos los mensajes publicados en Twitter por dichos candidatos durante la campaña electoral (entre los días 5 y 19 octubre). Los tweets fueron recogidos de la cuenta en Twitter oficial de cada candidato:

<https://twitter.com/Feijoo2012>
<https://twitter.com/pachivazquez>
<https://twitter.com/jorqueracaselas>
https://twitter.com/Xoan_Bascuas
<https://twitter.com/mariocondeconde>
<https://twitter.com/canedojose>
<https://twitter.com/xmbeiras>.

Decidimos desechar aquellos tweets que consistían en meras convocatorias así como aquellos que únicamente incluían links a otras webs. Por otra parte, recogimos la transcripción de sus discursos electorales de apertura y cierre de campaña (en el caso del candidato de SDC únicamente el de cierre, ya que no dio discurso de apertura). Contamos en suma con 2060 tweets (que incluían 37214 palabras) y 13 discursos (con 26328 palabras), por lo que el total de palabras que analizamos fue de 63542 (ver tabla 1, página siguiente).

Dado que no existe un diccionario del LIWC validado en gallego, los textos que estaban en dicha lengua fueron revisados ortográficamente con la ayuda de un filólogo especialista en lengua gallega y traducidos automáticamente al castellano (siguiendo el procedimiento empleado por Tumasjan *et al.* en 2010). Los tweets

de cada candidato fueron concatenados en un texto continuo que dividimos en cinco partes con similar número de palabras; en el caso de los discursos electorales, dividimos el conjunto de palabras de cada candidato en cuatro partes aproximadamente iguales para su análisis; obteniendo por tanto nueve muestras de texto para cada uno de los candidatos. Decidimos dividir los textos de esta forma para lograr un mayor número de muestras para cada candidato, contribuyendo así a mejorar el cálculo estadístico.

TABLA 1: NÚMERO DE TWEETS Y PALABRAS ANALIZADOS PARA CADA CANDIDATO EN EL CONJUNTO DE TWEETS Y EN SUS DISCURSOS, Y NÚMERO TOTAL DE PALABRAS ANALIZADAS CON EL LIWC.

Candidatos (partido)	<i>n° tweets analizados</i>	<i>n° total de palabras en tweets</i>	<i>n° palabras discurso apertura</i>	<i>n° palabras discurso cierre</i>	<i>n° total palabras en discursos</i>	<i>n° total de palabras analizadas</i>
<i>Bascuas (CxG)</i>	112	2139	1321	1331	2652	4791
<i>Beiras (AGE)</i>	671	10779	2726	2734	5460	16239
<i>Canedo (UPyD)</i>	62	1049	502	468	970	2019
<i>Conde (SCD)</i>	90	1705	0	5967	5967	7672
<i>Feijoo (PP)</i>	24	417	891	938	1829	2246
<i>Jorquera (BNG)</i>	357	7821	1585	1579	3164	10985
<i>P. Vázquez (PSdG)</i>	744	13304	3157	3129	6286	19590

Los textos obtenidos fueron analizados con la versión en español del programa LIWC 2001 (Ramírez-Esparza *et al.*, 2007) en 32 de sus categorías seleccionadas según la bibliografía previa sobre uso del LIWC en análisis de textos políticos (Pennebaker y Lay, 2002; Slatcher *et al.*, 2007; Tumasjan *et al.*, 2010). Las categorías analizadas fueron: SEISLTR (palabras de más de seis letras), PRONOM (total de uso de pronombres), YO (primera persona del singular), NOSOTROS (primera persona del plural), TU (uso total de la segunda persona), OTRO (total pronombres de tercera persona), NEGACIÓN (uso de negaciones), ARTÍCULO (artículos), PREPO (preposiciones), AFECTIVA (procesos afectivos), EMOPOS (emociones positivas), EMONEG (emociones negativas), ANSIEDAD, ENOJO, TRISTEZA, MECCOG (procesos cognitivos), CAUSA (causa y efecto), INSIGHT (entendimiento), DISCREP (discrepancias), INHIB (inhibiciones), TENTAT (tentativos), CERTEZA, SOCIAL (procesos sociales), COMU (comunicación), PASADO (verbos en tiempo pasado), PRESENTE (verbos en tiempo presente), FUTURO (verbos en tiempo futuro), INCL (palabras que indican inclusión), EXCL (exclusión), TRABAJO, LOGRO, y DINERO.

Análisis estadístico

Se realizó un análisis descriptivo de los datos, las variables continuas se mostraron como media, desviación típica, mediana, mínimo y máximo, y las categóricas como frecuencia y porcentaje. Para la comparación de las variables se utilizó la prueba no paramétrica Kruskal-Wallis (para la determinación de diferencias en el porcentaje de uso de categorías de palabras entre los textos de los siete candidatos) y la prueba U de Mann Whitney para determinar entre qué candidatos se daba esa diferencia. Se realizaron gráficos de líneas para las variables significativas. Se consideró estadísticamente significativos aquellos “*p*-valores” menores a 0,01.

Resultados

Hallamos diferencias significativas en el porcentaje de uso de 16 de las categorías analizadas: SEISLTR ($p < ,001$), YO ($p = ,004$), NEGACIÓN ($p < ,001$), ARTÍCULO ($p = ,001$), EMONEG ($p = ,001$), ANSIEDAD ($p = ,001$), ENOJO ($p = ,001$), TRISTEZA ($p = ,002$), INSIGHT ($p = ,003$), INHIB ($p = ,001$), TENTAT ($p = ,005$), CERTEZA ($p = ,002$), COMU ($p = ,010$), PRESENTE ($p < ,001$), FUTURO ($p = ,002$), y DINERO ($p = ,007$) (ver tablas 2 a 5 y figuras 1 a 4).

TABLA 2: PORCENTAJE DE USO DE LAS CATEGORÍAS LIWC “PALABRAS CON MÁS DE SEIS LETRAS”, “ARTÍCULOS” Y “VERBOS EN TIEMPO PRESENTE” PARA CADA CANDIDATO.

<i>LIWC categ. / Candidatos</i> <i>Media (DT)</i> <i>Mediana (min-max)</i>	<i>Palabras</i> <i>> 6 letras</i>	<i>Artículos</i> <i>(el, la, los)</i>	<i>Verbos tiempo</i> <i>presente (hago, voy)</i>
Bascuas	29,12 (5,04) 27,80 (22,85-35,29)	11,40 (1,43) 11,35 (9,58-13,41)	8,52 (1,90) 8,47 (6,21-12,23)
Beiras	27,84 (1,83) 28,12 (24,56-31,21)	13,44 (0,96) 13,05 (12,42-14,86)	7,34 (1,24) 7,58 (5,53-9,03)
Canedo	22,52 (3,57) 21,84 (17,99-27,54)	10,65 (2,37) 10,17 (7,95-14,49)	9,47 (2,21) 9,43 (7,25-14,45)
Conde	23,73 (3,18) 22,86 (18-27,38)	10,34 (1,39) 9,45 (8,95-12,80)	10,99 (1,34) 10,57 (9,32-13,18)
Feijoo	29,91 (3,85) 29,58 (25,26-36)	12,10 (3,61) 10,84 (8,42-20,21)	7,42 (2,78) 7,80 (2,82-11,03)
Jorquera	25,42 (1,32) 25,45 (23,31-27,50)	13,97 (1,28) 14,37 (11,65-15,37)	9,10 (0,94) 9,04 (7,49-10,31)
P.Vázquez	24,86 (1,87) 24,88 (22,13-27,48)	12,41 (1,10) 13,42 (9,74-15)	10,23 (1,23) 10,25 (8,37-11,73)
χ^2	25,979	22,691	25,288
<i>gl</i>	6	6	6
<i>Sig. Asintót.</i>	,000	,001	,000

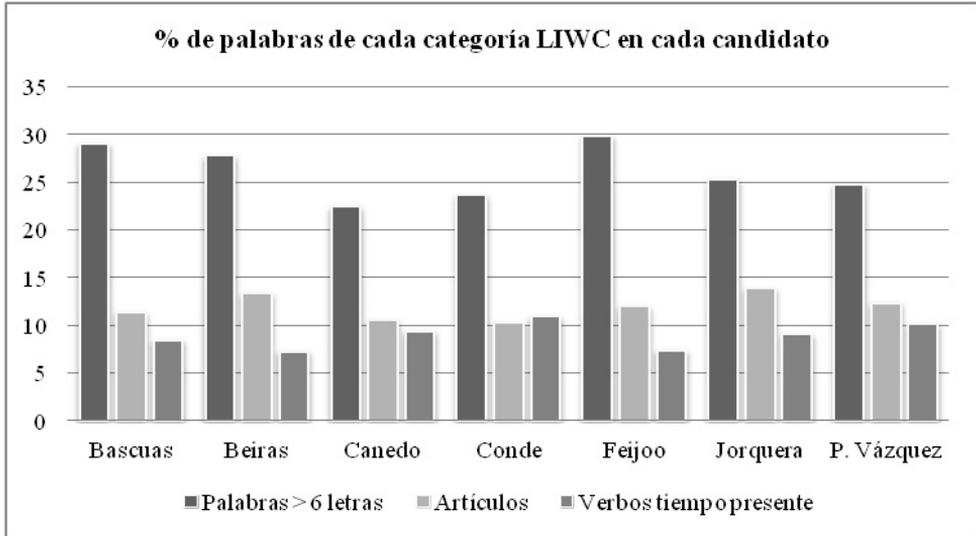


Figura 1: Gráfico con porcentaje de uso por cada candidato de las categorías LIWC “palabras con más de seis letras”, “artículos” y “verbos en tiempo presente”.

TABLA 3: PORCENTAJE DE USO DE LAS CATEGORÍAS LIWC “1ª PERSONA DEL SINGULAR”, “NEGACIONES”, “INHIBICIÓN”, “TENTATIVO” Y “CERTEZA” PARA CADA CANDIDATO.

<i>LIWC categ. / Candidatos Media (DT) Mediana (min-max)/</i>	<i>1ª persona sing. (yo, mi, mío)</i>	<i>Negaciones (no, nunca)</i>	<i>Inhibición (bloquear, prohibir)</i>	<i>Tentativo (tal vez, a lo mejor)</i>	<i>Certeza (siempre, nunca)</i>
Bascuas	1,28 (0,72) 1,62 (0-2,11)	0,98 (0,60) 0,92 (0,23-1,75)	0,32 (0,31) 0,23 (0-0,78)	1,44 (0,91) 1,38 (0,46-2,91)	0,92 (0,61) 0,73 (0,31-2,12)
Beiras	0,67 (0,43) 0,56 (0-1,28)	1,39 (0,64) 1,55 (0,28-2,46)	0,22 (0,15) 0,19 (0,07-0,52)	1,06 (0,25) 1,02 (0,70-1,38)	0,50 (0,20) 0,46 (0,14-0,85)
Canedo	2,25 (1,63) 2,12 (0,43-5,02)	2,31 (1,17) 1,93 (0,97-4,94)	0 (0) 0 (0-0)	0,80 (0,51) 0,76 (0-1,46)	1,24 (0,56) 0,97 (0,43-1,94)
Conde	1,25 (0,60) 1,41 (0,28-2,01)	3,33 (1,15) 3,57 (1,70-4,56)	0,39 (0,37) 0,28 (0-1,21)	1,43 (0,52) 1,49 (0,69-2,27)	1,35 (0,42) 1,49 (0,69-1,88)
Feijoo	0,36 (0,51) 0 (0-1,45)	2,36 (1,68) 1,43 (0,62-4,88)	0,33 (0,50) 0 (0-1,33)	0,36 (0,45) 0 (0-1,04)	0,89 (1,41) 0,25 (0-4,23)
Jorquera	0,76 (0,95) 0,43 (0,19-1,41)	2,37 (0,35) 2,33 (1,61-2,81)	0,51 (0,34) 0,51 (0,13-1,21)	1,02 (1,17) 1,03 (0,71-1,34)	0,80 (0,40) 0,75 (0,27-1,55)
P. Vázquez	1,01 (0,60) 0,75 (0,38-2,31)	2,14 (0,45) 2,10 (1,57-2,74)	0,39 (0,12) 0,38 (0,26-0,56)	0,88 (0,38) 0,79 (0,41-1,51)	0,57 (0,19) 0,58 (0,34-0,87)
χ^2	19,360	26,221	22,029	18,361	20,832
gl	6	6	6	6	6
Sig. Asintót.	,004	,000	,001	,005	,002

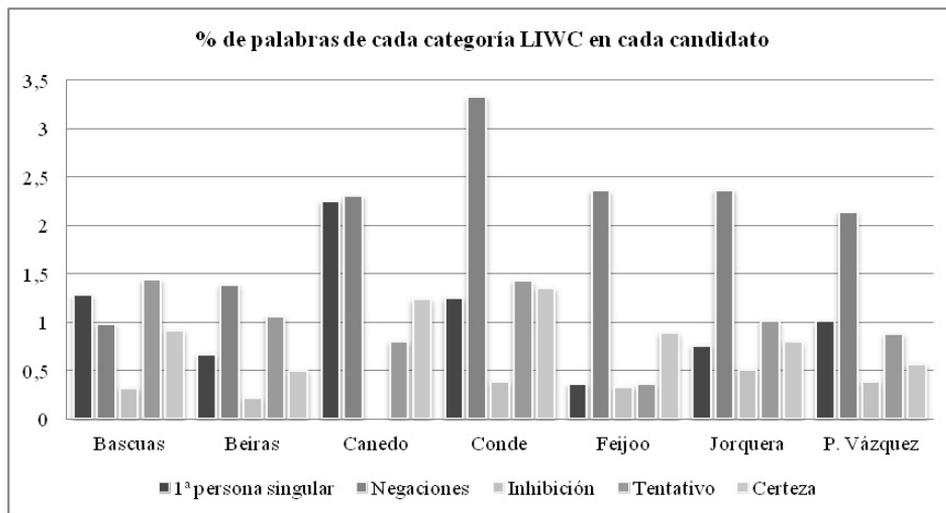


Figura 2: Gráfico con porcentaje de uso por cada candidato de las categorías LIWC “1ª persona del singular”, “negaciones”, “inhibición”, “tentativo” y “certeza”.

TABLA 4: PORCENTAJE DE USO DE LAS CATEGORÍAS LIWC “TOTAL EMOCIONES NEGATIVAS”, “ANSIEDAD”, “ENOJO”, Y “TRISTEZA” PARA CADA CANDIDATO.

<i>LIWC categ. / Candidatos Media (DT) Mediana (min-max)/</i>	<i>Emociones negativas (odio, enemigo)</i>	<i>Ansiedad (tenso, nervioso)</i>	<i>Enojo (odio, ira)</i>	<i>Tristeza (llorar, tristeza)</i>
Bascuas	0,38 (0,45) 0,16 (0-1,16)	0,06 (0,09) 0 (0-0,23)	0,02 (0,05) 0 (0-0,15)	0,19 (0,31) 0 (0-0,73)
Beiras	1,02 (0,38) 0,87 (0,49-1,85)	0,11 (0,12) 0,08 (0-0,32)	0,38 (0,21) 0,33 (0,14-0,69)	0,34 (0,19) 0,32 (0,14-0,61)
Canedo	0,91 (0,63) 0,93 (0-1,94)	0,15 (0,31) 0 (0-0,85)	0,55 (0,57) 0,47 (0-1,69)	0,05 (0,14) 0 (0-0,42)
Conde	1,65 (0,63) 1,68 (0,60-2,54)	0,46 (0,31) 0,41 (0-1,13)	0,56 (0,25) 0,51 (0,30-1,13)	0,40 (0,30) 0,41 (0-0,85)
Feijoo	0,78 (0,85) 0,82 (0-2,67)	0,18 (0,44) 0 (0-1,33)	0,29 (0,44) 0 (0-1,33)	0,15 (0,44) 0 (0-1,33)
Jorquera	1,44 (0,64) 1,53 (0,48-2,28)	0,34 (0,28) 0,19 (0,06-0,78)	0,58 (0,39) 0,51 (0,12-1,16)	0,24 (0,19) 0,19 (0-0,57)
P. Vázquez	1,29 (1,20) 1,32 (1,07-1,69)	0,21 (0,10) 0,19 (0,06-0,34)	0,45 (0,23) 0,38 (0,25-0,96)	0,19 (0,15) 0,19 (0-0,49)
χ^2	24,010	21,934	21,829	21,228
<i>gl</i>	6	6	6	6
<i>Sig. Asintót.</i>	,001	,001	,001	,002

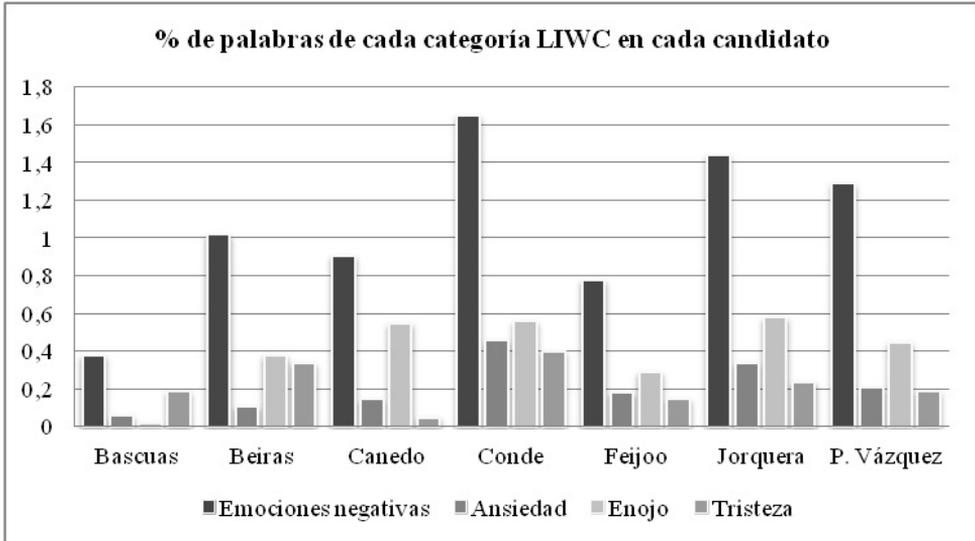


Figura 3: Gráfico con porcentaje de uso por cada candidato de las categorías LIWC “total emociones negativas”, “ansiedad”, “enojo”, y “tristeza”.

TABLA 5: PORCENTAJE DE USO DE LAS CATEGORÍAS LIWC “INSIGHT”, “COMUNICACIÓN”, “VERBOS EN TIEMPO FUTURO”, Y “DINERO” PARA CADA CANDIDATO.

<i>LIWC categ. / Candidatos Media (DT) Mediana (min-max)/</i>	<i>Insight (pensar, saber, conside- rar)</i>	<i>Comunicación (hablar, com- partir)</i>	<i>Verbos en tiempo futuro</i>	<i>Dinero (salario)</i>
Bascuas	1,69 (0,73) 1,83 (0,72-3,20)	2,30 (1,37) 1,89 (1,35-5,63)	1,26 (0,84) 1,15 (0,47-3,37)	0,35 (0,33) 0,23 (0-0,92)
Beiras	1,02 (0,28) 0,98 (0,80-1,70)	1,24 (0,22) 1,28 (1-1,65)	0,88 (0,46) 0,84 (0,07-1,60)	0,70 (0,35) 0,59 (0,22-1,34)
Canedo	1,75 (0,69) 1,87 (0,48-3,04)	1,88 (1,09) 1,40 (0-3,33)	1,75 (0,84) 1,90 (0-2,83)	0,60 (0,52) 0,76 (0-1,42)
Conde	2,53 (1,24) 2,29 (1,13-4,82)	2,34 (0,65) 2,11 (1,52-3,57)	1,38 (0,74) 1,13 (0,56-2,71)	1,11 (1,09) 0,85 (0-3,61)
Feijoo	1,90 (1,68) 1,06 (0-5,79)	2,32 (3,50) 1,33 (0-11,27)	1,15 (0,78) 1,23 (0-2,11)	0,31 (0,46) 0 (0-1,23)
Jorquera	1,63 (0,27) 1,60 (1,31-2,18)	1,33 (0,36) 1,47 (0,59-1,67)	0,79 (0,30) 0,65 (0,50-1,34)	0,67 (0,36) 0,59 (0,13-1,37)
P.Vázquez	1,30 (0,15) 1,31 (1,09-1,54)	1,71 (0,46) 1,63 (1,15-2,55)	1,81 (0,31) 1,77 (1,43-2,45)	1,12 (0,35) 1,22 (0,56-1,55)
χ^2	20,101	16,865	20,817	17,633
gl	6	6	6	6
Sig. Asintót.	,003	,010	,002	,007

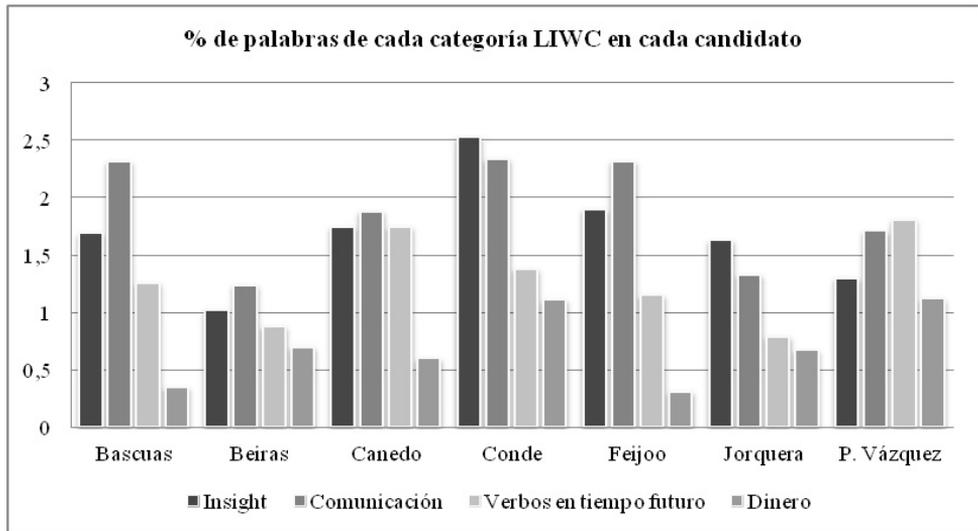


Figura 4: Gráfico con porcentaje de uso por cada candidato de las categorías LIWC “insight”, “comunicación”, “verbos en tiempo futuro”, y “dinero”.

Comparamos los resultados de cada candidato con los resultados de los demás candidatos uno a uno, para ver con cual/cuáles de ellos existían diferencias significativas en el porcentaje de uso de cada categoría lingüística (ver tabla 6, página siguiente).

Discusión

Al revisar los estudios sobre el LIWC que han analizado y explicado las categorías con diferencias significativas en nuestro estudio, hallamos que los tiempos verbales que elegimos al hablar nos dan información sobre el foco atencional en el que se centra el hablante. El tiempo verbal utilizado indicaría la distancia psicológica y el grado en que un tema se considera ya resuelto: con mayor uso del tiempo pasado a mayor distancia y cierre de los temas y del tiempo presente en temas pendientes o más cercanos (Tausczik y Pennebaker, 2010). Gunsch, Brownlow, Haynes y Mabe (2000) analizaron el lenguaje utilizado en anuncios políticos con contenido positivo, negativo o mixto; los anuncios positivos incluían más verbos en tiempo presente y futuro y más autoreferencias (con más pronombres en primera persona), mientras que los negativos utilizaban más verbos en pasado (y más palabras de emoción negativa y referencias al dinero y a otros).

TABLA 6: CATEGORÍAS LINGÜÍSTICAS CON DIFERENCIA SIGNIFICATIVA ($P \leq 0.010$) EN EL PORCENTAJE DE USO DE CADA CANDIDATO COMPARADO CON CADA UNO DE LOS DEMÁS CANDIDATOS.

	<i>Basctas</i>	<i>Beiras</i>	<i>Canedo</i>	<i>Conde</i>	<i>Feijoo</i>	<i>Jorquera</i>
<i>Beiras</i>	Artículo (.007) Enojo (.000) Comun. (.003)					
<i>Canedo</i>	Negación (.003) Enojo (.010) Inhibición (.004)	>6 letras (.002) Tristeza (.001) Insight (.009) Inhibición (.000) Certeza (.003) Futuro (.009)				
<i>Conde</i>	Negación (.000) Emoc. neg. (.002) Ansiedad (.002) Enojo (.000) Presente (.007)	>6 letras (.004) Negación (.001) Artículo (.001) Ansiedad (.008) Insight (.001) Certeza (.000) Comunicación (.000) Presente (.000)	Tristeza (.008) Inhibición (.002)			
<i>Feijoo</i>	Yo (.009)	Tristeza (.004) Tentativo (.004)	>6 letras (.003) Yo (.005)	>6 letras (.003) Yo (.005) Tentativo (.002) Presente (.005)		
<i>Jorquera</i>	Negación (.001) Artículo (.004) Emoc. neg. (.002) Enojo (.000)	>6 letras (.005) Negación (.003) Insight (.002)	Artículo (.007) Tristeza (.005) Inhibición (.000) Futuro (.009)	Artículo (.001) Comunicación (.001) Presente (.004)	>6 letras (.009) Tristeza (.009) Tentativo (.003)	
<i>Vázquez</i>	Negación (.002) Emoc. neg. (.001) Ansiedad (.009) Enojo (.000) Futuro (.005) Dinero (.001)	>6 letras (.005) Insight (.007) Presente (.001) Futuro (.001)	Tristeza (.007) Inhibición (.000) Certeza (.004)	Insight (.007) Certeza (.001)	>6 letras (.004) Tristeza (.009) Dinero (.003)	Insight (.006) Futuro (.000)

El uso elevado de palabras de emoción (positiva o negativa) se ha relacionado con una mayor inmersión e implicación en lo que se dice (Holmes *et al.*, 2007). Se ha vinculado el optimismo con el mayor uso de tiempos verbales presente y futuro, pronombres en primera persona del plural y emociones positivas, y especialmente con el uso reducido de palabras de emoción negativa; cuestión importante en este contexto ya que Zullo, Oettingen, Peterson y Seligman (1988) hallaron que el optimismo en los candidatos correlacionaba con más votos.

El alto estatus del hablante en un grupo correlaciona con que éste hable más a menudo y en nombre de todos (con un mayor uso de la primera persona del plural), mientras que los sujetos con bajo estatus y mayor inseguridad tendrían un discurso más tentativo y auto-focalizado (con mayor uso de la primera persona del singular) y tenderían a hacer más preguntas que afirmaciones (Kacewicz, Pennebaker, Davis, Jeon y Graesser, 2013; Sexton y Helmreich, 2000). El uso de la primera persona del singular se ha relacionado, además con una mayor honestidad e implicación en lo que se dice (Pennebaker, Slatcher y Chung, 2005). Las palabras tentativas podrían indicar inseguridad sobre el tema tratado o falta de integración de dicho tema dentro del discurso (Pasupathi, 2007). Por otra parte, el uso elevado de palabras de la categoría “certeza” se ha relacionado con una mayor estabilidad emocional y mejores habilidades sociales (Tausczik y Pennebaker, 2010)

Las palabras incluidas en la categoría de “*insight*” (como pensar, considerar, etc.) pertenecen a la categoría de mecanismos cognitivos y, junto con el uso de preposiciones, palabras de exclusión y palabras de más de seis letras, son indicativas de un lenguaje más complejo (Tausczik y Pennebaker, 2010).

El uso de palabras pertenecientes a la categoría de “negación” se ha interpretado como un marcador de auto-dominio e inhibición por parte del hablante; también las palabras de “inhibición” podrían indicar un intento de contener u ocultar pensamientos y sentimientos (Pennebaker y Lay, 2002). En ambos casos añadirían complejidad al discurso aunque le restarían espontaneidad. Las palabras de la categoría “comunicación” indicarían interés por las interacciones con los demás y el uso de artículos se ha vinculado al uso de un lenguaje más concreto (Tausczik y Pennebaker, 2010).

Teniendo en cuenta lo anterior expondremos algunas características del lenguaje de cada candidato en comparación con el resto.

Xoan Bascuas (CxG), es el que mostró un mayor porcentaje de uso de palabras tentativas, así como un porcentaje mayor que la media de palabras largas y relacionadas con la comunicación. Sin embargo, es el candidato que menor uso hizo de negaciones y de palabras de emoción negativa. El uso de palabras mayores de seis letras y de palabras tentativas se relaciona con una mayor complejidad en el discurso, aunque las palabras tentativas pueden ser percibidas como muestra de

inseguridad en lo que se dice. Se mostraría en contacto con sus electores, reflejando menor negatividad que la media de los candidatos.

Por su parte, Xose M. Beiras (AGE) mostró un alto porcentaje de uso de artículos en su lenguaje, y fue el candidato con menor uso de palabras referidas al insight, certeza y comunicación, así como de verbos en tiempo presente, siendo también bajo su porcentaje de uso de verbos en tiempo futuro y negaciones. Tales características indicarían un lenguaje más concreto y espontáneo en este candidato.

José Canedo (UPyD) utilizó el mayor porcentaje de pronombres en primera persona del singular, y fue superior a la media grupal su uso de palabras que indican enojo y de verbos en tiempo futuro. Sin embargo fue el que menos introdujo palabras largas y palabras pertenecientes a las categorías de tristeza e inhibición, siendo también menor su uso de artículos. Se trataría de un lenguaje más auto-focalizado y orientado hacia el futuro que la media; el mayor uso de la primera persona del singular y de palabras pertenecientes a la categoría de enojo transmitiría una mayor implicación en lo expresado, y el menor uso de palabras de inhibición podría transmitir mayor espontaneidad.

Mario Conde (SCD) destaca por ser el candidato que hizo el discurso más extenso, con mayor porcentaje de uso de negaciones y palabras de emoción negativa (ansiedad, tristeza), palabras de insight, certeza, comunicación y verbos en tiempo presente, con porcentajes superiores a la media en las categorías de enojo, dinero y tentativas. Por otra parte, fue el que menor uso hizo de artículos y fue bajo su uso de palabras extensas. Se trataría de un discurso centrado en el presente, conectado con su público y con mayor negatividad y defensividad que la media, que podría transmitir pesimismo. La combinación de palabras tentativas y de insight podría ser percibida como señal de que el hablante presenta sus dudas y reflexiones mientras se expresa.

Alberto Núñez Feijoo (PP) fue el que mayor porcentaje de palabras de más de seis letras utilizó, y también fue alto su uso de palabras relacionadas con la comunicación. Por otra parte, fue el que menos utilizó palabras tentativas, relacionadas con el dinero y pronombres en primera persona del singular, siendo también bajo su uso de verbos en tiempo presente. Ofrecería por tanto un lenguaje menos dubitativo; el uso de palabras largas indicaría mayor complejidad en el lenguaje y el menor uso de la primera persona del singular podría ser percibido como indicador de una menor implicación personal en lo expuesto. La elevación del uso de palabras de comunicación indicaría interés por interactuar con los demás.

Francisco Jorquera (BNG) fue el candidato que mayor uso hizo de artículos y de palabras que indican inhibición y enojo, así como el que menor porcentaje de ver-

bos en tiempo futuro utilizó. Se trataría de un lenguaje más concreto, el uso de palabras de emoción indicarían inmersión en lo que se expone pero puede ser percibido como más pesimista que sus competidores, dado el bajo uso de tiempos futuros y el uso de palabras de emoción negativa. Pareciera que el candidato realiza un análisis previo de las ideas a exponer (relacionado con el uso de palabras de inhibición), lo cual entendemos que propiciará que exprese mejor sus ideas pero puede dar la impresión de una menor espontaneidad.

Pachi Vázquez (PSdG) destaca por ser el que utilizó un mayor porcentaje de palabras relativas a temas de dinero y de verbos en tiempo futuro, siendo también superior a la media grupal su uso de verbos en tiempo presente, e inferior su uso de palabras de certeza. El mayor porcentaje de uso de verbos en tiempo presente y futuro transmitiría optimismo en este candidato, posiblemente centrado en exponer sus planes para el futuro.

Conclusión

Este estudio muestra que el programa LIWC resulta útil para detectar diferencias en el lenguaje utilizado en cada uno de los candidatos, lenguaje que influye en la impresión que nos formamos de la personalidad y fiabilidad de cada uno de ellos, y por tanto en nuestro voto.

Como limitaciones de nuestro estudio, señalar la necesidad de traducción de algunos de los textos, así como la imposibilidad de verificar la autoría de tweets y discursos de los candidatos; creemos sin embargo que este problema no invalida los resultados ya que no se pretende describir “la verdadera personalidad” de los analizados, sino únicamente el estilo lingüístico que muestran ante sus electores.

Otras limitaciones serían las características del programa utilizado, que no captura el sentido no literal de las expresiones ni la influencia del contexto en lo expresado, así como la escasez de estudios que utilicen la versión española del LIWC. Sin embargo se ha comprobado su equivalencia con la versión inglesa, cuya fiabilidad ha sido ampliamente contrastada.

El análisis de texto computarizado ofrece poder estadístico y consistencia en la codificación y supone una alternativa eficiente al análisis lingüístico basado en codificadores humanos, además de aportar nuevas vías para el análisis de la comunicación de los políticos y candidatos electorales. Por ello pensamos que se trata de un campo de estudio interesante, especialmente si seguimos ampliando el número de estudios en español y se analizan textos extensos tomados en diferentes ámbitos.

REFERENCIAS

- Borondo, J., Morales, A.J., Losada, J.C., & Benito, R.M. (2012). Characterizing and modelling an electoral campaign in the context of Twitter: 2011 Spanish Presidential Election as a case study. *Chaos*, 22(2), 023138.
- Carrera-Fernández, M.J., Guàrdia-Olmos, J. y Peró-Cebollero, M. (2013). Psicología y lenguaje en política: los candidatos a la Presidencia del Gobierno y su estilo lingüístico. *Anuario de Psicología*, 43(1), 39-52.
- Congosto, M. y Aragón, P. (2012). Twitter, del sondeo a la sonda: nuevos canales de opinión, nuevos métodos de análisis. *Más Poder Local*, 12, 50-56.
- Gayo-Avello, D., Metaxas, P.T., & Mustafaraj, E. (2011). Limits of electoral predictions using Twitter. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, 490-493.
- Gil de Zúñiga, H., Jung, N., & Valenzuela, S. (2011). Social media use for news and individual's social capital, civic engagement and political participation. *Journal of Computer Mediated Communication*, 17(3), 319-336.
- Gunsch, M.A., Brownlow, S., Haynes, S.E., & Mabe, Z. (2000). Differential linguistic content of various forms of political advertising. *Journal of Broadcasting & Electronic Media*, 44, 27-42.
- Holmes, D., Alpers, G.W., Ismailji, T., Classen, C., Wales, T., Cheasty, V., ... Koopman, C. (2007). Cognitive and emotional processing in narratives of women abused by intimate partners. *Violence Against Women*, 13, 1192-1205.
- Kacewicz, E., Pennebaker, J.W., Davis, M., Jeon, M., & Graesser, A.C. (2013). Pronoun Use Reflects Standings in Social Hierarchies. *Journal of Language and Social Psychology*. doi: 10.1177/0261927X13502654. Recuperado de <http://jls.sagepub.com/content/early/2013/09/18/0261927X13502654>.
- Newman, M.L., Groom, C.J., Handelman, L.D., & Pennebaker, J.W. (2008). Gender differences in language use: An analysis of 14.000 text samples. *Discourse Processes*, 45, 211-236.
- Newman, M.L., Pennebaker, J.W., Berry, D.S., & Richards, J.M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, 29(5), 665-675.
- Pasupathi, M. (2007). Telling and the remembered self: Linguistic differences in memories for previously disclosed and previously undisclosed events. *Memory*, 15(3), 258-270.
- Pennebaker, J.W., & Chung, C.K. (2008). Computerized text analysis of Al-Qaeda transcripts. In K. Krippendorff and M.A. Bock (Eds), *A content analysis reader* (pp 453-465). Thousand Oaks, CA: Sage
- Pennebaker, J.W., Francis, M.E., & Booth, R.J. (2001). *Linguistic Inquiry and Word Count (LIWC): LIWC2001*. Mahwah, NJ: Erlbaum Associates.
- Pennebaker, J.W., Mehl, M.R., & Nierderhoffer, K.G. (2003). Psychological aspects of natural language use: our words, our selves. *Annual Review of Psychology*, 54, 547-577.
- Pennebaker, J.W., & Lay, T.C. (2002). Language use and personality during crises: Analyses of Mayor Rudolph Giuliani's press conferences. *Journal of Research in Personality*, 36, 271-282.
- Pennebaker, J.W., Slatcher, R.B., & Chung, C.K. (2005). Linguistic markers of psychological state through media interviews: John Kerry and John Edwards in 2004, Al Gore in 2000. *Analyses of Social Issues and Public Policy*, 5, 1-9.
- Pennebaker, J.W., & Stone, L.D. (2003). Words of wisdom: Language use over the lifespan. *Journal of Personality and Social Psychology*, 85, 291-301.
- Ramírez-Esparza, N., Pennebaker, J.W., García, F.A. y Suriá, R. (2007). La psicología del uso de las palabras: Un programa de computadora que analiza textos en español. *Revista Mexicana de Psicología*, 24, 85-99.
- Rúas, J., Fernández-Cabana, M. y Puentes, I. (2013). Aplicación de la herramienta LIWC al análisis del discurso político. Los mítines de los candidatos en las elecciones al Parlamento de Galicia

- de 2012. En: M. Vicente, T. González y M. Pacheco (Coords.): *Investigar la comunicación hoy. Revisión de políticas científicas y aportaciones metodológicas*. Simposio Internacional sobre Política Científica en Comunicación. Segovia: Universidad de Valladolid.
- Sexton, J.B., & Helmreich, R.L. (2000). Analyzing cockpit communications: The links between language, performance, and workload. *Human Performance in Extreme Environments*, 5, 63-68.
- Slatcher, R.B., Chung, C.K., Pennebaker, J.W., & Stone, L.D. (2007). Winning words: Individual differences in linguistic style among U. S. presidential and vice presidential candidates. *Journal of Research in Personality*, 41, 63-75.
- Tausczik, Y., & Pennebaker, J.W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29, 24-54.
- Tumasjan, A., Sprenger, T., Sandner, P., & Welpe, I.(2010). Predicting elections with Twitter: What 140 characters reveal about political sentiment. *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media (ICWSM)*, Washington, S. 178-185.
- Zulrow, H.M., Oettingen, G., Peterson, C., & Seligman, M.E.P. (1988). Pessimistic explanatory style in the historical record: CAVing LBJ, presidential candidates, and East versus West Berlin. *American Psychologist*, 43, 673-682.