

DISCURSOS PRESIDENCIALES EN URUGUAY ENFOQUE DESDE EL ANÁLISIS ESTADÍSTICO DE TEXTO

Elena Vernazza Mañana*; José Luis Vicente Villardón**

*Instituto de Estadística, Facultad de Ciencias Económicas y de Administración, Universidad de la República, Uruguay. Eduardo Acevedo 1139 Montevideo, CP 11200. Uruguay

**Departamento de Estadística, Universidad de Salamanca, España. Alfonso X el Sabio s/n. Salamanca CP 37007. España.

evernazza@iesta.edu.uy; villardon@usal.es

Recibido 8 de mayo 2020, aceptado 1 de septiembre 2020

RESUMEN

El origen del análisis de datos textuales se remonta a los análisis realizados sobre obras literarias, destacándose un recuento de las palabras de la Biblia y el primer catálogo de clasificación de libros de una biblioteca. A principios de 1900 estas ideas se extienden dando lugar a la generación de resúmenes de largos textos, mediante análisis de frecuencias y medidas de significación relativa de cada término dentro de un texto. Desde una perspectiva estadística, el tratamiento de datos textuales se afianza cuando surge el Análisis Factorial de Correspondencias, desarrollado para el trato de datos lingüísticos. Actualmente, el análisis estadístico de textos se ha expandido y ha ido incrementando su implementación, dejando de lado las aplicaciones únicamente a obras literarias. Los principales ámbitos de aplicación son: análisis de mercado, búsquedas web, periodismo, psicología y educación, sociología y politología. Los textos analizados se estructuran de forma conjunta en un único elemento. Así, el análisis estadístico de datos textuales se define como el proceso de extraer información de dicho elemento. En este trabajo se realiza un análisis, desde esta perspectiva, de cuatro discursos presidenciales de Uruguay, correspondientes a Julio María Sanguinetti y Tabaré Vázquez. Ambos provienen de sectores ideológicos/políticos distintos, tradicionalmente opuestos/rivales y cada uno ha formado parte del gobierno como oposición, durante el mandato del otro. Los resultados son de carácter descriptivo multidimensional y se complementan con elementos de visualización. Los principales resultados obtenidos ponen de manifiesto la diferencia que existe en los discursos, a nivel tanto de candidato como de período.

Palabras clave: Análisis estadístico de datos textuales, discursos presidenciales

Códigos JEL: C18, H79, Z00

PRESIDENTIAL SPEECHES IN URUGUAY FROM A TEXTUAL DATA ANALYSIS APPROACH

Elena Vernazza Mañana*; José Luis Vicente Villardón**

*Instituto de Estadística, Facultad de Ciencias Económicas y de
Administración, Universidad de la República, Uruguay. Eduardo
Acevedo 1139 Montevideo, CP 11200. Uruguay

**Departamento de Estadística, Universidad de Salamanca, España.
Alfonso X el Sabio s/n. Salamanca CP 37007. España.

evernazza@iesta.edu.uy; villardon@usal.es

Received Mai 8th 2020, accepted September 1st 2020

ABSTRACT

The textual data analysis origin is based on the analyzes carried out mainly on literary works, which include a recount of all Bible's words and the presentation of the first catalog of book classification of a library. At the beginning of 1900 these ideas are extended, giving rise to summaries generation of long texts, based on the analysis of frequencies and measures of relative significance of each term within a given text. From a formal statistical perspective, the text data processing takes hold when the Factorial Correspondence Analysis arises, developed for linguistic databases treatment. Now a day, the textual data statistical analysis has been sold and has increased its implementation. Among the main fields of application are: market analysis, web searches and journalistic studies, studies of psychology, education and sociology. The analyzed texts are organized and structured together in a single element. Thus, the textual data statistical analysis is defined as the process of extracting information from that element. This paper analyzes, from this perspective, of four presidential speeches in Uruguay, corresponding to Julio María Sanguinetti and Tabaré Vázquez. Both come from different ideological / political sectors, traditionally opposed / rivals and each has been part of the opposition government, during the mandate of each other. The presented results are multidimensional descriptive and are complemented by visualization elements, a tool typically used in textual data statistical analysis. The main results obtained show the difference that exists in the speeches, both at the level of the candidate and the period.

Keywords: textual data statistical analysis, presidential speeches

JEL Code: C18, H79, Z00

INTRODUCCIÓN

El análisis de datos textuales surge de los análisis (de carácter cuantitativo) realizados fundamentalmente sobre obras literarias, entre los que se destacan un recuento de todas las palabras presentes en la Biblia y la presentación del primer catálogo de clasificación de libros de una biblioteca (Catálogo de Biblioteca Bodleiana de la Universidad de Oxford en el año 1876). Estos análisis estaban enfocados exclusivamente al recuento (estudio de frecuencia) de cada término sin tener en consideración, por ejemplo, el estudio de la distribución del vocabulario, ni la relación/asociación entre términos.

A principios de 1900 estas ideas se extienden dando lugar, en primer lugar, a la generación de resúmenes de largos textos. Tomando como punto de partida el análisis de frecuencias se crean medidas de significación relativa de cada término dentro de un texto determinado. A mediados de siglo, surgen nuevos desarrollos teóricos, que permiten entrar en el estudio de la distribución del vocabulario. Surgen, así, los primeros estudios de comparación léxica teniendo en consideración distintos autores e incluso un mismo autor en distintos momentos y obras.

El tratamiento de datos de texto desde una perspectiva estadística formal se afianza en la década del 70, cuando surge el Análisis Factorial de Correspondencias [2], desarrollado inicialmente para el trato de bases de datos lingüísticas.

Con el paso del tiempo y con el desarrollo de la informática (que avanza día a día), el análisis estadístico de datos textuales se ha expandido y ha ido incrementando su implementación en los más diversos ámbitos, dejando de lado las aplicaciones únicamente a obras literarias. Entre los principales ámbitos de aplicación se destacan: análisis de mercado, búsquedas y análisis web, estudios periodísticos, estudios de psicología y educación, sociología, politología, bibliotecas digitales, entre otros [1].

Los textos analizados (documentos, entrevistas, encuestas con preguntas abiertas, etc.) se organizan y estructuran de forma conjunta en un único elemento, denominado corpus. Así, el análisis estadístico de datos textuales (en inglés, *text data mining* o simplemente *text mining*) se define como el proceso de extraer información (estructurada y sistematizada) del corpus, a partir de su transformación en una matriz léxica o matriz de términos (*term matrix*).

En este trabajo se presentan un conjunto de indicadores utilizados como herramientas descriptivas y se ponen en práctica a través de una aplicación sobre los discursos de asunción de 2 expresidentes uruguayos, en el período post dictadura militar: 1985 – 2020.

El actual régimen electoral de Uruguay (vigente desde 1996) introdujo una reforma del sistema electoral previsto por la Constitución de 1967, estableciendo, entre otros aspectos:

- Elecciones internas abiertas y simultáneas de los candidatos a la Presidencia, Vicepresidencia de la República, así como de los órganos deliberativos de cada partido político.
- Candidatura única a la Presidencia de la República por cada partido.

Así, en el mes de junio de cada año electoral, se definen (a través de las elecciones internas) los candidatos oficiales a las elecciones presidenciales. En el mes de octubre del mismo año, son celebradas las elecciones nacionales, con la postulación de un único candidato por cada partido. En caso de que ninguno de los candidatos llegue a una adhesión del 50% +1 del total de votos emitidos se realiza una segunda vuelta, en el mes de noviembre. Cabe resaltar que, según la Constitución de Uruguay, el voto no es solamente un derecho, sino que es, sobre todo, un deber como ciudadano, además de una obligación.

En este trabajo se propone realizar un análisis, desde esta perspectiva, de cuatro discursos presidenciales de Uruguay, correspondientes a dos expresidentes. En particular, Julio María Sanguinetti (presidente en los períodos 1985-1990 y 1995-2000) y Tabaré Vázquez (presidente en los períodos 2005-2010 y 2015-2020). Ambos provienen de sectores ideológicos/políticos distintos, tradicionalmente opuestos/rivales y cada uno ha formado parte del gobierno como oposición, durante el mandato del otro. Los resultados presentados son de carácter descriptivo multidimensional y se complementan con elementos de visualización, herramienta típicamente utilizada en el ámbito del análisis estadístico de datos de naturaleza textual. Los principales resultados obtenidos ponen de manifiesto la diferencia que existe en los discursos, a nivel tanto de candidato como de período.

OBJETIVOS

El objetivo general de este trabajo es analizar en forma conjunta, complementaria y comparativa, cuatro discursos presidenciales de Uruguay, correspondientes a dos expresidentes. En particular, Julio María Sanguinetti (presidente en los períodos 1985-1990 y 1995-2000) y Tabaré Vázquez (presidente en los períodos 2005-2010 y 2015-2020).

A partir de este objetivo general, se plantean los siguientes objetivos específicos:

- Analizar por separado cada uno de los cuatro discursos.
- Analizar y comparar los discursos a nivel de candidato.
- Analizar y comparar los discursos a nivel de período.

MATERIAL Y MÉTODOS

Material

El objeto de estudio de este trabajo serán cuatro discursos de asunción de presidentes de Uruguay, correspondientes a dos expresidentes. En primer lugar, serán estudiados los cuatro discursos de forma independiente (analizando y comparando su contenido). Además, y de forma complementaria, los discursos serán analizados desde una perspectiva comparativa tanto a nivel individual de candidato, como a nivel de período.

En particular, los datos a analizar corresponden a los discursos de asunción de Julio María Sanguinetti y de Tabaré Vázquez, presidentes en los períodos 1985-1990 y 1995-2000, y 2005-2010 y 2015-2020, respectivamente.

Métodos

La metodología general a utilizar está basada íntegramente en la teoría desarrollada para el análisis estadístico de datos textuales. En particular, las estrategias utilizadas para analizar los datos en este trabajo son de carácter descriptivo y toman como insumo datos ordenados (tidy data, en particular: tidy text). Las características de este tipo de datos son [12][7]:

- Cada columna es una variable.
- Cada fila es una observación de esa variable.
- Cada tipo de variable corresponde a una tabla distinta.
- Debe ser posible enlazar dichas tablas, en caso de existir.

Al considerar los documentos analizados en formato de *texto ordenado* se identifican los siguientes elementos:

- Estructura general: una ficha por fila.
- Cadena (*string*): cadena de caracteres.
- Corpus: conjunto de textos con metadatos e información adicional.
- Matriz de términos: matriz tal que sus filas son los documentos analizados y las columnas los términos.

A partir de esta estructuración de los documentos a ser analizados, es posible estudiar cada uno de ellos de forma independiente y conjunta/comparada.

En lo que refiere a cada análisis independiente resulta pertinente considerar cada una de las palabras/términos que conforman cada documento, comenzando, por ejemplo, con un recuento absoluto (n_{palabra}) y/o relativo (frecuencia, de cada palabra (*tf*) [11]). Sin embargo, existen palabras presentes muchas veces en un documento (es decir, *tf* alta) que no son realmente importantes. Este tipo de palabras se conocen como palabras vacías (*stop words*) y suelen ser eliminadas en una primera instancia de depurado de los textos a ser analizados.

Desde otro enfoque, se propone observar la frecuencia inversa, dentro de un conjunto de documentos, de un término (*idf*). Así, se disminuye el peso de las palabras más comunes y se aumenta el de las menos comunes. Esta medida se define como:

$$\text{idf}(\text{palabra}) = \ln \left(\frac{n_{\text{documentos}}}{n_{\text{documentos que contienen palabra}}} \right)$$

Es posible combinar ambas medidas en una sola y obtener el *tf-idf* de un término, entendida como la frecuencia de un término dentro de un documento, ajustado por la frecuencia con la que se usa dentro de un conjunto de documentos.

Tomando como punto de partida estas medidas, y analizando que éstas presentan una distribución similar en cualquier tipo de texto analizado (libros, textos de sitios web, discursos, etc.), surge la Ley de Zipf [17] que establece que la frecuencia de una palabra en un texto es inversamente proporcional a su *ranking* (“posición” que ocupa una palabra, al ordenar todas las palabras por frecuencia), es decir:

$$\text{frecuencia}(\text{palabra}) \propto \frac{1}{\text{ranking}(\text{palabra})}$$

A partir del uso de estas medidas, será posible una primera comparación entre textos, en particular (en este trabajo) entre cuatro discursos de asunción presidencial.

Por último, se destaca que todas estas medidas de carácter descriptivo multidimensional suelen ser acompañadas de representaciones gráficas que complementan y facilitan su interpretación.

RESULTADOS

Análisis preliminar de los discursos

Para el procesamiento de los datos, se utilizó el software libre R Project [8], tomando como referencias las propuestas de Kwartler [6], y Silge y Robinson [11]. En particular los paquetes utilizados fueron: *tm*, *wordcloud*, *tidyverse*, *stringr*, *tidytext*, *textclean*, *ggplot2* [4], [5], [3], [14], [15], [10], [9], [13].

En cuanto a los principales resultados a presentar en este trabajo, y con el objetivo de abordar/presentar las principales estrategias (descritas en la sección Métodos) de tratamiento y análisis de este tipo de datos, se describen (en primera instancia) las principales características de los discursos brutos, es decir, sin depurar.

Así, el primer elemento a tener en consideración es la extensión, medida en cantidad de términos/palabras, de cada uno de los cuatro discursos analizados. Tal como se puede observar en la Tabla 1¹, los dos discursos de Tabaré Vázquez presentan una extensión sustancialmente menor que cualquiera de los dos discursos de Julio María Sanguinetti.

Presidente	Año	Cantidad de palabras
Sanguinetti	1985	4559
	1995	4682
Vázquez	2005	2532
	2015	2517

Tabla 1. Total de palabras por discurso.

Tomando como punto de partida estas diferencias, cabe comenzar a investigar cada uno de los discursos a la interna, palabra por palabra.

¹ Fuente: Esta, y todas las tablas del documento son de elaboración propia

En la Figura 1, se presentan gráficamente (nube de palabras) las palabras más dichas en términos absolutos ($n_{\text{palabra}} > 10$) de cada uno de los discursos y su análisis se complementa a partir de los indicadores presentados en la Tabla 2 (Ranking ≤ 10).

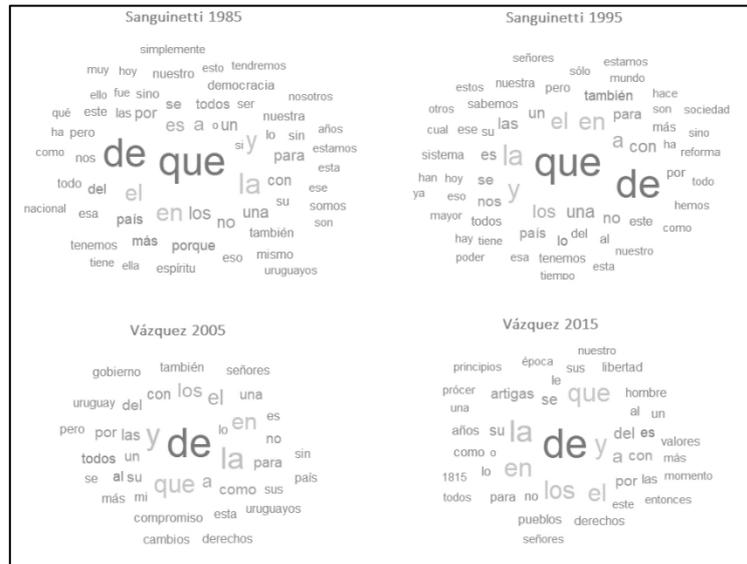


Figura 1: Nubes de palabras según discurso.

Presidente	Año	Palabra	n_{palabra}	n_{total}	Ranking	tf
Sanguinetti	1985	que	270	4559	1	0.05922
		de	255		2	0.05593
		la	167		3	0.03663
		y	132		4	0.02895
		en	125		5	0.02742
		el	111		6	0.02435
		a	82		7	0.01799
		es	73		8	0.01601

Tabla 2. Indicadores según discurso.

Presidente	Año	Palabra	n_{palabra}	n_{total}	Ranking	tf
Sanguinetti	1985	es	73	4559	8	0.01601
		los	67		9	0.01470
		un	58		10	0.01272
Sanguinetti	1995	que	276	4682	1	0.05895
		de	269		2	0.05745
		la	162		3	0.03460
		y	142		4	0.03033
		en	133		5	0.02841
		el	117		6	0.02499
		a	94		7	0.02008
		los	73		8	0.01559
		una	65		9	0.01388
		con	59		10	0.01260
Vázquez	2005	de	167	2532	1	0.06596
		y	109		2	0.04305
		la	106		3	0.04186
		que	81		4	0.03199
		en	75		5	0.02962
		el	59		6	0.02330
		los	48		7	0.01896
		a	47		8	0.01856
		las	29		9	0.01145
		con	28		10	0.01106

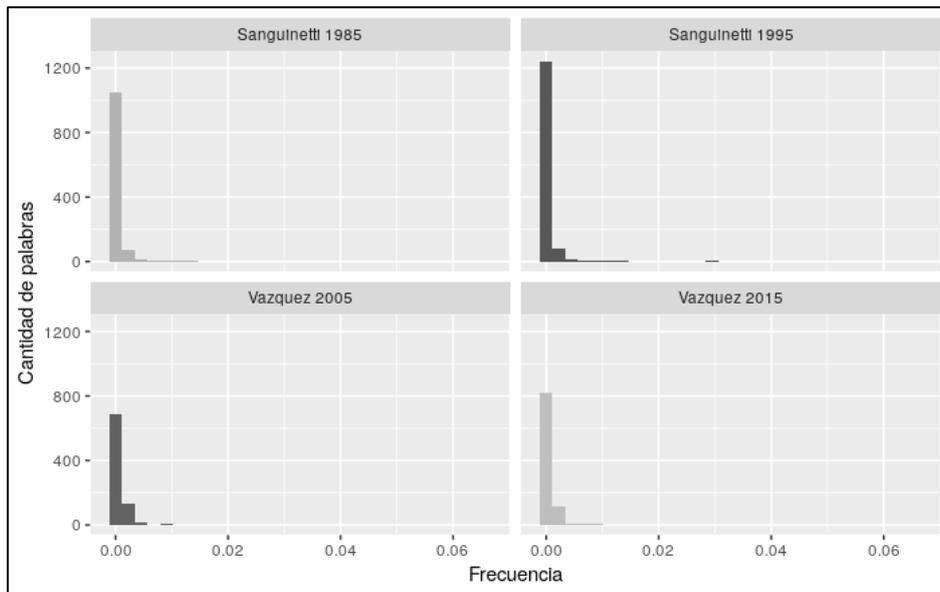
Tabla 2. Indicadores según discurso (cont.).

Presidente	Año	Palabra	n_{palabra}	n_{total}	Ranking	tf
Vázquez	2015	de	158	2517	1	0.06277

la	102	2	0.04052
y	95	3	0.03774
en	87	4	0.03456
que	81	5	0.03218
los	79	6	0.03139
el	74	7	0.02940
a	46	8	0.01828
del	29	9	0.01152
se	28	10	0.01112

Tabla 2. Indicadores según discurso (cont.).

Si, además, se considera la frecuencia relativa (tf) de cada palabra a la interna de cada discurso, es posible determinar que en todos los casos, existen: *muchas palabras dichas pocas veces y pocas palabras dichas muchas veces* (ver Figura 2).

Figura 2: Cantidad de palabras por tf según discurso.

Al analizar la información presentada, y tal como era de esperar, se detecta que las palabras más utilizadas, en cada uno de los discursos,

corresponden a palabras vacías (mencionadas previamente). Surge, por lo tanto, la necesidad de depurar los discursos analizados, a efectos de poder extraer información relevante de contenido.

Proceso de depurado

El proceso de depurado, realizado típicamente sobre este tipo de datos (de naturaleza textual), comienza teniendo en consideración las, ya mencionadas, palabras vacías.

En particular, en este trabajo se ha utilizado, inicialmente, un conjunto de 750 palabras.

A modo de ejemplo:

- a, acá, ahí, al, algo, algún, alguna, alguno, algunas, algunos, allá, allí, ambos, ante
- bien, bajo, bastante, breve, buen, buena, buenas, bueno, buenos
- cada, casi, como, con, cual, cuales, cualquier, cualquiera, cualesquiera, cuan, cuando
- de, del, demás, desde, donde
- ...
- tuvo, tuya, tuyas, tuyo, tuyos, tú
- último, un, una, unas, uno, unos, usa, usais, usamos
- va, vais, valor, vamos, van, varias, varios, vaya, veces, ver

Estas listas de palabras no son únicas y se actualizan, generalmente, en función de la temática en estudio.

Análisis de contenido de los discursos

Retomando los resultados expuestos en el apartado Análisis preliminar de los discursos, en la Tabla 3 se presentan las palabras totales por discurso depurado. El primer aspecto a destacar es que los dos discursos de J.M. Sanguinetti, aún poseen mayor extensión que los de Vázquez. De todas formas, en los cuatro casos analizados, se detecta un descenso de alrededor de un 60% de la longitud total del discurso (65%, 63%, 59% y 58%, respectivamente).

Presidente	Año	Cantidad de palabras
Sanguinetti	1985	1565

	1995	1722
Vázquez	2005	1026
	2015	1050

Tabla 3. Total de palabras por discurso depurado.

Sobre este nuevo escenario (discursos depurados), en la Figura 3 se presentan gráficamente las palabras más dichas en términos absolutos ($n_{\text{palabra}} > 10$) de cada uno de los discursos y su análisis se complementa a partir de los indicadores presentados en la Tabla 4 (Ranking ≤ 10). Es posible observar que las palabras con mayor tf resultan más informativas acerca del contenido general de cada discurso, con respecto al análisis realizado sobre los discursos sin depurar.

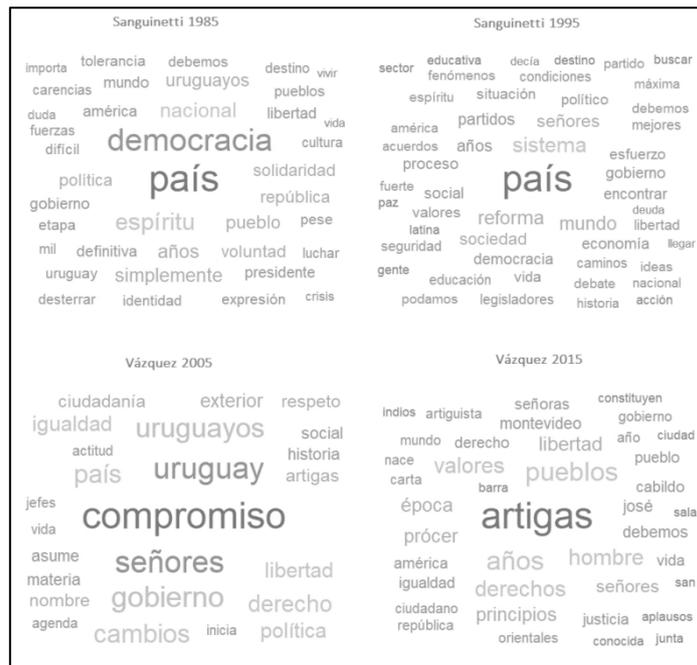


Figura 3: Nubes de palabras según discurso depurado.

Presidente	Año	Palabra	n_{palabra}	n_{total}	Ranking	tf
Sanguinetti	1985	país	31	1565	1	0.019808
		democracia	25		2	0.015974

		espíritu	17	3	0.010863
		nacional	12	4	0.007668
		años	11	5	0.007029
		simplemente	11	6	0.007029
		pueblo	10	7	0.006390
		uruguayos	10	8	0.006390
		política	8	9	0.005112
		república	8	10	0.005112
Sanguinetti	1995	país	36	1	0.020906
		sistema	16	2	0.009292
		reforma	13	3	0.007549
		mundo	12	4	0.006969
		señores	11	5	0.006388
		sociedad	10	6	0.005807
		años	9	7	0.005226
		partidos	9	8	0.005226
		democracia	8	9	0.004646
		economía	8	10	0.004646
Vázquez	2005	compromiso	17	1	0.016569
		señores	13	2	0.012671
		uruguay	13	3	0.012671
		gobierno	12	4	0.011696
		uruguayos	11	5	0.010721
		cambios	10	6	0.009747
		derechos	10	7	0.009747
		país	10	8	0.009747
		derecho	9	9	0.008772
		señoras	9	10	0.008772

Tabla 4: Indicadores según discurso depurado.

Presidente	Año	Palabra	$n_{palabra}$	n_{total}	Ranking	tf
Vázquez	2015	artigas	24	1050	1	0.022857
		años	15		2	0.014286

pueblos	15	3	0.014286
derechos	11	4	0.010476
hombre	11	5	0.010476
valores	11	6	0.010476
libertad	9	7	0.008571
época	8	8	0.007619
principios	8	9	0.007619
prócer	8	10	0.007619

Tabla 4: Indicadores según discurso depurado (cont.).

Por otra parte, comparando la estructura general/global de los cuatro discursos analizados, es posible determinar que todos ellos presentan un comportamiento similar (tal como se estableció en la sección Métodos). Todos contienen muchas palabras dichas pocas veces y pocas palabras dichas muchas veces y su distribución se ajusta a la establecida por la Ley de Zipf (ver Figuras 4 y 5).

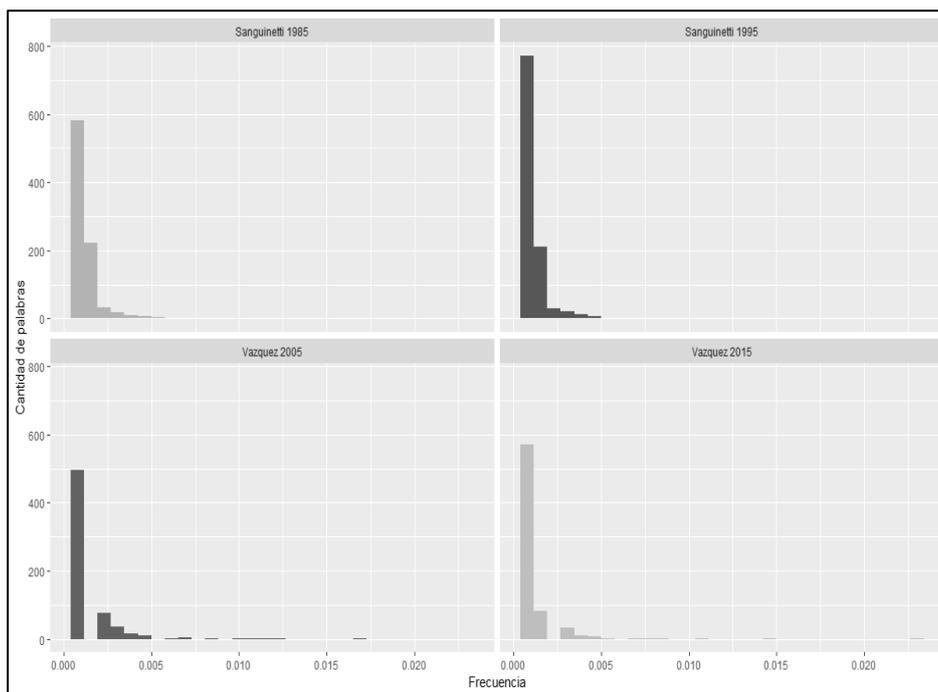


Figura 4: Cantidad de palabras por tf según discurso depurado.

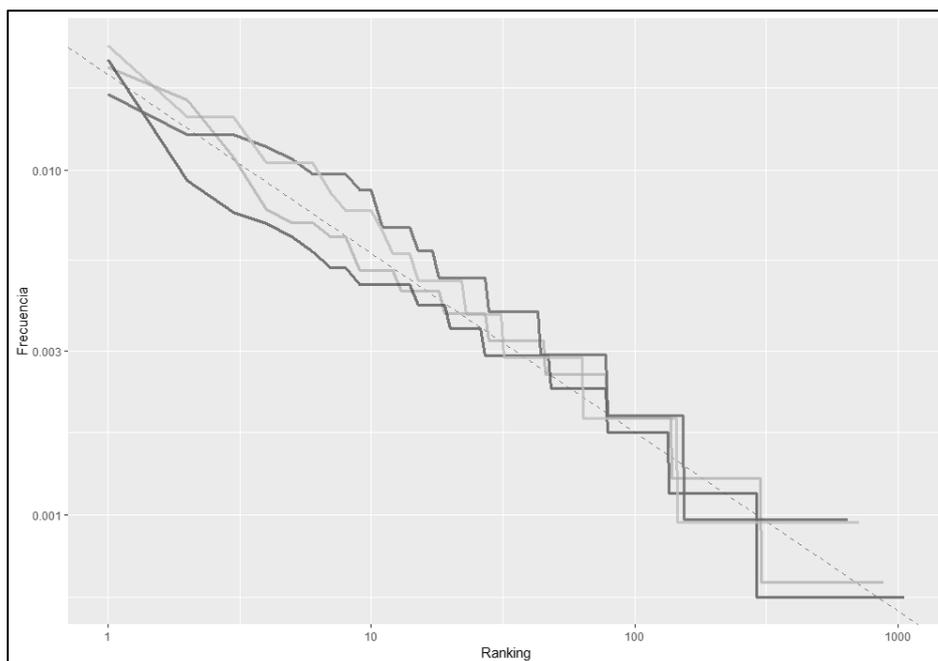


Figura 5: Ley de Zipf según discurso depurado.

En cuanto al análisis del contenido en sí de cada uno de los discursos, se propone un enfoque que permita establecer similitudes y diferencias entre ellos, teniendo en consideración las medidas *idf* y *tf-idf*, definidas en la sección Métodos. En la Tabla 5 se presentan las 10 palabras más frecuentes de cada discurso, considerando únicamente aquellas que tienen $tf-idf = 0$.

Por cómo se definen estos indicadores, las palabras tales que $idf(\text{palabra}) = 0$ (y $tf-idf(\text{palabra}) = 0$) son palabras “extremadamente” comunes en los documentos analizados, más específicamente, son palabras que aparecen en los cuatro documentos.

En cuanto a cada uno de los discursos, en la Tabla 5 se indican con un * aquellas palabras que no figuraban en la Tabla 4 ($Ranking \leq 10$). Es decir, son palabras que si bien no son las *más dichas* en su discurso, son dichas en todos los discursos.

Por otra parte, se destacan palabras que sí son frecuentes en el discurso al que pertenecen, pero no lo son en los demás, en particular:

- En el discurso de Sanguinetti 1985: espíritu, simplemente, uruguayos.
- En el discurso de Sanguinetti 1995: sistema, reforma.
- En el discurso de Vázquez 2005: compromiso, cambios, señoras.
- En el discurso de Vázquez 2015: artigas, hombre, época, principios, prócer.

Presidente	Año	Palabra	n_{palabra}	n_{total}	tf	idf	$tf-idf$
Sanguinetti	1985	país	31	1565	0.0198	0	0
		democracia	25		0.0160	0	0
		nacional	12		0.00767	0	0
		años	11		0.00703	0	0
		pueblo	10		0.00639	0	0
		política	8		0.00511	0	0
		república	8		0.00511	0	0
		solidaridad*	8		0.00511	0	0
		gobierno*	7		0.00447	0	0
		libertad*	7		0.00447	0	0
Sanguinetti	1995	país	36	1722	0.0209	0	0
		mundo	12		0.00697	0	0
		señores	11		0.00639	0	0
		años	9		0.00523	0	0
		democracia	8		0.00465	0	0
		gobierno*	8		0.00465	0	0
		valores*	7		0.00407	0	0
		vida*	7		0.00407	0	0
		libertad*	6		0.00348	0	0
		debemos*	5		0.00290	0	0

Tabla 5: Indicadores según discurso depurado – idf tf-idf.

Presidente	Año	Palabra	n_{palabra}	n_{total}	tf	idf	$tf-idf$
Vázquez	2005	señores	13	1026	0.0127	0	0
		uruguay	13		0.0127	0	0
		gobierno	12		0.0117	0	0
		derechos	10		0.00975	0	0
		país	10		0.00975	0	0
		libertad*	7		0.00682	0	0
		política*	7		0.00682	0	0
		república*	7		0.00682	0	0
		respeto*	5		0.00487	0	0
		historia*	4		0.00390	0	0
Vázquez	2015	años	15	1050	0.0143	0	0
		derechos	11		0.0105	0	0
		valores	11		0.0105	0	0
		libertad	9		0.00857	0	0
		señores*	7		0.00667	0	0
		debemos*	6		0.00571	0	0
		josé*	6		0.00571	0	0
		justicia*	5		0.00476	0	0
		pueblo*	5		0.00476	0	0
		vida*	5		0.00476	0	0

Tabla 5: Indicadores según discurso depurado – idf tf-idf (cont.).

Haciendo énfasis en estas palabras, que resultan “exclusivas” de un único discurso, en la Tabla 6 se presentan las 10 palabras con mayor $tf-idf$. Se puede observar que el 90% de ellas pertenecen a los discursos de Vázquez (4/10 a su primer discurso y 5/10 al segundo). Esto indicaría que Sanguinetti, en general, hace uso de palabras “más comunes”.

Presidente	Año	Palabra	n_{palabra}	n_{total}	tf	idf	$tf-idf$
Vázquez	2015	artigas	24	1050	0.0229	0.693	0.0158
Vázquez	2005	cambios	10	1026	0.00975	1.39	0.0135
Vázquez	2015	prócer	8	1050	0.00762	1.39	0.0106
Sanguinetti	1995	reforma	13	1722	0.00755	1.39	0.0105
Vázquez	2005	exterior	6	1026	0.00585	1.39	0.00811
Vázquez	2015	montevideo	6	1050	0.00571	1.39	0.00792
Vázquez	2015	hombre	11	1050	0.0105	0.693	0.00726
Vázquez	2015	cabildo	5	1050	0.00476	1.39	0.00660
Vázquez	2005	compatriotas	4	1026	0.00390	1.39	0.00540
Vázquez	2005	construcción	4	1026	0.00390	1.39	0.00540

Tabla 6: Indicadores según discurso depurado - Ordenados por $tf-idf$.

Por último, en cuanto a cada uno de los discursos y el uso de palabras “exclusiva” en cada uno de ellos, en la Tabla 7 se presentan las 10 palabras con mayor $tf-idf$, respectivamente.

Presidente	Año	Palabra	n_{palabra}	n_{total}	tf	idf	$tf-idf$
Sanguinetti	1985	simplemente	11	1565	0.00703	0.693	0.00487
		comprensión	5		0.00319	1.39	0.00443
		desterrar	5		0.00319	1.39	0.00443
		diversidad	5		0.00319	1.39	0.00443
		límites	5		0.00319	1.39	0.00443
		luchar	5		0.00319	1.39	0.00443
		voluntad	8		0.00511	0.693	0.00354
		administrar	4		0.00256	1.39	0.00354
		asechanzas	4		0.00256	1.39	0.00354
		crisis	4		0.00256	1.39	0.00354

Tabla 7: Indicadores según discurso depurado – idf $tf-idf$.

Presidente	Año	Palabra	n _{palabra}	n _{total}	tf	idf	tf-idf
Sanguinetti	1995	reforma	13	1722	0.00755	1.39	0.0105
		debate	5		0.00290	1.39	0.00403
		partido	5		0.00290	1.39	0.00403
		seguir	5		0.00290	1.39	0.00403
		soluciones	5		0.00290	1.39	0.00403
		partidos	9		0.00523	0.693	0.00362
		economía	8		0.00465	0.693	0.00322
		encontrar	8		0.00465	0.693	0.00322
		acción	4		0.00232	1.39	0.00322
		governabilidad	4	0.00232	1.39	0.00322	
Vázquez	2005	cambios	10	1026	0.00975	1.39	0.0135
		exterior	6		0.00585	1.39	0.00811
		compatriotas	4		0.00390	1.39	0.00540
		construcción	4		0.00390	1.39	0.00540
		humana	4		0.00390	1.39	0.00540
		permitanme	4		0.00390	1.39	0.00540
		compromiso	17		0.0166	0.288	0.00477
		igualdad	7		0.00682	0.693	0.00473
		naciones	6		0.00585	0.693	0.00405
		agenda	3	0.00292	1.39	0.00405	
Vázquez	2015	artigas	24	1050	0.0229	0.693	0.0158
		prócer	8		0.00762	1.39	0.0106
		montevideo	6		0.00571	1.39	0.00792
		hombre	11		0.0105	0.693	0.00726
		cabildo	5		0.00476	1.39	0.00660
		época	8		0.00762	0.693	0.00528
		carta	4		0.00381	1.39	0.00528
		orientales	4		0.00381	1.39	0.00528
		pueblos	15		0.0143	0.288	0.00411
		aplausos	3	0.00286	1.39	0.00396	

Tabla 7: Indicadores según discurso depurado – idf tf-idf (cont.).

DISCUSIÓN Y CONCLUSIONES

Teniendo en consideración los resultados presentados se puede apreciar, en primera instancia, una diferencia en la longitud general de los 4 discursos analizados: los 2 discursos de Sanguinetti resultan más largos que los dos de Vázquez. Resalta, además, que el primero de Sanguinetti (1985) es más largo que el segundo y que en el caso de Vázquez ocurre al revés.

En cuanto al análisis específico del contenido, a nivel general, se destaca que Vázquez utiliza más términos propios/exclusivos en sus discursos, mientras que Sanguinetti hace uso de palabras más comunes.

Del primer discurso de Sanguinetti, se destaca la utilización de las palabras: *país, democracia, nacional, uruguayos, política, república, libertad y gobierno*. Todas ellas podrían relacionarse con su condición de primer presidente democráticamente electo luego de 13 años de dictadura militar y con elementos que caracterizaron su política como primer mandatario, entre las que se destacan positivamente:

- La consolidación del régimen constitucional tras la dictadura militar.
- El carácter pacífico de la transición.
- Las medidas adoptadas en favor de las víctimas de la dictadura militar.
- La ley de amnistía dictada en marzo de 1985 en favor de las personas que todavía permanecían detenidas, condenadas por la justicia militar por delitos políticos.
- La restitución de funcionarios públicos destituidos por la dictadura, y la reparación de su carrera profesional.

Cabe resaltar, también, que su gobierno fue muy criticado, entre otras cosas, por:

- Mantenimiento de instituciones y prácticas de la dictadura militar que recortaban la vigencia de las libertades democráticas.
- Las limitaciones de la ley de amnistía mencionada, que no fue total.
- El sistemático bloqueo de la investigación de los crímenes cometidos durante la dictadura, la búsqueda de los detenidos desaparecidos y el juzgamiento de los responsables.

Por otra parte, el discurso de asunción de su segundo período presidencial se distancia del primero y se destaca por la reiterada utilización de palabras como *país, sistema y reforma*, lo que puede asociarse directamente con sus políticas implementadas.

En particular, el segundo mandato de Sanguinetti se caracterizó por la realización de las siguientes reformas:

- En Seguridad social: régimen mixto entre el Banco de Previsión Social (BPS) y un sistema complementario de ahorro individual (AFAPs).
- En Sistema electoral: reforma constitucional de 1996. Establece la elección presidencial por mayoría absoluta con posibilidad de segunda vuelta y el candidato único a Presidente por partido, entre otras modificaciones del sistema electoral y el régimen de gobierno.
- En Educación: a expansión de la oferta pública en educación inicial (universalización), creación de escuelas de tiempo completo en sectores pobres, reforma curricular de la educación media pública, desarrollo de la formación docente a través de la creación de seis centros estatales regionales, descentralizados.

En cuanto a los discursos de Vázquez se destaca, en el primero de ellos, la utilización de palabras de carácter progresista (ideología política asociada a su figura): compromiso, uruguay/os cambios, derecho/s. Estas palabras se relacionan, además con lo que fue su política como primer mandatario de izquierda en la historia del Uruguay. En particular, entre sus propuestas y proyectos, se destacan:

- Creación del Ministerio de Desarrollo Social (con varios planes sociales asociados).
- Creación del Plan CEIBAL (una PC por niño escolar).
- Creación de Consejo Ciudadano de Seguimiento de los Compromisos de Gobierno.
- Reconocimiento de los derechos políticos, civiles y sociales de todas y todos los uruguayos cualquiera sea su lugar de residencia.
- Profundización en los llamados derechos humanos "de segunda generación", promoviendo el acceso de la población a las prestaciones estatales que hacen a la calidad digna de la vida.
- Complementación en derechos laborales y seguridad social.
- Apoyo a las iniciativas internacionales tendientes a afirmar la vigencia y defensa de los Derechos Humanos.
- Recuperación de la tradición de la política exterior del Uruguay en la defensa y promoción de los derechos de los trabajadores y los sistemas de previsión social.

Por último, en 2015 Vázquez realiza, en su discurso de asunción, un repaso de la historia del Uruguay, con especial énfasis en la figura de

Artigas (prócer uruguayo) y lo cita como ejemplo de “hombre de principios”.

Es en este sentido que resaltan en su discurso palabras como: *Artigas, hombre, valores, libertad y principios*. Se destaca, además, la alta utilización de la palabra *derechos*, ya utilizada por el mandatario en su primer discurso.

BIBLIOGRAFÍA

- Aggarwal, C. (2018). *Machine Learning for Text*. Springer International Publishing AG. <https://doi.org/10.1007/978-3-319-73531-3>.
- Benzécri, J.P. (1973). *L'Analyse Des Dones. II L'Analyse Des Correspondances*. Paris. Dunod.
- Fellows, I. (2018). wordcloud: Word Clouds. R package version 2.6. <https://CRAN.R-project.org/package=wordcloud>.
- Feinerer, I. y Hornik (2018). tm: Text Mining Package. R package version 0.7-6. <https://CRAN.R-project.org/package=tm>
- Feinerer, I. Hornik, K. y David Meyer (2008). Text Mining Infrastructure in R. *Journal of Statistical. Software* 25(5): 1-54. URL: <http://www.jstatsoft.org/v25/i05/>.
- Kwartler, T. (2017). *Text Mining in Practice with R*. John Wiley & Sons Ltd.
- Leek, J. (2015). *The Elements of Data Analytic Style*, Leanpub, 2015-03-02.
- R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Viena, Austria. (<https://www.R-project.org/>).
- Rinker, T. W. (2018). textclean: Text Cleaning Tools version 0.9.3. Buffalo, New York. <https://github.com/trinker/textclean>.
- Silge, J. y Robinson, D. (2016). tidytext: Text Mining and Analysis Using Tidy Data Principles in R. *JOSS*, 1(3). doi: 10.21105/joss.00037 (URL: <https://doi.org/10.21105/joss.00037>) URL: <http://dx.doi.org/10.21105/joss.00037>.
- Silge, J. y Robinson, D. (2017). *Text mining with R: A tidy approach*. O'Reilly Media, Inc.
- Wickham, H. (2014). Tidy Data. *Journal of Statistical Software*, 59(10), 1 - 23. doi:<http://dx.doi.org/10.18637/jss.v059.i10>.
- Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Wickham, H. (2017). tidyverse: Easily Install and Load the *Tidyverse*. R package version 1.2.1. <https://CRAN.R-project.org/package=tidyverse>.
- Wickham, H. (2019). stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.4.0. <https://CRAN.R-project.org/package=stringr>.

Zhai, C. y Massung, S. (2016). Text data management and analysis: a practical introduction to information retrieval and text mining (First Edition). New York: ACM Books.

Zipf, G.K. (1936). The Psycho-Biology of Language: An Introduction to Dynamic Philology. A statistical study of vocabulary, The Modern Language Journal, Vol. 21, No. 2, pp. 125-127.