

PROPUESTA METODOLÓGICA DE ANÁLISIS DE COMUNICACIONES DEL PROGRESO EN BASE A LA LEY DE ZIPF

Emiliano Consoli

Universidad de Buenos Aires. Facultad de Ciencias Económicas. Instituto de Investigaciones en Administración, Contabilidad y Métodos Cuantitativos para la Gestión. Centro de Estudios Organizacionales.

consoli_emiliano@economicas.uba.ar

Resumen

Las Comunicaciones del Progreso (COP) son informes voluntarios que presentan las organizaciones adherentes al Pacto Global en busca de dar cuenta del impacto social, ambiental y económico de su operatoria. Las mismas, se basan en los principios de la gobernanza y buscan fortalecer el concepto de responsabilización.

Desde su origen, estos reportes han despertado el interés de la academia evidenciado en la gran cantidad de artículos centrados en su estudio. Es por ello que este trabajo busca aportar un método de análisis semiautomático de las COP basado en la ley de Zipf y sus derivados.

La ley de Zipf se trata de una ley empírica que describe un patrón de comportamiento de las palabras en el lenguaje. El descubrimiento de dicho patrón es la piedra angular que permite procesar la información para la extracción de términos que aproximen el contenido más importante de los textos.

Para contextualizar, en primer lugar se desarrolla brevemente la evolución conceptual del término gobernanza. Luego se explica la metodología propuesta para finalizar aplicando el método en cuarenta reportes de empresas de diferentes rubros.

Palabras Clave: Comunicaciones del Progreso – Frecuencia de palabras – Ley de Zipf

Abstract

Communications on Progress (COP) are voluntary reports submitted by organizations adhering to the Global Compact in order to account for the social, environmental and economic impact of their operations. They are based on the principles of governance and seek to strengthen the concept of accountability.

Since its inception, these reports have aroused the interest of the academy evidenced in the large number of articles focused on its study. That is why this work seeks to provide a method of semi-automatic analysis of POPs based on Zipf's law and its derivatives.

Zipf's law is an empirical law that describes a pattern of word behavior in language. The discovery of this pattern is the cornerstone that allows processing the information for the extraction of terms that approximate the most important content of the texts.

To contextualize, firstly, the conceptual evolution of the term governance develops briefly. Then the proposed methodology is explained to finish applying the method in forty reports of companies of different items.

Keywords: Communication on Progress - Word Frequency - Zipf Law

Introducción

La palabra "gobernanza" ha tomado notoriedad en los últimos tiempos (Mayntz, 2001). A pesar de la gran popularidad que ha logrado el tema en diferentes disciplinas de las ciencias sociales no se ha llegado aún al consenso generalizado sobre la utilización del término, por el contrario, prima la ausencia de coherencia conceptual (Hufty, 2008). Tal como lo indican Björk y Johansson (2001) *"One major problem of these new theoretical frameworks is that there are almost as many ideas of governance as there are researchers in the field"* (p. 2). Es necesario entonces revisar la bibliografía y marcar un hilo conductor que sirva para ordenar a los autores y sus aportes, para delimitar el alcance del término y para lograr mayor precisión en su uso. Para esta tarea es importante comenzar por la cuestión terminológica.

La raíz etimológica de la palabra "gobernanza" viene del latín *gubernare*, que significa pilotear (manejar) una nave; relacionado también al manejo de los asuntos públicos (Launay, 2005).

Entonces, originalmente la palabra "gobernanza" fue utilizada como sinónimo de "gobernar" asociada a los procesos de gobiernos de los estados nacionales de composición jerárquica con ejercicio del poder soberano sobre la sociedad civil. En la actualidad, la tendencia es hacia la utilización del término para hacer referencia a un nuevo modo de gobernar caracterizado por lazos cooperativos y participativos donde diferentes actores públicos y privados colaboran en la formulación de políticas públicas (Mayntz, 2001).

Al respecto, la Real Academia Española (RAE) esboza dos definiciones. Vinculada a la definición originaria RAE define: "*f. desus.* Acción y efecto de gobernar o gobernarse." (Rae.es, 2019). Como un acercamiento a la actual aplicación expone acerca de la gobernanza: "*f.* Arte o manera de gobernar que se propone como objetivo el logro de un desarrollo económico, social e institucional duradero, promoviendo un sano equilibrio entre el Estado, la sociedad civil y el mercado de la economía." (Rae.es, 2019).

1. Gobernanza: un sistema de coordinación no jerárquico

El concepto de gobernanza implica romper con los enfoques tradicionales de gobernar y gestionar donde las relaciones son jerárquicas con un estado que detenta el monopolio de coordinación de la sociedad y de la articulación política, para darle lugar a la situación donde la toma de decisiones se determina con la participación de múltiples actores. En tal sentido la gobernanza propone un nuevo modo de gobernar y gestionar donde la característica principal sea la coordinación entre distintos grupos de interés, la participación y la cooperación (Carmona, 2005).

Se observan tres enfoques claramente diferenciados cuya convergencia coincide fundamentalmente en 3 puntos, “los protagonistas y lugares de decisión son múltiples y distintos; las relaciones entre los protagonistas son horizontales más bien que verticales; las interacciones son autorreguladas” (Hufty, 2008, p. 3). En línea con el desarrollo de Hufty (2008) las tres corrientes dominantes en el debate actual son Gobierno Corporativo, Gobernanza Moderna y Gobernanza Global pudiendo vincular estos enfoques a distintos espacios decisorios según se muestra en la Figura I.

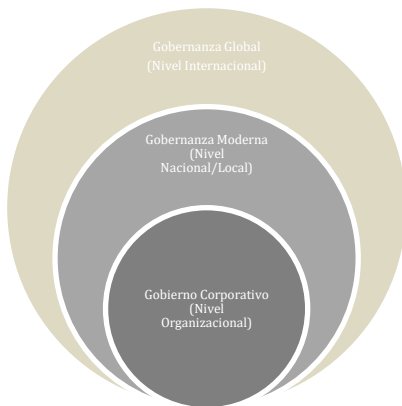


Figura 1: Niveles de la Gobernanza. Fuente: Elaboración propia

Gobernanza Global

El fenómeno de la globalización, la internacionalización de la economía y la aparición de actores privados que trascienden las barreras nacionales han aportado complejidad creciente a las sociedades generando fuertes cambios en las relaciones de poder (Carmona, 2005). La inexistencia formal de un organismo supraestatal capaz de intervenir y mediar en conflictos de carácter internacional, y la multiplicación de actores en el plano mundial (empresas, sindicatos transnacionales, organizaciones intergubernamentales, etc.) llevó a la aparición de convenios y acuerdos internacionales que regulan la vida mundial (Hufty, 2008).

En tal sentido, el concepto de gobernanza permite comprender la compleja elaboración de acuerdos cuyo fin es contribuir a solucionar problemas de coordinación por fuera del alcance de la autoridad de los Estados Nacionales (Rosenau & Czempiel, 1992). Este fenómeno donde existe coordinación y cooperación en el plano internacional entre diferentes actores (públicos y privados) pero en ausencia de una autoridad internacional articuladora Rosenau & Czempiel (1992) lo sintetizaron en el título de su libro: *"Governance without Government"*.

Gobernanza Moderna

Tal como se viene desarrollando, la Gobernanza Moderna trata de dar un enfoque alternativo al tradicional a las relaciones de poder en el seno del Estado. En tal sentido será utilizado el concepto como herramienta tanto de análisis como de acción/planificación. De esta forma el estado se ve en la situación de delegar una parte importante del poder que antiguamente ejercía soberanamente en favor de diferentes organismos tanto públicos como privados (Hufty, 2008). "Sin embargo, de lo que se trata en realidad en la gobernanza moderna no es tanto de la pérdida de control del Estado, sino más bien de que éste cambia de forma" (Mayntz, 2001).

La literatura especializada define la noción de Gobernanza Moderna como un nuevo estilo de gobierno caracterizado por un mayor grado de cooperación y vinculación entre el Estado y organizaciones no estatales con presencia de redes decisionales relacionadas a distintos niveles de gobierno, territorios, actores privados y a la sociedad civil (Carmona, 2005). Desde este punto de vista, la Gobernanza Moderna aparece como un esquema donde la formulación y aplicación de políticas públicas son

llevadas a cabo a través de la cooperación y la participación activa de organizaciones públicas y privadas, gubernamentales y no gubernamentales, de la sociedad civil, de organismos internacionales, entre otros grupos (Carmona, 2005).

Gobierno Corporativo

Por último, en el plano organizacional encontramos la vigencia del denominado Gobierno Corporativo vinculado a la gobernanza al interior de las organizaciones. Esta corriente aparece en Estados Unidos como mecanismo de control de dos tipos de organizaciones, las universidades y las empresas (Hufty, 2008, p. 4).

Hufty (2008) refiere al respecto,

Se trata por lo tanto de observar y teorizar los mecanismos concretos y las instituciones que permiten la coordinación entre las unidades de una organización el arbitraje de las divergencias, el mantenimiento de la cohesión del conjunto y la gestión de los costos de transacción implicados por un control más "horizontal" que "vertical".

En el año 1999 la Organización para la Cooperación y el Desarrollo Económicos (OCDE) ha aprobado los Principios de Gobierno Corporativo que, desde entonces, se convirtieron en una alternativa normativa de aplicación internacional de gran influencia para responsables políticos, inversores, empresas, etc. De esta manera se ha marcado la agenda del gobierno corporativo, con orientaciones específicas para la puesta en marcha de iniciativas de carácter legislativo y reglamentario tanto en países de la OCDE como en otros no-miembros (OCDE, 2004: 3).

2. Comunicaciones del Progreso (COP)

Con la finalidad de que las organizaciones adherentes al Pacto Global de la Organización de las Naciones Unidas (ONU) incorporen los principios universales de derechos humanos, trabajo, medio ambiente y lucha contra la corrupción, a través de acciones que permitan avanzar en los objetivos sociales surgen las Comunicaciones del Progreso. Se encuentran alineadas con los Objetivos del Desarrollo Sostenible, estipulados también por la ONU para el año 2030 (United Nations Global Compact, 2019).

Las COP son los informes que deben presentar las organizaciones que se incorporen a la iniciativa del Pacto Global. Su periodicidad es anual y dan cuenta de lo que han realizado en materia de sustentabilidad, así como también lo planificado para el siguiente año. Su formato puede variar entre (United Nations Global Compact, 2019):

- Un documento independiente;
- Parte de un informe de sostenibilidad o responsabilidad social corporativa;
- Parte de un informe financiero anual;
- Una plantilla COP básica (para los nuevos informes que ya no pueden publicar informe oficial);
- Una COP Express (disponible para las pequeñas y medianas empresas).

Cabe destacar que su estructura debe estar en consonancia con los diez principios del Pacto Global, que son los siguientes (United Nations Global Compact, 2019):

Cuadro 1: Principios del Pacto Global

Derechos humanos
Principio 1: Las empresas deben apoyar y respetar la protección de los derechos humanos fundamentales, reconocidos internacionalmente; y Principio 2: asegurarse de que no son cómplices en la vulneración de los derechos humanos.
Trabajo
Principio 3: Las empresas deben apoyar la libertad de afiliación y el reconocimiento efectivo del derecho de negociación colectiva; Principio 4: la eliminación de todas las formas de trabajo forzoso y obligatorio; Principio 5: la abolición efectiva del trabajo infantil; y Principio 6: la eliminación de la discriminación en materia de empleo y ocupación.
Ambiente
Principio 7: Las empresas deberán mantener un enfoque preventivo que favorezca el medio ambiente; Principio 8: fomentar las iniciativas que promuevan una mayor responsabilidad ambiental; y

Principio 9: fomentar el desarrollo y la difusión de las tecnologías respetuosas con el medio ambiente.

Anticorrupción

Principio 10: Las empresas deben trabajar contra la corrupción en todas sus formas, incluidas la extorsión y el soborno.
--

Elaboración propia

La primera COP se exige al año de haberse incorporado y tienen plazos para su publicación. En caso de vencerse los mismos, ya no pueden publicarse siendo pasibles de sanción. Actualmente el Pacto Global cuenta con 13.920 organizaciones miembros en todo el mundo, de las cuales 489 son argentinas y, dentro de ellas, 286 son empresas (United Nations Global Compact, 2019).

3. Metodología

Numerosas investigaciones buscan estudiar las Comunicaciones del Progreso presentadas por las empresas a fin de extraer conclusiones acerca de los impactos sociales, económicos y ambientales de las organizaciones.

Generalmente, el método utilizado por los investigadores consiste en la lectura atenta de los documentos en busca del contenido relevante. Esta manera de abordar el estudio de las COP presenta desventajas. En primer lugar, surge un problema asociado al tiempo disponible para su lectura que se ve agravado frente al aumento de empresas que adhieren al Pacto Global como así también por el crecimiento en volumen de las memorias.

Por otra parte, la diversidad en ramas industriales, tamaños y características de las organizaciones aporta complejidad dado que obliga al investigador a familiarizarse con terminologías distintas proveniente de la diversidad mencionada. Por último, además de ser un proceso que implica un costoso trabajo para el investigador, no minimiza la subjetividad propia de la persona.

Alternativamente, este trabajo busca aportar un método de estudio de las COP a través de técnicas de recuperación automática de palabras con alto contenido semántico a través de las cuales se pretende aproximar al contenido del documento. Para ello se parte de aportes esencialmente

provenientes de la bibliotecología y la indización automática cuyo punto de partida se encuentra en la conocida Ley de Zipf. (Urbizagástegui Alvarado y Restrepo Arango, 2011).

Mediante esta propuesta se espera una reducción drástica del tiempo y esfuerzo empleado para el estudio y seguimiento de las COP a la vez que se atenúa el componente subjetivo del investigador en la extracción de contenido relevante.

3.1. La ley de Zipf

El aporte de la ley de Zipf en la lingüística radica en haber detectado un patrón de comportamiento asociado al lenguaje que permite ordenar un arreglo distribucional de las palabras utilizadas en un documento cualquiera.

En concreto, esta ley empírica, relaciona la frecuencia de repetición de las palabras y su rango. Entendido este último como la longitud de las palabras (cantidad de letras). Las más utilizadas resultan ser las de menor longitud.

En base a este hallazgo, George Kingsley Zipf enuncia la teoría psicológica popularizada como ley del menor esfuerzo. La cual establece que las personas tendemos a realizar el mínimo esfuerzo posible para lograr resultados exitosos. Este comportamiento también se replica en el lenguaje y es por ello que las palabras más utilizadas son las de menor longitud. (Urbizagástegui Alvarado y Restrepo Arango, 2011)

Así, al ordenar las palabras según su frecuencia de aparición, cantidad de veces que se utiliza en un texto, donde la palabra de mayor frecuencia tiene rango uno, la segunda palabra de mayor frecuencia tiene rango dos, y así sucesivamente, llegamos al enunciado de la primera ley de Zipf que describe el comportamiento de las palabras de mayor frecuencia. (Lee Pao, 1977)

Logrando enunciarlo en la siguiente expresión matemática:

$$r \times f = c \quad (1)$$

donde,

r = rango

f = frecuencia

c = constante de Zipf

Hay que aclarar que este enunciado solo resulta válido para las palabras con mayor frecuencia. Para las palabras de menor frecuencia aplica la segunda enunciación de la ley de Zipf donde la representación de su comportamiento viene dada por la siguiente expresión (Lee Pao, 1977):

$$\frac{I_1}{I_n} = \frac{n(n+1)}{2} \quad (2)$$

Donde,

I_1 = Cantidad de palabras de rango 1

I_n = Cantidad de palabras con frecuencia n

n = cantidad de palabras

3.2. Punto de transición de Goffman

Siguiendo el desarrollo donde las palabras se ordenan de manera descendente según su frecuencia de aparición, William Goffman supuso la existencia de un punto donde se pasa de palabras de mayor frecuencia a palabras de menor frecuencia; es decir, el punto de transición de Goffman. (Lee Pao, 1977)

Adicionalmente, relaciona las expresiones de mayor frecuencia a términos generales y a las de menor frecuencia con términos específicos. Es decir que los primeros tienden a unificar el lenguaje siendo que se trata de artículos, conectores, pronombre, etc. Y los segundos tienden a aportar acerca del contenido del texto según el área del conocimiento a la que pertenecen.

Para localizar el punto de transición se parte de la expresión matemática enunciada para palabras de baja frecuencia (2). Considerando el comportamiento de las palabras más utilizadas, solo una tendrá frecuencia de aparición n , entonces:

$$\frac{I_1}{1} = \frac{n(n+1)}{2} \quad (2)$$

Igualando la expresión a cero llegamos a,

$$0 = n^2 + n - 2I_1 \quad (3)$$

Si observamos con atención, la expresión anterior responde a la forma de la ecuación cuadrática de segundo grado; por tanto, puede ser resuelta aplicando la conocida fórmula resolvente. Aunque como solo interesan para el caso los números naturales positivos, se considera únicamente su expresión positiva mediante la cual es posible conseguir el punto de transición. (Lee Pao, 1977)

$$n = \frac{-1 + \sqrt{1 + 8I_1}}{2} \quad (4)$$

3.3. Corte de Luhn

Hands Peter Luhn, en consonancia con la ley de Zipf y el punto de transición de Goffman, propone la existencia de dos puntos de corte (superior e inferior). El punto de corte superior descarta las palabras de alta frecuencia pues carecen de contenido. Es decir, que solo las palabras a su derecha son pasibles de tener significado. De la misma manera se establece un punto de corte inferior que descarta las palabras de baja frecuencia que tampoco poseen valor semántico para aproximar el

contenido del texto. Es entonces que las palabras de mayor contenido semántico útiles para describir de manera aproximada el contenido del texto se encuentran entre estos dos puntos de corte. (Urbizagástegui Alvarado y Restrepo Arango, 2011)

A continuación, se ofrece la representación gráfica del modelo completo (figura 1) considerando la distribución de palabras según la ley de Zipf, el punto de transición de Goffman y el corte de Luhn.

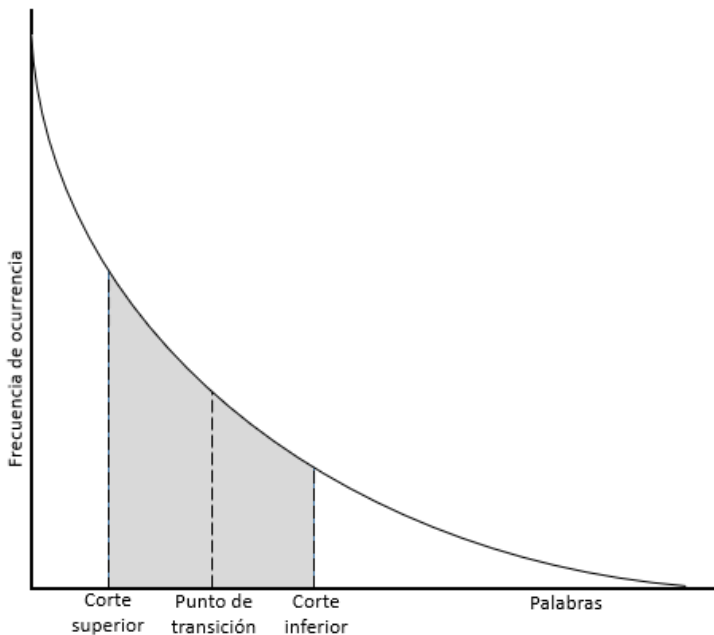


Figura 2: Distribución de palabras – Ley de Zipf. Elaboración propia en base a (Urbizagástegui Alvarado y Restrepo Arango, 2011)

3.4. Comparación de textos en base a la ley de Zipf

Como se vio hasta aquí, el reconocimiento del patrón de conducta de las palabras en el lenguaje permite su procesamiento. No solo para la extracción de palabras descriptivas, sino que para la comparación.

Para ello, es posible establecer un método predominantemente gráfico de comparación de textos como se muestra en la figura 2. De la misma manera que se ha desarrollado hasta aquí, se calcula la frecuencia de aparición de las palabras en ambos textos. El eje de ordenadas representa la frecuencia de aparición de las palabras en un texto A, mientras que el eje de abscisas la frecuencia de las palabras de un texto B. (Rudder, 2016)

Aquellas palabras cuya frecuencia sea similar en ambos textos tenderán a acercarse a la recta Z (Figura 2). Es decir que estas no son útiles para diferenciar los textos. De la misma manera, las que concentren mayor frecuencia en un texto y menor frecuencia en el otro tendrán tendencia a alejarse de la recta. Es decir que, las palabras que se separen de la línea punteada (Figura 2) serán aquellas que mejor diferencian el contenido de uno en comparación del otro.

Tal como se muestra en la figura 2, el punto *a* será una de las palabras que diferencien el contenido del texto A en relación con el texto B, quedando determinado según su frecuencia de aparición en ambos textos.

Esta técnica no solo permite comparar un texto con otro, sino que además permite comprar un grupo de textos contra otro, o contra un documento individual. Por ejemplo, si quisiera saber que palabras diferencian mejor a las COP emitidas por bancos de las elaboradas por automotrices podría utilizar este modelo agrupando las COP de bancos bajo la denominación texto A y las de automotrices como texto B. También resultaría útil, por ejemplo, para encontrar las diferencias entre los reportes que una misma organización emite en años diferentes.

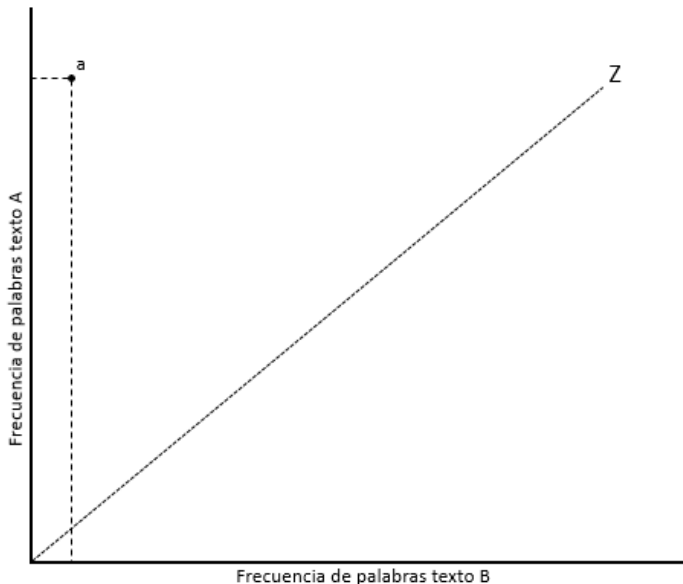


Figura 3: comparación de textos en base a la ley de Zipf.

3.5. Consideraciones generales

Este estudio cuenta con dos etapas. Una primera etapa exploratoria donde se determina las tendencias existentes en las COP y una segunda etapa descriptiva donde se busca determinar la relación entre las palabras.

Para la primera etapa, como unidad de análisis se tomaron las palabras que conforman los reportes de sustentabilidad o COP presentados por las empresas adherentes al Pacto Global según el relevamiento realizado por Emmanuel Oliverio et. al. (2018). Se ha tomado una muestra aleatoria sin reposición de tres compañías habiendo sido seleccionadas Banco Santander, Arcor y Mercedes Benz.

Para la segunda etapa, como unidad de análisis se tomaron las palabras que conforman los reportes de sustentabilidad o COP presentados por las

empresas argentinas adherentes al Pacto Global que han reportado y publicado su reporte para el ejercicio 2018 (United Nations Global Compact, 2019). Se ha tomado una muestra aleatoria sin reposición de cuarenta compañías.

Se ha utilizado el software RStudio, un entorno de desarrollo integrado (IDE) para el lenguaje de programación R.

Para la aplicación de los procedimientos descritos se ha seguido, en líneas generales, la metodología propuesta por Rubén Urbizagástegui Alvarado y Cristina Restrepo Arango (2011) con algunos agregados. Se expone a continuación.

- “Una palabra es considerada como una serie de caracteres tipográficos precedida y seguida de espacios en blanco en el texto.” (Urbizagástegui Alvarado y Restrepo Arango, 2011, p. 82)
- Las palabras unidas con guion fueran tratadas como una palabra.
- Las palabras fonéticamente diferentes se han tomado como distintas.
- Se eliminaron los nombres de autores citados en el texto como así también los nombres de miembros de las compañías.
- En cambio, los nombres de instituciones se han conservado pues para el caso se los considera relevantes para el contenido del texto dado que mayoritariamente se trata de organizaciones con algún tipo de articulación entre éstas y las compañías emisoras de reportes.
- Se retiraron los caracteres numéricos incluyendo montos, fechas, cantidades, mediciones, fórmulas, etc.
- Se retiraron las abreviaturas.
- Se retiraron del texto los links de páginas web.

4. Resultados

Como se ha mencionado en el apartado anterior, se han estudiado mediante la metodología propuesta los reportes emitidos por las compañías Arcor, Santander y Mercedes Benz. Al haber utilizado el relevamiento surgido del trabajo de Oliverio et. al (2018) se estudiaron las memorias correspondientes al ejercicio 2015. Adicionalmente, para el caso de Arcor se consideró, además, el reporte emitido para el año 2014 a fin de comparar ambos ejercicios.

Para el corte superior e inferior se han establecido cinco (5) palabras encima del punto de transición y diez (10) debajo. Con lo cual se han extraído quince (15) palabras de cada reporte. Además, se han eliminado para este procedimiento aquellas palabras cuyo contenido semántico resultara nulo.

Arcor

En el reporte Arcor 2015 se han contabilizado 32.106 palabras siendo que se detectan 4.652 palabras diferentes, su punto de transición se calculó en 71. El reporte de la misma empresa para el año 2014 presenta 30.447 palabras en total mientras que 4.148 son las palabras individuales encontradas, para este reporte el punto de transición se establece en 63.

Tabla 1. Palabras de mayor contenido semántico Arcor 2015

Rango (r)	Palabra	Frecuencia de Clase (fi)	Frecuencia relativa (fi/n)
29	EMPRESA	91	0,002834361
31	TRABAJO	82	0,00255404
32	COLABORADORES	81	0,002522893
32	DESARROLLO	81	0,002522893
34	SUSTENTABLE	80	0,002491746
38	PLANTAS	70	0,002180278
40	VIDA	67	0,002086837
42	MARCO	64	0,001993397
42	PROVEEDORES	64	0,001993397
44	PRODUCTOS	61	0,001899956
45	NIÑOS	60	0,00186881
46	DERECHOS	57	0,001775369
51	INFANCIA	52	0,001619635
51	PROYECTOS	52	0,001619635
55	CALIDAD	51	0,001588488

Fuente: Elaboración Propia

Tabla 2. Palabras de mayor contenido semántico Arcor 2014

Rango (r)	Palabra	Frecuencia de Clase (fi)	Frecuencia relativa (fi/n)
31	EMPRESA	77	0,002528985
32	COLABORADORES	75	0,002463297
33	DESARROLLO	74	0,002430453
35	TRABAJO	71	0,002331921
39	PROYECTOS	66	0,002167701
42	CONSUMO	60	0,001970638
42	PRODUCTOS	60	0,001970638
48	NIÑOS	53	0,00174073
49	INICIATIVAS	52	0,001707886
50	PROVEEDORES	51	0,001675042
53	FUNDACIÓN	49	0,001609354
53	VIDA	49	0,001609354
56	CAUDAD	48	0,00157651
56	COMUNIDAD	48	0,00157651
58	SUSTENTABLE	47	0,001543666

Fuente: Elaboración propia

Santander Río

De la misma manera, el Banco Santander Río ha publicado su reporte para el año 2015 con un total de 30.502 palabras de las cuales 4.570 son palabras individuales, en este caso el punto de transición corresponde a 69.

Tabla 3. Palabras de mayor contenido semántico Santander 2015

Rango (r)	Palabra	Frecuencia de Clase (fi)	Frecuencia Relativa (fi/n)
28	GESTIÓN	97	0,003180119
29	PROGRAMA	88	0,002885057
32	PERSONAS	79	0,002589994
38	SUCURSALES	60	0,001967084
39	COMITÉ	57	0,00186873
42	ARGENTINA	55	0,00180316
42	DESARROLLO	55	0,00180316
42	MILLONES	55	0,00180316
45	COLABORADORES	53	0,001737591
48	BANCA	52	0,001704806
48	RIESGOS	52	0,001704806
51	SERVICIOS	51	0,001672022
54	DURANTE	48	0,001573667
54	FORMACIÓN	48	0,001573667
54	GRUPO	48	0,001573667

Fuente: Elaboración propia

Mercedes Benz

Por último, la COP presentada por Mercedes Benz para el ejercicio 2015 cuenta con un total de palabras de 16.712 totales, siendo que 3.253 son las palabras diferentes en el documento. El punto de transición en este caso resulta ser 59.

Tabla 4. Palabras de mayor contenido semántico Mercedes Benz 2015

Rango (r)	Palabra	Frecuencia de Clase (fi)	Frecuencia relativa (fi/n)
18	MERCEDES BENZ	71	0,004248444
21	EMPLEADOS	64	0,003829584
22	ARGENTINA	63	0,003769746
24	DAIMLER	52	0,003111537
24	TRABAJO	52	0,003111537
29	COMPLIANCE	43	0,002573001
32	DESARROLLO	40	0,00239349
34	PROGRAMA	39	0,002333652
36	GESTIÓN	38	0,002273815
39	SEGURIDAD	36	0,002154141
40	CENTRO	35	0,002094303
41	RESIDUOS	34	0,002034466
42	CONCESIONARIOS	33	0,001974629
48	INTEGRIDAD	30	0,001795117
49	CAPACITACIÓN	29	0,00173528

Fuente: Elaboración Propia

El gráfico 1 (a continuación) es la representación gráfica surgida de la comparación de los reportes emitidos por Arcor para los ejercicios 2014 y 2015.

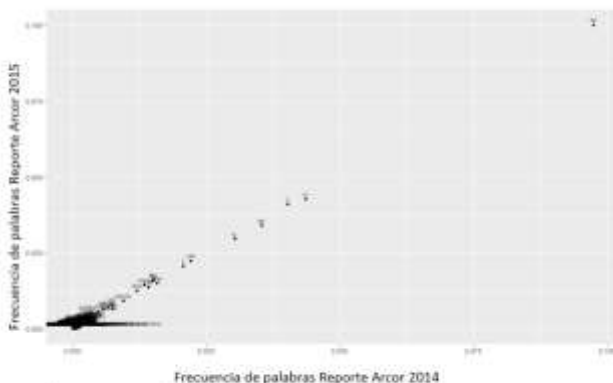


Gráfico 1: Comparación COP Arcor 2014 y 2015. Fuente: Elaboración propia

De manera complementaria la tabla 5 expone los resultados conseguidos de comprar los mismos documentos que el gráfico 1, aunque en esta ocasión se comparan las palabras según su frecuencia relativa, es decir, según el cociente entre la cantidad de veces que se repite la palabra y el total de palabras en el documento. El resultado de esta comparación deberá mostrar las palabras que mejor diferencian al reporte del año 2015 del confeccionado para el 2014.

Tabla 5. Comparación por frecuencias relativas Arcor 2014 y Arcor 2015

Orden	Palabras	Frecuencia relativa Arcor 2015	Frecuencia relativa Arcor 2014
1	CAJAS	3,74E+02	0
2	DEMANDA	3,74E+02	0
3	PERCEPCIÓN	2,80E+02	0
4	ESO	2,18E+02	0
5	OPERACIONAL	2,18E+02	0
6	FASE	1,87E+02	0
7	GENERANDO	1,87E+02	0
8	RIESGO	1,87E+02	0
9	KILOGRAMOS	1,56E+02	0
10	PRINCIPALMENTE	1,56E+02	0

Fuente: Eleboración Propia

La comparación anterior resulta de la misma organización en distintos momentos. A fin de comprar compañías de distintos rubros en un mismo marco temporal, se expone a continuación la representación gráfica de las diferencias surgidas del procesamiento de las memorias de Mercedes Benz y Santander Rio (gráfico 2).

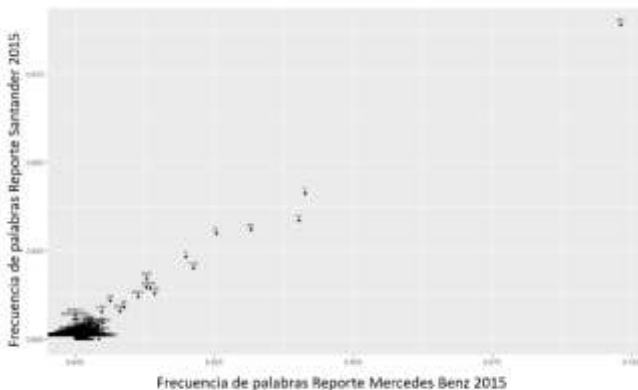


Gráfico 2: Comparación COP bancos e industria automotriz (Santander y Mercedes Benz)
Fuente: Elaboración propia

Tal como se hizo para la comparación de Arcor, también en este caso se comparan los documentos según la frecuencia relativa de sus documentos. La tabla 6 muestra las palabras que mejor diferencian el reporte de Mercedes Benz del de Santander Río. Mientas que la tabla 7 muestra las palabras que mejor describen al reporte de Santander Río en comparación con el reporte de Mercedes Benz.

Tabla 6. Comparación por frecuencias relativas Mercedes Benz 2015 y Santander 2015

Orden	Palabras	Frecuencia relativa Mercedes Benz 2015	Frecuencia relativa Santander 2015
1	MERCEDES BENZ	0.0042484442	0
2	DAIMLER	0.0031115366	0
3	CONCESIONARIOS	0.0019746290	0
4	INDUSTRIAL	0.0014959311	0
5	PLANTA	0.0011369076	0
6	TÉCNICA	0.0010172331	0
7	RUBRO	0.0009573959	0
8	CUANTO	0.0008377214	0
9	BUSES	0.0007778842	0
10	COMUNES	0.0007778842	0

Fuente: Elaboración propia

Tabla 7. Comparación por frecuencias relativas Santander 2015 y Mercedes Benz 2015

Orden	Palabras	Frecuencia relativa Mercedes Benz 2015	Frecuencia relativa Santander 2015
1	BANCO	0.0069175792	0
2	SANTANDER	0.0058684676	0
3	RÍO	0.0041636614	0
4	SUCURSALES	0.0019670841	0
5	BANCA	0.0017048062	0
6	CRÉDITO	0.0009507573	0
7	PROPUESTA	0.0007540489	0
8	PYME	0.0007212642	0
9	FOOD	0.0006229100	0
10	TARJETAS	0.0005573405	0

Fuente: Elaboración propia

Habiendo concluido la primera etapa exploratoria, se han identificado 3 tendencias marcadas. La primera responde a palabras de connotación positiva, la segunda tendencia a palabras relativas a la gestión de cada

compañía y una última tendencia observada apunta hacia la descripción de *stakeholders* o grupos de interés.

Es así que, para la muestra utilizada en la segunda etapa descriptiva, se han catalogado las palabras entre positivas, de gestión, referentes a *stakeholders* y neutrales. Esta última categoría agrupa a las palabras neutrales para el estudio. Es decir que, no significa que las palabras carecen de contenido, sino que no pueden ser catalogadas en las categorías anteriores. A fin de disminuir el potencial sesgo del autor, se han agrupado la totalidad de las palabras, se han catalogado y luego se vincularon a cada compañía.

Tabla 8. Palabras únicas

Categoría	Frecuencia
gestión	91
positiva	63
neutral	49
stakeholders	25

Fuente: Elaboración Propia

Tal como puede observarse en la tabla 8 se han extraído 228 palabras únicas. Siendo que la mayor cantidad de palabras únicas corresponde a la categoría gestión. Sin embargo, al considerar la cantidad de reportes en las que aparecen las palabras la mayor cantidad de repeticiones se concentra en las palabras de connotación positiva tal como puede observarse en la tabla 9.

Tabla 9. Repetición de Palabras

Categoría	Frecuencia	Frecuencia Relativa
positiva	267	0,434146341
gestión	178	0,303754266
neutral	95	0,169340463
stakeholders	75	0,139146568

Fuente: Elaboración propia

En la tabla 9 la frecuencia corresponde a la sumatoria de la cantidad de reportes en que aparece cada palabra. Mientras que la frecuencia relativa

es el cociente entre la frecuencia y la suma total de las frecuencias de las cuatro categorías.

Así mismo, si consideramos las 10 palabras de mayor repetición encontramos que 7 de fueron catalogadas como palabras de connotación positiva mientras que las 3 restantes refieren a la gestión.

A continuación, se presentan los resultados obtenidos del procesamiento según la metodología expuesta de cada uno de los reportes que componen la muestra.

Tabla 20. Tabla resumen palabras de mayor contenido semántico COP y 2018

ID	Compañía	Cantidad de palabras	Punto de Transición	Variedad	Frecuencia de Clase (f)			Frecuencia relativa (f/n)				
					Positivas	Gestión Stakeholders	neutral	Positivas	Gestión Stakeholders	neutral		
318	La pasada SRL	5129	38	1158	5	4	1	5	33%	27%	7%	13%
194	ESEF	5128	65	4411	5	4	3	3	32%	27%	26%	30%
177	EDESA SA	4348	52	2845	6	5	3	1	48%	43%	26%	7%
434	Barbana	2134	55	811	18	3	2	0	67%	10%	13%	0%
274	Grupo Luro	1988	55	2887	5	5	2	1	25%	66%	13%	7%
425	Sociedad Integral Emertense S.A.	9128	46	1312	6	3	3	3	48%	26%	26%	20%
322	Jetalia S.A.	5902	39	1217	6	3	2	4	48%	26%	13%	17%
117	IFF	46676	69	5014	8	4	2	5	40%	27%	13%	20%
381	Petrolera Australgas Energía S.A.	2791	9	538	9	5	0	5	80%	25%	0%	20%
352	Martelero Hermanos S.A.	18258	47	2134	5	7	1	2	13%	47%	7%	13%
409	Rensuik Argentina S.A.	18035	57	3058	7	4	1	3	47%	27%	7%	20%
25	Atlas Gráficas Molinas S.A.	4543	41	1379	9	3	1	2	60%	30%	7%	13%
297	Hotel Luro Group	5112	41	1461	6	4	2	5	40%	27%	13%	20%
81	Shogit S.A.	3117	54	2367	6	5	2	2	40%	33%	12%	12%
344	Hershey S.A.	7346	46	1854	7	5	1	2	47%	32%	7%	12%
458	Telmerado SRL	4452	39	1247	7	2	3	3	47%	13%	26%	20%
377	IB Cargo S.R.L.	14848	61	3451	6	4	2	3	40%	27%	13%	20%
189	Par-Jeamerlin Energy L.L.C.	12323	55	2765	7	4	3	1	47%	27%	26%	7%
339	La Voz del Interior S.A.	5809	40	1262	5	4	2	4	32%	27%	12%	17%
76	BBC+ Banco Páramo S.A.	18337	71	4935	5	8	1	1	33%	53%	7%	7%
349	Lupia SACIF	9436	45	2015	5	7	2	1	33%	47%	13%	7%
9	Servicios Argentinos 2010	33414	66	4043	6	4	2	3	40%	27%	13%	20%
4	Admiral Grupo Aeromilitar	59331	68	5068	9	3	2	1	60%	26%	12%	7%
80	Boston S.A.	9311	51	2381	8	4	3	0	53%	27%	26%	0%
413	Río Uruguay Cooperativa de Seguros Ltda.	14147	58	3012	7	5	0	5	47%	33%	0%	20%
413	GRUPO S.A.	2913	35	958	5	6	2	2	33%	40%	12%	12%
262	Frutandina S.A.	4941	40	3121	7	4	2	2	47%	27%	12%	12%
298	EDASA	6388	47	1785	6	5	3	1	48%	33%	26%	7%
112	Casera Puerto Santa Fe S.A.	20012	63	3719	7	3	2	5	47%	30%	13%	20%
438	Hershey S.A.	18894	54	5161	9	4	0	2	60%	27%	0%	13%
196	Elkanto Bricker Viente	2857	35	898	6	4	1	4	40%	27%	7%	12%
346	Marshall Midlat S.A.	1696	26	553	8	5	1	1	53%	32%	7%	7%
262	P.G.E. Tagari S.R.L.	1336	34	470	6	4	0	5	40%	27%	0%	13%
181	El Humero Igazú Hotel Thermal	4436	13	920	5	4	4	2	33%	27%	27%	13%
152	Compañía de Seguros La Mercedi Analisa S.A.	5099	57	1098	8	3	5	1	40%	26%	33%	7%
431	Securitas Argentina S.A.	13447	48	2167	3	6	2	4	20%	46%	13%	17%
157	COVIMAT S.R.L.	6251	11	1106	7	4	1	8	47%	27%	7%	20%
31	Banco Hertz S.A.	44638	74	5891	5	5	3	2	33%	33%	26%	13%
17	ATA Tecnología Administrativa S.A.	2730	30	729	8	4	0	5	53%	27%	0%	20%
326	Ryff S.A.	9722	49	2170	10	2	2	1	67%	13%	12%	7%
467	Refinera del Norte S.A.	15854	53	2408	8	4	1	2	53%	27%	7%	13%

Fuente: Elaboración Propia

Conclusiones

Habiendo aplicado la metodología desarrollada; podemos concluir que el procedimiento de textos, basado en la ley de Zipf y sus agregados, resulta ser una herramienta de gran potencia analítica.

Permite aproximar el contenido más importante de los documentos mediante técnicas no intelectuales, es decir, sin necesidad de leer el texto; a la vez que reduce drásticamente el tiempo necesario para el estudio de los reportes.

Particularmente, al considerar los datos extraídos se observa en primer lugar que las palabras destacadas como las de mayor contenido semántico para describir las memorias de sostenibilidad relevadas encontramos tres tendencias marcadas.

En primer lugar, encontramos predominio de palabras de connotación positiva. Considerando que se espera que las COP o los reportes de sostenibilidad reflejen el impacto de las empresas en el plano económico, social y ambiental; este predominio evidencia que en este tipo de documento no informan los impactos mencionados sino que se centran en detallar las acciones de corte filantrópicas que las organizaciones emisoras impulsan o financian. Ejemplos de este tipo de palabras son colaboradores, desarrollo, sustentables, sostenible, niñez, integridad, iniciativa, entre otras.

En segundo lugar, encontramos una tendencia menos marcada hacia la presencia de palabras orientadas al negocio o la operatoria de la empresa. Este deja entrever que los reportes de sostenibilidad son utilizados por las organizaciones como informes de gestión. Por ejemplo: concesionaria, riesgo, banca, entre otras.

Por último, podemos observar una tercera tendencia que nos permite inferir que los reportes dedican parte de su contenido a informar acerca de su relación con distintos grupos de interés. Por ejemplo, proveedores, comunidad.

Además, se ha realizado una comparación entre las memorias publicadas por la empresa Arcor para el año 2014 y 2015. Tal como se esperaba, las palabras tienden a agruparse en el centro del gráfico dado que existe una gran porción de las mismas que unifican el lenguaje. En tal sentido pueden distinguirse claramente los artículos, conectores, preposiciones que confirman la utilidad de la herramienta. Aunque no se observa ninguna palabra que marcadamente se aleje del grupo y sea útil para diferenciar los documentos.

Adicionalmente, se realizó la misma comparación, pero a través de la frecuencia relativa de las palabras en ambos textos (tabla 5). Se extrajeron las diez palabras que mejor diferencian al ejercicio 2015 del 2014. El resultado conseguido redunda en palabras de bajo o nulo contenido semántico a la vez que su frecuencia relativa resulta muy baja y por tanto poco representativo. Esta evidencia permite concluir que no existen diferencias sustanciales entre los reportes presentados por la empresa en los dos años consecutivos.

Finalmente, al comprar con la misma herramienta las COP de Mercedes Benz y de Banco Santander Río si se han logrado palabras cuya frecuencia relativa las muestra como capaces de diferenciar los documentos. Aunque responden a particularidades asociadas a los negocios de cada compañía confirmando la tendencia anteriormente mencionada donde se observa que las COP son utilizadas como informes públicos de gestión.

Cabe mencionar que, tras el análisis estadístico del contenido de las COP, no encontramos diferencias sustanciales entre las distintas memorias. Es decir, que se evidencia una clara tendencia hacia la similitud de los diferentes reportes con particularidades aportadas por la operación y el tipo de negocio de cada una.

Referencias bibliográficas

- Björk, P., y Johansson, H. (2001). Toward Governance Theory: In Search for a Common Ground. *IPSA Papers*.
- Carmona, R., (2005), Nuevas formas de gobierno y gestión pública en el escenario local: Elementos y perspectivas para el estímulo de procesos concertados de desarrollo, *III Congreso Argentino de Administración Pública*. Congreso llevado a cabo en Tucuman, Argentina.
- Hufty, M. (2008). Una propuesta para concretizar el concepto de gobernanza: El Marco Analítico de la Gobernanza. *Actes & Mémoires de l'Institut Français d'Études Andines*. 25. 77-100.
- Launay, C. (2005). La gobernanza: Estado, ciudadanía y renovación de lo político. Origen, definición e implicaciones del concepto en Colombia. *Revista Controversia*. 185. 91-105. Centro de investigación y educación popular (CINEP).
- Lee Pao, M. (1978). Automatic text análisis based on transition phenomena of word occurrences. *Journal of the American society for information science*, 14 (40).121-124.
- Mayntz, R. (2001). El estado y la sociedad civil en la gobernanza moderna. *Revista del CLAD Reforma y Democracia*. 21. 1-8.
- OCDE (2004), *Principios de Gobierno Corporativo de la OCDE*, Paris Francia, OCDE.
- OCDE (2015), *Directrices de la OCDE sobre el Gobierno Corporativo de las Empresas Públicas*, Paris Francia, OCDE.
- Oliverio, E., Consoli, E., Colacioppo, P. (2018). Análisis exploratorio sobre las políticas de diversidad en las empresas argentinas. V Simposio de Responsabilidad Social de las Organizaciones (SIRSO). Universidad Autónoma de Chile (UA), Chile.

- Rae.es. (2017). Real Academia Española. [en línea] disponible en: <http://www.rae.es/> [consultado 16 Sep. 2019].
- Rosenau, J., Czempiel, E. (1992), *Governance without Government: Order and Change in World Politics*, Cambridge Reino Unido, Cambridge University Press.
- Rudder, C., (2016), *Dataclismo: Amor, sexo, raza e identidad; lo que nuestra vida online cuenta de nosotros*, Barcelona, España , Aguilar.
- United Nations Global Compact, (2019), United Nations Global Compact. Recuperado de <https://www.unglobalcompact.org/>
- Urbizagástegui Alvarado, R., Restrepo Arango, C. (2011). La ley de Zipf y el punto de transición de Goffman en la indización automática. *Revista Investigación Bibliotecológica*, 25 (54). 71-92.