

# 'El discurso de Gastón': Un acercamiento *computacional* al análisis del discurso

Luis Josué Nuñez Guerrero

## 1. Introducción

El presente trabajo analiza, desde los presupuestos teóricos de los Estudios Críticos del Discurso (ECD), el discurso de apertura del año académico 2006-1 realizado por el chef y empresario peruano Gastón Acurio en la Universidad del Pacífico (UP). Desde este marco teórico, “el discurso de Gastón”<sup>1</sup> se entiende como una práctica discursiva y social que constituye y refleja concepciones compartidas a partir del uso de diversas estrategias lingüístico-discursivas. Identificar e interpretar estas estrategias es el objetivo principal de este trabajo. Como se evidencia en el título, se hará uso de herramientas computacionales para el pre-procesamiento, análisis exploratorio y visualización de la data, lo cual se detallará en la sección correspondiente a la metodología. A continuación, se planteará el problema específico y presentarán algunos los conceptos que guiarán el análisis y las estrategias a identificar e interpretar.

### 1.1. Planteamiento del problema

En el marco de los ECD, se tiene una concepción tridimensional del *discurso*: “es *texto*, es *práctica discursiva* y, fundamentalmente, es *práctica social*” (Arrunátegui 2010, p. 438). Lo anterior busca dar cuenta que el discurso es una práctica que significa, constituye y construye la realidad: tiene “efectos de corto, mediano y largo plazo en dimensiones simbólicas como los conocimientos, las creencias y las identidades de las personas, pero también producen cambios en las prácticas y relaciones sociales, así como en el mundo material” (Navarro & Tromben 2019: 297).

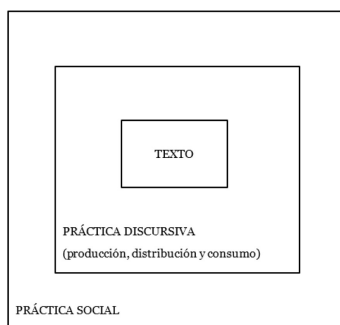


Figura 1: tomada de Fairclough (1992, p. xx)

En la figura 1 se puede visualizar –también– un modelo teórico-analítico que, en este trabajo, permitirá analizar “el discurso de Gastón” y los efectos que tuvo en la construcción de la idea

---

<sup>1</sup>Puede verse en link

de nación y la *peruanidad* como identidad que siguen teniendo relevancia a día de hoy. En este trabajo se hará énfasis en el análisis textual: se prestará atención a aspectos formales como el vocabulario, la gramática, la cohesión y la organización estructural del discurso en tanto texto. Sobre esto último, es importante mencionar que este trabajo se enmarca dentro de un estudio más amplio sobre la construcción de la idea de nación y sus vínculos con el proyecto *gastropolítico* (García, 2023).

El análisis textual del 'discurso de Gastón' tomará como punto de partida el vocabulario, prestando atención a las palabras y combinación de palabras más significativas y relevantes del texto. Lo anterior también nos permitirá identificar los “actores sociales” o participantes. Seguido de esto, tomaremos a la cláusula u “oración simple” como unidad de análisis, pues es la unidad estructural a partir de la cual construimos significados. Dentro de la cláusula nos centraremos en los distintos procesos propuestos por Halliday (2004) dentro del sistema de transitividad, los cuales pueden observarse en la figura 2.

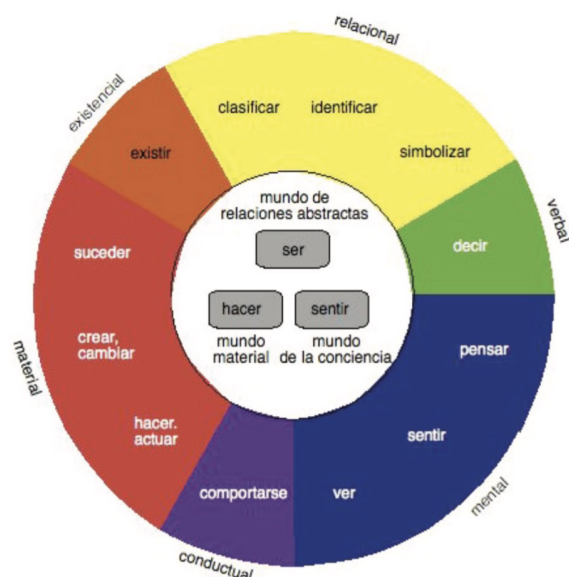


Figura 2: Halliday (2004)

Esta categorización será utilizada para clasificar los verbos de las distintas cláusulas del discurso y, así, dar cuenta de “la realidad discursiva”<sup>2</sup> construida por el autor y las estrategias que ha utilizado para hacerlo. Veamos algunos ejemplos:

- (1) a. “Somos una gran nación [...]”
- b. “Todos los estudios de mercado hechos fuera del Perú por empresas internacionales indican que [...]”

El uso de un proceso relacional en (1a) permite vincular el atributo “gran nación” al participante “nosotros”, o sea, los peruanos. Por otro lado, en (1b) observamos un proceso verbal que le permite hacer uso de una estrategia intertextual, esto es, vincular su mensaje con el de otros textos. Lo anterior, en un nivel interpretativo nos permitiría evidenciar como se construye la idea de nación a partir de intereses empresariales externos.

<sup>2</sup>Vich, 2010

## 2. Justificación

El uso de técnicas computacionales y *software* no es ajeno al análisis del discurso como disciplina y metodología (Fajardo, 2023, pp. 26-30), pero se han priorizado los análisis de variación lingüística, de contenido o de sentimiento. Son pocos<sup>3</sup> los trabajos guiados por los principios de los ECD que han utilizado herramientas de la lingüística computacional o de la lingüística de corpus, la cual ha aprovechado estas técnicas para facilitar el trabajo con grandes cantidades de datos. Como adelante en la sección anterior, este trabajo, como parte de una investigación más amplia, busca ser un primer acercamiento a las técnicas computacionales desde los ECD y, además, se presenta como un estudio 'piloto', que busca luego poder ampliarse para abarcar un corpus heterogéneo de textos relevantes en la construcción *gastropolítica* de la idea de nación. A continuación se presentará en detalle la data elegida y su relevancia.

### 2.1. Data: 'El discurso de Gastón'

La data a analizar en este trabajo es el llamado “discurso de Gastón”, el cual tuvo una duración de 25 minutos aproximadamente y un total 3875 palabras transcrito. Como se mencionó en la introducción este fue el discurso de orden en la ceremonia de apertura del año académico 2006-1 en la Universidad del Pacífico (UP). El texto en cuestión ha sido abordado desde distintas disciplinas: en algunos casos ha sido el objeto de estudio (Zuñiga 2007)<sup>4</sup>, en otros casos ha sido analizado o estudiado en función a su relevancia en procesos como la construcción de la Marca Perú (Palacios 2017)<sup>5</sup> o su importancia para el estudio y comprensión del llamado Boom Gastrómico (Matta 2019)<sup>6</sup> y García 2023). Además, este discurso tuvo un impacto más allá del ámbito simbólico, pues produjo cambios considerables: es considerado un texto fundacional no solo para la consolidación del proyecto de Marca Perú, sino para la fundación de la Sociedad Peruana de Gastronomía (APEGA) e inspiró a Fernando Villarán, ex ministro de Trabajo, a escribir el también celebrado artículo “Visión estratégica de la culinaria peruana” (2006). Teniendo lo anterior en cuenta, la relevancia de este discurso está fuera de duda y que, a pesar de los diversos estudios que lo han abordarlo, analizarlo desde los ECD permitirá señalar y analizar mediante qué recursos y estrategias una de las “más importantes declaraciones del nacionalismo peruano” (García 2023: 91) constituye y es el reflejo de concepciones hasta hoy en día compartidas bajo la idea de peruanidad.

## 3. Metodología

La metodología utilizada en la elaboración de este trabajo se puede visualizar en la figura 3. Como se puede observar, y ya se ha venido mencionando, se ha hecho uso de herramientas computacionales para la limpieza, pre-procesamiento, análisis exploratorio, visualización y organización de la data y los resultados. Se utilizó el lenguaje de programación Python y librerías para el manejo de texto, procesamiento del lenguaje natural y de visualización (histogramas y redes de palabras).

---

<sup>3</sup>véase Navarro & Tromben (2019)

<sup>4</sup>Zuñiga Lossio, M. (2007). Límites y posibilidades de lo multicultural e intercultural en el discurso del chef peruano Gastón Acurio. Construyendo Nuestra Interculturalidad, 4.

<sup>5</sup>Palacios, M. (2017). Marca Perú: Perú Nebraska, subjetividad neoliberal y nueva narrativa nacional. Tesis de maestría, Pontificia Universidad Católica del Perú.

<sup>6</sup>Matta, R. (2019). La revolución gastronómica peruana: una celebración de “lo nacional”. Entrevista a Mirko Lauer. *Anthropology of food*, (14).

Se utilizó el modelo de español 'es\_dep\_news\_trf' de spaCy, el cual prioriza exactitud a eficiencia. Esto último no fue de mucha relevancia para este trabajo, pues se trata de un solo documento, pero si será relevante si se trabaja con un corpus amplio.

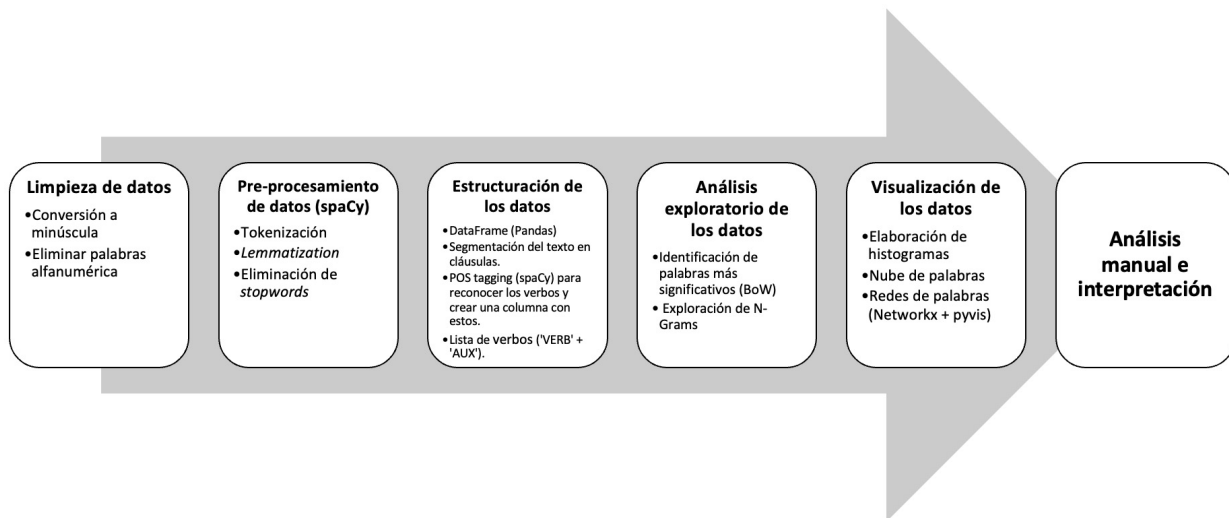


Figura 3: Elaboración propia

Además, la clasificación de los verbos de acuerdo a las tipos de procesos del sistema de transitividad, presentados secciones atrás, se realizó manualmente. Esto podría automatizarse, en un futuro trabajo, haciendo uso de técnicas de *machine learning*. Lo anterior también podría extender el aporte de este tipo de tecnologías ya no solo al análisis del discurso como texto –retomando la concepción tridimensional–, sino que también podría aportar más directamente en el análisis del discurso como práctica discursiva y social. Por ejemplo, automatizar la identificación de vínculos intertextuales no explícitos<sup>7</sup>. De todas formas, el factor humano es vital para llevar a cabo los principios teóricos y analíticos de los ECD. Todo lo anterior puede observarse en detalle en el Jupyter Notebook alojado en el repositorio de GitHub.

## 4. Análisis y discusión

En esta sección se presentarán los hallazgos producidos tras el análisis exploratorio de los datos presentado en la sección anterior. También se discutirán estos hallazgos en relación con los fundamentos teóricos e interpretativos de los ECD.

### 4.1. Palabras más frecuentes y significativas

En el histograma de la figura 4 se puede observar las 10 palabras (unigramas) más significativas y frecuentes del discurso. Estas fueron obtenidas mediante la función *get\_topgrams*, la cual extrae características por medio de *CountVectorizer*, de la librería *scikit-learn*, y permite construir un modelo *bag of words* o BoW. Lo anterior proporciona información sobre el tema del discurso.

<sup>7</sup>Similar al algoritmo que utiliza Turnitin para detectar plagios

Se puede observar que 'Peruano' es la palabra más utilizada, ya sea como gentilicio o categorización<sup>8</sup>. Además, la presencia del pronombre 'yo' es relevante, pues muestra no solo el lugar de enunciación del discurso (primera persona singular), sino que también hace referencia a la construcción de un 'nosotros' (primera persona plural). La construcción de la idea de nación se plantea desde la construcción de un nosotros 'peruano'.

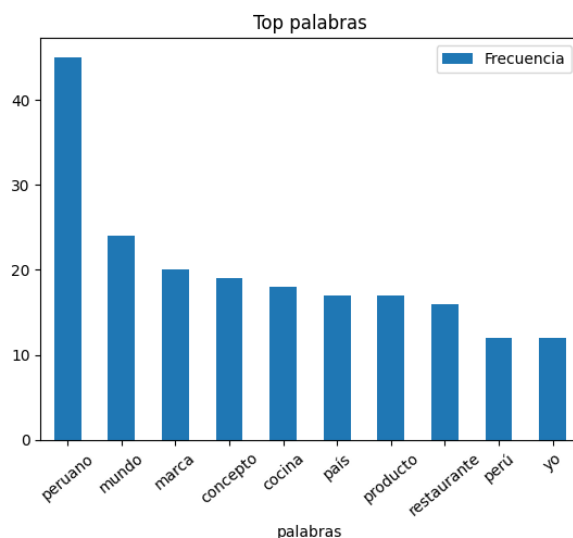


Figura 4: Elaboración propia

## 4.2. Exploración de n-gramas: Bi-gramas más frecuentes

Los n-gramas son secuencias continuas de palabras o *tokens* en un corpus. Se les llama bi-gramas cuando constan de dos palabras y tri-gramas cuando se trata de tres. Prestar atención a estas secuencias nos ayuda a tener un mejor contexto del uso de palabras en el corpus, pues las palabras no suelen usarse en aislamiento, sino que toman nuevos significados en su relación con otras. En la la figura 5 se observa una red de palabras que muestra los bi-gramas más recurrentes en 'el discurso de Gastón'.

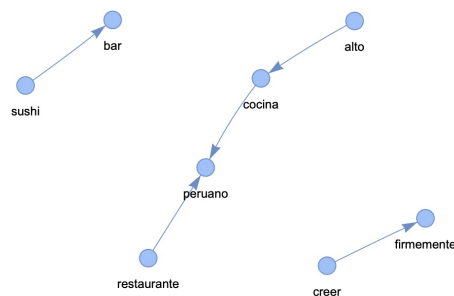


Figura 5: Elaborados con las librerías NetworkX y Pyvis

Podemos observar como la mayoría de bi-gramas surgen en combinación con palabras presen-

<sup>8</sup>según la clasificación Representación de Actores sociales de Van Leeuwen (1995)

tadas en el histograma de la figura 4, por ejemplo: 'cocina', 'peruano' y 'restaurante'. Llama la atención –también– la aparición del verbo 'creer' seguido, como indica la flecha, de 'firmemente' en cláusulas como 'creemos firmemente que todo tiene su momento', algo que no solo evidencia un acto de habla asertivo, sino que muestra el uso de un nosotros exclusivo, pues hace referencia no a los peruanos, sino a los chefs y cocineros peruanos. Finalmente, es importante mencionar que no se encontraron tri-gramas en la data.

### 4.3. Verbos y procesos del sistema de transitividad

En el discurso se identificaron un total de 193 verbos diferentes (501 en total), considerando también los verbos considerados auxiliares por el *part-of-speech tagger* (POS-tagger) de la librería spaCy. Los verbos que spaCy considero auxiliares fueron: 'ser', 'haber', 'poder', 'ir' y 'querer'. Lo anterior es problemático, pues dificultó la identificación de procesos relacionales, los cuales abundan en el discurso y cumplen un rol importante en términos de la construcción de una realidad discursiva particular. También se complicó la detención de modalizadores<sup>9</sup>, pues el verbo poder o haber suelen cumplir esta función, como se evidencia en la cláusula con la que inicia el discurso: 'Si bien podríamos pensar que [...]' o en '[...] habría unos 500 restaurantes mejicanos'.

En la figura 6 se muestra un histograma con la frecuencia de aparición de los verbos (incluido los auxiliares). En el el se muestra que los 4 verbos más frecuentes son 'ser', con más de 60 apariciones, 'haber', 'estar', 'poder'.

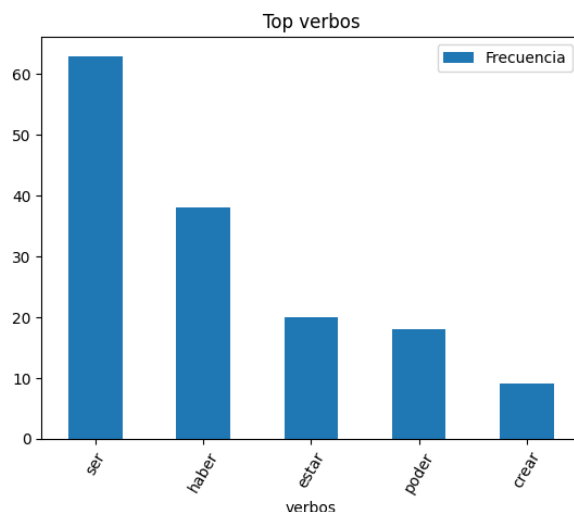


Figura 6: Elaboración propia

Lo anterior nos da pistas no solo de una primacía de procesos relacionales en el discurso, sino también la presencia de construcciones pasivas como 'La historia nos ha enseñado siempre lo contrario [...]' o 'Dentro del Perú hemos vivido una revolución editorial y educativa', las cuales representan a "nosotros", peruanos, como carentes de agencia y receptores de las acciones de otros.

<sup>9</sup>'La modalización representa el grado de probabilidad y validez que el escritor asigna a las propuestas' (Achugar, 1999)

Finalmente, se realizó una red de palabras que vincula las palabras más significativas, como se evidencia en la figura 4, y los verbos que los acompañan en las distintas cláusulas del discurso. Para la elaboración de esta red de palabras se hizo la función *noun.chunks* del *parser* (analizador sintáctico) de spaCy, la librería NetworkX para la elaboración del gráfico de redes y Pyvis. En la figura 4 se puede observar una versión estática de la red de palabras. La versión interactiva puede observarse en el repositorio.

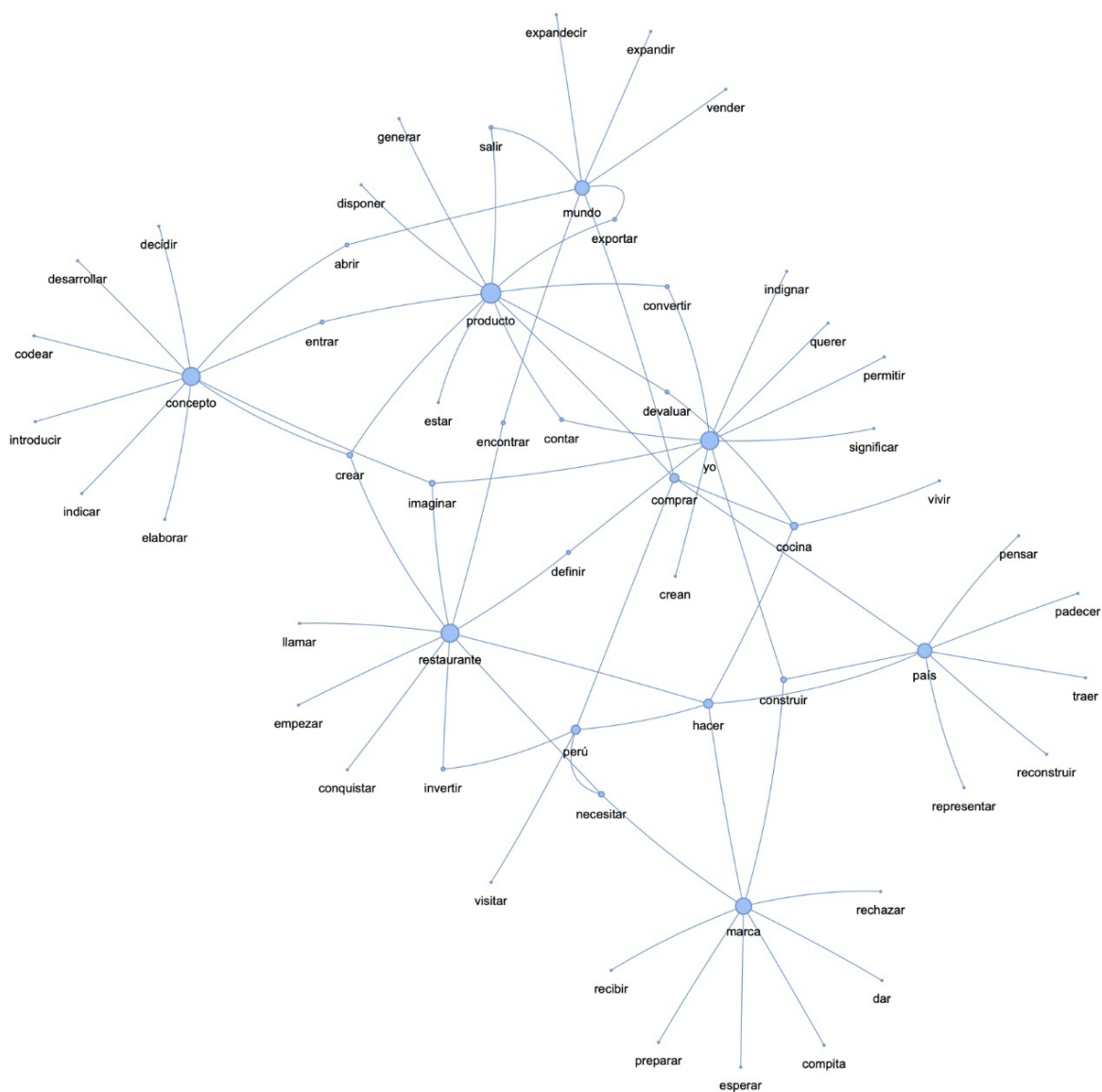


Figura 7: Elaboración propia

El gráfico anterior nos da mucha información relevante para analizar, pues se visualiza una gran cantidad de verbos que, aunque no son cuantitativamente relevante, son importantes para explicar cómo se está tratando y representando el tema principal del discurso. Por ejemplo, observar procesos materiales ('vender', 'invertir', 'exportar'). Esto último es interesante en términos de elección léxica, la vinculación semántica que estos verbos y nombres (por ej.

'producto', 'riqueza', 'marca') tienen con el campo de la economía y el mercado. Lo anterior se vincula con uno de los argumentos principales y la idea de fondo de plantear la idea de nación en paralelo y sintonía con la construcción de una marca-nación: La habilidad de mercantilizar y vender –conceptualizados– los recursos nacionales es la vía para un desarrollo sostenible.

## 5. Conclusión

En conclusión, en el presente trabajo ha realizado un análisis del discurso desde los presupuestos de los ECD mediado por herramientas y métodos computacionales que han facilitado la organización y pre-procesamiento de la data a analizar. Se ha logrado dar cuenta de los temas principales del discurso e identificar estrategias léxicas y discursivas utilizadas por el autor para construir una realidad discursiva caracterizada por la idea de nación, la peruanidad y la posibilidad de prosperidad a partir de la construcción de una marca nacional que se venda al mercado internacional y al mundo.

## 6. Referencias bibliográficas

Achugar, M. (1999). Construcción de la memoria: análisis de la confesión de un represor. *Revista iberoamericana de discurso y sociedad*, 1(4), 2.

Navarro, F., & Tromben, C. (2019). 'Estamos en guerra contra un enemigo poderoso, implacable': los discursos de Sebastián Piñera y la revuelta popular en Chile. *Literatura y lingüística*, (40), 295-324.

Fajardo Becerra, D. P. (2023). Definición de una metodología para análisis de discurso basado en lingüística computacional y técnicas de aprendizaje de máquina (Doctoral dissertation, Universidad EAFIT).

Loscalzo, J. R. (2019). Análisis de texto con n-grams: relaciones entre discursos presidenciales y cotizaciones de divisas. In *XXII Concurso de Trabajos Estudiantiles (EST)-JAIIO 48* (Salta, 2019).

Lozano, E. S., & Nakayama, A. (2023). Análisis semántico del discurso sobre el cambio climático en social media: El enfoque de la opinión pública acerca de Greta Thunberg, incendios en Australia y COP-25. *TECHNO REVIEW. International Technology, Science and Society Review/Revista Internacional de Tecnología, Ciencia y Sociedad*, 14(1), 1-20.

García, M. (2023). *Gastropolítica: Una mirada alternativa al auge de la cocina peruana*. Instituto de Estudios Peruanos.

Van Leeuwen, T. (1995). Representing social action. *Discourse & society*, 6(1), 81-106.

Vich, V. (2010). El discurso sobre la sierra del Perú: la fantasía del atraso. El discurso sobre la sierra del Perú: la fantasía del atraso, 253-265.