

## Análisis sociolingüístico del discurso académico argumentativo: un enfoque metodológico desde la lingüística de corpus

Sociolinguistic Analysis of Argumentative Academic Discourse: A Methodological Approach from Corpus Linguistics

Karime Vargas Cáceres<sup>1</sup>

**Copyright:** © 2023

Revista Internacional de Cooperación y Desarrollo. Esta revista proporciona acceso abierto a todos sus contenidos bajo los términos de la [licencia creative commons](#) Atribución–NoComercial–SinDerivar 4.0 Internacional (CC BY-NC-ND 4.0)

**Tipo de artículo:** Resultado de Investigación

**Recibido:** agosto de 2023

**Revisado:** septiembre de 2023

**Aceptado:** noviembre de 2023

### Autora

<sup>1</sup> Doctora en Lingüística (Universidad de Antioquia). Magíster en Semiótica (Universidad Industrial de Santander). Licenciada en educación básica con énfasis en Lengua Castellana (Universidad Industrial de Santander). Miembro del grupo de investigación Cultura y Narración en Colombia (CUYNACO) Profesora Escuela de Idiomas, Universidad Industrial de Santander  
Correo electrónico: [kvargasc@uis.edu.co](mailto:kvargasc@uis.edu.co)  
Orcid: <https://orcid.org/0000-0002-6719-292X>

### Cómo citar:

Vargas Cáceres, K. (2023). Análisis sociolingüístico del discurso académico argumentativo: un enfoque metodológico desde la lingüística de corpus. *Revista Internacional de Cooperación y Desarrollo*. 10(2), 12-24  
DOI: [10.21500/23825014.6670](https://doi.org/10.21500/23825014.6670)

 OPEN ACCESS



### Resumen

Este artículo muestra cómo se llevó a cabo el proceso de recolección y análisis sociolingüístico de un corpus de lengua escrita en un ambiente académico universitario. Se tienen en cuenta variables como el estrato, el género y la edad, el índice de nebulosidad, la densidad léxica y la fuerza argumentativa. Para la recopilación de información, se solicitó la redacción de un texto argumentativo a partir de una pregunta genérica alrededor de la situación de profesionalización docente; también, se implementó un cuestionario sociocultural. Los datos se analizaron bajo un enfoque mixto de investigación. La caracterización del corpus y análisis se hizo mediante los programas WordSmith Tools, TermoStat Web 3.0 y SPSS. Los resultados permitieron consolidar el corpus de análisis de un trabajo doctoral y explicar vínculos sociales y textuales en la producción de textos argumentativos. Aunque las variables sociales no influyeron significativamente en los resultados, el tiempo de escritura sí. Este estudio hace énfasis en la necesidad de tener en cuenta las variables sociales en la producción de textos en la universidad para generar alternativas que contribuyan a orientar estas prácticas escriturales.

**Palabras clave:** argumento; lingüística textual; sociolingüística; fuerza argumentativa; variables sociales.

### Abstract

This article shows how the process of collecting and sociolinguistic analysis of a corpus of written language was carried out in a university academic environment. Variables such as stratum, gender and age, nebulosity index, lexical density and argumentative strength are taken into account. To collect information, an argumentative text was requested to be written based on a generic question about the situation of teacher professionalization; Also, a sociocultural questionnaire was implemented. The data were analyzed under a mixed research approach. The char-

acterization of the corpus and analysis was done using the WordSmith Tools, TermoStat Web 3.0 and SPSS programs. The results allowed us to consolidate the corpus of analysis of a doctoral work and explain social and textual links in the production of argumentative texts. Although social variables did not significantly influence the results, writing time

did. This study emphasizes the need to take into account social variables in the production of texts at the university to generate alternatives that contribute to guiding these writing practices.

**Keywords:** argumentation; textual linguistics; sociolinguistics, argumentative force; social variables.

## 1. Introducción

La lingüística de corpus ofrece la posibilidad de trabajar con datos reales que permite hacer un acercamiento a la manera como se usa una lengua (Leech, 1997). Diversos campos de estudio han acogido la lingüística de corpus como principal herramienta metodológica tanto de recopilación como de análisis de datos lingüísticos (Tolchisky, 2014; Quiroz, 2003; Leech, 1997). En el campo de la educación, resulta útil para la enseñanza en primera y segunda lengua porque recopila expresiones auténticas y naturales y evita que la enseñanza de lengua se haga de manera aislada y fuera de contexto (Tolchisky, 2014; Quiroz, 2003). Estos ejemplos artificiales no reflejan la lengua vida de usuarios reales y situaciones reales de uso (Leech, 1997; Parodi, 2008).

El objetivo de este artículo es explicar cómo se recopilaron los datos, desde una lingüística de corpus, de un trabajo doctoral e indicar algunas correlaciones entre variables sociales y textuales en un caso específico de estudio dentro de una universidad pública de Bucaramanga. Se analizan variables como el estrato, el género, la edad, el índice de nebulosidad, la densidad léxica y la fuerza argumentativa.

Se considera que un estudio de tipo correlacional puede ser una prueba apropiada para relacionar variables sociales y textuales dado que las variables no son manipuladas. Por el contrario, permite observar y medir relaciones existentes en entornos naturales para ayudar a identificar patrones y tendencias en los datos recopilados. Es importante destacar que los estudios correlacionales no establecen causalidad. Solo muestran que hay una relación estadística entre las variables, pero no determinan la dirección de la relación ni la causa y efecto (Hernández et al., 2014).

Se parte de la idea según la cual las variables sociales condicionan las maneras como los sujetos hablan y escriben (López, 2004; Almeida, 1999). Desde los estudios sociolingüísticos, se ha indicado que las variables sociales muestran atributos de los grupos y permiten configurar identidad (Hernández y Almeida, 2005). Una de las investigaciones pioneras en el campo educativo que asoció las variables sociales con las lingüísticas fue desarrollada por Basil Bernstein (1975). Este sociólogo británico definió las características lingüísticas y discursivas de niños de clase obrera y niños de clase media. Los primeros evidenciaron el uso de un código restringido que se caracterizó por una escasa gama de recursos sintácticos: frases cortas, simples, sin concluir; uso simple y repetitivo de estructuras: conjunciones, órdenes, preguntas cortas, muletillas; y bajo nivel de simbolismo y de pensamiento lógico abstracto. En contraposición, los niños de clase media, se diferenciaron por el empleo de un código elaborado, pues contaron con mayores herramientas discursivas para comunicar sus ideas (Almeida, 1999; Caicedo, 1997).

Las desigualdades socioeconómicas en Latinoamérica y específicamente en Colombia se relacionan con el precio de las materias primas y los gobiernos que no han permitido una mirada más equitativa de las clases sociales, los ingresos tributarios y los gastos sociales con los demás países del mundo. Como consecuencia, los servicios básicos de bienestar como la salud, la educación o las pensiones no reciben los recursos necesarios para un adecuado funcionamiento (Riccardi, Agudelo & Bossio, 2022). Estas diferencias repercuten en los modos de hablar y escribir, tarea que explica muy bien la sociolingüística.

De acuerdo con lo anterior, partiendo de evidencia empírica, se recopila información textual para su posterior análisis de un grupo de estudiantes universitarios concreto. En el campo de la educación, la sociolingüística aporta nociones y referentes útiles para lograr el encuadre teórico y metodológico que permita el diseño de estrategias pedagógicas y didácticas acordes con las necesidades de la educación de manera equilibrada. La sociolingüística aplicada hace su labor señalando a las autoridades responsables lo que sucede y por qué sucede, y ayuda, según sus posibilidades, a solucionar problemas comunitarios (López, 2004). Este artículo pretende ser un aporte para el uso de lingüística de corpus y los estudios sociolingüísticos en entornos académicos. En un primer momento se hace alusión al proceso de recolección y en segunda medida se analizan correlaciones entre variables sociales y textuales.

## 2. Breve estado del arte y aproximación teórica

### 2.1 Una definición de corpus

Un corpus es una colección o conjunto de textos que está formado por al menos dos o más textos que comparten ciertos rasgos definitorios, limitado sólo por características inherentes a la naturaleza de los mismos; se constituye en una pieza comunicativa única que se define por su cierre semántico y su coherencia (Parodi, 2008). Esta colección de textos se considera representativa de una lengua o dialecto porque ocurre en ambientes naturales y se usa para identificar la variación (Quiroz, 2003). Algunas de las primeras consideraciones al construir un corpus tienen que ver con el tipo, el número, la selección de textos particulares, la selección de muestras dentro de los textos y la longitud de la muestra. Cada una de estas etapas implica una decisión de muestreo, consciente o no (Biber, 1999).

Siguiendo los criterios del Corpus Inglés Hablado y Escrito de Longman (LSWE), los criterios para compilar un corpus se sintetizan en preguntas como las siguientes: ¿cómo se seleccionan las muestras de texto?, ¿cuántas muestras de texto deben incluirse?, ¿cuánto deben durar las muestras de texto?, ¿qué tan grande debe ser todo el corpus?, ¿debe el corpus distinguir entre registros? Si es así,

¿qué registros debería incluir?, ¿debería el corpus incluir registros hablados y escritos?, ¿qué tiempo y recursos hay disponibles para la construcción del corpus?, ¿cuánta atención se debe prestar a la revisión y edición del corpus? (Biber, Johansson, Leech, Contad & Finegan, 2007).

Según Biber (1999) el tamaño de la muestra no es la consideración más importante al seleccionar una muestra representativa; más bien, lo es una completa definición de la población objetivo y las decisiones sobre el método de muestreo. La representatividad se refiere a la medida en que una muestra incluye la totalidad del rango de variabilidad en una población. En diseño de corpus, la variabilidad puede considerarse desde la situación y desde las perspectivas lingüísticas, y ambas son importantes para determinar la representatividad.

Aunque el trabajo con lingüística de corpus se ha extendido en los últimos años, su uso en educación ha sido menos difundido, de aquí que se encuentren escasos trabajos que describan los procesos de recolección y análisis de estos conjuntos de textos y su incidencia en las prácticas académicas. En la enseñanza de lenguas específicas es más complejo porque los materiales son pocos, cubren pocas áreas y no satisfacen las expectativas de los estudiantes y docentes, además, los textos son artificiales (Quiroz, 2003). El trabajo de Quiroz (2003) permite argumentar sobre la relevancia de los trabajos de corpus en el contexto colombiano para obtener una comprensión profunda de la lengua al precisar los patrones lingüísticos y comunicativos en diferentes contextos.

En el caso de los ambientes académicos, el análisis de corpus en relación con la enseñanza se puede implementar de diversas maneras. La lingüística de corpus permite acceder a piezas textuales auténticas ya sea para el trabajo de ejemplificación en el aula o para mostrar las especificidades en las prácticas de escritura académica, en este caso, universitaria (Quiroz, 2003; Tolchisky, 2014). De igual modo, es un instrumento útil para entender mejor cómo los estudiantes utilizan su lengua pues se centra en el estudio de producciones lingüísticas concretas en

situaciones comunicativas concretas (Quiroz, 2003; Tolchisky, 2014).

## 2.2 Una mirada sociolingüística de la argumentación en la universidad

La sociolingüística, como disciplina científica, establece la relación entre el lenguaje y la sociedad (Hernández y Almeida, 2005; Caicedo, 1997). Más que la descripción de una comunidad de habla, analiza los rasgos lingüísticos que distinguen unos estilos de otros dentro de cada sociolecto; estudia la motivación de tales distinciones e indica las variables sociales que impulsan el cambio de una variedad de lengua a otra (López, 2004; Almeida, 1999). Estas diferencias se encuentran condicionadas por variables de tipo social como el estrato, el género o la edad que influyen en la manera como se usa el lenguaje (Silva-Corvalán, 2001; Blas, 2005; Hernández y Almeida, 2005). Las investigaciones sociolingüísticas pretenden responder: ¿hablan diferente las personas que conforman los diversos grupos sociales? (Almeida, 1999). Los rasgos del lenguaje aportan información sobre atributos sociales, permiten identificar a los integrantes de los grupos y formar identidad social (Hernández y Almeida, 2005).

Por otra parte, en el ámbito universitario, la producción de textos demanda la adquisición y el dominio de unos conocimientos relacionados tanto con la producción como con la interpretación (Olave, Rojas y Cisneros, 2013; Navarro, 2018). La lectura, la escritura y la oralidad son los ejes fundamentales del trabajo intelectual, a través de esas habilidades los estudiantes se apropian de los conocimientos y los socializan a través de géneros discursivos específicos de cada comunidad sociodiscursiva (Henao y Castañeda, 2002; Navarro, 2018).

Los discursos académicos comparten rasgos prototípicos que permiten agruparlos bajo el término de macrogénero o colonia de géneros (Bhatia, 1993; Parodi, 2008). Así, un estudiante de licenciatura se enfrenta a una cantidad de géneros académicos propios de su especialidad: ponencias, artículos, informes, ensayos, entre otros (Perea, 2015; Galindo, 2012). Por eso, en la universidad es un imperativo el dominio de las características de los géneros discursivos que se producen en cada disciplina (Carlino,

2005; Olave, Rojas y Cisneros, 2013; Henao y Castañeda, 2002; Morales y Cassany, 2008; Carlino, 2013).

## 2.3 Variación léxica, índice de nebulosidad y fuerza argumentativa en el discurso académico argumentativo

Argumentar es una actividad de dar cuenta o razón de algo a alguien con el fin de lograr su comprensión y su asentimiento (Vega, 2013); es un fenómeno socioinstitucional que tiene lugar dentro de grupos sociales, en espacios públicos de discursos y bajo diversas modalidades (Vega, 2013). La argumentación cumple diversos propósitos, entre ellos: a) abogar, b) examinar, c) transformar d) enfrentar y e) afirmar. Solo en dos o tres de estos casos estaríamos propiamente defendiendo una tesis, aunque todos ellos involucran algún tipo de respaldo (Tindale y Barrientos, 2021). En el discurso cotidiano, es entendida como una práctica que integra diferentes campos que incluyen componentes cognitivos, lingüísticos, sociales y que permiten las interacciones cooperativas y polémicas (Plantin, 2005).

La argumentación es uno de los principales modos discursivos a partir de los cuales se difunde el conocimiento en la universidad (Carlino, 2003; Molina, 2017; Ruiz, Márquez, Badillo y Rodas, 2018; Carlino, 2013). Es un tipo de texto tanto oral como escrito con el que el estudiante se enfrenta o debería enfrentarse tanto en su actividad académica como profesional (Pardo y Baquero, 2001). Por eso, es necesario que se estudie la manera como se producen estas prácticas discursivas en entornos académicos universitarios, inicialmente, para describirlas desde un enfoque lingüístico y, posteriormente, para proponer alternativas de solución a los problemas identificados.

En cuanto al análisis de esas producciones discursivas, existen índices que brindan datos acerca de la calidad de los textos y las diferencias en la producción. Uno de ellos tiene que ver con la densidad léxica o grado de variación de palabras que un usuario emplea para comunicar una idea (Thomas, 2005; López-Pérez, 2008). La relación *type/token* (TTR) es un coeficiente que indica qué tan variado es el vocabulario empleado en un texto (Thomas, 2005; López-Pérez, 2008; Lancaster University, 2019).

Un *token* es cada una de las palabras individuales en un texto, sin importar las veces que ocurra. Los *types* son palabras únicas o diferentes (Lancaster University, 2019). El valor TTR (*Type-Token Ratio*) se obtiene dividiendo los *types* por sus *tokens*. Un TTR alto indica un alto grado de variación léxica o variedad de vocabulario (Thomas, 2005; López-Pérez, 2008). La relación *type/token* es una manera objetiva y simple de cuantificar la riqueza léxica (López-Pérez, 2008). Este índice da información acerca de la disponibilidad léxica, entendida como el caudal léxico utilizable en una situación comunicativa dada (López, 2004). Desde el punto de vista sociolingüístico, tener información sobre el léxico permite establecer estratificaciones de comunidades de habla. Sin embargo, esta idea puede resultar polémica porque no siempre el conocimiento y el dominio del vocabulario se asocia a estratos socioeconómicos altos. Es por esta razón que para analizar el estrato en el trabajo doctoral se propuso una posestratificación de acuerdo con datos obtenidos en el cuestionario. Aquí se incluyeron datos como la cantidad de libros que leían al mes, la frecuencia de lectura, las clases de libros de mayor preferencia, el grado de escolaridad de los padres y la cantidad de actividades culturales a las que asistían en los últimos seis meses. Con esta nueva información, se valoraron situaciones asociadas al acceso a recursos educativos, culturales y financieros de los participantes.

A su vez, la relación *type/token* presenta limitaciones que sesgan la observación. El resultado está influido por la longitud del texto (cuanto más largo sea un texto más probable es que no se manifiesten palabras diferentes) (Fundéu DLE, 2021). Cuando se escribe se usan unidades cruciales en la construcción de significado global de un texto (conjunciones, artículos, preposiciones). Cumplen una función más gramatical que semántica y se repiten significativamente, por lo tanto, disminuyen el valor de la variable *types* al calcular TTR. Además, otro de los inconvenientes es que se consideran como palabras distintas las formas de un mismo lema. Finalmente, TTR es solo un aspecto a tenerse en cuenta para determinar la riqueza léxica, por ejemplo, para este último criterio se pueden incluir el grado de dificultad tanto del vocabulario como de las estructuras

sintácticas utilizadas (Fundéu DLE, 2021). En este estudio se tuvo en cuenta la longitud de los textos (400 palabras o *tokens*) y la complejidad se determinó gracias al tipo de estructura argumentativa y al índice de nebulosidad de cada muestra del corpus.

En los análisis textuales, también se puede determinar el nivel de lecturabilidad o nebulosidad de un texto, es decir, la falta de claridad en la expresión de las ideas (Barrio-Cantalejo et al., 2008). Casi todas las lenguas han desarrollado procedimientos de evaluación de la legibilidad o lecturabilidad de un texto (Barrio-Cantalejo). Estos índices señalan cuantitativamente qué tan compleja es la lectura y comprensión de un texto escrito de acuerdo con diversos parámetros como la cantidad de palabras y oraciones. Un texto es más fácil de leer si está integrado por oraciones y palabras cortas (Ávila, 2003; Barrio-Cantalejo et al., 2008).

Se destacan en lengua inglesa tres escalas: Szigriszt, RES de Flesch y INFLESZ, con valores que van desde 0 hasta 100; las puntuaciones inferiores indican grados de dificultad mayor. Los valores de normalidad se encuentran entre los 55 y los 65 puntos (Barrio-Cantalejo et al., 2008). En España, Fernández-Huerta adaptó la Fórmula RES de Flesch en 1959 y la llamó Fórmula de Lecturabilidad, con una escala de interpretación de 7 tramos (Barrio-Cantalejo et al., 2008). En Colombia, Fernando Ávila propone una serie de parámetros para determinar el índice de nebulosidad o medir la dificultad de lectura de un texto (Ávila, 2003). Se tienen en cuenta el total de palabras del texto, el número de oraciones separadas por puntos y el número de palabras largas (aquellas que superan los tres grupos silábicos). Con base en esta última fórmula, se calculó el índice de nebulosidad en las muestras del corpus que se analiza. Tanto la relación TTR como el índice de nebulosidad se obtienen a partir de cálculos numéricos y están condicionados por la disponibilidad de recursos léxicos que se construyen debido a la exposición de factores externos como la cultura y las experiencias acumuladas.

En cuanto a la fuerza argumentativa, no existe un procedimiento matemático para determinar qué

tan fuerte es un texto en el plano argumentativo. La fuerza argumentativa es un concepto que se determina por el ajuste de un texto a una situación de comunicación específica. Tiene que ver con la aceptabilidad del texto argumentativo a un contexto particular (Santibáñez-Yáñez, 2015; Noemi & Rossel, 2017). Esto quiere decir que la fuerza argumentativa es contextual, un texto puede tener diferente fuerza dependiendo de los escenarios y los públicos a los que se presente. En la universidad, quien determina la aceptabilidad, la suficiencia y la pertinencia de un producto escrito es el profesor. La evaluación se hace basada en la experiencia y dependiendo del criterio del evaluador, cuando la evaluación cuenta con mayor objetividad se recurre a rúbricas o rejillas.

Perelman y Olbrechts-Tyteca (1989) no hablan directamente del concepto de fuerza argumentativa, pero enumeran algunos aspectos de los que dependen la fuerza de un argumento. Un argumento A1 es más fuerte que un argumento A2 si la adhesión o grado de aceptación de las premisas es mayor, si es considerado más pertinente o relevante, si tiene menos contraargumentos y si el auditorio lo considera más valioso (Apostel, 2007; Perelman y Olbrechts-Tyteca 1989). El auditorio determina la validez, la pertinencia, la relevancia y la aceptación del planteamiento.

Desde el enfoque pragmatialéctico, un discurso argumentativo puede ser defectuoso porque presenta contradicciones como un todo o sus argumentos inaceptables o erróneos. Para ver si la argumentación es sólida, se deben evaluar esas debilidades. Para evaluar la solidez en este enfoque, toda la argumentación debe descomponerse en argumentos individuales, pero antes se debe evaluar si la argumentación es un todo consistente (Van Eemeren et al, 2006).

Para que un argumento único sea considerado como consistente o fuerte debe comprender tres requerimientos: i) cada enunciado que conforma el argumento debe ser aceptable; ii) el razonamiento subyacente al argumento debe ser válido y iii) el esquema argumentativo debe ser apropiado y empleado correctamente (Van Eemeren et al., 2006).

Estos tres criterios coinciden con la propuesta de Johnson y Blair (2006) quienes indican que para que un argumento sea sólido o fuerte debe cumplir con los criterios de relevancia, aceptabilidad y suficiencia. El argumento es irrelevante si la ley de paso no se aplica al dato o cuando el respaldo no es conocido por el público. Cuando los datos suministrados no cumplen con las exigencias en el campo específico de cada argumentación, se dice que ese argumento es inaceptable porque no está basado en soportes confiables.

Una de las preguntas cruciales es si se puede cuantificar la fuerza argumentativa. Para Marraud (2022) la cuestión fundamental en cualquier teoría de los argumentos que incorpore el concepto de fuerza es si lo trata como un concepto cualitativo, comparativo o métrico. Según el autor, entre las partes de un argumento figuran todas las consideraciones relevantes para la evaluación de sus propiedades lógicas y las premisas explícitas pueden ser insuficientes (Marraud, 2022). Por su parte, Apostel (2007) indica que a) el sistema de pesos y medidas se puede hallar, b) que de hecho es aplicado a muchos agentes en muchas circunstancias, pero que c) el sistema de pesos y medidas es variable y se pueden encontrar leyes que describan su variación.

Para Tindale y Barrientos (2021) la adhesión, la pertinencia, los vínculos de proximidad, la objetividad y la resistencia a la refutación, son características definitorias de los argumentos fuertes. Estas características las establecen basados en el trabajo de Perelman y Olbrechts-Tyteca (1989). El criterio de adhesión se refiere a la relación entre el auditorio y las premisas; en la pertinencia o relevancia, el argumento es proyectado hacia el contexto; el vínculo de proximidad, se da por la relación de las premisas con la tesis; en la objetividad, los contraargumentos u objeciones son determinados por el contexto del auditorio; en último lugar, resistir a la refutación es lo ideal para evitar afectar el grado de fuerza del argumento.

En este trabajo, la fuerza argumentativa se determinó cuantitativamente a partir de una alternativa matemática que se propuso como trabajo doctoral

para calcular este criterio en un texto académico producido en un ámbito universitario. Se propone calcular la fuerza argumentativa teniendo en cuenta la sumatoria de factores que permiten asegurar la solidez de un texto a partir de los siguientes criterios: a) el tipo de estructura argumentativa, b) la clase de argumentos, c) el uso de menos falacias argumentativas y c) el empleo de cualificadores modales.

### 3. Metodología

Bajo un enfoque de investigación mixto, se trabajó con estudiantes universitarios de quinto semestre de un programa de licenciatura de una universidad pública de Santander. 27 mujeres y 7 hombres con edades comprendidas entre los 18 y los 27 años y ubicados en estratos socioeconómicos de 1 a 4. Para el ejercicio de escritura se solicitó la redacción de un texto argumentativo en un espacio institucional que permitía acceder a las bases de datos o a diversos recursos electrónicos para seleccionar las fuentes pertinentes para apoyar los argumentos, si así lo deseaban. La redacción se efectuó en presencia del investigador, se aclaró la finalidad del proyecto y se otorgó un incentivo académico por la participación.

Por tratarse de ambientes académicos, los grupos ya se encontraban definidos. La selección se realizó mediante muestreo no probabilístico (muestreo incidental no aleatorio). Dentro de este tipo se opta por el muestreo por conveniencia (Niño 2011; Hernández et al., 2014), debido a la facilidad de acceso a los participantes, dado que la docente investigadora labora en esta institución educativa y ha orientado esta asignatura durante seis semestres consecutivos. Según Ñaupás et al. (2018) estos métodos son altamente utilizados por parte de profesores de universidad. Este grupo constituye una comunidad de práctica, pues se agrupan por una actividad social o circunstancia en común (Serrano, 2011). Conforman un grupo relativamente homogéneo respecto del tipo de educación, el nivel de escolaridad y la lengua dominante.

El desencadenante escritural, que permitió la redacción de los textos, se organizó en un documento en Microsoft *Power Point* y estuvo integrado por un conjunto de cinco *memes* publicados en una red

social acerca de la labor docente y un conjunto de siete comentarios de algunos lectores de un diario regional con ocasión de un paro nacional programado por las centrales obreras colombianas. Si bien la noticia no fue relevante en esta ocasión, sí lo fueron los comentarios. El documento enviado a los correos y la instrucción para la redacción rezó del siguiente modo: “A partir de las situaciones presentadas anteriormente, redacte un texto en el que responda las siguientes preguntas: ¿Crees que lo que dicen los memes y los comentarios en Facebook tienen algo de verdad? ¿Cuál es tu opinión al respecto?”. Para la redacción se solicitó usar el procesador de texto de *Microsoft Word*, nombrar el archivo de la forma Texto + apellidos, y cuidar que la extensión no superara las 400 palabras.

Las variables controladas durante la fase de escritura fueron, por un lado, el tema de escritura, pues no se trató de un tema libre y espontáneo porque la situación construida tuvo que ver con la percepción docente que circula en la sociedad bajo la forma de memes y comentarios virtuales en redes sociales; por otro lado, se controló la extensión de los textos pues todos tuvieron una extensión de 400 palabras.

El tiempo de escritura estuvo entre los 37 minutos hasta dos horas con 34 minutos. El ejercicio se había previsto para dos horas de trabajo en el aula. A partir de la hora de envío del texto al correo del investigador, se codificaron las muestras del corpus. Además, el tiempo es una variable independiente relevante no solo para el trabajo de codificación sino porque ofrece información sobre el desempeño argumentativo y la calidad de los textos.

Los datos también fueron recopilados a través de un cuestionario con el que se recabaron los datos socioculturales para hacer las relaciones de las variables.

### 4. Resultados

Se recopilaron 33 textos escritos en lengua materna, específicamente texto de no ficción y argumentales-posiciones (Calsamiglia & Tusón, 2001). Para el análisis, los textos fueron convertidos en formato txt para ser procesados por los programas *WordSmith*

*Tools y TermoStat Web 3.0.* El uso de estos dos programas es complementario y permite hacer un trabajo más eficiente en la descripción y caracterización del corpus. A continuación, se enuncian los principales resultados después de correlacionar las variables sociales y textuales. Estas correlaciones se realizaron con SPSS.

#### 4.1 Análisis descriptivo de las variables

Se calculó el índice de nebulosidad de cada muestra del corpus basados en los trabajos de [Ávila \(2003\)](#) quien indica que el valor ideal deberá estar alrededor de 15. El primer dato relevante en el corpus analizado fue el número de oraciones: máximo 22 y mínimo 6. Este dato determina el número de palabras de cada proposición, el número de palabras largas y el total de palabras de todo el texto. Casi el 40% de los textos presentaron índices de nebulosidad por debajo de 18 puntos. Dichos valores son relativamente bajos para la muestra estudiada. Sin embargo, también se observa un porcentaje alto de textos con índices superiores a 21 puntos (45%). Se revisaron los casos de los estudiantes con mayor y menor índice de nebulosidad. Para ejemplificar las diferencias en redacción entre estos dos casos, se muestran dos fragmentos discursivos:

No debíamos leer las guías ni hacer síntesis de ellas para evidenciar que habíamos entendido las ideas planteadas en el texto, no se ejercía ningún tipo de pensamiento crítico, era algo así como una taller de caligrafía (sic), ahora que lo pienso, sin embargo y afortunadamente, mi profesora de sociales fue aquella fuente de la cual abastecimos nuestros deseos de forjar una posición crítica frente a cualquier ámbito del que pudiéramos hablar, ya que continuamente se abrían mesas de debate puesto que, siempre debíamos llevar un bagaje de temas o noticias de actualidad para reflexionar sobre ellos, de esta forma no solo asumíamos una postura también nos hacía cada vez más conscientes y receptivos de aquellos aspectos que no marchaban bien (T16).

El párrafo anterior presenta una sola idea en el texto. La cantidad de palabras por cláusula influyen en los altos valores en índices de nebulosidad. En el segundo ejemplo se evidencia

una mejor organización de las ideas alrededor de tres oraciones.

Igualmente, la docencia es una carrera que nunca finaliza. Esta labor exige al profesional renovar sus conocimientos, iniciar nuevas especializaciones y educarse durante toda la vida. Lo anterior debido a que, para enseñar, primero es necesario aprender (T6).

En la relación *type/token*, un valor alto indica alto grado de variación de vocabulario, se encontraron cifras entre 45,65 (T24) y 57,25 (T23). Aunque estos valores no están muy lejanos uno del otro, las muestras se clasificaron en cuatro rangos. El mayor grado de variación léxica que puede tener un texto es 1 (de acuerdo con los datos de WordSmith, es 100). Al dividir el número de palabras diferentes sobre la cantidad total se va a cercar más a la unidad.

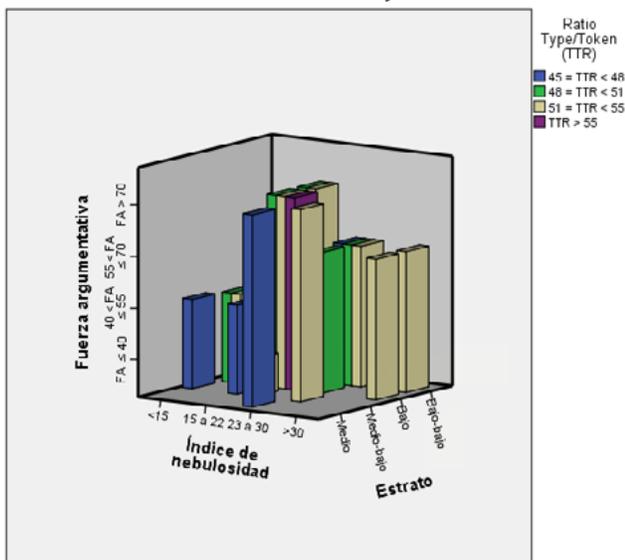
En cuanto a la fuerza argumentativa, este valor se calculó a través del algoritmo matemático propuesto, aunque la fórmula no se presenta en este artículo, básicamente este dato se calculó a través de la sumatoria de tres fuerzas independientes: el tipo de estructura argumentativa, la clase y el número de argumentos y el uso de cualificadores modales. Según la fórmula propuesta, se calculó la fuerza argumentativa del corpus. Se hallaron índices inferiores a 30 puntos y algunos superiores a los 70 (sobre 100, máximo valor). El mayor porcentaje de textos se encuentra en rangos de fuerzas argumentativas entre los 55 y los 70 puntos (39,4%). El 21,2% se hallan en un rango superior a los 70 puntos, un valor que se considera alto. Se analizaron de modo individual los textos con mayores índices en la fuerza argumentativa y se encontró un mayor número de argumentos basados en datos y hechos, menos falacias y estructuras argumentativas subordinadas y evidenciales. Estos criterios son los que más aportan valor para el cálculo de la fuerza argumentativa teniendo en cuenta la fórmula propuesta.

#### 4.2 Fuerza argumentativa, índice de variación léxica, índice de nebulosidad y estrato

Los análisis realizados con la población descrita evidenciaron que el estrato no determina mayores diferencias con respecto a la FA. Aunque las

relaciones *type/token*, que indican el grado de variación léxica, y el índice de nebulosidad fueron tenidas en cuenta en esta investigación como variables independientes, se quiso saber cuál era la relación entre la FA, la cantidad de palabras diferentes y el grado de nebulosidad que se evidenciaron en las producciones discursivas. Esta información se detalla en la Figura 1.

**Figura 1.** Fuerza argumentativa, índice de variación léxica, índice de nebulosidad y estrato



Fuente: elaboración propia a partir de los datos obtenidos

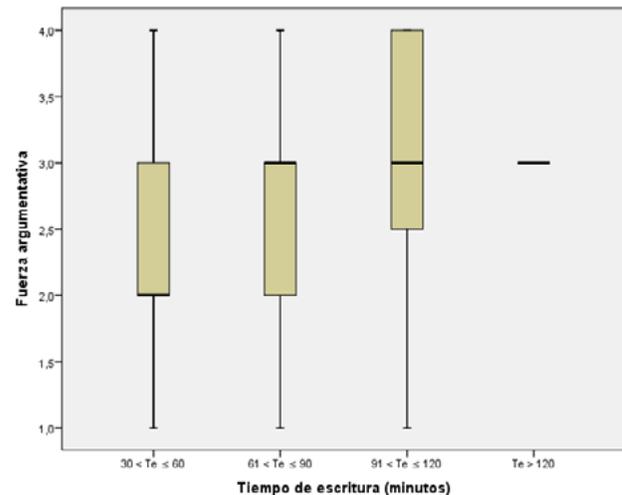
La gráfica anterior muestra que la fuerza argumentativa podría ser independiente del grado de dificultad en la redacción y la cantidad de palabras variadas que se empleen para construir los argumentos. Esto se puede afirmar en razón a que cada color de barra indica diferentes grados de variación léxica, pero la mayoría se presentan a la misma altura con respecto a la fuerza argumentativa. Este dato corrobora el hecho de que la fuerza argumentativa depende más de factores intrínsecos a la argumentación como la clase de argumento, la estructura argumentativa y los cualificadores modales en lugar de la cantidad de palabras diferentes. Es decir, se podría encontrar un único argumento, pero expresado de diversas formas en el plano superficial.

### 4.3 Fuerza argumentativa y tiempo de escritura

También se quería saber si la fuerza argumentativa era directamente proporcional al tiempo de es-

critura que utilizaban los estudiantes, es decir, que si a mayor tiempo, mejores serían los puntajes en esta variable. En la Figura 2 se presentan estas relaciones.

**Figura 2.** Cruce de variables fuerza argumentativa y tiempo de escritura



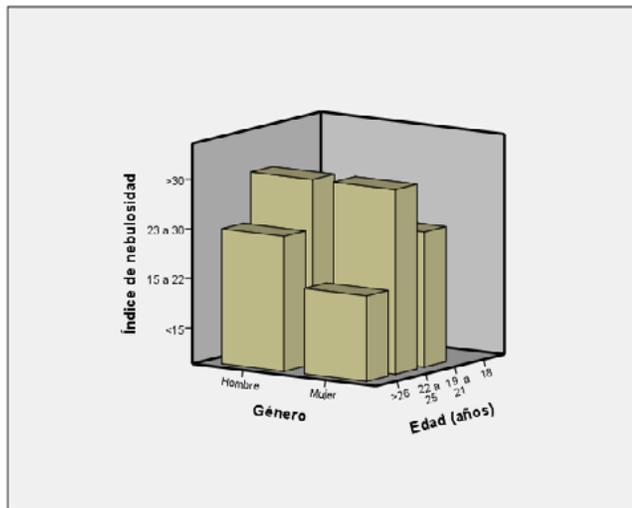
Fuente: elaboración propia a partir de los datos obtenido

Homogéneamente, se encontraron textos con fuerzas argumentativas similares en dos rangos de tiempo, entre 30 y 90 minutos, pero también se evidencia un aumento en la fuerza con respecto al rango de más de 90 minutos. Dado que la población son estudiantes universitarios, se muestra que tienen una fuerza argumentativa mínima de 2 (entre 40 y 55 puntos) siendo consecuente con su conjunto de experiencias sobre la argumentación. La variable procedimental tiempo de escritura sí marca una diferencia con respecto al criterio en discusión. Posiblemente, tendría que ver con la elaboración de esquemas previos para planificar el trabajo escritural o con el hecho de buscar los argumentos de autoridad y los datos en las fuentes disponibles que tuvieron para la redacción.

### 4.4 Edad y el género con respecto al índice de nebulosidad

En los datos, las distancias no son siempre tan claras. Se debe recordar que cuanto menor es el índice de nebulosidad más fácil de comprender es el texto construido. En la Figura 3 se indican las relaciones entre las variables género y edad con respecto al índice de nebulosidad.

**Figura 3.** Cruce de variables género y edad con respecto al índice de nebulosidad



Fuente: elaboración propia a partir de los datos obtenidos

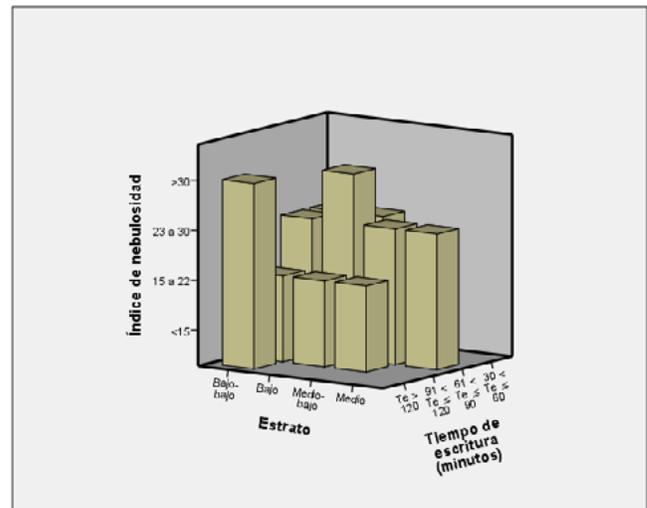
Se observa que en el rango de 22 a 25 años los puntajes en la variable índice de nebulosidad son homogéneos en hombres y mujeres. Sin embargo, el índice de nebulosidad disminuye en mujeres y hombres con mayor edad dentro del grupo, pero la disminución es más marcada en las mujeres. Este hecho podría indicar que a mayores edades las mujeres redactan más claramente que los hombres. La sociolingüística ha explicado estas diferencias entre hombres y mujeres y entre diferentes grupos etarios, y aunque esta muestra es reducida, también indica evidencia frente a ese comportamiento (Silva-Corvalán, 2001; Blas, 2005; Hernández y Almeida, 2005).

#### 4.5 Estrato y tiempo de escritura con respecto al índice de nebulosidad

En esta correlación se tiene en cuenta si el índice de nebulosidad se ve afectado por el estrato del estudiante y por el tiempo que tarda en la escritura. Las relaciones entre el estrato y el tiempo de escritura con respecto al índice de nebulosidad se muestran en la Figura 4.

De acuerdo con el analizador empleado en este trabajo, basado en Ávila (2003), los índices cercanos a 15 son más accesibles en su redacción, valores más altos indican mayor dificultad de lectura en razón a la extensión de las cláusulas (debido a

**Figura 4.** Cruce de variables estrato y tiempo de escritura con respecto al índice de nebulosidad



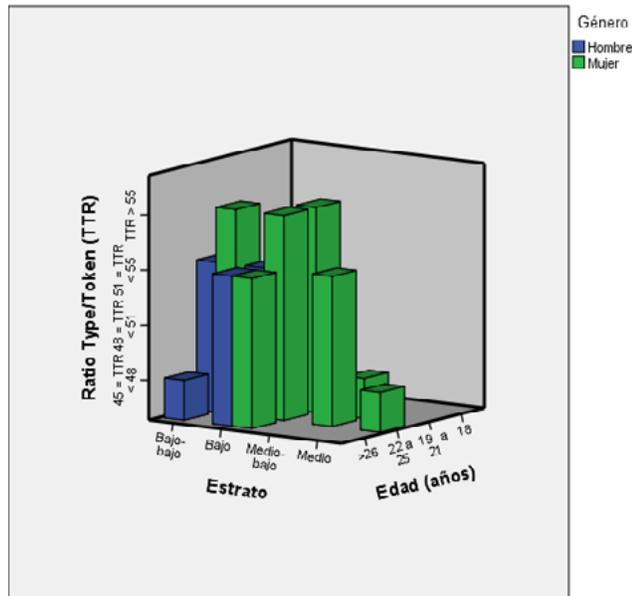
Fuente: elaboración propia a partir de los datos obtenidos

la subordinación) y el número de palabras largas. Entonces, el índice de nebulosidad es un indicador de la dificultad de lectura que puede tener un texto. Los datos indican cierta homogeneidad en el tiempo de escritura entre los 60 y 90 minutos entre los estratos. Sin embargo, entre los 90 y 120 minutos estos índices también disminuyen. Nuevamente, es el tiempo el que determina mejores desempeños en la construcción de los textos, quizá relacionados en esta ocasión con la revisión del texto. Dos datos llaman la atención, índices de nebulosidad superiores a 30 en los dos estratos más bajos, pero con tiempos de escritura muy diferentes: entre 60 y 90 minutos y más de 120 minutos.

#### 4.6 Estrato, edad y género con respecto al TTR

Al contrario del índice la nebulosidad en el que menores valores indican mejor redacción, el grado de variación léxica es directamente proporcional: cuanto mayor sea este valor más palabras diferentes contiene el texto. En contrapartida, un valor bajo indicará un número alto de repeticiones, aspecto que se podría traducir en que el texto es menos rico o variado desde el punto de vista del vocabulario. En este criterio es importante considerar que las comparaciones se deben dar entre textos con el mismo tamaño. Los cruces entre estas variables se detallan en la Figura 5.

**Figura 5.** Cruce de variables estrato, edad y género con respecto a TTR



Fuente: elaboración propia a partir de los datos obtenidos

Aunque los datos no se pueden generalizar en este caso de estudio por la distribución heterogénea de la muestra, las mujeres manifiestan mayor variación léxica que los hombres en los estratos bajos. También se ven bajos grados de variación en hombres y mujeres principalmente en estratos superiores.

## 5. Conclusiones y recomendaciones

Se analizó la relación entre las variables sociales y variables textuales en la producción de textos argumentativos escritos. Se empleó un cuestionario sociocultural y una prueba de escritura en el aula que tuvo como desencadenante las diversas opiniones disfóricas que se construyen en la sociedad alrededor de la profesión docente y que circulan de modo libre en internet a través de la forma de memes. Con la información del cuestionario sociocultural se evidenció que los participantes pertenecían en su mayoría a estratos socioeconómicos bajos.

Con los estudios bivariados no se halló evidencia que apoye la hipótesis de que las variables sociales como la edad, el género y el estrato sociocultural influyan en la redacción de textos argumentativos con respecto a la fuerza argumentativa, el tipo de disputa, la relación *type/token* y el índice de nebulosidad.

Las correlaciones en este contexto específico y concreto, con este diseño y con estos participantes en este momento determinado, no reflejaron vínculos entre las variables sociales y textuales. Para mostrar de modo más eficaz la influencia de una variable sobre otra, se necesita una muestra más amplia y comparar un mayor número de sujetos que cumplan las mismas condiciones.

En ese sentido, surgen hipótesis emergentes derivadas de los datos. No son tanto las variables sociales las que influyen en la redacción de los textos, sino factores relacionados con la lectura de textos: la frecuencia, la clase de textos, la cantidad de documentos leídos. Para llegar a esta conclusión, se compararon los resultados de los participantes que obtuvieron los puntajes más altos y más bajos en los textos y se contrastaron con los datos recabados en el cuestionario.

También se encontró que la variable que mayormente presentaba incidencia fue la variable procedimental tiempo de escritura. Se evidenció que los participantes que tardaron más tiempo (entre 60 y 90 minutos) produjeron textos con mayores fuerzas argumentativas.

Sin importar cualquiera de las perspectivas que se adopte, la argumentación permite el desarrollo del pensamiento y el replanteamiento de los conocimientos. La argumentación en la universidad es un modo discursivo con el cual se construye, codifica y transmite el conocimiento de acuerdo con una serie de convenciones cuyo dominio es indispensable para la formación y actuación profesional. Por tanto, el diagnóstico como la atención de la escritura académica universitaria son temas de una urgencia inaplazable.

A partir de los datos empíricos recopilados y analizados en esta investigación, se contribuye al conocimiento de las competencias argumentativas de un grupo de estudiantes. Los datos permiten indicar la necesidad de orientar y construir material didáctico en las aulas universitarias para construir producciones discursivas idóneas en el plano argumentativo.

## 6. Referencias

- Almeida, M. (1999). *Sociolingüística*. Universidad de la Laguna.
- Apostel, L. (2007). ¿Cuál es la fuerza de un argumento?: Algunos problemas y sugerencias. *Praxis Filosófica*, (25), 129-138. [http://www.scielo.org.co/scielo.php?script=sci\\_arttext&pid=So120](http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=So120)
- Ávila, F. (2003). *Cómo se escribe*. Norma.
- Barrio-Cantalejo, I. M., Simón-Lorda, P., Melguizo, M., Escalona, I., Marijuán, M.I., y Hernando, P. (2008). Validación de la Escala INFLEZ para evaluar la legibilidad de los textos dirigidos a pacientes. *Anales del Sistema Sanitario de Navarra*, 31 (2), 135-152. [http://scielo.isciii.es/scielo.php?script=sci\\_arttext&pid=S1137](http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1137)
- Bernstein, B. (1975). *Lengua y clase social*. Ediciones Minuit.
- Bhatia, V. (1993). *Analysing genre: Language use in professional settings*. Longman.
- Biber, D; Johansson, S; Leech, G; Contad, S., & Finegan, E. (2007). *Logman Grammar of Spoken and Written English* (Sixth impr). England: Pearson Education Limited.
- Biber, D. (1999). Representativeness in corpus design. *Literary and Linguistic Computing*, 8(4), 243-257. <https://doi.org/10.1093/lc/8.4.243>
- Blas, J. L. (2005). *Sociolingüística del español: desarrollos y perspectivas en el estudio de la lengua española en contexto social*. Cátedra Lingüística.
- Caicedo, M. (1997). *Introducción a la sociolingüística*. Universidad del Valle.
- Calsamiglia, H., & Tusón, A. (2001). *Las cosas del decir. Manual de análisis del discurso*. Barcelona: Ariel.
- Carlino, P. (2003). Leer textos científicos y académicos en la educación superior: Obstáculos y bienvenidas a una cultura nueva. *Uni-Pluri/Versidad*, 3 (2) 17-23. <https://aprendeonline.udea.edu.co/revistas/index.php/unip/article/viewFile/12289/>
- Carlino, P. (2005). *Escribir, leer y aprender en la universidad. Una introducción a la alfabetización académica*. Fondo de Cultura Económica.
- Carlino, P. (2013). Alfabetización académica diez años después. *Revista Mexicana de Investigación Educativa* (RMIE), 18 (57), 355-381. <http://www.re-dalyc.org/pdf/140/14025774003.pdf>
- Fundación del Español Urgente (Fundéu DLE). (2021). *Consideraciones teóricas en torno a la riqueza lingüística*. <https://www.fundeu.es/consideraciones-teoricas/>
- Galindo, A. (2012). Producción argumentativa escrita en lengua materna de estudiantes en formación universitaria bilingüe y tradicional en la Universidad del Quindío, Colombia. *Forma y Función*. 25 (2), 115-137. <https://revistas.unal.edu.co/index.php/formayfuncion/article/view/39835>
- Henao, J. I., y Castañeda, L. S. (2002). *Ciencia y pedagogía: El papel del lenguaje en la apropiación del conocimiento*. Arfo.
- Hernández, J. M., y Almeida, M. (2005). *Metodología de la investigación sociolingüística*. Comares.
- Hernández, R; Fernández, C., y Baptista, M. P. (2014). *Metodología de la Investigación*. (6a ed.). Mc Graw-Hill
- Johnson, R., y Blair, J. A. (2006). *Logical Self-Defense*. International Debate Education Association
- Lancaster University. (2019). *Uso de WordSmith para la investigación basada en Corpus*. [https://www.lancaster.ac.uk/fss/courses/ling/corpus/blue/wordsmith\\_top.htm](https://www.lancaster.ac.uk/fss/courses/ling/corpus/blue/wordsmith_top.htm)
- Leech, G. (1997): Teaching and language corpora: a convergence, in WICHMANN, A., et al. (Eds.): *Teaching and language corpora*. (pp. 1-23). Longman.
- López, H. (2004). *Sociolingüística*. (3 ed.). Gredos.
- López-Pérez, M. V. (2008). Estudio sobre la lengua de instrucción (LI): un índice de frecuencias léxicas lematizado a partir de un manual escolar. *Resla*, 21, 221-229. <https://dialnet.unirioja.es/servlet/articulo?codigo=2926037>
- Marraud, H. (2022). La fuerza lógica de los argumentos a la luz del extraño caso de los comedores de ajo crudo: Un comentario a “Argumentos fuertes” de Ch. Tindale y R.M. Barrientos (RIA 22, 2021). *Revista Iberoamericana De Argumentación*, (24), 72-84. <https://doi.org/10.15366/ria2022.24.005>
- Molina, M. E. (2017). Escritura académica, argumentación y prácticas de enseñanza en el primer año universitario. *Enunciación*, 22(2), 138-153. DOI: <http://doi.org/10.14483/22486798.11929>

- Morales, O. A., y Cassany, D. (2008). Leer y escribir en la universidad: Hacia la lectura y la escritura crítica de géneros científicos. *Memorialia*, (5) 69-82. <http://www.saber.ula.ve/handle/123456789/16457>
- Navarro, F. (2018). Didáctica basada en géneros discursivos para la lectura, la escritura y la oralidad académicas. Navarro, F., y Aparicio, G. (comp). *Manual de lectura, escritura y oralidad académicas para ingresantes a la universidad*. (1ª ed). Universidad Nacional de Quilmes.
- Niño, V. M. (2011). *Metodología de la investigación. Diseño y ejecución*. (1ª ed.). Ediciones de la U.
- Noemi, C., & Rossel, S (2017). Competencia argumentativa psicosocial: esquemas, estructura y tipos de argumentos en estudiantes universitarios chilenos. *Lenguaje*, 45(1), 11-33. <http://www.scielo.org.co/pdf/leng/v45n1/0120-3479-leng-45-01-00011.pdf>
- Ñaupas, H., Valdivia, M. R., Palacios, J. J., y Romero, H. E. (2018). *Metodología de la investigación cuantitativa-cualitativa y redacción de la tesis*. Quinta edición. Ediciones de la U.
- Olave, G., Rojas, I., y Cisneros, M. (2013). Leer y escribir para no desertar en la universidad, *Folios*. 38, 45-59. <http://www.scielo.org.co/pdf/folios/n38/n38a04.pdf>
- Pardo, F., y Baquero, J. (2001). La estructura argumentativa: base para la comprensión y producción de textos científicos y argumentativos. *Forma y Función*. 14, 98-118. <https://revistas.unal.edu.co/index.php/formayfuncion/article/view/17218>
- Parodi, G. (2008). Lingüística de corpus: Una introducción al ámbito. *Rla*, 46(1), 93-119.
- Perea, F. J. (2015). Gramática y producción textual: bases lingüísticas para la intervención didáctica. *Tejuelo* (22), 94-119. <https://dialnet.unirioja.es/servlet/articulo?codigo=5317964>
- Perelman, C., y Olbrechts-Tyteca, L. (1989). *Tratado de la argumentación. La nueva Retórica*. Gredos.
- Plantin, C. (2005). *La argumentación*. Ariel.
- Quiroz, G. (2003). *Tendencias profesionales de las industrias del lenguaje y la formación de profesionales del lenguaje*. 131-149. [https://www.researchgate.net/publication/270453146\\_Preparacion\\_y\\_procesamiento\\_de\\_un\\_corpus\\_para\\_la\\_creacion\\_de\\_materiales\\_en\\_clase\\_de\\_espanol\\_para\\_propositos\\_especificos/link/54ab26550cf2ce2df668da32/download](https://www.researchgate.net/publication/270453146_Preparacion_y_procesamiento_de_un_corpus_para_la_creacion_de_materiales_en_clase_de_espanol_para_propositos_especificos/link/54ab26550cf2ce2df668da32/download)
- Riccardi, D., Agudelo Taborda, J., y Bossio Blanco, V. del C. (2022). Desigualdad socioeconómica en América Latina y Colombia. Panorama de las políticas públicas para la redistribución de la riqueza. *Revista Internacional de Cooperación y Desarrollo*, 9(2), 18-33. <https://doi.org/10.21500/23825014.6093>
- Ruiz, F. J., Márquez, C., Badillo, E., y Rodas, J. M. (2018). Desarrollo de la mirada profesional sobre la argumentación científica. *Revista Complutense de Educación*, 29(2), 559-576. <http://dx.doi.org/10.5209/RCED.53452>
- Santibáñez-Yáñez, C. (2015). Robustez como categoría para el análisis de la cognición: el caso de la competencia argumentativa. *Cinta Moebio*. 52, 60-68. Doi: <https://doi.org/10.4067/S0717-554X2015000100005>
- Serrano, M. J. (2011). *Sociolingüística*. Ediciones del Serbal.
- Silva-Corvalán, C. (2001). *Sociolingüística y pragmática del español*. Georgetown University Press.
- Tindale, Ch. y Barrientos, R.M. (2021) "Argumentos fuertes / Strong Arguments. - Una discusión sobre "La fuerza de los argumentos y la perspectiva retórica" de Corina Yoris- Villasana (RIA 21, 2020). *Revista Iberoamericana de Argumentación*, 22:140-150.
- Thomas, D. (2005). *Type-token Ratios in One Teacher's Classroom Talk: An Investigation of Lexical Complexity*. 1-23.
- Tolchisky, L. (2014). El uso de corpus lingüísticos como herramienta pedagógica. *Textos de Didáctica de la Lengua y de la Literatura*. 65, 9-17. <https://diposit.ub.edu/dspace/bitstream/2445/66774/1/653407.pdf>
- Van Eemeren, F., Grootendorst, R., y Snoeck, F. (2006). *Argumentación: análisis, evaluación, presentación*. Biblos
- Vega, L. (2013). *La fauna de las falacias*. Trotta.