



Discursos con inclinación de género en *Twitter*, Ciudad de México: un estudio de corpus

Gendered Discourses in Twitter, Mexico City: A Corpus-Based Study

Carlos Núñez Mercado 

UNIVERSIDAD VERACRUZANA
MÉXICO
carnunez@uv.mx

César Antonio Aguilar 

UNIVERSIDAD VERACRUZANA
MÉXICO
ceaguilar@uv.mx

Recibido: 29-4-2024 / **Aceptado:** 8-8-2025

DOI: 10.4151/S0718-09342025011901314

Resumen

En la actualidad, los discursos con inclinación de género continúan siendo un tema relevante en los estudios de lenguaje y género. Se han analizado dichos discursos en sus formas orales y escritas; sin embargo, existe un vacío en la literatura con respecto a la dimensión digital. En este sentido, la presente investigación explora los patrones lingüísticos producidos en Twitter (ahora X) a través de un corpus piloto (*Corpus Twitter CDMX*) con el fin de encontrar frecuencias que verifiquen discursos con inclinación de género en la red social. Para generalizar los patrones y discursos, se llevó a cabo una comparación con un corpus de lengua general del español de *SketchEngine* (*esTenTen2018 Mexican domain .mx*), el cual compila textos tomados a partir de la web. Los hallazgos sugieren verificación de este tipo de discursos tales como diferencias de género, violencia de género, oportunidades, igualdad y discriminación en el trabajo, paternidad y heterosexualidad normativa. A su vez, el resultado estadístico con mayor significancia consistió en la verificación de la colocación de mujer con el adjetivo *bello* y *hermoso* con LogDice 6.0 y 5.1, respectivamente, lo cual resultó en un discurso de consumo de femineidad.

Palabras clave: lingüística de corpus, lenguaje y género, redes sociales

Abstract

Gendered discourses remain a central topic within the field of language and gender. Numerous studies have examined this type of discourses in both written and oral forms. However, there is a gap in the literature when it comes to their digital dimension. In this sense, this corpus-based study employs a pilot corpus (*Corpus Twitter CDMX*) to explore language patterns produced in Twitter (now X) in order to identify frequencies that verify gendered discourses in social media. To ensure data generalizability, the *esTenTen2018 Mexican domain (.mx)* was used as a reference

corpus. Findings suggest the articulation of gendered discourses on Twitter, including femininity consumption, gender differences, gender violence, employment opportunities, fatherhood, and compulsory heterosexuality. Notably, the femininity consumption discourse, through the co-occurrence of linguistic patterns ‘beautiful-women’ and ‘pretty-woman’ scored the highest statistical significance values in the reference corpus (LogDice 6.0 and 5.1, respectively).

Keywords: corpus linguistics, language and gender, social media

INTRODUCCIÓN

En los últimos años se han realizado estudios para comprender y describir el lenguaje del Internet o discurso digital (Androutsopoulos, 2006, 2013; Barton & Lee, 2013; Herring, 2007, 2019; Herring & Androutsopoulos, 2015; Weller, 2014). Esto ha llevado a investigar las formas electrónicas de comunicación y el discurso producido a través de estos medios, puesto que se han presentado características únicas no antes vistas en los discursos hablados y escritos producidos digitalmente (Crystal, 2011). Dentro de tales exploraciones, se han llevado a cabo diferentes estudios de lenguaje y género (Baker, 2005; Bamman et al., 2014; Ferré-Pavia & Zaldívar, 2022; Herring & Paolillo, 2006; Kendall & Tannen, 2015; Pérez-Sabater, 2019; Rodrigues et al., 2017; Vásquez & Sayers China, 2019). Empero, en el caso de México, no se ha abordado todavía los discursos con inclinación de género presentes en la red social *Twitter*, la cual ya se considera como un objeto de estudio para realizar diferentes tipos de investigaciones (Rogers, 2014). Los discursos de esta clase se definen como aquellos que expresan algo acerca de las mujeres y hombres, niños y niñas, y sobre sus acciones, comportamientos, opiniones o posturas, predilecciones, relaciones e identidades (Litosseliti, 2014). Así, en este trabajo, se presentan los resultados de una aproximación a este fenómeno.

El objetivo de este estudio *corpus-based* consistió en verificar discursos con inclinación de género a partir de patrones lingüísticos hallados en un corpus piloto de *Twitter/X*. Para explorar los patrones lingüísticos se utilizaron los marcadores de género ‘hombre’, ‘mujer’, ‘niño’, ‘niña’, ‘chico’, ‘chica’, ‘caballero’ y ‘dama’, los cuales fueron adaptados del inglés con base en Baker (2016). Se utilizaron herramientas y técnicas de la metodología lingüística de corpus (Baker, 2018; Březina, 2018; McEnery & Hardie, 2012; Sinclair, 1991). La plataforma de análisis empleada para esta investigación fue *Sketch Engine* (Kilgarriff et al. 2014). Para generalizar los hallazgos, se comparan los patrones del corpus piloto con un corpus de referencia del grupo *Ten Ten Corpus Family* (Jakubíček et al. 2013), específicamente, el corpus en español *esTenTen2018 Mexican domain.mx*, y se explican o interpretan los discursos según el contexto y sociedad en la que se producen.

A continuación, se expone el apartado del marco teórico, el cual aborda un panorama histórico del enfoque de las investigaciones de lenguaje y género, seguido de

la definición de discurso desde el estudio de discursos con inclinación de género. Luego, se procede a describir diversos estudios relacionados con el discurso, el género y los medios digitales. Posteriormente, se describen las herramientas y técnicas de la lingüística de corpus utilizadas como marco metodológico para esta investigación. Finalmente, se exponen los resultados, la discusión y las conclusiones.

1. Marco teórico

1.1 Lenguaje y género

Coates (2012) señala que el estudio de lenguaje y género puede clasificarse en dos ramas principales: la primera está relacionada con las investigaciones de carácter cuantitativo y sociolingüístico, es decir, las diferencias de uso del lenguaje entre hombres y mujeres pertenecientes a distintas clases sociales tomando en cuenta la variable género como binaria (Labov, 2006; Lakoff, 1973; Kathon, 2023; Kerswill, 2003; Tannen, 1994; Trudgill, 1984). A modo de ejemplo, el estudio de Trudgill (1984) encuentra que en promedio las mujeres hablan con un dialecto más cercano al inglés estándar mientras que los hombres usan en mayor proporción la forma vernácula de la lengua. La segunda rama se refiere al estudio de la lengua como discurso, concretamente, el análisis de conversaciones auténticas tomando en cuenta sus contextos sociales (Coates, 2012). Aunado a esto, el término ‘género’ como categoría binaria hombre-mujer ha transformado su significado a partir de los años noventa, concibiéndose a partir de esa época como una construcción social y flexible a partir del lenguaje (Coates, 2012; Eckert & McConnell-Ginet, 2013; Sunderland, 2015; Talbot, 2019). Coates (2012) afirma que durante los últimos 30 años existen tres enfoques principales en los estudios de lenguaje y género: dominio, diferenciación y construccionismo social, siendo este último el más utilizado en las investigaciones recientes de lenguaje y género.

La presente investigación se centra en un estudio de género y lenguaje desde el construccionismo social a través de la metodología lingüística de corpus. De manera específica, esta investigación adopta el concepto de género definido por Sunderland (2004). La autora argumenta que, en los discursos con inclinación de género, dicho concepto adquiere un significado relacionado con los seres humanos y cualquiera de las diferencias entre hombres y mujeres son social o culturalmente aprendidas, mediadas o construidas. Esta forma de concebir el género contrasta con el esencialismo biológico del término sexo, lo cual conlleva a generar un cambio de enfoque que va desde un determinismo hacia las posibilidades de desaprender y reaprender, de resistencia al orden establecido (i.e. las estructuras, normas o prácticas dominantes en una sociedad) y al cambio tanto a nivel personal como social (Sunderland, 2004). Sin embargo, a pesar de que lo anterior puede ser una solución al determinismo biológico, el género sigue siendo algunas veces percibido y usado de una manera social determinista, por lo que, un significado de género se convierte,

especialmente en un discurso popular de diferencias de género (Sunderland, 2004) o mantiene relaciones desiguales entre géneros (Talbot, 2019). Pese a que dichas diferencias expresadas en el discurso son un fenómeno global –o casi global– es posible encontrarlas en comportamientos específicos (e. g., patrones lingüísticos), los cuales, por lo general, pertenecen a un contexto particular de una comunidad de habla (Sunderland, 2004).

1.2 El concepto de ‘discurso’ desde los discursos con inclinación de género

Bucholtz (2003) afirma que el discurso se puede abordar desde dos perspectivas. La primera refiere al estudio del discurso desde el punto de vista lingüístico, es decir, cómo se divide en áreas de estudio tales como fonología, morfología, y sintaxis. Las combinaciones de sonidos forman palabras, las palabras forman enunciados, y finalmente, la combinación de enunciados resulta en discurso. El otro concepto de discurso está basado en su función. Discurso, desde este punto de vista, es el lenguaje dicho en contexto tal como es usado en situaciones sociales, y es precisamente esta definición la que el estudio de lenguaje y género ha adoptado debido a su enfoque en el contexto donde el lenguaje es usado (Bucholtz, 2003). En la práctica, sin embargo, las dos definiciones son compatibles debido a que la primera también puede partir del contexto en el que es usada. Litosseliti (2014) menciona que el término discurso se utiliza en varias disciplinas de diferente forma, lo que ocasiona imprecisión y confusión en su definición. En este sentido, Litosseliti (2014) afirma que dicho término se puede abordar desde un nivel básico; es decir, como un conjunto de oraciones o expresiones. Sin embargo, a partir de este punto, los analistas del discurso comienzan a diferir, por ejemplo, desde el punto de vista lingüístico, el discurso es un texto con patrones y reglas de cohesión; desde el enfoque sociológico, consiste en conversaciones que poseen ciertas funciones sociales; también se puede abordar desde la perspectiva de una teoría social crítica, como las manifestaciones culturales de una sociedad; es decir, las formas de pensar y actuar de un determinado grupo de personas.

Para van Dijk (1990, como se citó en Litosseliti, 2014) consiste en una forma específica del uso del lenguaje y una forma específica de interacción social. Finalmente, para Foucault (1972, como se citó en Litosseliti, 2014) el discurso es una construcción social de la realidad desde una perspectiva en particular, una forma de conocimiento o prácticas ideológicas. Para ilustrar la conexión entre discurso y situaciones sociales, se puede considerar el debate sobre la brecha salarial entre hombres y mujeres en el ámbito profesional (Litosseliti, 2014). Las discusiones relacionadas con dicho tema usualmente se respaldan en discursos de oportunidades iguales o feministas, en donde las situaciones o contextos sociales son las interacciones entre políticos, debates en los medios de comunicación, protestas por grupos de mujeres, entre otros; de manera que no es posible resaltar únicamente un discurso

feminista homogéneo porque cualquier discurso sobre las mujeres incluye y hace referencia a otros discursos acerca de los hombres, empleo, lugares de trabajo (Litosseliti, 2014).

Los lingüistas interesados en la teoría crítica social señalan que es más apropiado hablar múltiples *discursos* (Litosseliti, 2014). Algunas características que los discursos comparten, según Litosseliti (2014), son: *a) identificables y significativos*: por ejemplo, los discursos con inclinación de género en términos de la fortaleza del hombre y la belleza de la mujer, las emociones femeninas, heterosexualidad normativa –la cual se define en Sauntson (2021) como un sistema que prescribe, promueve y normaliza la monogamia, la reproducción y roles de género tradicionales– se apoyan en discursos relacionados con el género y son identificables en los usos del lenguaje de los hablantes; *b) competitivos, opuestos o de apoyo*: existen discursos que pueden considerarse que apoyan a hombres o mujeres, sin embargo, también pueden llegar a ser opuestos o paradójicos; *c) representan y constituyen formas de pensamiento y actuar*: desde un punto de vista posestructuralista y la teoría de la construcción social, los discursos construyen o dan sentido a como se concibe el mundo; *d) contienen ideologías y el poder social se representan a través de ellos*: los discursos son intrínsecamente ideológicos dado que postulan valores y puntos de vista en detrimento de otros; *e) existen en relación con otros discursos*: por ejemplo, los discursos feministas existen en relación con los discursos sobre el patriarcado, la dominación del hombre en la vida pública, discursos de violencia doméstica, entre otros. Finalmente, Sunderland (2004) afirma que un discurso tiene la característica de ser descriptible; por ejemplo, en sus posibles rastros lingüísticos, se les puede otorgar un nombre, tales como discurso académico o discurso feminista y, dicho proceso, es extremadamente interpretativo.

1.3 Discursos con inclinación de género

Para Mackenzie (2021), el estudio de ‘discursos con inclinación de género’ o *gendered discourses* es de gran relevancia en las investigaciones de lenguaje, género y sexualidad. El autor sostiene que las exploraciones acerca de los discursos con inclinación de género ofrecen una forma de conceptualizar y dirigir la atención hacia la naturaleza del género en nuestro mundo social, al nombrar maneras específicas con las cuales las personas se posicionan como sujetos de género. Es decir, el género se construye y se expresa mediante prácticas sociales, lenguaje, comportamientos y normas culturales. Litosseliti (2014) menciona que tales discursos representan, reconstruyen, sostienen y confrontan las prácticas sociales de género; por ejemplo, la concepción del hombre como agresivo y la cosificación de la mujer son prácticas sociales presentes en los discursos. Los discursos con inclinación de género son aquellos que son articulados por ambos, hombres y mujeres, focalizan sus similitudes y diferencias, en diferentes maneras y situaciones, de manera implícita (Baker, 2016; Litosseliti, 2014).

En su análisis acerca de la identificación y denominación o etiquetado de los discursos con inclinación de género, Sunderland (2004) reconoce una variedad, tales como: *a) discurso de diferencias de género*, los cuales sugieren una temática masculina como norma o una incompreensión mutua entre sexos, que pueden ser abordados desde una perspectiva crítica; *b) discurso de heterosexualidad*, en donde tal tópico se configura como lo normativo, conveniente, y hegemónico, por ejemplo, regular la conducta sexual masculina, propugnar la tolerancia y la heterosexualidad obligatoria; *c) discurso de género y oportunidades de trabajo*, los cuales atienden aspectos laborales y sus limitaciones prácticas; *d) discurso relacionado con la menopausia*, comprenden la rivalidad entre las farmacéuticas y la terapia alternativa, pero también la emancipación feminista; y, finalmente, *e) discurso de paternidad*, destacan el discurso de paternidad compartida, la paternidad en la familia tradicional, la nueva paternidad, y la diversidad.

Los ejemplos anteriores, según Litoselitti (2014), delimitan actitudes hacia el género, ofreciendo una visión particular, explícita o implícita, de las posturas de hombres y mujeres. Así como respuestas a indagaciones acerca de las identidades creadas como resultado de diferentes posturas, en contextos y situaciones, así como también reflejan oportunidades y desigualdades de género que se crean y mantienen como resultado de dichos discursos. Por ejemplo, Holmes (2008) encuentra que en los entornos laborales los discursos de género se utilizan para establecer liderazgo y revelan aspectos interpersonales. Por su parte, Toffoletti y Starr (2016) identifican el uso de discursos en profesoras universitarias e investigadoras que abordan roles de género respecto al trabajo remunerado y al trabajo doméstico, en los cuales se identifican patrones en detrimento de su carrera profesional y como son silenciados en el ámbito laboral. En contextos deportivos, Adams et al. (2010) hallan el rol de los discursos para la construcción y reafirmación de la masculinidad, a saber, el uso de lenguaje tóxico relacionado con la guerra, el género y la sexualidad para fomentar la agresividad y violencia en los jugadores con el fin de potenciar su desempeño atlético. Namatende-Sakwa (2019) revela discursos implícitos en libros de texto de primaria que mantienen la construcción de roles binarios de género y la hegemonía masculina, así como también un desconocimiento por parte de los maestros en la manera en que los discursos son empleados en el salón de clases.

1.4 El discurso digital y el género

El estudio de discurso a través de medios electrónicos ha estado creciendo y desarrollándose rápidamente en los últimos años entre académicos. Si bien algunas investigaciones comparten los mismos resultados ya vistos en las dimensiones de discurso hablado y escrito, también se han hallado singularidades a través de estos medios (Kendall & Tannen, 2015). Por ejemplo, se dice que, en conversaciones (síncrona o asíncrona) a través de mensajes de textos, el género femenino tiende a usar más signos de puntuación que los hombres, y esto puede derivar de diferentes factores, tales como edad, educación, gusto lingüístico, personalidad o una

combinación de ellos (Crystal, 2011). Pérez-Sabater (2019) en un estudio sobre el uso de emoticones en grupos de WhatsApp encontró que, en conversaciones para coordinar salidas o llegar a acuerdos entre amigos o amigas, los hombres incluyen emoticones con poca frecuencia, en comparación con las mujeres que hacen un uso considerable en este tipo de conversaciones. Además, los hombres usan un estilo sucinto para organizar salidas sin saludos iniciales ni despedidas. De hecho, uno de los emoticones más frecuentemente utilizado por los hombres es el *thumbs-up* y es comúnmente usado para cerrar un acuerdo. Por otra parte, las mujeres frecuentemente los utilizan al final de los enunciados con el objetivo de intensificar sus emociones después de expresarlas digitalmente escritas (Sampietro, 2016, como se citó en Pérez-Sabater, 2019).

Vásquez y Sayers China (2019) analizaron las reseñas escritas en Amazon con el fin de observar cómo a través del discurso se construye y orienta una identidad de género. Además, observan cómo se representa el género en las descripciones de ciertos productos. Acerca del discurso analizado en los comentarios sobre las bolsas para pañales, Vásquez y Sayers China (2019) descubrieron que este producto es de uso femenino, por lo que un hombre cargando una bolsa de este tipo afecta las normas establecidas para el género masculino. Baker (2016), en su estudio mayormente cualitativo, de un corpus creado a partir del foro *Q+A* (preguntas y respuestas), examinó los términos con una mayor frecuencia de visualización en líneas de concordancia para identificar los discursos con inclinación de género presentes alrededor de los marcadores como: *man*, *woman*, *girl*, *boy*, *guy*, *gentleman*, y *lady*. Finalmente, Baker (2016) identificó cuatro grupos de discursos con inclinación de género: Marte y Venus (*Mars and Venus*), caballerosidad (*Male Chivalry*), estándares sexuales (*Sexual Standards*) e igualdad de las mujeres (*Women's Equality*).

El estudio de identidad de género y variación léxica en *Twitter* de Bamman et al., (2014) está dividido en dos partes, el primero es un análisis cuantitativo tradicional binario, en donde se encontró que los pronombres, términos que denotan emociones y los emoticones y términos que refieren a la familia, amistad o parentesco (*mom*, *mommy*, *sister*, *daughter*, *aunt*, *auntie*, *grandma*, *kids*, *child*, *dad*, *husband*, *hubs*), son marcadores de género generalmente asociados a usuarios del sexo femenino. De estos últimos, solo unos cuantos son asociados a usuarios del sexo masculino (*wife*, *wife's bro*, *bruh*, *bros*, and *brotha*). En la segunda parte de la investigación, Bamman et al. (2014) afirman que identificaron usuarios que revierten los patrones anteriormente descritos, por ejemplo, las dimensiones del uso estándar de lengua o la expresividad que se relacionan con el discurso femenino se revirtió al encontrar que ciertos grupos de mujeres usan con más frecuencia groserías y palabras tabú. Para finalizar esta sección, Lane (2022) menciona que los discursos con inclinación de género que influyen a las mujeres en la construcción de feminidad ‘estratégica’ a través de identidades digitales

en *Facebook*, son los relacionados con la belleza, heteronormatividad, y profesionalismo.

2. Marco metodológico

El objetivo de esta investigación es encontrar discursos con inclinación de género que representen prácticas sociales mediante la exploración de patrones lingüísticos en el *Corpus Twitter CDMX* –construido para esta investigación– que deriven del uso de marcadores de género: ‘hombre’, ‘mujer’, ‘chica’, ‘chico’, ‘niña’, ‘niño’, ‘caballero’ y ‘dama’. Dicho corpus compila textos (*tweets*) en tres subcorpus o grupos de usuarios: Hombres, Mujeres y Comunidad LGTTBIQ+. Las preguntas de investigación que este estudio busca responder son:

- ¿Qué patrones lingüísticos se encuentran en el *Corpus Twitter CDMX* con respecto al uso de marcadores de género?
- ¿Cuáles discursos con inclinación de género emergen de dichos patrones lingüísticos?
- ¿Cuáles patrones lingüísticos son estadísticamente significativos en un corpus de referencia de mayor tamaño?
- ¿Qué interpretación, significado y explicación social tienen dichos discursos o patrones?

En este estudio se utilizó a la lingüística de corpus como método para analizar los datos (McEnery & Wilson, 2001; McEnery & Hardie, 2012; Parodi, 2008; Rojo, 2021). En la lingüística de corpus se establecen cuatro niveles o dimensiones de análisis. La primera dimensión es, como Březina (2018) propone, la exploración de los datos y sus principales tendencias, la segunda se refiere a la estadística inferencial y la cantidad de evidencia; es decir, si se tiene suficiente evidencia para refutar la hipótesis nula. A su vez, en este nivel también se verifica si los patrones lingüísticos ocurrieron por casualidad o si reflejan un uso significativo en la población muestra. La tercera alude al tamaño del efecto; en otras palabras, la magnitud de las diferencias encontradas al comparar dos corpus. Finalmente, la interpretación lingüística o el significado social que los resultados tienen o pueden tener en contexto del cual provienen (Březina, 2018; Březina et al. 2015).

Sin embargo, las investigaciones de corpus no se basan únicamente en un análisis cuantitativo (McEnery & Wilson, 2001). Los estudios cualitativos pueden estar relacionados con el tamaño de la muestra y cabe destacar que los corpus de tamaño reducido brindan una mayor facilidad al investigador para analizar los textos de manera más detallada y exhaustiva (Koester, 2022). Aunado a esto, Baker (2018) propone un análisis de líneas de concordancias, el cual combina criterios tanto cuantitativos como cualitativos. Para llevar a cabo una exploración de tal clase, Sinclair (1991, como se citó en Baker, 2018) sugiere tomar treinta líneas aleatoriamente, examinarlas para verificar que patrones y prosodias presentan. Posteriormente

examinar otras treinta líneas y así sucesivamente hasta agotar todos aquellos patrones que sean de interés.

Con respecto a los procedimientos y métodos de la lingüística de corpus, para explorar patrones del lenguaje, el análisis de concordancias se ha consolidado de manera estandarizada por los investigadores (Jones, 2022; McEnery & Hardie, 2012). Para esta investigación, se decidió optar por utilizar dos dimensiones de Březina (2018) y el análisis de líneas de concordancia de Sinclair (1991) y Baker (2018). En la primera etapa del análisis se inicia directamente con la exploración de las principales tendencias de los marcadores de género (*hombre, mujer, chico, chica, niño, niña, caballero y dama*) y los adjetivos positivos y negativos con los que colocan en el corpus piloto de *Twitter/X*. Consecuentemente, para llevar a cabo la generalización de los patrones lingüísticos hallados, se utilizó un corpus de referencia lengua de general de español denominado *esTenTen Mexican domain .mx* desarrollado por Jakubíček et al., (2013).

2.1 Corpus empleados

El *Corpus Twitter CDMX* (no publicado aún) representa el discurso digital producido en esta red social por usuarios geolocalizados con una dirección IP en la Ciudad de México. Contiene textos de tipos tres de usuarios: Hombres, Mujeres, y Comunidad LGTBIQ+. Para el caso de los dos primeros, se incluyeron tweets con los perfiles de los usuarios que contengan nombre(s) masculino o femenino, ya sea en el perfil o en la mención (@) de la cuenta (los nombres neutros tales como @Ariel, no fueron incluidos en la muestra). Para el tercer grupo, se incluyeron los tweets de los usuarios que abiertamente mencionan en la descripción de su perfil que pertenecen a la Comunidad LGTBIQ+, ya sea a través del símbolo de la bandera o se describen como pertenecientes a dicha comunidad (e. g., bisexual, *non-binary*).

La cantidad de evidencia total del corpus consistió en 4.998 *tweets*. La temporalidad de este corpus digital comprende 3 años desde el 1 de junio de 2018 al 30 de junio de 2021. Los archivos del corpus se transformaron a formato .txt UTF8 para mayor confiabilidad y utilidad en la plataforma (Březina, 2018; Reppen, 2022) que se utilizó en el análisis de los datos. Los emoticones no fueron incluidos, ya que el proceso de anotación requería mucho tiempo y el costo de creación del corpus se incrementaría y la efectividad del análisis se reduciría (Berber Sardinha, 2022). El tamaño total del *Corpus Twitter CDMX* es de 138.361 palabras (165.756 tokens), el cual se subdivide en:

- subcorpus Hombres Usuarios: 67.970 palabras
- subcorpus Mujeres Usuarios: 55.201 palabras
- subcorpus Comunidad LGTBIQ+ Usuarios: 15.190 palabras

Con respecto al corpus de referencia, dentro de la familia de corpus *TenTen* (Jakubíček et al., 2013) de varias lenguas del mundo, se encuentra el corpus *Spanish Web 2018* o *esTenTen18* que comprende diferentes variantes del español, clasificándolas

en dos principalmente: la variante del español de Europa y la variante del español de América. El *esTenTen18* tiene una extensión de 16.9 billones de palabras provenientes de diferentes sitios *web* y de diferentes variantes del español (Kilgarriff et al., 2014). La generalización de los hallazgos consistió en explorar la variante del español de México; es decir, el subcorpus del *esTenTen18 Mexican domain .mx* para realizar la comparación. Dicho corpus tiene una extensión de 1.4 billones de tokens, lo cual comprende un 7,193% de todo el corpus (Kilgarriff et al., 2014).

3. Exploración de los datos y principales tendencias: *Corpus Twitter CDMX*

Para comenzar con la exploración de los patrones lingüísticos, se analizaron los marcadores de género adaptados de la investigación de Baker (2016). Los marcadores en su idioma original fueron *man, woman, girl, boy, gentleman* y *lady*. Las equivalencias buscadas en el *Corpus Twitter CDMX* fueron ‘*hombre*’, ‘*mujer*’, ‘*chica*’, ‘*chico*’, ‘*niña*’, ‘*niño*’, ‘*caballero*’ y ‘*dama*’ a través de la herramienta *CONCORDANCE* de *Sketch Engine* (Kilgarriff et al., 2014), el cual visualiza las líneas de concordancia.

Al emprender este tipo de investigaciones, se sugiere explorar adjetivos positivos y negativos visualizados a través de líneas de concordancias para comprender su uso real (Baker & Brookes, 2021). Para evitar el sesgo en la clasificación de modificadores positivos y negativos, Březina (2018) sugiere realizar una medida de concordancia entre evaluadores (*inter-rater agreement*). Para este estudio se utilizaron dos evaluadores y la prueba estadística Cohen’s Kappa para análisis de contenido en textos (Březina, 2018):

Fórmula 1

Cálculo para fijar la concordancia entre evaluadores

$$\text{Cohen's } k = \frac{\text{grado de concordancia} - \text{concordancia por casualidad}}{1 - \text{concordancia por casualidad}}$$

$$\text{Cohen's } k = \frac{0.88 - 0.48}{1 - 0.48} = 0.76$$

Para interpretar este resultado, se sugiere utilizar la escala de Krippendorff (2012), pues contiene valores o puntos de corte recomendados para evaluar que el grado de concordancia sea confiable y no por casualidad; es decir, entre 0,67 y 0,8 para establecer un grado de concordancia confiable o muy confiable, respectivamente. Mientras que el número 1, indica un grado de concordancia absoluto. Dicho esto, se observa un grado de concordancia confiable debido a que el valor Cohen’s K equivale a 0,76 para la clasificación de modificadores positivos y negativos de los marcadores de género.

El marcador de género ‘mujer’ obtuvo un total de frecuencias de 119 (717,46, por millón de tokens / 0,07175%, en el corpus total de 165.756 tokens) como sustantivo. Los modificadores positivos más frecuentes para el marcador son: ‘inteligente’, ‘bella’, ‘hermosa’, ‘biológica’, ‘brillante’, ‘maravillosa’, ‘autónoma’, ‘sincera’, ‘chingo’, y los negativos ‘fatal’. La coocurrencia ‘mujer’ ‘brillante’ parece a primera vista un adjetivo positivo. Empero, al analizar el contexto de la línea de concordancia, se puede apreciar un reclamo social, ya que no se reconocen dichas características positivas cuando se habla de una mujer siendo inteligente o brillante, se les demerita con los adjetivos ‘mamona’ y ‘loca’, respectivamente, lo que conlleva un discurso de diferencias de género. El modificador positivo ‘autónoma’, denota una lucha por la igualdad y equidad, en este sentido, se infiere que hay rasgos del movimiento feminista, por lo que se deduce un discurso de diferencias de género. Otro modificador positivo fue ‘bella’; sin embargo, al analizar las líneas de concordancia y el contexto del texto, se puede apreciar una crítica a sus capacidades profesionales e incluso intelectuales. Se deduce un uso negativo del adjetivo y un discurso de oportunidades, igualdad y discriminación en el lugar de trabajo y de consumo de femineidad. Finalmente, el patrón mujer *biológica* sugiere un discurso heteronormativo, en el cual se observa un disenso hacia las personas trans.

El marcador ‘hombre’ obtuvo una frecuencia de 64 (385,86, por millón de tokens / 0,03859%, en el corpus total de 165.756 tokens) y los modificadores positivos más frecuentes son: ‘guapo’, ‘inteligente’, ‘trans’, ‘decente’, ‘valiente’, ‘brillante’, ‘hermoso’, ‘libre’, ‘verdadero’; únicamente se presentó un modificador negativo ‘irresponsable’. Con base en el contexto de las líneas de concordancia, ‘hombre irresponsable’ se refiere a un ‘padre’, por lo que podemos inferir que existe un discurso con inclinación género relacionado con la paternidad. Referente al rol familiar, es muy común en la sociedad ver el rol del padre menospreciado o de menos importancia en comparación con el rol de la maternidad, lo cual pone en desventaja a ambos roles. Con respecto a los modificadores positivos se reconoce la existencia de un discurso de diferencias de género ya que hacen referencia a las cualidades intelectuales del hombre (inteligente, brillante) en comparación con el uso en el marcador ‘mujer’. El modificador ‘trans’ es el único relacionado con los usuarios de la comunidad LGBTIQ+ que hace referencia al trato hacia los hombres trans, lo cual podría verificar un discurso de diferencias de género. Otros modificadores hacen referencia a la apariencia física (hermoso, guapo); a sus modales (decente); y otros modificadores relacionados con masculinidades (valiente, libre, verdadero).

En relación con los marcadores de género ‘chica’, ‘chico’, ‘niña’, ‘niño’, la búsqueda en la herramienta *Word Sketch* se realizó únicamente con la palabra clave en género (gramatical) masculino a causa de que la herramienta ejecuta ambas al mismo tiempo y no es posible realizarla por separado. El marcador de género ‘niño(a)’, el número total de frecuencias es 68 (410,09, por millón de tokens / 0,04101%, en el corpus total de

total de 165.756 tokens) y el marcador de género ‘chico(a)’ como sustantivo, el número total de frecuencias es 33 (199,01, por millón de tokens / 0,01990%, en el corpus total de 165.756 tokens). Del marcador de género ‘dama’, no se encontró ningún uso en el corpus, a diferencia de ‘caballero’, el cual su número total de frecuencias es 4 (24,12, por millón de tokens / 0,002412%, en el corpus total de 165.756 tokens), el único resultado corresponde al adjetivo *amable* (un caballero muy amable).

El marcador de género ‘chica’ coloca con dos modificadores que hacen referencia a la nacionalidad de un país extranjero ‘asiática’ y ‘rusa’, por otro lado, ‘chico’ únicamente coloca con ‘cabro’. El marcador ‘niño’ tiene tres modificadores ‘jotito’, ‘migrante’ y ‘chiquito’, mientras que ‘niña’ ‘linda’ y ‘llena’. El modificador ‘jotito’, es un término que puede referir a una persona homosexual o afeminada, al analizar la línea de concordancia del subcorpus de usuarios LGTBIQ+; se entiende que hace referencia a su infancia y cómo vivió una situación difícil con respecto a su género durante esa etapa de su vida, ya que también utiliza palabras como ‘inútil’ y ‘vulnerable’. Aquí es posible derivar un discurso de heterosexualidad normativa, la cual dicta a través de un sistema de creencias, normas culturales y sociales a la heterosexualidad como la única orientación sexual (Cruz, 2020). Del marcador de género ‘dama’, no se hallaron ocurrencias en el Corpus *Twitter* CDMX. Por otra parte, el marcador ‘caballero’, únicamente colocó con el modificador positivo ‘amable’. Dicho patrón lingüístico puede relacionarse con el tipo de discurso del contexto de Baker (2018) *Male Chivalry* o caballerosidad, un rol que se espera del género masculino.

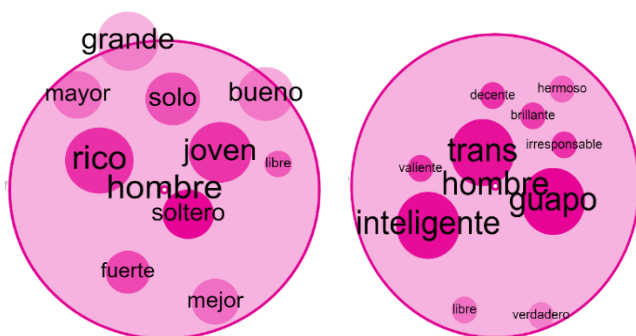
3.1 Comparación con corpus de referencia: Spanish Web 2018 Mexican domain .mx

Según Březina (2018), el corpus de referencia debe ser más extenso con el fin de proporcionar una mayor cantidad de evidencia sobre las frecuencias de las palabras. Debido a que los corpus de *Twitter* son usualmente contruidos con base en tendencias a partir de etiquetas o *hashtags*, se eligió el *Spanish Web Mexican domain .mx* como corpus de referencia debido a que contenía dos similitudes: el discurso digital en español y el contexto mexicano de los textos. Aunado a esto, al ser un corpus de lengua general del español de gran tamaño es posible generalizar los hallazgos al compararlos con los usos del lenguaje de un corpus de menor tamaño (Březina, 2018; McEnery & Brookes, 2022). El primer paso para la comparación entre ambos corpus fue realizar la exploración de las principales tendencias del *Spanish Web 2018 Mexican domain .mx* (*esTenTen18 Mexican domain .mx*, de ahora en adelante), la cual consistió en la búsqueda de los 10 modificadores más frecuentes que colocan con los marcadores de género ya antes mencionados y, consecuentemente, compararlos con las tendencias en el *Corpus Twitter* CDMX. Lo anterior, con el fin de verificar si existe alguna coincidencia en el uso de dichos marcadores de género entre ambos corpus. El marcador ‘hombre’ obtuvo un resultado de 454, 624 ocurrencias (322,59 por millón de tokens) en el

esTenTen18 Mexican domain .mx. Los 10 modificadores más frecuentes fueron ‘rico’, ‘joven’, ‘grande’, ‘bueno’, ‘solo’, ‘soltero’, ‘mayor’, ‘mejor’, ‘fuerte’, ‘libre’ y ‘mismo’. A continuación, se comparan los 10 modificadores de ambos corpus:

Figura 1

Comparación de uso de modificadores en el esTenTen18 Mexican domain .mx (izquierda) y el Corpus Twitter CDMX (derecha)



Nota: Visualizado por *SketchEngine* (Kilgarriff et al. 2014)

La Figura 1 visualiza diferentes tipos de datos, por ejemplo, mientras más grande sea el círculo, mayor el número de frecuencias. En otras palabras, en el *esTenTen18 Mexican domain .mx*, los adjetivos ‘rico’, ‘joven’ y ‘grande’ son los que obtuvieron mayor número de frecuencias; en cambio, en el *Corpus Twitter CDMX*, los tres adjetivos más frecuentes son ‘guapo’, ‘inteligente’ y ‘trans’. Podemos deducir que, en cuanto a los 10 adjetivos más frecuentes, existe únicamente la similitud del adjetivo ‘libre’ con una frecuencia de 17,551 con valor *LogDice* 7.2 (Rychlý, 2008). Después de analizar las líneas de concordancia no se halló un discurso con inclinación de género similar al del *Corpus Twitter CDMX*. Por lo tanto, no es posible generalizar de la coocurrencia *hombre-libre* un discurso de diferencias de género.

Referente al marcador ‘mujer’ obtuvo un resultado de 676.297 ocurrencias (479,88 por millón de tokens) en el corpus *esTenTen18 Mexican domain .mx*. Los 10 modificadores más frecuentes fueron ‘joven’, ‘indígena’, ‘mexicana’, ‘primera’, ‘mayor’, ‘soltera’, ‘hermosa’, ‘trabajadora’, ‘bella’, ‘migrante’. Se encontraron dos similitudes en los marcadores ‘hermoso’ (1,789 / *LogDice* 6.0) y ‘bello’ (1,586 / *LogDice* 5.7), como se muestra a continuación:

Figura 2

Comparación de uso de modificadores en el esTenTen18 Mexican domain .mx (izquierda) y el Corpus Twitter CDMX (derecha)



Nota: Visualizado por SketchEngine (Kilgarriff et al., 2014)

Ahora bien, el siguiente paso es determinar los discursos con inclinación de género que se derivan de las colocaciones ‘mujer bella’ y ‘mujer hermosa’. Para realizar el análisis de líneas de concordancias, se optó por la herramienta de muestra aleatoria de *Sketch Word*. Según Kilgarriff et al. (2014) cuando se trabaja con una muestra amplia de líneas de concordancia, se utiliza una muestra aleatoria para reducir la muestra y, al mismo tiempo, se mantiene la representatividad de la muestra. Para el análisis cualitativo de las líneas de concordancia aleatorias en el *esTenTen18 Mexican domain .mx*, se seleccionó el número 300, el cual es un número realizable en términos de limitaciones de tiempo y cantidad para dicho tipo de análisis.

Figura 3

Extracto de líneas de concordancia: colocación mujer + hermosa

	Details	Left context	KWIC	Right context
1	de10.com.mx	s><s>Estoy estudiando en Inglaterra y hay cada	mujer	hermosa aquí que parecen caídas del cielo insp
2	de10.com.mx	ación que llevan.</s><s>En holanda no vi tantas	mujeres	hermosas como hubiera imaginado, hay mucho
3	de10.com.mx	as.</s><s>En la europa del este te encuentras a	mujeres	muy hermosas en la calle o en comercio, lo únic
4	de10.com.mx	lo claro),y los que dicen que "en todos lados hay	mujeres	hermosas ", pues el tema no va por allí, NO MA
5	de10.com.mx	o a muchos europeos,que en este continente las	mujeres	hermosas se concentran en EUROPA ORIENTA
6	de10.com.mx	Malinchista pero nada que ver.</s><s>Aquí hay	mujeres	hermosas , miden en promedio entre 1.70 a 1.9
7	foroactivo.mx	<s>- Ya mi amor – intervino la rubia y hermosa	mujer	- nuestro hijo es así, le gustan mucho las chicas
8	foroactivo.mx	l ahora 28 rango:ninguno Descripción: Hermosa	mujer	rubia de ojos azules es rápida ágil e inteligente.<
9	sinergiaradio.m...	nás grande de Países Bajos, también tiene a las	mujeres	más hermosas , así que si tienes pensado viajar
10	de10.com.mx	oy de acuerdo no quiero que en mexico no haya	mujeres	hermosas pero hay que ser realista se ven men

Fuente: SketchEngine (Kilgarriff et al., 2014)

Después de analizar las líneas de concordancia, se encontró un discurso relacionado con estereotipos de belleza femenina, el cual define a la mujer europea como el estándar de belleza. Las dos colocaciones verifican un discurso de consumo de femineidad debido a que usualmente la industria de la moda y el entretenimiento resaltan su figura estética.

La comparación del resto de los marcadores se realizó siguiendo el mismo procedimiento para ‘hombre’ y ‘mujer’. El marcador ‘niño(a)’ obtuvo un número de ocurrencias de 563.222 (399,65 por millón de tokens). A continuación, se visualiza la comparación del uso del marcador ‘niño(a)’ en ambos corpus:

Figura 4

Comparación de uso de modificadores en el esTenTen18 Mexican domain .mx (izquierda) y el Corpus Twitter CDMX (derecha)



Nota: Visualizado por *SketchEngine* (Kilgarriff et al. 2014)

Como se destaca en la Figura 4, la única similitud entre el corpus *esTenTen18 Mexican domain .mx* y el *Corpus Twitter CDMX* fue el modificador ‘migrante’, el cual coloca con el marcador ‘niño’ un total de 1.307 coocurrencias (*LogDice* 8.1). Las prácticas sociales reflejadas a través del discurso se inclinan hacia salvaguardar la integridad de los niños y las niñas migrantes debido a que pertenecen a grupos vulnerables. El lexema más común que se encuentra en las líneas de concordancia del corpus *esTenTen18 Mexican domain .mx* es ‘proteger’, ‘protección’ y ‘proteja’. De la misma forma, al analizar su texto completo, otras concordancias corroboran la vulnerabilidad de los niños y las niñas migrantes con expresiones relacionadas con la violencia física y sexual, fragilidad jurídica y desastres naturales. Finalmente, la vulnerabilidad está ligada a ambos sexos, lo que nos lleva a decir que no existe un discurso con inclinación hacia un género en específico de dicho patrón lingüístico.

4. Discusión

Esta investigación reconoce al lenguaje producido en medios digitales como la cuarta dimensión del discurso (Crystal, 2011). Se construyó el *Corpus Twitter CDMX*

con el fin de hallar usos del lenguaje que resulten en discursos con inclinación de género (Litosseliti, 2014; McKenzie, 2021; Sunderland, 2004). El propósito de dividir el *Corpus Twitter CDMX* en los subcorpus Hombres Usuarios (HU), Mujeres Usuarios (MU), y Usuarios de la comunidad LGTBIQ+ fue para verificar si los discursos encontrados son exclusivos o compartidos tanto de los usuarios como de los marcadores de género. La siguiente tabla visualiza esta información:

Figura 5

Tabla de usuarios del Corpus Twitter CDMX y discursos con inclinación de género

Discursos con inclinación de género en el <i>Corpus Twitter</i>	Líneas de concordancia	Marcador
<i>Diferencias de género</i>	MU_cdmxTwitter Un hombre inteligente es un	hombre inteligente, una mujer inteligente es una mamona.
	MU_cdmxTwitter una mujer inteligente es una mamona. Un	hombre brillante es un genio, una mujer brillante es una loca. (Gracias
	LGTBIQ+U_cdmx gineco-endocrino dijo que nos quiere enseñar a tratar	hombres trans de forma estructura y función. Ahora si tiene toda mi
	MU_cdmxTwitter las declaraciones de libertad (como ser	mujer autónoma, independiente, que disfruta de su sexualidad
<i>Violencia de género</i>	HU_cdmxTwitter Cómo se le llama a un	hombre que es controlador, manipulador, grosero, infiel, mentiroso
	HU_cdmxTwitter violencia derivada de la masculinidad es muy (...)	hombres decentes " como Biden quienes mandan a la guerra hasta a
	MU_cdmxTwitter de chicos hablar de cómo vieron la marcha "de las	chicas " y "estuvo chingón, si tienen que cambiar las cosas" casi lloro.
	HU_cdmxTwitter Ojalá y este #9M (...) no por la inseguridad (...) si no por	mujeres más chingonas del mundo!
<i>Oportunidades, igualdad y discriminación en el trabajo</i>	MU_cdmxTwitter Nada peor que enredar con brillos falsos a una	niña llena de sueños para luego meterse entre sus piernas.
	MU_cdmxTwitter capacidad para gobernar? Una bella	mujer perfecta para un concurso de
	HU_cdmxTwitter amigos. Se adquiere a una	mujer comprometida con la transparencia y
	MU_cdmxTwitter Yo no estoy ni cerca de ser la	niña más linda del mundo, (...) siempre que salgo llevo mi dinero y
<i>Paternidad</i>	MU_cdmxTwitter el significado de bruja es	Mujer Empoderada .
	MU_cdmxTwitter ¿Pues qué tipo de padre es? ¿Dónde estaba? Que	hombre tan irresponsable .
	HU_cdmxTwitter Ángela Ponce, para mí sólo es un	hombre que parece mujer pero NO LO ES.
	HU_cdmxTwitter hombre que se percibe como mujer/peleó en MMA con	mujer biológica: Tamikka Brents, a la cual le rompió el cráneo en los
<i>Consumo de femineidad</i>	LGTBIQ+U_cdmx mes del orgullo y salir a marchar es la forma de decirle: niño	jotito que él puede ser lo que quiera ser, que el cielo es el
	HU_cdmxTwitter el corazón repleto que hermosa	mujer. Qué hermosa coincidencia. Qué hermosa
	MU_cdmxTwitter concurso de belleza pero sin experiencia para atender l	mujeres bellas para que los calenturientos voten por Morena
	MU_cdmxTwitter Y la capacidad para gobernar? Una bella	mujer perfecta para un concurso de belleza pero sin experiencia
	HU_cdmxTwitter Chicos qué suerte la mía. Me escribió una	chica rusa , solitaria, que se quiere venir a casar conmigo. Adios

Como se observa, existen diferentes discursos que comparten y que a su vez son exclusivos en relación con su aparición en los subcorpus o el uso de los marcadores. A saber, el discurso de *diferencias de género* fue reconocido por los usuarios mujeres y de la comunidad LGTBIQ+, en la mayoría de los casos aludiendo al marcador *hombre*, lo cual puede suponer la tendencia de que los hombres tienen más privilegios que las mujeres en nuestra sociedad mexicana. En relación con el discurso de *violencia de género*, su uso fue diversificado en los usuarios mujeres y hombres y los marcadores utilizados tales como *hombre*, *chica*, *mujer*, *niña*, fueron variados. Por otra parte, un patrón similar en el contexto de las líneas de concordancia analizadas consistió en que son las mujeres quienes son víctimas de la violencia (e.g. marcha en México del #9M “un día sin mujeres” en protesta con la violencia, falta de justicia y derechos). El discurso de *oportunidades, igualdad y discriminación en el trabajo* obtuvo un uso exclusivo del marcador *mujer*, pero distribuido en los subcorpus de hombres y mujeres usuarios. Esto verifica la lucha del feminismo por la igualdad de los derechos de las mujeres en el ámbito laboral. Por otra parte, también se crítica la inserción en dicho ámbito únicamente por sus cualidades físicas (e. g., *bella mujer*: sin capacidad para gobernar o para atraer votantes) y no profesionales. El discurso de *paternidad* se derivó de las mujeres usuarios y el marcador *hombre* como una crítica al rol del padre. En referencia a el discurso de *heterosexualidad normativa*, se usaron tres marcadores distintos (*hombre*, *mujer*, *niño*) y se

puede inferir dos posicionamientos en la red social de Twitter, el primer en contra de las personas trans encontrado en el subcorpus de hombres usuarios; es decir, si crítica su participación en competencias femeninas tanto deportivas como de belleza. El segundo, a favor de la diversidad de género, encontrado en el subcorpus de los usuarios de la comunidad LGBTQ+. Por último, el discurso de *consumo de femineidad* fue utilizado en los subcorpus de mujeres y hombres usuarios, con los marcadores *mujer* y *chica*.

El corpus de referencia seleccionado para generalizar los hallazgos fue el *esTenTen18 Mexican domain .mx*, de lengua general del español. El marcador ‘mujer’, en el cual se encontraron dos usos de adjetivos similares (e. g., mujer ‘hermosa’ y mujer ‘bella’) con un resultado estadístico *LogDice* 6,0 y 5,7, respectivamente. Esto quiere decir que existe fuerza colocacional entre los adjetivos ‘hermoso’ y ‘bello’ con el marcador ‘mujer’ apenas significativa. De dicho patrón, es posible deducir un discurso de consumo de femineidad (Litosseliti, 2014) hacia la belleza de la mujer. En este sentido, Perdomo Colina (2021, p. 610) analiza la construcción de la femineidad en los certámenes de belleza en Colombia, llegando a la conclusión que dicha construcción “estuvo mediada por el discurso estético como expresión normativa hetero dominante de la estructura patriarcal.” Estos resultados sugieren, en síntesis, que el patrón lingüístico ‘mujer’ ‘bella’ o ‘hermosa’ es exclusivo al sexo femenino y conlleva a prácticas sociales relacionadas con sus características fisionómicas. De igual manera, Lane (2022) identifica dicho discurso en la construcción de la feminidad en *Facebook*.

En conjunto, estos resultados responden a las preguntas de investigación e indican los discursos con inclinación de género en la red social *Twitter/X* en la Ciudad de México que pueden ser generalizables con base en un corpus de lengua general del español.

CONCLUSIONES

La presente investigación verifica en la red social *Twitter/X* que los discursos con inclinación de género parten de las posturas y actitudes hacia el género de hombres y mujeres (Litosseliti, 2014) y, en consecuencia, emergen por parte de los tres tipos de usuarios: hombres, mujeres y Comunidad LGTBIQ+. Lo anterior es un primer acercamiento para la inclusión de los usuarios de la comunidad mencionada en estudios de corpus. Bamman et al. (2014) señalan que investigaciones cuantitativas previas en género por lo regular tratan esta variable social como binaria (mujer/hombre), por lo tanto, lo que sugieren los autores es que en lugar de preguntarnos como mejorar los algoritmos para esclarecer la realidad sobre el verdadero género de la muestra (hombre o mujer), podemos dar un paso atrás y preguntarnos cómo sus elecciones lingüísticas participan en la construcción de las identidades de género, de modo que, dichos patrones de lenguaje lejos de ser un caso aparte puedan encajar coherentemente dentro de un panorama más amplio del

comportamiento en redes sociales. Por ejemplo, es posible observar que los discursos de *heterosexualidad normativa* y *diferencias de género* son utilizados por usuarios hombres, mujeres y de la comunidad LGBTQ+. Es decir, podemos incluir en los discursos con inclinación de género de Litosseliti (2014) y Sunderland (2004), las posturas y actitudes hacia el género de hombres y mujeres de la comunidad LGBTQ+.

REFERENCIAS BIBLIOGRÁFICAS

- Adams, A., Anderson, E., & McCormack, M. (2010). Establishing and Challenging Masculinity: The Influence of Gendered Discourses in Organized Sport. *Journal of Language and Social Psychology*, 29(3), 278-300. <https://doi.org/10.1177/0261927X10368833>
- Androutsopoulos, J. (2006). Introduction: Sociolinguistics and computer-mediated communication. *Journal of Sociolinguistics*, 10(4), 419-438.
- Androutsopoulos, J. (2013). Online data collection. En C. Mallinson, B. Childs, & G. V. Herk (Eds.), *Data Collection in Sociolinguistics: Methods and Applications* (pp. 236-249). Routledge. <https://www.languageatinternet.org/articles/2008/1609>
- Baker, P. (2005). *Public discourses of gay men*. Routledge.
- Baker, P. (2016). Gendered discourses. En P. Baker, & J. Egbert (Eds.), *Triangulating Methodological Approaches. Corpus-Linguistic Research* (pp. 143-156). Routledge.
- Baker, P. (2018). Corpus Methods in Linguistics. En L. Litosseliti (Ed.), *Research Methods in Linguistics* (pp. 167-191). Bloomsbury.
- Baker, P., & Brookes, G. (2021). Lovely nurses, rude receptionists, and patronising doctors: Determining the impact of gender stereotyping on patient feedback. En J. Angouri, & J. Baxter (Eds.), *The Routledge Handbook of Language, Gender, and Sexuality* (pp. 694-716). Routledge.
- Bamman, D., Eisenstein, J., & Schnoebelen, T. (2014). Gender identity and lexical variation in social media. *Journal of Sociolinguistics*, 18(2), 135-160. <https://doi.org/10.1111/josl.12080>
- Barton, D., & Lee, C. (2013). Language online: investigating digital texts and practices. *Choice Reviews Online*, 52(01), 52-0110. <https://doi.org/10.5860/choice.52-0110>
- Berber Sardinha, T. (2022). Corpus Linguistics and the Study of social media: A case study using Multidimensional analysis. En A. O'Keeffe, & M. J. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 1385-1417). Routledge.

- Březina, V. (2018). *Statistics in Corpus Linguistics: A Practical guide*. Cambridge University Press. <https://eprints.lancs.ac.uk/id/eprint/131100/>
- Březina, V., McEnery, T., & Wattam, S. (2015). Collocations in context. *International Journal of Corpus Linguistics* (pp. 139-173). <https://doi.org/10.1075/ijcl.20.2.01bre>
- Bucholtz, M. (2003). Theories of Discourse as Theories of Gender: Discourse Analysis in Language Gender Studies. En J. Holmes, & M. Meyerhoff (Eds.), *The Handbook of Language and Gender* (pp. 43-68). Blackwell Publishing.
- Coates, J. (2012). Gender and Discourse Analysis. En J. P. Gee, & M. Handford (Eds.), *The Routledge Handbook of Discourse Analysis* (pp. 90-103). Routledge.
- Cruz, R. J. (2020). Heteronormatividad y diversidad sexual en la formación del profesorado: Estudio etnográfico en una escuela Normal de la Ciudad de México. *Diálogos sobre Educación*, 11(21). <https://doi.org/10.32870/dse.v0i21.678>
- Crystal, D. (2011). *Internet Linguistics: A Student Guide*. Routledge. <http://ci.nii.ac.jp/ncid/BB04830816>
- Eckert, P., & McConnell-Ginet, S. (2013). *Language and Gender*. Cambridge University Press.
- Ferré-Pavia, C., & Zaldívar, G. (2022). El feminismo trans excluyente en Twitter: un monólogo sesgado en #ContraElBorradoDeLasMujeres. ICONO 14. *Revista de Comunicación y Tecnologías Emergentes*, 20(2).
- Herring, S. C. (2007). A faceted Classification Scheme for Computer-Mediated Discourse. *Language@Internet*, 4(1). https://www.languageatinternet.org/articles/2007/761/Faceted_Classification_Scheme_for_CMD.pdf
- Herring, S. C. (2019). The Coevolution of Computer-Mediated Communication and Computer-Mediated Discourse Analysis. En P. Bou-Franch, & P. Garcés-Conejos Blitvich (Eds.), *Analyzing Digital Discourse New Insights and Future Directions* (pp. 25-67). Palgrave MacMillan.
- Herring, S. C., & Androutsopoulos, J. (2015). Computer-Mediated Discourse 2.0. En D. Tannen, H. Hamilton, & D. Schiffrin, D. (Eds.) *The Handbook of Discourse Analysis* (pp. 127-151). Wiley Blackwell.
- Herring, S. C., & Paolillo, J. C. (2006). Gender and genre variation in weblogs. *Journal of Sociolinguistics*, 10(4), 439-459. <https://doi.org/10.1111/j.1467-9841.2006.00287.x>

- Holmes, J. (2008). Gendered Discourse at Work. *Language and Linguistics Compass*, 2(3), 478-495 <https://doi.org/10.1111/j.1749-818X.2008.00063.x>
- Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2013). *The TenTen Corpus Family. Proceedings of the 7th International Corpus Linguistics Conference CL 2013* (pp. 125-127). https://www.sketchengine.co.uk/wp-content/uploads/The_TenTen_Corpus_2013.pdf
- Jones, C. (2022). What are the basics of analysing a corpus? En A. O'Keeffe, & M. J. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 298-323). Routledge.
- Kathon, M.N.S. (2023). Gender and evaluation in contemporary American English: A corpus study based on pronominal and nominal expressions with male and female reference. *ICAME Journal, Sciendo*, 47(1), 39-61. <https://doi.org/10.2478/icame-2023-0003>
- Kendall, S., & Tannen, D. (2015). Discourse and gender. En D. Tannen, H. Hamilton, & D. Schiffrin (Eds.), *The Handbook of Discourse Analysis* (pp. 639-660). Wiley Blackwell.
- Kerswill, P. (2003). Dialect levelling and geographical diffusion in British English. En D. Britain, & J. Cheshire (Eds.), *Social Dialectology: In honour of Peter Trudgill* (pp. 223-243). John Benjamins.
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). *The Sketch Engine: ten years on. Lexicography*, 1(1), 7-36. <https://doi.org/10.1007/s40607-014-0009-9>
- Koester, A. (2022) Building small specialised corpora. En A. O'Keeffe, & M. J. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 143-173). Routledge.
- Krippendorff, K. (2012). *Content analysis: an introduction to its methodology*. Sage.
- Lane, L. (2022). Strategic femininity on Facebook: women's experiences negotiating gendered discourses on social media. *Feminist Media Studies*, 23(5), 2440-2454. <https://doi.org/10.1080/14680777.2022.2056754>
- Labov, W. (2006). *The Social Stratification of English in New York City (2a ed.)*. Cambridge University Press.
- Lakoff, R. T. (1973). Language and woman's place. *Language in Society*, 2(1), 45-79. <https://doi.org/10.1017/s0047404500000051>
- Litosseliti, L. (2014). *Gender and Language Theory and Practice*. Routledge.

- Mackenzie, J. (2021). Analysing gendered discourses online: child-centric motherhood and individuality Mumsnet Talk. En J. Angouri, & J. Baxter (Eds.), *The Routledge Handbook of Language, Gender, and Sexuality* (pp. 694-716). Routledge.
- McEnery, T., & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press.
- McEnery, T., & Wilson, A. (2001). *Corpus Linguistics: An Introduction*. Edinburgh University Press.
- Namatende-Sakwa, L. (2019). Gendering the text through implicit citations of gendered discourses. *Gender and Language*, 13(1) 72-93. <https://doi.org/10.1558/genl.34847>
- Parodi, G. (2008). Lingüística de corpus: una introducción al ámbito. *Revista de Lingüística Teórica y Aplicada*, 46(1). <https://doi.org/10.4067/s0718-48832008000100006>
- Pérez-Sabater, C. (2019). Emoticons in Relational Writing Practices on WhatsApp: Some Reflections on Gender. En P. Bou-Franch, & P. Garcés-Conejos Blitvich (Eds.), *Analyzing Digital Discourse New Insights and Future Directions* (pp. 173-199). Palgrave Macmillan.
- Perdomo Colina, H. E. (2021). Los discursos sobre la construcción de la feminidad durante el periodo de inicio y consolidación del Concurso Nacional de Belleza en Colombia: 1934-1955. *Cuestiones de Género: de la igualdad y la diferencia*, 16, 610-632. <https://doi.org/10.18002/cg.v0i16.6947>
- Reppen, R. (2022). Building a corpus: what are key considerations? En A. O'Keeffe, & M. J. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 71-86). Routledge.
- Rodrigues, D. L., Lopes, D., Prada, M., Thompson, D., & Garrido, M. (2017). A frown emoji can be worth a thousand words: Perceptions of emoji use in text messages exchanged between romantic partners. *Telematics and Informatics*, 34(8), 1532-1543. <https://doi.org/10.1016/j.tele.2017.07.001>
- Rogers, R. (2014). Debanalising Twitter. The Transformation of an Object of Study. En K. Weller, A. Bruns, J. Burgess, M. Mahrt, & C. Puschmann (Eds.), *Twitter and Society* (pp. 9-27). Peter Lang.
- Rojo, G. (2021). *Introducción a la lingüística de corpus en español*. Routledge.
- Rychlý, P. (2008). A Lexicographer-Friendly association score. *RASLAN* (pp. 6-9). Masarykova Univerzita. <https://nlp.fi.muni.cz/raslan/2008/papers/13.pdf>

- Sauntson, H. (2021). Applying queer theory to language, gender, and sexuality research in schools. En J. Angouri, & J. Baxter (Eds.), *The Routledge Handbook of Language, Gender, and Sexuality* (pp. 338-353). Routledge.
- Sinclair, J. McH. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.
- Sunderland, J. (2004). *Gendered Discourses*. Springer.
- Sunderland, J. (2015). Researching Language and Gender. En B. Paltridge, & A. Phakiti (Eds.), *Research Methods in Applied Linguistics* (pp. 505-525). Routledge.
- Talbot, M. (2019). *Language and Gender* (3ra ed.). Polity Press.
- Tannen, D. (1994). *Gender and Discourse*. Oxford University Press.
- Toffoletti, K., & Starr, K. (2016). Women Academics and Work-Life Balance: Gendered Discourses of Work and Care. *Gender, Work & Organization*, 23(5), 489-504. <https://doi.org/10.1111/gwao.12133>
- Trudgill, P. (1984). Sex and covert prestige. En P. Trudgill (Ed.), *On Dialect* (pp. 169-177). Blackwell.
- Vásquez, C., & Sayers China, A. (2019). My Manly Husband... to... Sitting Down to Take a Pee: The Construction and Deconstruction of Gender in Amazon Reviews. En P. Bou-Franch, & P. Garcés Conejos (Eds.), *Blitvich, Analyzing Digital Discourse New Insights and Future Directions* (pp. 193-218). Palgrave Macmillan.
- Weller, K. (2014). Twitter and Society. *Peter Lang GmbH*. Verlag.