

La integración del análisis de sentimientos en la investigación sociolingüística: dimensiones afectivas del discurso espontáneo

A integração da análise de sentimentos na pesquisa sociolinguística: dimensões afetivas do discurso espontâneo

Renata ENGHELS 

Ghent University
Ghent, Belgium
renata.enghels@ugent.be

Resumen: Este estudio explora la integración del análisis de sentimientos (AS) en la investigación sociolingüística, con el objetivo de examinar cómo las dimensiones afectivas se manifiestan en el habla coloquial. A partir de una revisión teórica de los enfoques cognitivos y funcionales del significado, se plantea la posibilidad de cuantificar aspectos relacionales y evaluativos del lenguaje mediante herramientas de AS, en particular el software Lingmotif. Se aplicó esta herramienta a dos corpus orales del español de Madrid (CORMA y COLAm), permitiendo analizar el tono emocional de las interacciones en función de variables sociolingüísticas como género, edad y contexto comunicativo. Los resultados muestran que las conversaciones en contextos institucionales (como la atención al cliente) presentan perfiles afectivos más positivos e intensos, mientras que el habla juvenil se caracteriza por un uso más frecuente de expresiones irónicas, lenguaje tabú y estrategias de *banter*. Además, se observa una evolución microdiacrónica: las diferencias de género en la expresión emocional tienden a atenuarse con el tiempo. El estudio concluye que el AS, pese a sus limitaciones, constituye una herramienta valiosa para detectar patrones afectivos en el habla espontánea y para ampliar las metodologías de análisis del significado social en la sociolingüística contemporánea.

Palabras Clave: análisis de sentimientos; variación social; discurso espontáneo; Corpus Oral de Madrid.

Resumo: Este estudo investiga a integração da análise de sentimentos (AS) na pesquisa sociolinguística, com o objetivo de examinar como as dimensões afetivas se manifestam na fala coloquial. A partir de uma revisão teórica dos enfoques cognitivos e funcionais do significado, propõe-se a possibilidade de quantificar

aspectos relacionais e avaliativos da linguagem por meio de ferramentas de AS, em particular o software Lingmotif. Essa ferramenta foi aplicada a dois corpora orais do espanhol falado em Madri (CORMA e COLAm), permitindo analisar o tom emocional das interações com base em variáveis sociolinguísticas como gênero, idade e contexto comunicativo. Os resultados mostram que as conversas em contextos institucionais (como atendimento ao cliente) apresentam perfis afetivos mais positivos e intensos, enquanto a fala juvenil se caracteriza pelo uso frequente de expressões irônicas, linguagem tabu e estratégias de *banter*. Além disso, observa-se uma evolução microdiacrônica: as diferenças de gênero na expressão emocional tendem a se atenuar com o tempo. Conclui-se que a AS, apesar de suas limitações, é uma ferramenta promissora para detectar padrões afetivos na fala espontânea e para ampliar as metodologias de análise do significado social na sociolinguística contemporânea.

Palavras-chave: análise de sentimentos; variação social; discurso espontâneo; Corpus Oral de Madrid.

1 INTRODUCCIÓN

Desde hace varias décadas, la investigación sociolingüística ha mostrado un interés sostenido por la relación entre categorías sociales — como género, edad o clase social — y las formas en que los hablantes configuran sus estilos comunicativos y rutinas interaccionales (Gumperz, 1982; Eckert, 2000; Holmes, 1995; Tannen, 1984; Fairclough, 1992, entre otros). Estos trabajos han puesto de relieve que el significado en la interacción no se limita a la transmisión de contenido proposicional, sino que incluye dimensiones relacionales, identitarias y afectivas, construidas de manera dinámica en el discurso. Así, las elecciones lingüísticas se convierten en recursos para negociar relaciones sociales, proyectar identidades (cf. el concepto de la ‘indexicalidad social’; Silverstein, 2003 entre otros) y expresar actitudes, en estrecha conexión con el contexto y las normas socioculturales que enmarcan la comunicación. Partiendo de esta tradición, este artículo plantea contribuir a la cuestión central de saber qué aspectos contextuales intervienen cuando hablamos de significado en la lengua. El significado no es una propiedad única ni fija de las palabras o de las oraciones, sino una realidad multidimensional. Desde enfoques cognitivos y funcionales, se entiende como algo que construimos e interpretamos a través del uso lingüístico. La perspectiva cognitiva subraya que el significado está anclado en la experiencia y la cognición (Langacker, 1991), mientras que la funcional,

como el modelo de Halliday y Matthiessen (2004), destaca los distintos papeles que desempeña el significado: expresar ideas, gestionar relaciones sociales y estructurar el mensaje en contexto. Estos enfoques integrados abren paso a una visión dinámica y estratificada del significado que incorpora dimensiones conceptuales, contextuales, relacionales y afectivas.

Sin embargo, surge una segunda pregunta crucial: si el significado es tan abstracto, dependiente del contexto y a menudo no perceptible de manera directa, ¿cómo podemos capturarlo? Y, más aún, ¿cómo analizarlo de forma sistemática y, en la medida de lo posible, cuantitativa? En los últimos años, la semántica y la pragmática han experimentado lo que se ha denominado un 'giro empírico y cuantitativo'. Esto implica un desplazamiento de los métodos puramente introspectivos hacia el uso de datos lingüísticos reales — principalmente de corpus — y el desarrollo de procedimientos de análisis más sistemáticos. Como señalan Geeraerts (2010) y Glynn (2010), este cambio no convierte el significado en un objeto totalmente objetivo, pero sí permite definirlo de forma más fundamentada y transparente.

En este marco, el análisis de sentimientos emerge como una herramienta complementaria prometedora. Se trata de una técnica de análisis computacional del texto que detecta y cuantifica el tono emocional o la actitud evaluativa expresada en el lenguaje, transformando aspectos subjetivos del significado en datos medibles. Permite visualizar la polaridad (positiva o negativa) y la intensidad (grado de fuerza) de la expresión, abriendo nuevas pistas para explorar el papel de la emoción y la postura en el discurso. A partir de estas premisas, este artículo se propone responder a las siguientes preguntas de investigación:

- a) ¿Puede integrarse el análisis de sentimientos en la investigación sociolingüística para revelar patrones afectivos, significado social y variación estilística en el habla espontánea?
- b) ¿Cómo se manifiestan estas dimensiones a través de variables como el género, la edad y el contexto discursivo?

Para ello, el artículo se estructura de la siguiente manera: en primer lugar, se revisan los fundamentos conceptuales y metodológicos del análisis de sentimientos, así como su potencial para el estudio del significado social. A continuación, se presenta un análisis aplicado a dos corpus de habla

coloquial en español — el *Corpus del español hablado en Madrid* (CORMA) y el *Corpus Oral de Lenguaje Adolescente*, Madrid (COLAm) — que incluye una perspectiva microdiacrónica. Finalmente, se ofrecen reflexiones sobre los desafíos, implicaciones y posibles soluciones derivadas de la integración de este enfoque en la investigación sociolingüística.

2 BASES CONCEPTUALES Y TÉCNICAS DEL ANÁLISIS DE SENTIMIENTOS

Antes de poder aplicar el análisis de sentimientos (de aquí en adelante, AS) a preguntas lingüísticas, es fundamental precisar qué es, qué tipo de significado captura y cómo funciona. El objetivo de este apartado no es ofrecer una descripción técnica exhaustiva — pues la perspectiva de este artículo es la de una lingüista y no la de una especialista en computación —, sino proporcionar una visión general de sus principios y enfoques, así como de las razones por las que puede resultar útil en sociolingüística. El AS, también conocido como opinion mining o emotion recognition, es un área de investigación en rápido crecimiento dentro del Procesamiento del Lenguaje Natural (PLN, de Natural Language Processing, NLP). Su finalidad principal es la detección y la clasificación automáticas de elementos subjetivos del lenguaje, como opiniones, actitudes o emociones (Pang; Lee, 2008). Moreno-Ortiz lo define como el reconocimiento de “datos subjetivos, tales como juicios, opiniones y sentimientos hacia personas, objetos y sus características” (Moreno-Ortiz, 2024, p. 141-142). En las últimas décadas, el AS ha despertado un interés creciente tanto académico como comercial, intensificado por el desarrollo de la IA generativa y de modelos de aprendizaje profundo (deep learning). Se ha aplicado en ámbitos tan diversos como el seguimiento de marcas (brand monitoring), el análisis de reseñas de productos, el estudio del discurso político o la monitorización de redes sociales. Más recientemente, también ha comenzado a explorarse en terrenos más cualitativos, como los estudios literarios (p.ej., Dejaeghere et al., 2024).

En términos metodológicos, se distinguen dos enfoques principales. Primero, el enfoque basado en *Machine Learning*, que se entrena a partir de grandes conjuntos de datos anotados, utilizando modelos estadísticos de clasificación para determinar la orientación sentimental de un texto. En la actualidad, este enfoque incluye arquitecturas neuronales avanzadas (o *Transformers*) como BERT (Devlin et al., 2019) o GPT de OpenAI (Brown et al.,

2020), que han mejorado notablemente la precisión en tareas de clasificación de sentimientos. Sin embargo, estos sistemas funcionan a menudo como 'cajas negras' (Moreno-Ortiz, 2023, p. 18, p. 48): aunque su rendimiento sea muy alto, ofrecen poca transparencia sobre las pistas lingüísticas que motivan sus clasificaciones, lo que limita su aplicabilidad para la investigación lingüística.

Segundo, el enfoque basado en léxicos (*lexicon-based approach*) calcula una puntuación de sentimientos a partir de (a) listas predefinidas de palabras con carga evaluativa, y (b) reglas contextuales que pueden modificar su perfil semántico según el uso real. Dado que este proceso implica codificación experta por parte de lingüistas — por ejemplo, al elaborar un léxico de sentimientos a partir de diccionarios, léxicos computacionales o corpus —, este método permite un seguimiento más explícito del lenguaje evaluativo en contexto. Aunque tiende a obtener resultados menos precisos que los modelos de *machine learning* en tareas de referencia (*benchmark*), sobre todo ante usos figurados o irónicos, su transparencia lo convierte en una opción especialmente atractiva para aplicaciones lingüísticas. En concreto, permite identificar la clasificación y polaridad exactas de cada ítem léxico que contribuye al resultado global del análisis. Además, al no requerir infraestructuras complejas ni conocimientos técnicos avanzados, resulta más accesible para investigadores sin formación en PLN, lo que lo hace muy adecuado para estudios exploratorios como el presente. En suma, aunque el análisis de sentimientos presenta desafíos, este trabajo parte de la hipótesis de que puede constituir una pista metodológica prometedora para explorar de forma sistemática las dimensiones evaluativas y afectivas del discurso, en particular en el estudio de estilos comunicativos diferentes según varias dimensiones sociales.

Para el análisis empírico que presentamos en este artículo hemos optado por emplear una herramienta de análisis de sentimientos basada en léxico que se ajusta particularmente bien a los objetivos de nuestro estudio, a saber Lingmotif. Este software, desarrollado en la Universidad de Málaga bajo la dirección de Antonio Moreno-Ortiz (Moreno-Ortiz, 2017, 2023, 2024), ha sido concebido para investigadores que desean analizar el componente evaluativo de textos sin necesidad de conocimientos avanzados de PLN. Para los fines de esta investigación, agradecemos la colaboración del Prof. Moreno-Ortiz, quien nos facilitó el análisis detallado de sentimientos de

todas las conversaciones del corpus CORMA, que constituye la base de nuestro examen de las dimensiones afectivas en el discurso espontáneo.

Lingmotif es una herramienta de AS basada en léxico que identifica automáticamente en un texto las palabras y expresiones con carga evaluativa y aplica reglas contextuales para ajustar su interpretación en función de elementos como la negación o los intensificadores. El sistema se apoya en un conjunto de elementos léxicos núcleo que incluyen palabras de polaridad frecuente — adjetivos, verbos, sustantivos y adverbios —, así como en reglas diseñadas para detectar *sentiment shifters*, es decir, elementos que modifican el valor afectivo de un término. Una de las características más útiles para la investigación del discurso es que Lingmotif descompone el texto en oraciones, localiza las unidades portadoras de sentimiento y les asigna un índice de valencia o polaridad. A partir de este procesamiento, el software genera dos puntuaciones principales: el *Text Sentiment Score* (TSS), que refleja la polaridad global del texto en un continuo que va de negativo a positivo, y el *Text Sentiment Intensity* (TSI), que indica la densidad o concentración de sentimiento expresada. El valor global de un texto se calcula a partir de la suma acumulada de puntuaciones positivas, negativas y neutras, lo que permite obtener un perfil de sentimiento detallado.

En términos interpretativos, el TSS se expresa en una escala normalizada de 0 a 100, en la que el valor 50 funciona como punto de referencia para una orientación global neutra o predominantemente factual. Valores ligeramente superiores o inferiores a 50 (aproximadamente entre 45 y 55) pueden interpretarse como cercanos a la neutralidad, mientras que valores por encima de este intervalo indican una orientación progresivamente positiva y valores por debajo una orientación progresivamente negativa. Por su parte, el TSI también se sitúa en una escala de 0 a 100 y cuantifica la fuerza evaluativa del discurso con independencia de su polaridad. De manera orientativa, valores por debajo de 40 se asocian a textos con baja carga afectiva y un tono mayoritariamente descriptivo o narrativo; valores intermedios (en torno a 40–60) reflejan una implicación evaluativa moderada; y valores elevados corresponden a discursos con una expresión emocional más intensa y explícita.

Por su transparencia, facilidad de uso y adaptabilidad a distintos dominios discursivos, Lingmotif resulta especialmente adecuado para

estudios sociolingüísticos que requieren una interpretación contextualizada de las dimensiones afectivas del discurso.

3 EL ANÁLISIS DE SENTIMIENTOS COMO HERRAMIENTA PARA EL ESTUDIO DEL SIGNIFICADO SOCIAL

Como ya señalaba Mills (2012), existe una relación estrecha entre la identidad de un hablante y su uso lingüístico. Esta identidad no es una característica estática, sino que se desarrolla durante la interacción, lo que da lugar a sociolectos o estilos comunicativos específicos. Los signos lingüísticos adquieren un significado social en el seno de una comunidad determinada, moldeados por su contexto social y cultural, y la contextualización de estos signos en el uso se considera parte integrante de su significado. Durante la conversación, diversos aspectos de la persona social — género, clase, edad, etnicidad — se ‘ponen en escena’ (*staged*), contribuyendo a procesos de estilización discursiva (Gumperz, 1982; Eckert, 2000; Holmes, 1995; Tannen, 1984; Fairclough, 1992). Esta perspectiva conecta con la noción de *acomodación* entre miembros de una comunidad, que puede favorecer la inclusión de miembros del *in-group* y, en algunos casos, la exclusión de hablantes del *out-group*. Tales dinámicas se inscriben en la noción de ‘indexicalidad social’ (Blommaert, 2005; Coupland, 2007; Silverstein, 2003, 2009), que más recientemente Ghezzi (2024) ha reformulado comparando el estilo comunicativo con el estilo general de un grupo social, que incluye elementos visibles como la vestimenta o el peinado.

Esta tradición de estudios ha permitido establecer ciertas afirmaciones generalmente aceptadas sobre la relación entre estilo discursivo y categorías sociales como género, edad o contexto comunicativo. En lo que respecta al género, investigaciones previas han mostrado que las mujeres tienden a emplear más estrategias de atenuación (*hedging*) y disculpa, asociadas a menudo con empatía y cercanía (Holmes, 1995; Bucholtz, 1999). En términos de expresividad, los hombres suelen preferir estilos más directos y de alta activación emocional (*high-arousal*), mientras que las mujeres tienden hacia formas más moderadas de afecto positivo. Estas tendencias pueden influir en cómo se expresa y se percibe el sentimiento en distintos grupos de hablantes (Mulac *et al.*, 2001).

La edad también desempeña un papel relevante en la expresión del significado afectivo. Los hablantes jóvenes suelen usar un estilo más informal, a menudo cargado de marcadores de postura afectiva (*affective stance markers*) que transmiten solidaridad o entusiasmo (Tagliamonte, 2016). Por el contrario, los hablantes de mayor edad tienden a recurrir a rutinas más formales que transmiten respeto y deferencia (Coupland; Coupland, 1993). Estas diferencias generacionales reflejan cambios más generales en las normas y expectativas comunicativas.

El contexto discursivo es otro factor clave. En conversaciones informales entre pares suelen observarse turnos de palabra rápidos, expresiones emocionales e intensificadores que señalan intimidad o entusiasmo. En cambio, los entornos formales e institucionales tienden a seguir rutinas más estructuradas de pregunta-respuesta, y posturas afectivas marcadas por la cortesía y el respeto (Tannen, 1984; Gumperz, 1982).

El análisis de sentimientos, aplicado con cautela, puede resultar útil para explorar estas dimensiones del significado social, ya que permite identificar y cuantificar patrones afectivos asociados a estilos comunicativos, relaciones interpersonales y contextos discursivos, aportando una perspectiva empírica que complementa los enfoques cualitativos tradicionales. Efectivamente, en los enfoques sociolingüísticos tradicionales sobre el estilo y el discurso, pueden identificarse varias limitaciones que el AS podría contribuir a superar. En primer lugar, existe el desafío de la anotación manual: el análisis detallado de rutinas afectivas suele ser muy laborioso y, por ello, se restringe a conjuntos de datos reducidos. Las herramientas computacionales de análisis de sentimientos permiten procesar automáticamente grandes volúmenes de datos, lo que facilita la detección de patrones a mayor escala. En segundo lugar, la comparabilidad entre estudios es limitada. Comunidades distintas pueden expresar una misma postura mediante recursos muy diferentes. Al revés, el AS introduce métricas que posibilitan la comparación del estilo afectivo en distintos contextos. Un tercer aspecto problemático es que muchas etiquetas estilísticas suelen ser binarias (p. ej., formal vs. informal, cortés vs. descortés). La puntuación continua que ofrece el AS aporta mayor gradación y permite captar matices sutiles en el tono y la postura. Asimismo, las rutinas interaccionales se han tratado a menudo como tipos estáticos, cuando en realidad el estilo es un proceso dinámico que se desarrolla a lo largo de una

conversación. El seguimiento secuencial de sentimientos a lo largo de una interacción posibilita observar cómo varía la carga afectiva durante el transcurso del intercambio. Finalmente, los métodos más tradicionales pueden pasar por alto afectos implícitos o no léxicos. Si bien este estudio se centra en un enfoque léxico, cabe señalar que los modelos más recientes basados en Transformers (cf. *supra*) son capaces de detectar tonos emocionales incrustados en estructuras complejas o patrones discursivos más sutiles.

En suma, el AS no sustituye el trabajo interpretativo detallado de la sociolingüística, pero sí amplía el repertorio metodológico disponible. En las secciones que siguen, aplicaremos Lingmotif al corpus CORMA con el fin de examinar patrones afectivos en el discurso espontáneo y, posteriormente, realizaremos un análisis comparativo de carácter microdiacrónico entre CORMA y el corpus COLAm. Este planteamiento permitirá explorar cómo se distribuyen y varían las dimensiones evaluativas en distintos contextos y comunidades de habla, ofreciendo así una perspectiva empírica sobre la integración del análisis de sentimientos en la investigación sociolingüística.

4 ANÁLISIS DEL DISCURSO AFECTIVO EN CORPUS ORALES

4.1 Análisis afectivo en el corpus CORMA

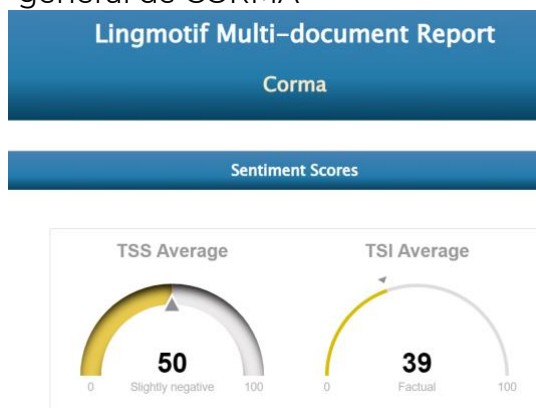
Para llevar a cabo el análisis de los perfiles de sentimiento, emplearemos el CORMA (*Corpus Oral del español hablado en Madrid*), un recurso desarrollado por nuestro equipo de investigación en la Universidad de Gante y colegas de la UNED, a partir de un trabajo de campo realizado en Madrid entre 2016 y 2019 (véase Enghels *et al.*, 2020, así como el portal: corma.ugent.be). Se trata de un corpus directo, obtenido en situaciones comunicativas reales, que documenta interacciones conversacionales auténticas en la capital española durante la segunda década del siglo XXI. El corpus reúne unas cien conversaciones informales que reflejan una amplia gama de situaciones comunicativas, desde intercambios entre amigos, familiares o conocidos hasta interacciones en contextos de atención al cliente. La amplitud de contextos representados confiere al corpus una gran variabilidad situacional y sociolingüística, lo que garantiza su representatividad del español coloquial espontáneo. En términos cuantitativos, CORMA contiene aproximadamente 58 horas de grabaciones de audio, equivalentes a unas 460.000 palabras producidas por 485

hablantes. La recopilación de datos se diseñó para lograr un equilibrio razonable entre distintas categorías sociolingüísticas, especialmente edad, género y clase social. En lo que respecta a la edad, se establecieron cuatro grupos generacionales: generación 1 (0-11 años), generación 2 (12-25 años), generación 3 (26-55 años) y generación 4 (más de 55 años). Aunque la generación 2 presenta la mayor proporción de participantes, las generaciones 3 y 4 cuentan también con una representación significativa. En cuanto al género, las mujeres están sobrerrepresentadas, lo que refleja su mayor participación en todos los grupos de edad. La variable de clase social, de difícil definición precisa, como mencionado antes, se aproximó mediante el estatus socioeconómico de los barrios de residencia de los participantes, diferenciando dos grandes categorías: media-alta/alta y baja¹.

Una de las principales ventajas de trabajar con este corpus de elaboración propia es el control absoluto sobre los procesos de recogida y tratamiento de los datos, lo que no solo asegura su calidad, sino que también facilita la incorporación de metodologías complementarias y el uso de herramientas específicas, como el software Lingmotif, en etapas posteriores del análisis. La Figura 1 presenta el perfil global de sentimientos del corpus CORMA, elaborado a partir de los informes generados por Lingmotif para cada una de las conversaciones. En promedio, el *Text Sentiment Score* del conjunto se sitúa en torno a 50, lo que indica una orientación global neutra o ligeramente negativa (ver sección 2). El *Text Sentiment Intensity* medio es de 39, valor asociado a un tono predominantemente factual, característico de un discurso anclado en experiencias reales, narraciones y descripciones concretas.

¹ La clase social, aunque constituye una variable clave en numerosos estudios sociolingüísticos, plantea unas dificultades metodológicas cuando se trata de operacionalizarla en análisis basados en corpus orales espontáneos. Su definición es compleja y multidimensional, e implica la consideración de factores como nivel educativo, ocupación, ingresos, redes sociales o nivel cultural, cuya medición precisa no siempre es viable. Además, en muchos corpus no se dispone de datos fiables y homogéneos para determinar estos indicadores de manera consistente. Por estas razones, la variable de clase social no será considerada en el presente estudio.

Figura 1 – Perfil de sentimientos general de CORMA



Fuente: datos extraídos de Enghels (2025, p. 16).

Más allá de este perfil general, la Figura 2 pone de manifiesto la variación que se observa entre conversaciones concretas y en diferentes contextos discursivos. Por ejemplo, las interacciones registradas en una tienda de ropa (véase el identificador a la izquierda: atención al cliente, ropa – atropa) presentan puntuaciones de sentimiento muy positivas (*very positive*) y se clasifican como de alta intensidad (*very intense*). En cambio, las conversaciones entre amigas adolescentes (*ir_am2_f_03/04/05*) tienden a mostrar un tono más negativo. Las charlas entre amigas de mayor edad (*am_gen4_f.01b*) suelen exhibir una orientación más positiva, pero con menor intensidad, lo que refleja un estilo más factual y mesurado.

Figura 2 – Perfil de sentimientos de diferentes contextos comunicativos

ID	TSS CAT2	TSS CAT3	TSS CAT9	TSI CAT
atropa.01a	POS	POS	Very positive	Extremely intense
atropa.01b	POS	POS	Very positive	Very intense
atropa.02_1	POS	POS	Fairly positive	Fairly intense
ir_am2_f_03	NEG	NEG	Very negative	Factual
ir_am2_f_04	NEG	NEG	Fairly negative	Factual
ir_am2_f_05	NEG	NEG	Fairly negative	Factual
am_gen4.f.01b_1	POS	POS	Slightly positive	Factual
am_gen4.f.01b_2	POS	POS	Fairly positive	Factual
am_gen4.f.01b_3	POS	POS	Fairly positive	Factual

Fuente: datos extraídos de Enghels (2025 p. 17).

Estas alternancias, tanto de naturaleza sociolingüística como discursiva, evidencian que el perfil de sentimientos no es uniforme, sino que varía según el contexto comunicativo y las características de los hablantes. En la sección siguiente, agruparemos las conversaciones que muestren perfiles consistentes en estas dimensiones, con el fin de examinar con mayor detalle los patrones emergentes.

Comenzaremos analizando los patrones de variación en función del contexto discursivo. Tal como se observa en la Tabla 1, las interacciones en entornos de atención al cliente obtienen las puntuaciones más altas tanto en el índice de sentimiento (TSS) como en la intensidad (TSI). Estos datos están en línea con la idea general de que este tipo de conversaciones suele caracterizarse por un tono predominantemente positivo, acompañado de una expresividad elevada y fórmulas de cortesía, probablemente con el objetivo de mantener un clima comunicativo agradable. Desde una perspectiva cuantitativa, estos resultados se traducen en valores de TSS claramente situados por encima del intervalo de neutralidad ($\approx 45-55$), como en el caso de la atención al cliente (TSS = 61,20), lo que indica una orientación positiva marcada. Asimismo, el TSI medio de 43,63 apunta a una intensidad evaluativa moderada-alta, coherente con la presencia recurrente de estrategias expresivas y de cortesía.

Tabla 1 – AS según el contexto discursivo de la conversación

	amigos	familia	atención al cliente
TSS (promedio)	42,38	52,9	61,20
TSI (promedio)	39,45	38,1	43,63

Fuente: creada por la autora, 20252.

En cambio, las conversaciones entre amigos tienden a situarse en valores más neutros o incluso ligeramente bajos en promedio. Estas conversaciones presentan un TSS medio de 42,38, valor situado por debajo del umbral de neutralidad y compatible con una orientación global

² Dado que los valores presentados corresponden a promedios agregados por contexto discursivo, el análisis se limita en este punto a una comparación descriptiva de tendencias generales, sin aplicar pruebas inferenciales. La aplicación de inferencias estadísticas basadas en datos agregados podría ocultar la variabilidad interna de los grupos y conducir a interpretaciones poco robustas.

ligeramente negativa o no evaluativa, mientras que su TSI medio (39,45) se mantiene en un rango bajo, asociado a un tono predominantemente factual.

Este patrón podría explicarse por la mayor relajación de las normas conversacionales en contextos de confianza, donde la expresión emocional puede incluir recursos como la ironía, las quejas o las bromas. Cabe subrayar, no obstante, que estos mismos recursos pragmáticos pueden adquirir valores sociopragmáticos opuestos según el tipo de relación entre los interlocutores: mientras que, en interacciones entre desconocidos, la ironía puede indexar distancia interpersonal o actitudes negativas, en contextos de familiaridad y complicidad puede funcionar precisamente como marcador de cercanía y alineamiento. Este hecho pone de relieve una limitación central de los enfoques léxicos de análisis de sentimientos, que tienden a neutralizar o reinterpretar negativamente fenómenos cuyo valor afectivo es altamente dependiente del contexto interactivo. Así, la comparación entre ambos tipos de interacción ilustra no solo cómo el entorno comunicativo influye en la orientación y la intensidad del sentimiento expresado, sino también la necesidad de interpretar los resultados cuantitativos a la luz de las prácticas discursivas específicas de cada contexto, cuestión que se aborda con mayor detalle en la sección 5.

Consideremos los siguientes fragmentos. En el ejemplo (1), los datos extraídos de una conversación (por teléfono) en el contexto de una tienda de ropa muestran una elevada densidad de expresiones positivas y afectuosas, como *muy bien*, *gracias*, *un besito* o *hola cariño*, que contribuyen a un perfil de sentimiento alto tanto en polaridad como en intensidad. Este patrón es coherente con la naturaleza de la interacción de tipo atención al cliente, donde la cortesía y la expresividad positiva desempeñan un papel importante para mantener un clima comunicativo agradable.

(1)

ROPAj 3F1: Vale.

ROPAj 3F1: Muy bien, muy bien, gracias Virginia.

ROPAj 3F1: Venga a ti un besito, hasta mañana.

ROPAj 3F1: Bueno, otra.

ROPAj 3F1: Pues es una pena, porque hemos tenido un ratito aquí, y además había—Como cuatro o cinco chicas jóvenes probando ropa

ROPAj 3F1: Hola cariño, sin problemas. Es que ya me han traído tu vestido para que te lo pruebes.

ROPAj 3F1: Así que nada, cuando tengas un rato, que ya sabes qué prisa ya sabes que no hay.

ROPAj 3F1: Venga, un besito.

ROPAj 3F1: Adiós Mari Carmen, adiós.

ROPAj 3F1: Sí, y además las de un probador intercambiaban opiniones con el otro.

ROPAj 3F1: O mejor un alfiler, iban intercambiando–

ROPAj 3F1: Qué pena.

En contraste, las conversaciones entre amigos, tal como ilustran los fragmentos (2a-b), presentan un perfil menos positivo en los valores de Lingmotif, debido a la presencia de quejas típicas de las charlas informales (2a) (p.ej., *me parece muy caro*) o al uso recurrente de palabras tabú (p.ej., *cojones, jodida*) (2b). Sin embargo, en este segundo caso, el contenido no responde necesariamente a una intención negativa, sino que forma parte de una estrategia de *banter* o juego verbal propio de contextos de confianza, como ocurre en bromas entre amigos. Este tipo de recursos, aunque pueden ser evaluados como negativos en un análisis léxico automático, cumplen una función lúdica y relacional que matiza su interpretación.

(2a)

AM3F1: Pues no–, además que no deja de ser, las las sueltas mes a mes, treinta euros al mes.

AM3F2: A mí me parece muy caro.

AM3F1: Porque si incluyes los–, y y cuarenta de la matrícula, aquí pagas matrícula.

AM3F1hi: Una mariquita.

AM3F1: Pagas siempre matrícula eh, veinte euros aquí, veinte euros aquí, aquí cuarenta, pagas cuarenta y luego treinta al mes, para– para yy que me sabes, qué barato es, yo no lo veo– AM3F1hi: Una mariquita.

AM3F2h: Tráela a ver.

AM3F1hi: Una mari– AM3F1hi: Quique, hemos visto una mariquita.

(2b)

AMF1: Que me apuntas a mi, que cojones voy a saber yo.
 AM4M9: Pues la más gorda que tengo.
 AM4M6: Esa sí que es jodida, dos cuatro, vamos a condenar una que yo sé.

AM4M6: Yo me lo he condena'o aposta pero porque sabía lo que hacía.

AMM10: Pue' tú pierde' veintiocho con aquella, treinta y nueve y cuarenta y cuatro, cinco pa ti.

Estos ejemplos ilustran la importancia de contextualizar los resultados del análisis de sentimientos y de considerar las normas discursivas y relacionales propias de cada contexto, para evitar lecturas simplificadas de los valores de polaridad e intensidad.

Pasamos a la variación por edad. Los datos presentados en la Tabla 2 muestran diferencias sistemáticas entre generaciones en términos tanto de orientación como de intensidad del sentimiento. Desde el punto de vista del TSS, la Generación 3 (adultos) presenta valores claramente por encima del intervalo de neutralidad (TSS = 55,33), lo que permite caracterizar su estilo comunicativo como globalmente positivo. Esta orientación se ve reforzada por un TSI elevado (TSI = 48,33), indicativo de una mayor implicación afectiva y expresividad. En el extremo opuesto, la Generación 2 obtiene el TSS más bajo del conjunto (TSS = 40), valor situado por debajo del umbral de neutralidad y compatible con una orientación global neutra-baja o ligeramente negativa. No obstante, este grupo presenta un TSI intermedio (38,21), lo que sugiere que la menor positividad no se traduce necesariamente en una menor expresividad, sino más bien en un uso de recursos evaluativos con polaridad menos claramente positiva. Las Generaciones 1 y 4 se sitúan en posiciones intermedias.

Tabla 2 – AS según la edad del hablante

	GEN1	GEN2	GEN3	GEN4
TSS (promedio)	49,75	40	55,33	47,14
TSI (promedio)	25,75	38,21	48,33	44,12

Fuente: creada por la autora, 2025.

No obstante, esta aparente negatividad en la Generación 2 debe matizarse: gran parte de su producción incluye el uso frecuente de léxico tabú y estrategias de *banter* que, como comentamos antes, aunque puntuadas negativamente por un análisis léxico automático, no implican necesariamente una carga emocional negativa en el sentido pragmático. En general, la mayoría de los grupos generacionales tienden a un perfil más bien factual. La mayor intensidad (TSI) observada en la Generación 3 se explica, en parte, por la presencia de conversaciones especialmente expresivas en contextos como tiendas de ropa o peluquerías, donde predominan fórmulas de cortesía y manifestaciones de afecto. Estos resultados subrayan que la edad no actúa de forma aislada, sino en interacción con el contexto social y situacional, que determinan conjuntamente el tono emocional de la interacción.

En cuanto a la variación según el género, los resultados presentados en la Tabla 3 muestran diferencias limitadas tanto en la orientación como en la intensidad del sentimiento. Desde el punto de vista del TSS, las conversaciones entre mujeres presentan un valor medio ligeramente superior (TSS = 40,04) al observado en las conversaciones entre hombres (TSS = 37,17). En lo que respecta a la intensidad, el TSI es moderado en ambos casos, aunque ligeramente más elevado en las conversaciones entre hombres (TSI = 39,33) que en las mantenidas entre mujeres (TSI = 36,14). Este patrón indica que, si bien la polaridad media tiende a ser algo más positiva en las interacciones femeninas, las conversaciones masculinas presentan una mayor densidad de recursos evaluativos, sin que ello se traduzca necesariamente en una orientación afectiva más positiva.

Tabla 3 – AS según el género del hablante

	mujeres	hombres
TSS (promedio)	40,04	37,17
TSI (promedio)	36,14	39,33

Fuente: creada por la autora, 2025.

De manera interesante, la idea extendida de que el habla femenina sea más intensa emocionalmente no encuentra respaldo en estos datos: las interacciones entre hombres muestran, de hecho, una intensidad media algo más alta. No obstante, en ambos casos las diferencias son pequeñas y no permiten sostener patrones robustos vinculados al género, lo que sugiere

que las variaciones observadas responden más a factores contextuales y situacionales que a una distinción sistemática entre el habla masculina y femenina.

En la sección siguiente, nos centraremos en el habla juvenil, examinando específicamente las diferencias de género para determinar si su perfil afectivo ha experimentado cambios a lo largo del tiempo. Este análisis permitirá valorar en qué medida las tendencias observadas en la expresión emocional de jóvenes hablantes reproducen patrones previos o, por el contrario, reflejan transformaciones recientes vinculadas a dinámicas socioculturales emergentes.

4.2 Comparación microdiacrónica: análisis de sentimientos del habla juvenil

Si centramos la atención en la Generación 2, el análisis revela una diferencia notable en la orientación global del sentimiento (Tabla 4). En términos de TSS, las chicas alcanzan un valor medio de 54, situado dentro del intervalo de neutralidad-alta y en el umbral de una orientación claramente positiva, mientras que los chicos presentan un TSS notablemente inferior (34), valor claramente por debajo del punto de referencia neutro y compatible con una orientación global negativa. Esta diferencia cuantitativa permite afirmar que el discurso de las adolescentes se caracteriza, en promedio, por un tono significativamente más positivo que el de sus pares masculinos. En cambio, en lo que respecta a la intensidad del sentimiento, los valores de TSI son relativamente próximos (33 en chicos frente a 38 en chicas) y se sitúan en un rango bajo-medio, asociado a una expresividad afectiva moderada. La ausencia de diferencias relevantes en este parámetro sugiere que, si bien la polaridad del discurso difiere de forma clara entre ambos grupos, la fuerza con la que se expresan los contenidos evaluativos es comparable en chicos y chicas.

Tabla 4 – AS de GEN2: comparación de género (CORMA)

	chicos	chicas
TSS (promedio)	34	54
TSI (promedio)	33	38

Fuente: creada por la autora, 2025.

Los dos fragmentos (3a-b) muestran perfiles conversacionales bastante diferentes en función del género en el habla juvenil. En la conversación entre chicos (3a), el intercambio gira en torno a bromas y comentarios provocadores, con referencias a chistes machistas, violencia y conductas delictivas ficticias, acompañadas de un uso abundante de palabras tabú. Este uso lingüístico busca provocar reacciones, reforzar vínculos mediante humor y exhibir una cierta dureza discursiva. El contenido, aunque negativo en apariencia, no necesariamente refleja una intención hostil, sino una estrategia lúdica habitual en la interacción masculina adolescente. En cambio, en el diálogo entre las chicas, el eje temático se centra en la comida, y en concreto en elecciones percibidas como saludables, con evaluaciones positivas marcadas (como *buena, guais*), intensificadores y expresiones de gusto. La conversación tiene un tono colaborativo y orientado al consenso, con mayor uso de adjetivos valorativos positivos.

(3a)

AM2M7: ¿Quién es el mejor portero del mundo?

AM2M7: La mujer, porque es la única capaz de parar dos bolas a la vez.

AM2M3: Tú, chistes machistas no, tío, chistes machistas no.

AM2M7: Pero que no es machista eso.

AM2M3: Que eso es micromachismo

AM2F2: () una calle.

AM2M3: Del chungo y del malo.

AM2M3: Machismo, no.

AM2M4: Del que matas, ¿sabes?

AM2M3: De que lo asesino.

AM2F1: ¿Tú?

AM2M4: () es un violador y un-

AM2F1: La verda' es que normalmente () que te enseñen, ¿sabes?

AM2M3: Un pederasta también ().

AM2F1: El móvil.

AM2M3: () vaya mierda que van a escuchar esta gente.

(3b)

IR2F2: Tiene buena pinta eh.

IR2F1: Sí, ya pero no sé. Es que aquí pedir un chino o algo así no me apetece.

IR2F2: No. Teriyaki.

IR2F1: Buah una ensalada César qué buena.

IR2F2: A lo healthy. A lo healthy.

IR2F1: Las ensaladas de aquí son como mazo de guais.

También resulta de interés examinar el cambio microdiacrónico, es decir, evoluciones lingüísticas sutiles o de corto plazo, analizando cómo interactúan el género y la edad a lo largo de distintos marcos temporales en dos corpus. En este sentido, CORMA puede considerarse representativo de datos más recientes, mientras que el corpus COLAm (*Corpus Oral de Lenguaje Adolescente de Madrid*) recoge interacciones orales grabadas a comienzos del siglo XXI (Stenström *et al.*, 2002). COLAm se compone de conversaciones espontáneas entre adolescentes madrileños en entornos informales y documenta prácticas lingüísticas propias de esa época. Así ofrece un valioso punto de comparación para detectar posibles cambios en las dinámicas conversacionales, los temas abordados y la carga valorativa del discurso juvenil a lo largo de las últimas dos décadas (ver también Enghels *et al.*, 2020). Es una pregunta especialmente relevante porque no solo han evolucionado las etiquetas tradicionales sobre lo que se considera masculino o femenino, sino que las mujeres han ganado mayor empoderamiento y los roles sociales clásicos se han ido debilitando progresivamente. Como resultado, las diferencias entre hombres y mujeres parecen hoy menos marcadas que en el pasado (Acuña Ferreira, 2009; Hancock *et al.*, 2015). En el plano lingüístico, este cambio social puede tener dos consecuencias posibles: (i) que empiecen a observarse rasgos nuevos o distintos característicos del habla femenina juvenil, o (ii) que se produzca una neutralización general de las diferencias entre los estilos discursivos masculinos y femeninos.

La Tabla 5 presenta los valores en el corpus COLAm. Los resultados muestran un contraste muy acusado entre chicos y chicas en ambos parámetros. En términos de orientación del sentimiento, las chicas alcanzan un TSS medio de 60, claramente por encima del intervalo de neutralidad ($\approx 45-55$) y compatible con una orientación global marcadamente positiva. En cambio, los chicos presentan un TSS de 37, situado muy por debajo del punto de referencia neutro. En cuanto a la intensidad, las diferencias son igualmente pronunciadas. El TSI de las chicas (54) se sitúa en un rango alto, asociado a una fuerte implicación afectiva, mientras que el TSI de los chicos (30) corresponde a un nivel bajo de intensidad, característico de un discurso más factual³.

Tabla 5 – AS de GEN2: comparación de género (COLAm)

	chicos	chicas
TSS (promedio)	37	60
TSI (promedio)	30	54

Fuente: creada por la autora, 2025.

Ahora bien, la Tabla 6 incorpora ahora una dimensión temporal al comparar dos periodos, es decir principios de los 2000 (COLAm) y finales de la década de 2010 (CORMA), para los adolescentes. Los resultados muestran un patrón constante en ambos momentos: las chicas tienden a expresarse con mayor positividad que los chicos. Sin embargo, con el paso del tiempo esta diferencia se reduce ligeramente, especialmente en lo que respecta a la intensidad. En concreto, las chicas presentan una disminución notable en la carga emocional de su discurso, mientras que los chicos muestran un ligero incremento. Este patrón sugiere una tendencia hacia un equilibrio mayor en la expresión afectiva, posiblemente vinculada a la transformación de las normas sociales. Así, las chicas podrían estar adoptando estilos más moderados socialmente, mientras que los chicos parecen sentirse cada vez más cómodos expresando emociones. En otras palabras, aunque las diferencias de género se mantienen, la forma de expresar la emoción parece evolucionar hacia modelos más equilibrados y atenuados.

³ Para ambos valores, las diferencias son estadísticamente significativas según las pruebas *t* realizadas ($t = -3.47, p < 0.05$ para TSS; $t = -3.52, p < 0.05$ para TSI).

Tabla 6 – microdiacronía de AS de GEN2: comparación COLAm vs. CORMA

	chicos		chicas	
	COLAm	CORMA	COLAm	CORMA
TSS (promedio)	37	34	60	54
TSI (promedio)	30	33	54	38

Fuente: creada por la autora, 2025.

5 REFLEXIONES FINALES: DESAFÍOS, IMPLICACIONES Y PROPUESTAS

El análisis que precede permite concluir provisionalmente que el análisis de sentimientos ofrece un complemento valioso a los enfoques sociolingüísticos tradicionales para el estudio del significado. Permite realizar comparaciones sistemáticas entre grupos sociales y seguir la evolución de patrones lingüísticos a lo largo del tiempo, como hemos mostrado en la comparación de los corpus CORMA y COLAm. Asimismo, proporciona indicadores cuantificables de las actitudes de los hablantes, un aspecto fundamental para comprender fenómenos pragmáticos y estilísticos en el discurso.

No obstante, para aprovechar plenamente su potencial, es necesario abordar una serie de desafíos en estudios futuros. Uno de ellos es el tratamiento de expresiones propias de comunidades concretas (*in-group*), que pueden tener un significado afectivo muy marcado para los miembros de ese grupo, como el de los adolescentes, pero ser opacas para quienes no forman parte de él. Esto incluye argot escolar, memes o expresiones internas que rara vez aparecen en los recursos léxicos estándar. A ello se suma la naturaleza cambiante del lenguaje juvenil, con una incorporación constante de neologismos y significados adaptados. Términos tabúes (p.ej. *puto* en español) pueden invertir su carga evaluativa (de negativa a neutra o positiva) en pocos años, lo que provoca errores de clasificación si los modelos no se actualizan periódicamente. Otra dificultad importante que hemos comprobado es la presencia de *banter*, ironía y sarcasmo, que pueden invertir el sentido literal de las palabras: expresiones como *la puta profesora* o *joder* en español pueden transmitir desde crítica y frustración hasta complicidad o admiración, dependiendo del contexto. La polisemia, moldeada por las normas internas del grupo, refuerza esta complejidad.

Además, el habla juvenil y coloquial se caracteriza por turnos muy breves o fragmentarios, un ritmo acelerado y la abundancia de muletillas, exclamaciones y vacilaciones. Estos elementos pueden desempeñar un papel clave en la gestión de la interacción y la expresión de actitudes, pero suelen ser ignorados por los sistemas automáticos.

Superar estas limitaciones requiere el desarrollo de metodologías más sensibles al contexto del discurso concreto. Entre las estrategias posibles destacan:

- a) la actualización dinámica de los léxicos: es necesario incorporar de manera continua nuevos términos de argot y expresiones comunitarias mediante técnicas de *slang mining*;
- b) el desarrollo de modelos conscientes del contexto: sería provechoso integrar el historial conversacional y afinar los modelos con datos procedentes de corpus específicos de habla juvenil y comunitaria;
- c) la integración de léxicos especializados por dominios: conviene crear complementos (*plug-ins*) para ámbitos concretos (como anteriormente para el turismo o la economía) y adaptarlos al lenguaje juvenil, de modo que reflejen con precisión sus expresiones y recursos lingüísticos.

Estas mejoras solo serán posibles mediante una colaboración estrecha entre lingüistas computacionales por un lado, y teóricos y sociolingüistas por el otro. Mientras los primeros pueden diseñar modelos adaptativos y léxicos dinámicos, los segundos aportan un conocimiento profundo de las rutinas discursivas, las normas comunitarias y las funciones pragmáticas del estilo. Esta sinergia permitirá construir herramientas de análisis de sentimientos que capten de manera precisa y matizada las complejas y cambiantes formas en que se codifica el significado social en la interacción humana.

REFERENCIAS

ACUÑA FERREIRA, V. Gender and expletives as discourse markers: some uses of *Joder* in young women's interactions in Spanish and Galician. **Feminismo/s**, v. 38, p. 53-83, 2021.

BLOMMAERT, J. **Discourse: A Critical Introduction**. Cambridge: Cambridge University Press, 2005.

BROWN, T. B. *et al.* Language models are few-shot learners. **arXiv**, 2020. Disponível em: <<https://arxiv.org/abs/2005.14165>>. Acesso em: 31 mar. 2025.

BUCHOLTZ, M. Why be normal? Language and identity practices in a community of nerd girls. **Language in Society**, v. 28, n. 2, p. 203-223, 1999.

COUPLAND, N. **Style: Language variation and identity**. Cambridge: Cambridge University Press, 2007.

COUPLAND, N.; COUPLAND, J. Discourses of ageism and anti-ageism. **Journal of Aging Studies**, v. 7, n. 3, p. 279-301, 1993.

DEJAEGHERE, T.; SINGH, P.; LEFEVER, E.; BIRKHOLZ, J. Exploring aspect-based sentiment analysis methodologies for literary-historical research purposes. *In*: SPRUGNOLI, R.; PASSAROTTI, M. **Proceedings of the Third Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA) @ LREC-COLING-2024**, 2024. Torino: ELRA and ICCL. Disponível em: <<https://aclanthology.org/2024.lt4hala-1.16/>>. Acesso em: 31 mar. 2025.

DEVLIN, J.; CHANG, M.; LEE, K. ; TOUTANOVA, K. BERT: pre-training of deep bidirectional transformers for language understanding. *In*: **Proceedings of the 2019 Conference of the North**. Minneapolis: Association for Computational Linguistics, 2019. Disponível em: <<http://aclweb.org/anthology/N19-1423>>. Acesso em: 31 mar. 2025.

ECKERT, P. **Linguistic Variation as Social Practice: The Linguistic Construction of Identity in Belten High**. Oxford: Blackwell, 2000.

ENGHELS, R. Pragmatic markers and sentiment analysis: toward an integrated model of meaning in contemporary colloquial Spanish. **Revista Diadorim**, v. 26, n. 3, e67788, 2024. Disponível em: <<https://doi.org/10.35520/diadorim.2024.v26n3a67788>>. Acesso em: 21 jan. 2026.

ENGHELS, R.; De Latte, F.; Roels, L. El Corpus Oral de Madrid (CORMA): materiales para el estudio (socio)lingüístico del español coloquial actual. **Zeitschrift für Katalanistik**, v. 33, p. 45-76, 2020.

FAIRCLOUGH, N. **Discourse and Social Change**. Cambridge: Polity Press; Blackwell, 1992.

GEERAERTS, D. **Theories of Lexical Semantics**. Oxford: Oxford University Press, 2010.

GHEZZI, C. Social indexicality and pragmatic change in the late 20th century: *Cioè* and *tipo* in Italian radio phone-ins. *In*: PONS BORDERÍA, S.; SALAMEH JIMÉNEZ, S. **Language Change in the 20th Century: Exploring**

Microdiachronic Evolutions in Romance Languages. Amsterdam: John Benjamins, 2024. p. 158-187.

GLYNN, D. Corpus-driven cognitive semantics: quantitative approaches to synonymy. *In*: GLYNN, D.; FISCHER, K. **Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches**. Berlin: Mouton de Gruyter, 2010. p. 77-138.

GUMPERZ, J. **Discourse Strategies**. Cambridge: Cambridge University Press, 1982.

HALLIDAY, M. A. K.; MATTHIESSEN, C. **An Introduction to Functional Grammar**. London: Edward Arnold, 2004.

HANCOCK, A.; WILDER STUTTS, H.; BASS, A. Perceptions of gender and femininity based on language: implications for transgender communication therapy. **Language and Speech**, v. 58, n. 3, p. 315-333, 2015.

HOLMES, J. **Women, Men and Politeness**. Harlow: Addison Wesley Longman, 1995.

LANGACKER, R. W. **Foundations of Cognitive Grammar: Descriptive Application**. Stanford: Stanford University Press, 1991.

MILLS, S. **Gender Matters: Feminist Linguistic Analysis**. London; Bristol: Equinox, 2012.

MORENO-ORTIZ, A. Lingmotif: una herramienta de análisis de sentimiento enfocada en el usuario. **Procesamiento de Lenguaje Natural**, v. 58, p. 133-140, 2017.

MORENO-ORTIZ, A. **Making Sense of Large Social Media Corpora: Keywords, Topics, Sentiment, and Hashtags in the Coronavirus Twitter Corpus**. Cham: Springer Nature Switzerland, 2024.

MORENO-ORTIZ, A. The linguist's role in sentiment analysis: from knowledge provider to data annotator. *In*: MACI, S.; GAROFALO, G. **Investigating Discourse and Text: Corpus-Assisted Analytical Perspectives**. Bern: Peter Lang, 2023. p. 25-54.

MULAC, A. *et al.* Empirical support for the Gender-as-Culture Hypothesis: An intercultural analysis of male/female language differences. **Human Communication Research**, v. 27, p. 121-152, 2001.

PANG, B.; LEE, L. Opinion mining and sentiment analysis. **Foundations and Trends® in Information Retrieval**, v. 2, n. 1-2, p. 1-135, 2008.

SILVERSTEIN, M. Indexical order and the dialectics of sociolinguistic life. **Language and Communication**, v. 23, n. 3-4, p. 193-229, 2003.

SILVERSTEIN, M. Pragmatic indexing. *In*: MEY, J. L.; BROWN, K. **Concise Encyclopaedia of Pragmatics**. Amsterdam: Elsevier, 2009. p. 756-759.

STENSTRÖM, A.-B.; ANDERSEN, G.; HASUND, I. K. **Trends in Teenage Talk: Corpus Compilation, Analysis and Findings**. Amsterdam: John Benjamins, 2002.

TAGLIAMONTE, S. A. **Teen Talk: The Language of Adolescents**. Cambridge: Cambridge University Press, 2016.

TANNEN, D. **Conversational Style: Analyzing Talk Among Friends**. Norwood: Ablex Publishing, 1984.

ENGHELS, Renata. La integración del análisis de sentimientos en la investigación sociolingüística: dimensiones afectivas del discurso espontáneo. **Entrepalavras**, Fortaleza, v. 16, e97204, 2026. DOI: 10.36517/ep16.97204