



Análisis multicapa del habla conversacional conflictiva a partir de la herramienta computacional OralStats Furious

Multilayer Analysis of Conflictive Conversational Speech Using the Computational Tool OralStats Furious

ADRIÁN CABEDO NEBOT

UNIVERSITAT DE VALÈNCIA

<https://orcid.org/0000-0002-3881-9308>

Artículo recibido el / *Article received*: 2024-02-01

Artículo aceptado el / *Article accepted*: 2024-04-29

RESUMEN: Esta investigación explora algunas características prosódicas del conflicto discursivo, a partir de la comparación entre el análisis manual y automático de 1371 grupos entonativos procedentes de 7 conversaciones de una misma familia. De estos, 738 fueron designados manualmente como no conflictivos, mientras que 633 fueron etiquetados como conflictivos. Herramientas estadísticas, incluyendo ANOVA y *boxplots*, han permitido esclarecer singularidades inherentes al discurso conflictivo. Por ejemplo, el examen del tono revela disparidades significativas tanto entre hablantes como en contextos de conflicto. A pesar de la consistencia general en los datos, se observan variaciones notables, especialmente en el caso de algún hablante, cuyo tono muestra una notable variabilidad entre secuencias conflictivas y no conflictivas. De manera significativa, la velocidad de habla también resultó ser un marcador prosódico distintivo, mostrando diferencias significativas tanto entre hablantes como en escenarios de conflicto. En contraste, variables como la inflexión tonal, intensidad y duración no mostraron diferencias significativas basadas en niveles de conflicto o identidades de hablantes. La creación y desarrollo del programa Oralstats Furious para la categorización automática introdujo un elemento adicional al estudio. La visualización interactiva de datos ha resultado muy importante para una validación precisa, por lo que se permite a los investigadores refinar tanto las clasificaciones manuales como las automáticas. Las discrepancias entre las categorizaciones manuales y automáticas revelaron disparidades significativas, insinuando posibles sobreestimaciones o subestimaciones del conflicto en análisis manuales.

Palabras clave: conflicto, prosodia, clasificación manual, clasificación automática.

ABSTRACT: This research explores some prosodic features of discursive conflict, based on the manual and automatic analysis of 1371 intonational groups from 7 conversations within the same family. Out of these, 738 were manually designated as non-conflictive, while 633 were labeled as conflictive. Statistical tools, including ANOVA and boxplots, have helped clarify singularities inherent to discursive conflict. The research focuses on the examination of tone, revealing significant disparities both among speakers and in conflict contexts. Despite general consistency in the data, notable variations are observed, especially in the case of a particular speaker whose tone shows remarkable variability between conflictive and non-conflictive sequences. Importantly, speech rate emerged as a distinctive prosodic marker, showing significant differences both among speakers and in conflict scenarios. In contrast, variables such as tonal inflection, intensity, and duration did not show significant differences based on conflict levels or speaker identities, suggesting the need for precise and contextualized analysis. The creation and development of the Oralstats Furious program for automatic categorization introduced an additional layer to the study. Discrepancies between manual and automatic categorizations revealed significant disparities, hinting at possible overestimations or underestimations of conflict in manual analyses. Interactive data visualization proved invaluable for accurate validation, allowing researchers to refine both manual and automatic classifications.

Key words: conflict, prosody, manual tagging, automatic tagging.

1. INTRODUCCIÓN

Este artículo presenta el diseño y ejecución de OralStats Furious (<https://github.com/acabedo/furious>), un programa de análisis computacional del habla, diseñado con el lenguaje de programación R (R Core Team, 2023), que se enfoca en identificar desviaciones en la expresión prosódica de los hablantes con respecto a su uso lingüístico habitual, así como a catalogar de manera automática fragmentos conversacionales como conflictivos o no conflictivos según el cumplimiento de un conjunto de normas de restricción prosódica (por ejemplo, presencia de unos registros inferiores o superiores de tono en relación con un valor de corte establecido previamente).

El objetivo principal es observar secuencias de habla en las que se detecta habla emocional relacionada con el conflicto (enfado, ira, etc.), dado que, según algunos autores de la bibliografía (Roth y Tobin 2006, 2010; Wichmann, 2012), los valores prosódicos de variables como el tono o la intensidad podrían variar según el nivel de armonía o polémica en la conversación. Así pues, efectuando un análisis prosódico del discurso conflictivo, con varias capas de análisis fónico, este trabajo se encuadra en el marco del proyecto ESPRINT, «Estrategias pragmático-retóricas en la interacción conversacional conflictiva entre íntimos y conocidos: intensificación, atenuación y gestión interaccional» (ref. PID2020-114805GB-I00, coordinado por Marta Albelda y Maria Estellés).

En general, la investigación del presente estudio plantea cómo, a partir de un corpus de conversaciones espontáneas, un programa de computación y visualización de datos permite analizar los fragmentos de conversación armónica y los fragmentos

etiquetados como conflictivos por los propios investigadores para cumplir los siguientes objetivos:

1. obtener resultados a partir de datos parametrizables de forma automática; es el caso del análisis prosódico, que permite la obtención de valores como la media tonal, de intensidad, de velocidad de habla o duración.
2. añadir un análisis discursivo para detectar posibles cambios en la dinámica de intercambio de turnos.
3. cotejar (1) y (2) con el etiquetado manual por parte de los investigadores que, presentes en el acto comunicativo, discriminan los fragmentos conflictivos y no conflictivos; de esta manera, se puede comprobar si existen regularidades en estos últimos fragmentos frente a los primeros que permitan, eventualmente, una detección automática del conflicto en nuevos corpus.

En relación con el anterior punto número tres, la adopción de la técnica del observador participante implica la inmersión directa del investigador en el entorno objeto de estudio (Watts, 2003). Este enfoque permite capturar de manera más precisa y contextualizada las complejidades de las interacciones, especialmente en situaciones conflictivas, donde la interpretación de las dinámicas puede ser fundamental (Briz, en prensa; Estellés 2023; Estellés, en prensa). Al participar activamente en el escenario, el investigador no solo observa, sino que también experimenta las complejidades emocionales, los matices lingüísticos y las sutilezas contextuales que caracterizan las interacciones conflictivas.

2. LA PROSODIA COMO MARCA PRAGMÁTICA DIFERENCIAL

La prosodia emerge como un componente esencial en la comunicación oral, adquiriendo un valor pragmático significativo en la interacción conversacional (Couper-Kuhlen y Selting, 1996; Hidalgo, 2019). Este elemento no solo involucra la entonación y el ritmo, sino también la variación en la intensidad, velocidad y pausas al hablar. En la conversación, por tanto, la prosodia actúa como una herramienta dinámica que contribuye a la expresión de significados más allá de las palabras utilizadas.

Prosody is among the above-mentioned contributors to the communicative process. It is the perceptual pattern of intonation, stress, and pause; the physical correlates of which are fundamental frequency (F0), amplitude, and duration, respectively (...) As a multifaceted phenomenon, prosody falls under both categories of linguistic and nonlinguistic cues. As a linguistic cue, a hypothesized independent prosodic representation, similar to phonological or syntactic representations, becomes an integral part of speakers' linguistic competence (Cohen, Douaire y Elsabbagh, 2001: 74).

La prosodia, que aborda modulaciones del tono, ritmo, intensidad y duración del habla, desempeña un papel esencial en la comunicación y no opera de manera aislada. En español, por ejemplo, la modificación del tono se erige como el procedimiento más común para transmitir significados y afecta desde segmentos mínimos (vocales, sílabas...) hasta elementos más extensos en la expresión lingüística (grupos entonativos,

intervenciones...), como señala parte de la bibliografía (Hidalgo, 1996, 2019, 2020, 2013; Hübscher, Borràs y Prieto, 2017; Prieto, 2003).

En general, una creencia ampliamente aceptada en la comunicación lingüística es la de que un significado destacado va de la mano de una forma destacada (Levinson, 1983). Esta premisa fundamental en la teoría de la comunicación lingüística postula que cuando se desea expresar un significado especial, fuera de lo común o resaltado, la forma en que se comunica esa idea también debe ser distintiva y destacada. En otras palabras, existe una relación intrínseca entre la carga semántica que se desea transmitir y la forma en que se elige estructurar y expresar esa información en el lenguaje. En estos casos, la prosodia es un actor fundamental.

The point that we wish to make here is that prosody can be seen as one of the orderly ‘details’ of interaction, a resource which interlocutors rely on to accomplish social action and as a means of steering inferential processes. Prosodic features, we suggest, can be reconstructed as members’ devices, designed for the organization and management of talk in social interaction. They can be shown to function as part of a signalling system which - together with syntax, lexico-semantics, kinesics and other contextualization cues - is used to construct and interpret turn-constructual units and turns-at-talk. (Couper-Kuhlen y Selting, 1996: 37)

La relación entre fenómenos pragmáticos y su expresión prosódica ha sido objeto de atención por la bibliografía más reciente. En algunos de estos estudios, se ha demostrado que la intuición nativa a veces no coincide completamente con la realidad cuando se enfrenta a datos de corpus. Por ejemplo, en la expresión de un valor pragmático como la (des)cortesía, aunque se ha sugerido que tonos bajos pueden estar vinculados a la cortesía, esta relación no se sostiene consistentemente en los datos recopilados, donde puede encontrarse contrariamente marcación mediante tonalidad alta (Orozco 2010; Rojas, Blondet y Álvarez 2014; Álvarez, Blondet y Avendaño 2011). La variabilidad es evidente y algunas investigaciones muestran resultados dispares: algunos indican que un tono alto se asocia con la cortesía, mientras que otros sugieren que expresiones corteses se manifiestan con tono bajo (Hidalgo y Cabedo, 2014; Wilson y Wharton, 2006; Brown y Prieto, 2017).

La complejidad aumenta, en realidad, al explorar cualquier fenómeno pragmático (cortesía, atenuación, humor, ironía...). Aunque la intuición sugiere un comportamiento prosódico marcado en estos casos, algunos autores revelan que no hay correlatos acústicos constantes (Cabedo, 2021). La expresión de estos elementos pragmáticos puede manifestarse tanto con prominencia prosódica como sin ella, dependiendo del contexto. Como ejemplo concreto de lo que se está señalando, en diferentes idiomas se observa una tendencia hacia un tono más bajo, menor intensidad y mayor duración al dirigirse a una figura de autoridad; sería el caso, por ejemplo, del japonés (Idemaru, Winter y Brown, 2019). No obstante, la bibliografía también presenta escenarios opuestos. Estos hallazgos sugieren que la relación entre tono y pragmática no sigue un patrón fijo y puede variar según contextos comunicativos diferentes (Nadeu y Prieto, 2011).

En última instancia, asignar comportamientos prosódicos específicos a elementos del habla presenta un desafío significativo. La prosodia afecta los fragmentos del discurso de manera diversa y no siempre existe una correlación unívoca con

significados específicos. Se hace necesario considerar factores adicionales junto con las configuraciones fónicas para una comprensión más completa de la interacción entre prosodia y pragmática.

Por todo ello, el conflicto discursivo, como manifestación de desalineación o desacuerdo sostenido en el tiempo de manera consecutiva, sugiere una marcación formal más allá de las palabras y puede implicar, de manera directa, una alteración prosódica que vaya muy por encima (o muy por debajo, según el caso) de los valores prosódicos de los hablantes (Estellés, 2023).

3. SOBRE EL CONFLICTO DISCURSIVO Y LA PROSODIA

El estudio de la relación entre conflicto discursivo y prosodia ha recibido atención en la bibliografía, aunque no ha sido abordado de manera amplia como puedan haber sido tratados fenómenos pragmáticos derivados de situaciones comunicativas similares, como la (des)cortesía, el humor o la ironía. En relación con el conflicto discursivo, autores como Roth y Tobin han desarrollado investigaciones particulares y conclusiones cualitativas en algunas publicaciones (Roth y Tobin, 2006, 2010).

Generalmente, como se ha dicho anteriormente, la prosodia desempeña un papel crucial en la comunicación humana, influyendo en la interpretación y desarrollo de las interacciones verbales. Cuando los hablantes aumentan su tono, intensidad y velocidad respecto al interlocutor anterior, según Roth y Tobin (2006: 29), azuzan la situación comunicativa y se muestran ostensivamente a sí mismos al querer superar la intervención previa. En contraste, al utilizar un volumen, tono y velocidad más bajos, logran generar una situación de calma. Esta situación, aceptada y extendida también por otros autores, se genera «independent of the power and status differentials, speakers in non-conflictual situations use lowered pitch registers to express difference in content and reluctance to submit to the normal turn-taking routines (e.g., the question–answer sequence)» (Hothorn, Hornik y Zeileis 2006: 44). Finalmente, en el ámbito educativo, los mismos Roth y Tobin (2010: 40) observan que un aumento significativo en indicadores prosódicos, como un tono más alto y rápido, crea un entorno más acalorado, generando recursos para la asincronía en términos de emoción. A diferencia de la creencia generalizada, los profesores podrían ser alentados a no elevar la voz para afirmar su poder sobre los estudiantes.

Precisamente en cuanto al poder y la jerarquía, Wichmann (2011: 203) expone que, en situaciones de conflicto entre participantes con igualdad de poder, el aumento de la voz puede indicar cooperación, mientras que, en relaciones de desigualdad, la prosodia puede quedar desalineada. En esos contextos de desigualdad, Wichmann (2012) destaca que, durante momentos de conflicto, el participante más poderoso habla por debajo del otro; en el ámbito de las ciencias naturales, la configuración del tono se asocia al volumen del ser vivo que lo emplea, así «it suggests that pitch may be used iconically. According to the so-called Frequency Code, a high pitch signals “small animal and a low pitch signals” large (and therefore more powerful) animal» (Wichmann, 2012: 343).

Más recientemente, Estellés (2023) se ha acercado a la visualización de la información prosódica en el conflicto discursivo en el español hablado de Valencia. En línea similar a lo ya expuesto con anterioridad, sus conclusiones inciden en que «no todo el mundo se comporta fónica e interactivamente igual con todos los interlocutores, y a veces se emplean parámetros fónicos distintos con unos y con otros dependiendo de la

relación mutua, de la desalineación con el contenido del turno precedente» (Estellés, 2023: 245). Resultados similares se habían encontrado ya para el inglés, donde las respuestas desalineadas o de desacuerdo pueden presentar tonos más altos o, paradójicamente, más bajos, según la disposición estratégica del intercambio comunicativo (Zellers y Ogden, 2014: 304).

Un aspecto similar y vinculado a la expresión del conflicto es también la expresión emocional, sobre todo, aspectos como el enfado, la sorpresa o el asco. En cuanto a la identificación de emociones en la prosodia durante la interacción verbal, se destaca la importancia de la cautela hermenéutica, incluso cuando los participantes interpretan ciertas emociones en el habla (Szczepek, 2010: 866). La interpretación de las emociones exhibidas y los estados emocionales de los demás requiere precaución, dado que cualquier conjetura por parte del investigador es solo una aproximación al estado emocional real del hablante.

En cualquier caso, se ha abierto incluso la posibilidad de que la expresión emocional y su correlato prosódico puedan explicarse desde un punto de vista neurológico:

Our data suggest distinct neural pathways for happy and angry prosody when presented in incongruent emotional semantic context. This evidence possibly reflects functionally different ways to handle positive and negative information with regard to potential “threats” for the participant. Concerning valence specificity, we found that the processing of happy prosody while ignoring negative semantic content engaged a left-sided neural network of middle temporal and inferior frontal areas, which have been previously associated with the processing of prosodic information per se. (Wittfoth et al., 2010: 7)

Lo cierto es que los trabajos realizados sobre español, emociones y prosodia sí han encontrado valores altos de rango tonal para la expresión de la sorpresa y la alegría, mientras que, a diferencia quizá de lo esperado, el valor del enfado queda en un registro tonal moderado, similar a otros valores como el asco y el miedo (Garrido y Chica, 2018: 32). De cualquier manera, los datos de laboratorio estudiados en esta última investigación contrastan con los encontrados por Padilla (2023) para el español hablado espontáneo, en los que se apuntan tendencias como el nivel tonal neutro en la manifestación del enfado, si bien también se señala que «el análisis de los datos no parece apoyar la hipótesis de huellas dactilares prosódicas excluyentes adscritas a cada emoción concreta» (Padilla, 2023: 165).

En resumen, el estudio de la relación entre conflicto discursivo y prosodia revela la influencia crucial de factores paralingüísticos como el rango tonal, la intensidad o la velocidad de habla en la comunicación humana. Aunque los resultados varían en cuanto a la expresión emocional, el reparo, la cautela y la precaución son muy importantes en la interpretación y valoración investigadora.

4. METODOLOGÍA

La metodología de esta investigación es cuantitativa, dado que se sirve de una gran cantidad de datos y los procesa de modo computacional y estadístico para poder luego llegar a interpretaciones relacionadas sobre el comportamiento particular del conflicto discursivo y, muy especialmente, de su relación con la prosodia.

4.1. DATOS ANALIZADOS

Esta investigación utiliza como corpus de análisis base el corpus llamado ESPRINT-conversación VLC (Valencia), el cual comprende trece interacciones que presentan un total de 221 minutos, aproximadamente tres horas y cuarenta minutos. De este conjunto, se seleccionaron específicamente las conversaciones *casa1*, *casa4*, *casa6*, *casa10*, *casa11*, *casa12* y *casa13*, escogidas por contar con transcripciones completas y una calidad de audio superior, sumando en conjunto 56 minutos de grabación.

Figura 1. Distribución de los datos analizados según el procesamiento con Oralsats. Ips (intonational phrases)



Según la Figura 1, puede observarse que hay un total de 21 hablantes y 8176 alófonos, aunque no es realmente la distribución por cada una de las 7 conversaciones analizadas. El programa *Oralsats* discrimina entre hablante por conversación, por lo que *casa11_A*, *casa1_A*, *casa4_A*, etc., por ejemplo, corresponden a la misma hablante A, aunque se han etiquetado como diferentes porque están vinculados a la conversación de la que proceden; de esta manera, en la fase de investigación y análisis, pueden observarse los patrones de conducta de un modo más ágil. Los 1371 grupos entonativos del análisis han sido procesados para disponer de su configuración prosódica, como se detallará en la Sección 4.2 y en toda la Sección 5.

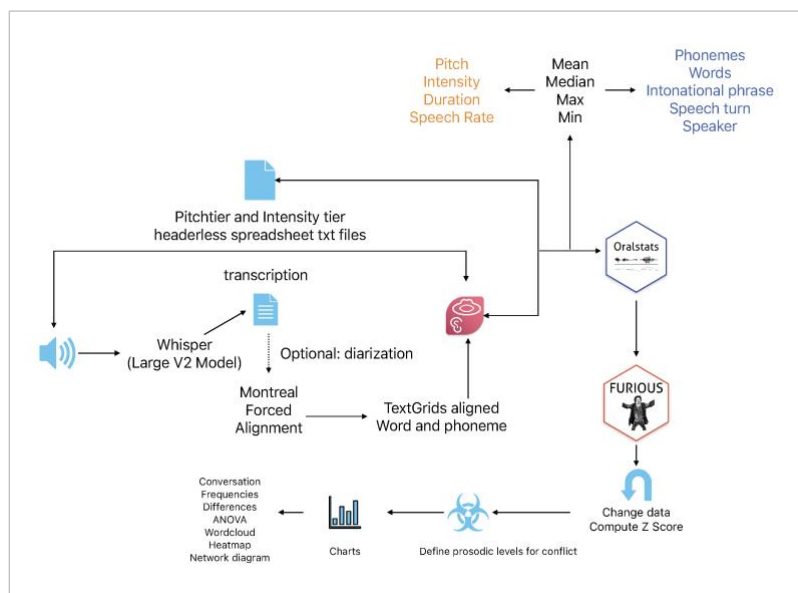
Los hablantes del corpus ESPRINT-Valencia son cuatro (A, B, C y D); A y B forman un matrimonio en el que A es la esposa (61 años de edad) y B el marido (61 años de edad), mientras que C (22 años de edad) y D (25 años de edad) son las hijas de los anteriores. Todos ellos tienen un nivel de instrucción alto y son bilingües en castellano y valenciano. La familia fue informada previamente de que iba a ser grabada en algún momento próximo por parte de C, que es la investigadora participante en la recogida de esta muestra. Ella, por tanto, es la parte implicada en el entorno de conflicto y la que, posteriormente, etiqueta los fragmentos en conflictivos y no conflictivos. Destacamos en este caso la técnica del investigador participante, ya que esta figura puede recuperar información contextual difícilmente recuperable para quien no estuvo presente en el acto comunicativo (proxémica, gestualidad, etc.).

4.2. HERRAMIENTA DESARROLLADA

En la Figura 2, se observa un diagrama detallado con las principales características del programa *Oralsats Furious* (código fuente disponible en <https://github.com/acabedo/furious>); se trata de un programa de análisis prosódico y de visualización estadística que permite recoger material transcrito y vincularlo con

elementos prosódicos como el tono, la intensidad, la duración o la velocidad de habla. Esta funcionalidad, basada en las propiedades del Oralstats (Cabedo, 2022), se amplía con un criterio de delimitación y etiquetado de los grupos entonativos en *furiosos* o *no furiosos*, según si entran en los márgenes de unos valores prosódicos fijados por el investigador.

Figura 2. Esquema de desarrollo del programa Oralstats Furious



Los niveles prosódicos definidos por el investigador permiten determinar qué se considera constitutivo de la categoría *furioso* o, por el contrario, se articulan como un comportamiento general y esperable. Posteriormente, el propio programa permite comparar este etiquetado automático con cualquier etiquetado manual que se hay realizado previamente.

Tomando como base la idea expuesta en Estellés (2023), se propone utilizar sobre todo los valores Z-Score, es decir, valores estandarizados de las variables que permiten no solo comparar estas entre ellas, aunque tengan factores de medición distintos, sino también diferentes hablantes, independientemente de sus características sociolingüísticas (sexo, edad...).

Al mismo tiempo, los Z-Score permiten observar valores muy por encima o muy por debajo de la media del hablante; acorde con lo expuesto en la Sección 5, la manifiesta marcación formal de valores extremos puede estar relacionada con la expresión de valores discursivos extremos y, entre ellos, el conflicto podría ser uno.

Finalmente, Furious permite generar distintos modos de visualización, entre los que se incluyen mapas de calor, gráficos de barras, nubes de palabras y diagramas de red.

5. PERFIL PROSÓDICO DEL CONFLICTO

En el análisis estadístico y cuantitativo de este artículo pretendemos acercarnos a un conjunto de objetivos concretos:

1. Observar si las variables prosódicas de tono, inflexión tonal, intensidad, duración y velocidad de habla manifiestan unos valores notablemente distintos según la adscripción del investigador a las categorías de conflictivo o no conflictivo.
2. Utilizar Oralstats Furious para establecer automáticamente secuencias *furiosas* o *no furiosas* a partir de valores Z y, concretamente, de aquellos que superen valores de 1.96 por arriba o por abajo en relación con la media del hablante (con un intervalo de confianza del 95% significa que el 95% de los datos de la distribución normal estándar se encuentran dentro de 1.96 desviaciones estándar de la media; en otras palabras, lo superior o inferior a este número son datos extremos en la variable).
3. Comparar 1 y 2 para observar si hay o no relación estadística significativa. En el caso de haberla, se abre la posibilidad a que un etiquetado automático con Oralstats Furious pueda etiquetar (y localizar) en una conversación no analizada previamente fragmentos conflictivos con un grado de acierto notable.

Es crucial destacar que, aunque el enfoque del estudio cuantitativo es de naturaleza inferencial, los datos examinados se limitan exclusivamente a los cuatro hablantes descritos, quienes fueron registrados en diversas conversaciones en el español hablado de Valencia. Por lo tanto, los resultados que se presentarán en las secciones siguientes sobre el perfil prosódico de los datos analizados son aplicables inicialmente a estos hablantes específicos. No obstante, se presume que, al ampliar el corpus a otras ciudades y a distintos hablantes, los resultados sean similares, dado que se cumplen las características de aleatoriedad azarosa.

5.1. PROSODIA DEL CONFLICTO SEGÚN ETIQUETADO MANUAL

En la investigación realizada, la catalogación manual distinguió 738 grupos entonativos no conflictivos y 633 conflictivos de un total de 1371.

En esta sección, a partir de la variable independiente *conflicto*, se realiza un análisis de la varianza de las distintas variables numéricas computadas; la variable *conflicto* se combina además con las variables *conversación* y *hablante*, para certificar que el papel situacional o idiolectal no esté relacionado con la variabilidad estudiada o, en caso de estarlo, en qué modo lo está.

La combinación de ANOVA (Análisis de la Varianza) y *boxplot* (gráfico de caja) es relevante desde la investigación de los datos porque estas dos herramientas de base estadística ofrecen perspectivas complementarias para analizar la distribución de las variables y las diferencias entre grupos.

ANOVA evalúa la hipótesis nula de que no hay diferencias significativas entre las medias de los grupos y, en caso de rechazar esta hipótesis, proporciona información sobre qué grupos son significativamente diferentes. Los datos estadísticos de la prueba

ANOVA suelen combinarse con los llamados *boxplots*. Un *boxplot* es una representación gráfica que muestra la distribución en cuartiles de un conjunto de datos y resalta la presencia de valores atípicos.

5.1.1 Tono

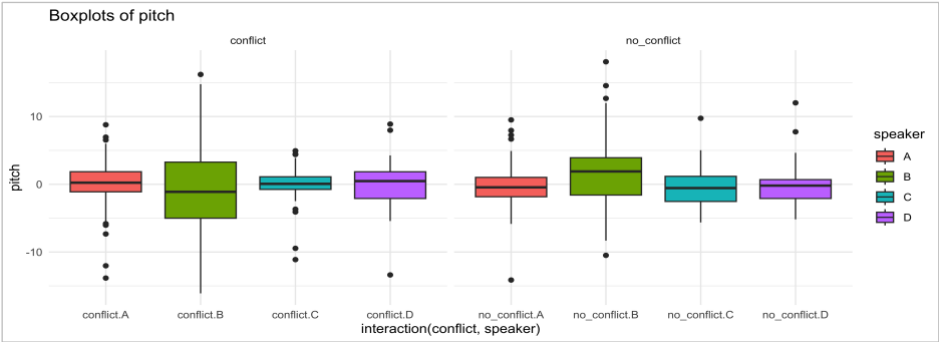
Los datos de tono¹ muestran una variación considerable en las grabaciones, con un valor mínimo de -16.10 y un máximo de 18.08 desviaciones estándar. La mediana, que se encuentra en 0.00, y un valor promedio, también muy próximo a cero (0.01), indican que hay una distribución simétrica de los datos

Tabla 1. Valores estadísticos de la prueba ANOVA para el contraste entre el tono y las variables de hablante, conflicto y conversación

<i>term</i>	<i>df</i>	<i>sumsq</i>	<i>meansq</i>	<i>statistic</i>	<i>p.value</i>
conflict	1	1.83	1.83	0.14	0.71
speaker:conflict	3	267.19	89.06	6.75	0.00
conflict:source	4	153.34	38.34	2.91	0.02
speaker:conflict:source	3	111.58	37.19	2.82	0.04

Los resultados del análisis de varianza (ANOVA) revelan que existen interacciones significativas en las conversaciones analizadas. Estas interacciones (hablante y conflicto; conversación y conflicto; y conversación, hablante y conflicto) sugieren una influencia notable en el uso de valores de tono que dependen, como ya ha observado la bibliografía (Estellés, 2023; Padilla, 2023), del contexto comunicativo concreto. Cabe resaltar, en tal sentido, que la diferencia entre las categorías *conflicto* y *no conflicto* de la variable *conflicto* no se diferencian sin tener en cuenta factores situacionales o individuales como los que se han comentado anteriormente. Para mejorar la visualización de estos datos, puede observarse el *boxplot* de la Figura 3:

Figura 3. Diagrama de caja de la variable tono según conflicto y hablante



¹ La medición tomada para el tono ha sido la del valor Z, dado que los valores absolutos en Hz extraídos de PRAAT podían estar afectados por factores como el sexo o la edad. Se ha descartado tomar en este punto valores como semitonos porque deberían haber sido computados igualmente en relación con una medida de comparación, como la media o mediana del hablante.

La Figura 3 refleja consistencia en los datos para las distribuciones de cada hablante, aunque los registros de tono para segmentos no conflictivos son ligeramente inferiores a los respectivos segmentos conflictivos. Solo en el caso de B, el marido, se observa una ligera inconsistencia, ya que el valor de tono es, contrariamente al observado para otros hablantes, más alto en secuencias no conflictivas que en las conflictivas.

5.1.2. Inflexión tonal o tonema

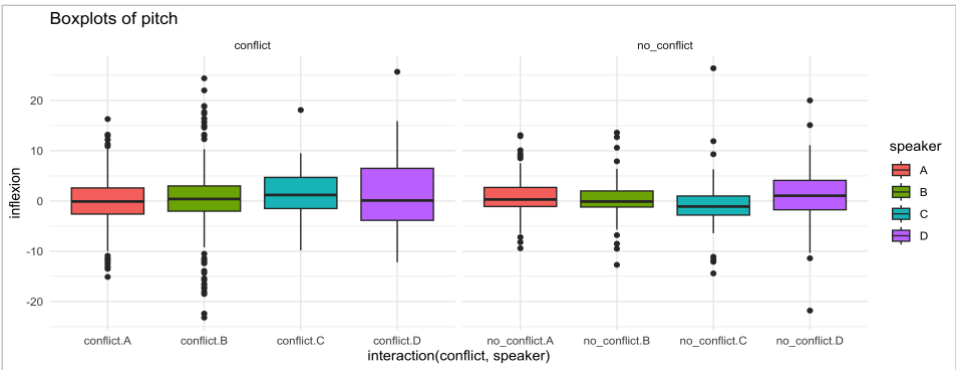
La variable de inflexión tonal final o tonema, medida en semitonos, dado que hay posibilidad de relativizar la distancia tonal entre dos puntos, presenta una distribución amplia, con valores que varían desde -23.2 hasta 26.4 semitonos, que constituyen valores extremos. La mediana, ubicada en 0.2 semitonos, indica que el 50% de los datos se encuentra por encima y por debajo de este punto, mientras que la media, que es de 0.33 semitonos, sugiere una leve asimetría hacia valores superiores. Los cuartiles revelan que el 25% de los datos está por debajo de -2.1 semitonos (primer cuartil), y el 75% está por debajo de 2.7 semitonos (tercer cuartil).

Tabla 2. Valores estadísticos de la prueba ANOVA para el contraste entre la inflexión tonal y las variables de hablante, conflicto y conversación

<i>term</i>	<i>df</i>	<i>sumsq</i>	<i>meansq</i>	<i>statistic</i>	<i>p.value</i>
conflict	1	8.99	8.99	0.32	0.57
speaker:conflict	3	142.24	47.41	1.69	0.17
conflict:source	4	105.60	26.40	0.94	0.44
speaker:conflict:source	3	90.20	30.07	1.07	0.36

No hay evidencia significativa de diferencias entre los niveles de *speaker*, *conflict* o *source*, ni de interacciones significativas entre estos factores. Los p-valores altos sugieren que no hay efectos significativos en ninguno de los factores ni en sus interacciones.

Figura 4. Diagrama de caja de la variable tono según conflicto y hablante



Los datos de la Figura 4 muestran que el comportamiento de la inflexión tonal es similar por hablante; D muestra inflexiones tonales más elevadas para secuencias

conflictivas, aunque la diferencia con el conflicto es irrelevante, apenas unas décimas de semitono.

5.1.3. Intensidad

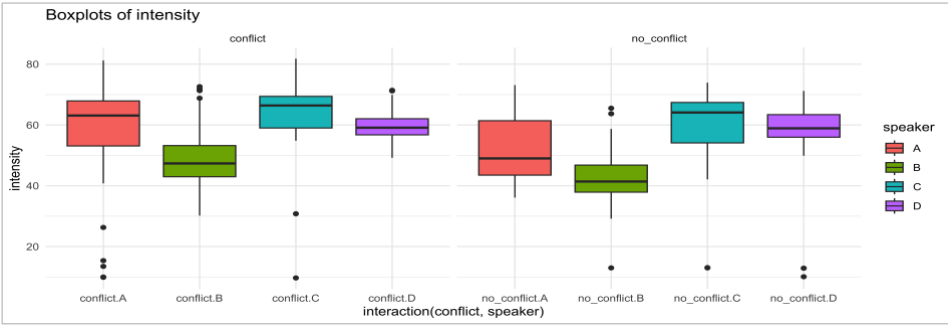
Los datos de intensidad en decibelios presentan una amplia gama de valores, con un mínimo de 9.70 dB y un máximo de 81.80 dB, como valores extremos. La mediana, situada en 56.75 dB, y el valor promedio, de 55.57 dB, indican una distribución relativamente simétrica. Los cuartiles muestran que la mayoría de las observaciones se establecen en un rango de 46.50 dB a 65.40 dB.

Tabla 3. Valores estadísticos de la prueba ANOVA para el contraste entre la intensidad y las variables de hablante, conflicto y conversación

<i>term</i>	<i>df</i>	<i>sumsq</i>	<i>meansq</i>	<i>statistic</i>	<i>p.value</i>
conflict	1	12.025.36	12.025.36	230.56	0.00
speaker:conflict	3	145.17	48.39	0.93	0.43
conflict:source	4	907.17	226.79	4.35	0.00
speaker:conflict:source	3	138.18	46.06	0.88	0.45

Los resultados del análisis de varianza (ANOVA) para la variable de intensidad revelan patrones distintivos. Se observan efectos altamente significativos para las variables de conflicto y conversación, pero no existen diferencias individuales; la interacción hablante-conflicto no es significativa, sugiriendo que la influencia del conflicto en la intensidad no varía de manera significativa entre los hablantes si no se vincula a conversaciones específicas.

Figura 5. Diagrama de caja de la variable intensidad según conflicto y hablante



Los valores de la Figura 5 muestran semejanzas entre hablantes y la expresión del conflicto; solo la hablante A se distancia de los otros hablantes, ya que alza generalmente mucho más la voz en secuencias conflictivas, mientras que baja significativamente la voz en fragmentos no conflictivos. Curiosamente, el hablante B es el que muestra valores de intensidad más bajos tanto para conflicto como para no conflicto.

En cualquier caso, los resultados de esta variable deben tomarse con cautela, dado que se trata de la variable que puede presentar mayor variación según la distancia que los hablantes interpongan con respecto al dispositivo de grabación.

5.1.4. Duración

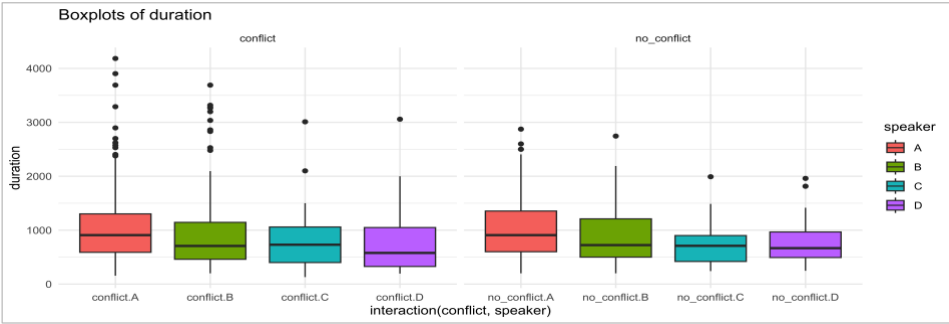
Los datos de duración de los grupos entonativos, en milisegundos, indican que los grupos entonativos varían en tiempo desde un mínimo de 130 ms hasta un máximo de 4184 ms. La mediana se encuentra en 830 ms y la media en 959.5 ms.

Tabla 4. Valores estadísticos de la prueba ANOVA para el contraste entre la duración y las variables de hablante, conflicto y conversación

<i>term</i>	<i>df</i>	<i>sumsq</i>	<i>meansq</i>	<i>statistic</i>	<i>p.value</i>
conflict	1	1.562.91	1.562.91	0.00	0.95
speaker:conflict	3	344.662.65	114.887.55	0.34	0.80
conflict:source	4	2.649.045.20	662.261.30	1.96	0.10
speaker:conflict:source	3	381.277.35	127.092.45	0.38	0.77

Los resultados del análisis de varianza (ANOVA) para la variable de *duración* señalan que no existe un impacto significativo en la duración para ninguno de los cruces analizados.

Figura 6. Diagrama de caja de la variable duración según conflicto y hablante



De la Figura 6, se pueden deducir algunas conclusiones generales:

1. Las secuencias conflictivas y las no conflictivas suelen durar más o menos lo mismo.
2. Si se tiene en cuenta al hablante, algunos hablantes emiten grupos entonativos más largos, independientemente del conflicto que transmiten.
3. La duración, al menos en estas conversaciones, no presenta significación estadística. Aun así, llama la atención el alto número de *outliers* que presentan, por ejemplo, los hablantes A y B.
4. Los hablantes A y B, que conforman el matrimonio de la familia, presentan duraciones de grupos entonativos mayores que sus hijas.

5.1.5. Velocidad de habla

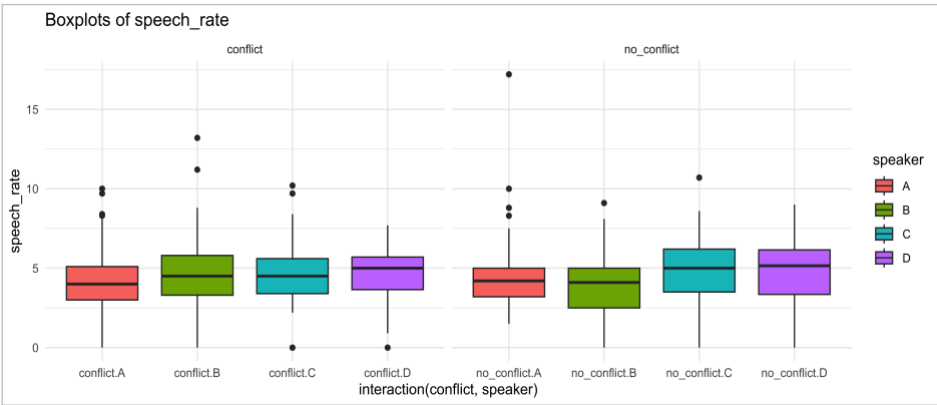
La velocidad de habla, expresada en palabras por segundo, presenta una gama de variaciones. La mínima velocidad registrada es de 0.9 palabras por segundo. El primer cuartil revela que el 25% de las observaciones tienen una tasa de habla de 3.3 palabras por segundo o menos, mientras que la mediana, ubicada en 4.3 palabras por segundo, y la media, de 4.4 palabras por segundo, sugieren una tendencia central en los datos. El tercer cuartil, con un valor de 5.4 palabras por segundo, muestra que el 75% de las observaciones tienen una velocidad de habla de 5.4 palabras por segundo o menos. La velocidad de habla máxima registrada es de 17.2 palabras por segundo.

Tabla 5. Valores estadísticos de la prueba ANOVA para el contraste entre la velocidad de habla y las variables de hablante, conflicto y conversación

<i>term</i>	<i>df</i>	<i>sumsq</i>	<i>meansq</i>	<i>statistic</i>	<i>p.value</i>
conflict	1	2.44	2.44	0.80	0.37
speaker:conflict	3	45.91	15.30	5.03	0.00
conflict:source	4	16.51	4.13	1.36	0.25
speaker:conflict:source	3	10.09	3.36	1.10	0.35

Los resultados del análisis de varianza (ANOVA) para la variable de *velocidad de habla* arrojan información significativa. El conflicto y el hablante parecen diferenciarse significativamente independientemente de la conversación en la que se integran.

Figura 7. Diagrama de caja de la variable velocidad de habla según conflicto y hablante

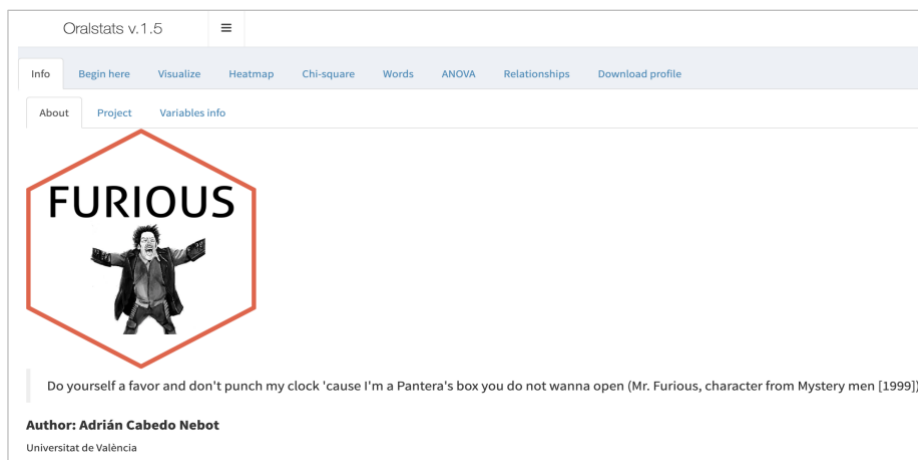


Los valores de la Figura 7 muestran similitud en la velocidad de habla emitida en secuencias de *conflicto* o de *no conflicto*. Los valores más altos pertenecen en ambos casos a la hablante D y, al revés de los observado en variables como tono o intensidad, la hablante A, cuya modulación tonal o de intensidad resultaba mayor, presenta la velocidad de habla más baja.

5.2. CATALOGACIÓN AUTOMÁTICA CON ORALSTATS FURIOUS

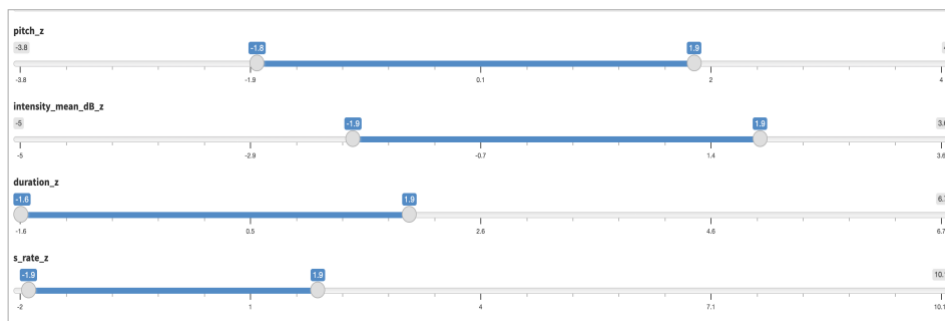
En la Figura 8, en la que se observa la configuración visual del programa, se observan una serie de pestañas que cobrarán más relevancia en esta sección del artículo.

Figura 8. Pantalla inicial del programa Oralstats Furious



Concretamente, en la pestaña *Begin here*, se visualiza un filtro con todas las variables de la base de datos y sus valores. Al modificar los filtros, como se observa en la Figura 9, el sistema etiqueta todo lo que queda fuera de los límites del selector como *conflictivo/furioso* y aquello que queda dentro como *no conflictivo/no furioso*.

Figura 9. Ejemplo de filtro sobre la base de datos para configurar la etiqueta de furioso



Con los parámetros de superioridad e inferioridad de 1.96 en las variables estandarizadas con valores Z (tono, inflexión tonal, intensidad, duración y velocidad de habla), 932 de 1371 grupos entonativos se han identificado como *no furiosos*, mientras que el resto, 439 grupos entonativos, han sido catalogados como *furiosos* y, por tanto, debe comprobarse con los manualmente etiquetados como conflictivos.

Inicialmente, hay una clara disonancia con la catalogación manual, que establecía 194 grupos conflictivos más.

Figura 10. Representación gráfica de la correspondencia entre valores manuales y automáticos

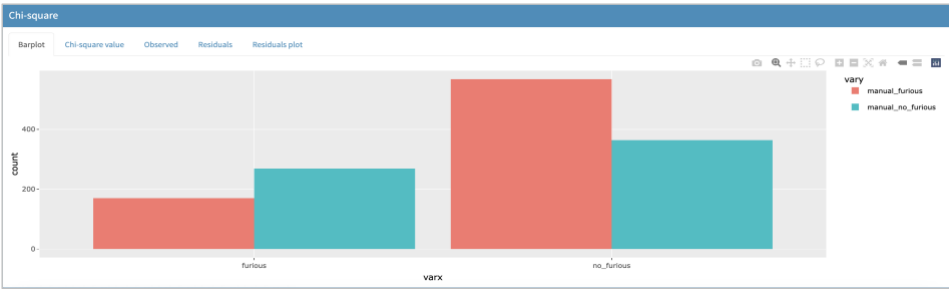


Tabla 6. Distribución de correspondencias y residuos estandarizados entre la clasificaciones manual y automática del conflicto

	<i>manual_furious</i>	<i>Residuos</i>	<i>manual_no_furious</i>	<i>Residuos</i>
furious	170	-4.31	269	4.65
no_furious	568	2.96	364	-3.19

Con un valor de Chi cuadrado de 58.394 (1 grado de libertad, valor p de 2.146e-14), se rechaza la hipótesis de independencia de las variables, es decir, la catalogación manual y la automática, basada en datos fónicos, corresponden de manera significativa, aunque no en la dirección que cabía esperar, sino más bien en la contraria. Mientras hay muchas secuencias identificadas como conflictivas que no coinciden con valores significativamente marcados en el apartado prosódico (568), hay 269 casos en los que la investigadora no había etiquetado como conflictivas y que, sin embargo, sí que presentan valores prosódicos extremos.

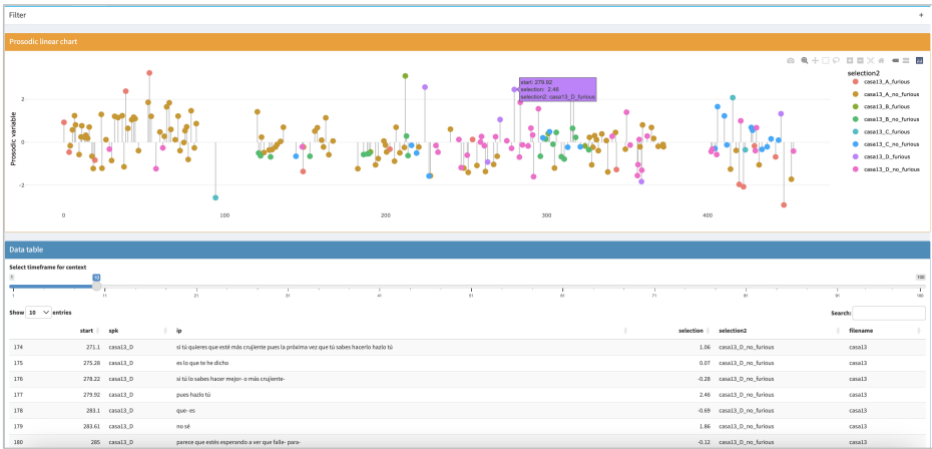
Los escenarios que se plantean en este análisis ya han sido detectados y establecidos por Estellés (Estellés, en prensa) y, en general, corresponden a alguna de las siguientes casuísticas:

1. La investigadora pudo catalogar como secuencias conflictivas, con indicación incluso en la transcripción de marcas de énfasis prosódico, fragmentos donde, acústica y objetivamente, no se aprecian diferencias tonales, de intensidad o duración con enunciados anteriores o posteriores.
2. Algunas secuencias que presentaban conflicto discursivo se extendieron por inercia a grupos entonativos colindantes en la fase de transcripción y etiquetado.
3. La investigadora dejó de etiquetar como conflictivas secuencias que se enuncian claramente marcadas desde un punto de vista prosódico. En estos casos, Estellés (en prensa) ha detectado que, continuamente, la investigadora no marca como conflictivas las desavenencias y desacuerdos que se observan

cuando el hablante B, el marido, no está presente en las interacciones, es decir, las conversaciones entre A, B y C (madre e hijas) suelen definirse desde un punto de vista neutro.

Para corroborar lo anterior, una de las visualizaciones de Oralsats Furious, tomado de Oralsats nativo (Cabedo, 2022) y adaptado ya por Estellés (2023) para el conflicto discursivo, permite comprobar los etiquetados realizados, tanto los manuales como los automáticos, de manera interactiva, de tal manera que cuando el investigador desliza el ratón por encima de los puntos tonales más llamativos y pulsa con el cursor, se despliega un fragmento de la conversación y se escucha el audio correspondiente; de este modo, el investigador, formado en el análisis crítico del conflicto discursivo, puede perfilar mejor tanto las catalogaciones manuales como las realizadas por el algoritmo automático de Oralsats Furious.

Figura 11. Visualización sobre la línea de tiempo de los enunciados, su tono y la adscripción a un hablante con manifestación o no de conflicto discursivo



Así pues, en la Figura 11 se ha seleccionado la conversación 13 y el cursor se ha pulsado encima de uno de los puntos tonales, catalogados automáticamente como conflictivos, y que se asocian, en este caso, a la hablante D. Concretamente, en el segundo 279.92, la hablante D, conversando con A, dice *pues hazlo tú*, con un valor tonal que se sitúa 2.46 desviaciones estándar por encima de la media. En este caso, por ejemplo, la catalogación manual no había contemplado este enunciado como conflictivo.

6. REFLEXIONES FINALES

En el marco de esta investigación, se ha realizado el análisis de una amplia catalogación manual de 1371 grupos entonativos, de los cuales 738 fueron clasificados como no conflictivos y 633 como conflictivos. Este análisis, respaldado por herramientas estadísticas como ANOVA y *boxplots*, ha proporcionado una visión detallada de las complejidades en la expresión prosódica durante interacciones marcadas por el conflicto discursivo.

El estudio del tono, central en el análisis prosódico, ha revelado diferencias significativas tanto entre hablantes como en contextos de conflicto. Aunque la consistencia en los datos sugiere patrones comunes, se observan algunas inconsistencias notables, particularmente en el caso del hablante B, cuyo tono ha destacado por su variabilidad en secuencias conflictivas y no conflictivas. La velocidad de habla también ha surgido como una variable distintiva, mostrando diferencias significativas entre hablantes y contextos de conflicto. Este hallazgo subraya la relevancia de la velocidad como un marcador prosódico clave junto al tono.

En relación con variables como la inflexión tonal, intensidad y duración, el análisis estadístico no ha evidenciado diferencias significativas entre niveles de conflicto ni entre hablantes. Esto sugiere que estas variables pueden no ser tan sensibles a los matices contextuales en las conversaciones analizadas.

La implementación de Oralstats Furious para la catalogación automática ha añadido una dimensión adicional al estudio. La comparación entre la catalogación manual y automática ha revelado discrepancias significativas, sugiriendo que el análisis manual podría incurrir en sobreestimaciones o subestimaciones del conflicto. La visualización interactiva de datos ha proporcionado una herramienta valiosa para la validación precisa de las catalogaciones, permitiendo al investigador refinar tanto las clasificaciones manuales como las automáticas.

A pesar de las variaciones observadas, las generalizaciones indican que la expresión prosódica en situaciones de conflicto discursivo está influenciada por factores contextuales y de hablante. El empleo de herramientas estadísticas y de visualización ha contribuido a una comprensión más profunda de la complejidad en la comunicación prosódica en contextos conflictivos.

7. BIBLIOGRAFÍA

- Álvarez, Alexandra, María Alejandra Blondet y Darcy Rojas (2011). (Des)cortesía y prosodia: Una relación necesaria. *Oralia: Análisis del discurso oral*, 14, 437-450.
- Briz, Antonio (en prensa). Los conflictos en la conversación coloquial entre familiares, amigos o conocidos.
- Brown, Lucien y Prieto, Pilar (2017). *(Im)politeness: Prosody and Gesture*. En J. Culpeper, M. Haugh, y D. Z. Kádár (Eds.), *The Palgrave Handbook of Linguistic (Im)politeness* (pp. 357-379). Palgrave Macmillan UK. https://doi.org/10.1057/978-1-137-37508-7_14
- Cabedo, Adrián (2021). Prosodic modulation as a mark to express pragmatic values: The case of mitigation in Spanish. *Journal of Pragmatics*, 181, 196-208. <https://doi.org/10.1016/j.pragma.2021.05.028>
- Cabedo, Adrián (2022). *Oralstats*. <https://github.com/acabedo/oralstats>
- Cohen, Henri, Josée Douaire y Mayada Elsabbagh (2001). The role of prosody in discourse processing. *Brain and cognition*, 46(1-2), 73-82.
- Couper-Kuhlen, Elizabeth y Margaret Selting (1996). Towards an interactional perspective on prosody and a prosodic perspective on interaction. *Prosody in conversation: Interactional studies*, 11.

- Estellés, Maria (2023). Visualizando el conflicto discursivo a través de la expresión fónica: Un estudio a partir de dos conversaciones. *Normas*, 13(1), 224-247. <https://doi.org/10.7203/Normas.v13i1.27986>
- Estellés, Maria (en prensa). La identificación del conflicto en conversación espontánea: Participantes vs. Analistas. *Cultura, Lenguaje y Representación*.
- Garrido, Juan Maria y Juan Antonio Chica Sabariego (2018). Pitch range and identification of emotions in Spanish speech: A perceptual study. *Estudios de Fonética Experimental*, 27, 13-36.
- Hidalgo, Antonio (2019). *Sistema y uso de la entonación en español hablado*. Universidad Andrés Hurtado.
- Hidalgo, Antonio (1996). *Entonación y conversación coloquial: Sobre el funcionamiento demarcativo-integrador de los rasgos suprasegmentales* [PhD Thesis].
- Hidalgo, Antonio (2013). La fono(des)cortesía: marcas prosódicas (des)cortes en español hablado. Su estudio a través de corpus orales. *RLA: Revista de lingüística teórica y aplicada*, 51, 127-150.
- Hidalgo, Antonio (2020). Rasgos melódicos de la emoción: Estudio de un corpus conversacional. *Phonica*, 16, 36-53.
- Hidalgo, Antonio y Adrián Cabedo (2014). On the importance of the prosodic component in the expression of linguistic im/politeness. *Journal of Politeness Research*, 10(1), 5-27.
- Hothorn, Torsten, Kurt Hornik y Achim Zeileis (2006). Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics*, 15(3), 651-674. <https://doi.org/10.1198/106186006X133933>
- Hübscher, Iris, Joan Borràs y Pilar Prieto (2017). Prosodic mitigation characterizes Catalan formal speech: The Frequency Code reassessed. *Journal of Phonetics*, 65, 145-159. <https://doi.org/10.1016/j.wocn.2017.07.001>
- Idemaru, Kaori, Bodo Winter y Lucien Brown (2019). Cross-cultural multimodal politeness: The phonetics of Japanese deferential speech in comparison to Korean. *Intercultural Pragmatics*, 16(5), 517-555. <https://doi.org/10.1515/ip-2019-0027>
- Levinson, Stephen (1983). *Pragmatics*. Cambridge University Press.
- Nadeu, Marianna y Pilar Prieto (2011). Pitch range, gestural information, and perceived politeness in Catalan. *Journal of Pragmatics*, 43(3), 841-854. <https://doi.org/10.1016/j.pragma.2010.09.015>
- Orozco, Leonor (2010). *Estudio sociolingüístico de la cortesía en tratamientos y peticiones. Datos de Guadalajara* [PhD Thesis].
- Padilla, Xose (2023). Cómo construimos las emociones en la entonación coloquial. *Estudios de Fonética Experimental*, 32, 155-168.
- Prieto, Pilar (2003). *Teorías de la entonación*. Ariel.
- R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rojas Avendaño, Darcy, María Alejandra Blondet y Álvarez, Alexandra (2014). Configuración tonal de la atenuación en el habla de Mérida. *Lengua y Habla*, 18, 93-106.
- Roth, Wolff-Michael y Kenneth (2010). Solidarity and conflict: Aligned and misaligned prosody as a transactional resource in intra-and intercultural communication involving power differences. *Cultural Studies of Science Education*, 5, 807-847.

- Roth, Wolff-Michael y Kenneth Tobin (2006). *International Conference on Conversation Analysis*. <https://api.semanticscholar.org/CorpusID:235260731>
- Szczepek, Beatrice (2010). Prosody and alignment: A sequential perspective. *Cultural Studies of Science Education*, 5, 859-867.
- Watts, Richard (2003). *Politeness*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511615184>
- Wichmann, Anne (2011). Prosody and pragmatic effects. *Pragmatics of society*, 181-214.
- Wichmann, Anne (2012). Prosody in context: The effect of sequential relationships between speaker turns. En G. Elordieta & P. Prieto (Eds.), *Prosody and meaning* (pp. 329-348). De Gruyter. <https://doi.org/10.1515/9783110261790.329>
- Wilson, Deirdre y Tim Wharton (2006). Relevance and prosody. *Journal of Pragmatics*, 38(10), 1559-1579. <https://doi.org/10.1016/j.pragma.2005.04.012>
- Wittfoth, Matthias, Christin Schröder, Dina Schardt, Reinhard Dengler, Hans-Jochen Heinze y Sonja Kotz (2010). On emotional conflict: Interference resolution of happy and angry prosody reveals valence-specific effects. *Cerebral Cortex*, 20(2), 383-392.
- Zellers, Margaret y Richard Ogden (2014). Exploring interactional features with prosodic patterns. *Language and Speech*, 57(3), 285-309.