



Encuentro
de JÓVENES
INVESTIGADORES

CRITERIOS DE SEGMENTACIÓN EN TRANSCRIPCIONES DE ENTREVISTAS A ESCOLARES CON DESARROLLO TÍPICO DEL LENGUAJE Y CON SOSPECHA DE TRASTORNO DE DESARROLLO DEL LENGUAJE (TDL)

Valentina Trossero Ghiberto

Departamento de Letras, FHUC, UNL;

Director: Horacio Dotti.

Área: Humanidades

Palabras claves: segmentación, TDL, corpus.

INTRODUCCIÓN

El presente trabajo se enmarca dentro del proyecto CAI+D *Estudios gramaticales del español y lenguas relacionadas: variación lingüística y problemas de adquisición del lenguaje* y surge a partir de la producción de un corpus de escolares con desarrollo típico y atípico del lenguaje. Entendemos por desarrollo típico aquel que no ha presentado dificultades o retrasos, el que ocurre en la mayoría de los individuos. En esta línea, el tema a desarrollar es la pregunta por los distintos criterios de segmentación que podemos adoptar ante un determinado corpus. Dicho tema se enmarca institucionalmente en una adscripción en investigación que comenzó este año. El corpus consiste en transcripciones de una serie de entrevistas en lengua materna a escolares, niños y adolescentes, con desarrollo típico del lenguaje y con sospecha o diagnóstico clínico de Trastorno Específico del Lenguaje (TEL) o, como actualmente es denominado, Trastorno de Desarrollo del Lenguaje (TDL). A partir del tratamiento del corpus mencionado, surge la pregunta de cuál sería la mejor manera de segmentar las transcripciones, es decir, qué criterios se utilizan para la segmentación en función de los objetivos de investigación. Eso nos permite reflexionar sobre qué propiedades hacen segmentable a los enunciados lingüísticos y sobre las teorías lingüísticas que sustentan un criterio de segmentación. Y, en el caso del corpus del proyecto en curso, también habilita la pregunta acerca de cómo conviene segmentar transcripciones de sujetos en riesgo de TDL o con diagnóstico clínico y de sujetos de control con desarrollo típico.

La importancia de dicha tarea de segmentación reside en que la adopción de criterios y normas, como la elección de las unidades con las que trabajamos, afectan los resultados de las métricas de complejidad sintáctica. La proliferación de los criterios de segmentación alude, entonces, a dar cuenta de esa complejidad considerando las características del corpus que se trabaja, si fue producido con aprendices de una segunda lengua o no, si tiene un origen oral u escrito, entre otros.

Título del proyecto: CRITERIOS DE SEGMENTACIÓN EN TRANSCRIPCIONES DE ENTREVISTAS A ESCOLARES CON DESARROLLO TÍPICO DEL LENGUAJE Y CON SOSPECHA DE TRASTORNO DE DESARROLLO DEL LENGUAJE (TDL)

Instrumento: CAI+D *Estudios gramaticales del español y lenguas relacionadas: variación lingüística y problemas de adquisición del lenguaje*.

Año convocatoria: 2023

Organismo financiador: UNL, FHUC.

Director/a: Dotti, Horacio.



Marco teórico de referencia y metodología

Seguiremos una metodología cualitativa al momento de comparar los supuestos teóricos de las diferentes perspectivas, y se sumará la metodología cuantitativa en la comparación de resultados a partir de la aplicación de los diversos criterios.

El TDL es un trastorno de neurodesarrollo que afecta la adquisición – generando retraso – y el procesamiento del lenguaje en la niñez, persistiendo incluso en la edad adulta. Como exponen Ana Campos y Lorna Halliday (2020):

Diversos términos se han usado para las dificultades que algunos niños experimentan en la adquisición de su idioma nativo, incluyendo Disfasia, Trastorno del Desarrollo del Lenguaje (TDL) y Trastorno Específico del Lenguaje (TEL) (Leonard, 2020). En general, todos se refieren a la misma entidad clínica: dificultades significativas en el desarrollo del lenguaje oral sin una causa médica conocida (Leonard, 2014).” (2020:80)².

La diferencia más clara entre el TEL y el TDL como constructos es que el primero para ser diagnosticado requiere la exclusión de otras afecciones, como la hipoacusia o el Trastorno del Espectro Autista. El TDL, en cambio, puede ser diagnosticado a individuos que además tengan otra afección. Esto responde a las necesidades clínicas prácticas de dar un diagnóstico y de brindar terapias para aquellos afectados por múltiples dificultades. Como el TDL es una reformulación del TEL, se dice que se refieren a una misma entidad clínica.

Con respecto a este proyecto en específico, el corpus que nos compete contiene muestras de entrevistas transcritas con individuos menores de edad en escolarización a los que se les aplicó un test de la batería MABILIN para medir su performance sintáctica. MABILIN es una batería de tests lingüísticos de origen brasilero, pero que ha sido adecuado al español rioplatense (Dotti et al., 2018). Ello posibilita investigar acerca del desarrollo de la actuación sintáctica tardía tanto de los individuos típicos como de aquellos con sospecha de TDL o TDL diagnosticado. Para abordar dichas transcripciones es necesario establecer criterios que nos permitan la segmentación de dicho corpus. En nuestro caso, el estudio sintáctico exige una segmentación en “enunciados” de longitud aproximadamente oracional, y no en párrafos – demasiado extensa –, ni en frases – demasiado breve. Los criterios de segmentación, de esa forma, deberían generar objetos mensurables con una misma métrica para todas las muestras. De esa forma, se obtendrían indicios de la complejidad sintáctica de un texto oral u escrito, pudiendo compararse con otras muestras textuales.

La segmentación de un corpus en unidades requiere de criterios que la organicen y justifiquen. Tanto los intereses teóricos como la naturaleza de los datos impactan en las decisiones a tomar. En ese sentido, una problemática a estudiar es la adecuación de criterios originados en investigaciones con textos escritos con respecto a su aplicación a textos orales transcritos, como acontece en el corpus de escolares con el que trabajamos.

Dicho corpus es producido para investigaciones léxicas y sintácticas, pues se desarrolla partiendo de intereses propios de una psicolingüística generativista, en tanto se analiza la actuación que los hablantes despliegan a partir de sus competencias adquiridas. La *psicolingüística* generativista postula que los hablantes nacen con una facultad lingüística que luego especializan según la lengua en que se crían, y que diferentes hablantes pueden diferir con respecto al desarrollo y uso de dicha facultad. La facultad del lenguaje sería una condición necesaria para que el uso del lenguaje ocurra. La psicolingüística se enfoca, entonces, en las capacidades de uso de la facultad. Por su parte, la *lingüística* generativista atiende a la facultad en sí y su relación interfásica, de límite, con la expresión sonora o gestual y con los

1 Por “conocida”, entiéndase “puntual”.



conceptos e intenciones de los humanos. Está centrada en la sintaxis y, en algunas corrientes, también en la morfología. Por otro lado, las áreas de la lingüística en las cuales se utilizan corpus segmentados corresponde no sólo a la psicolingüística, sino también al aprendizaje de segundas lenguas, en la lingüística computacional y también en el análisis del discurso.

Con este marco disciplinar presente, la unidad de medida que sirve a la segmentación del corpus de escolares es la clásica “unidad-T”, creada por Hunt (1965) para observar el desarrollo de la escritura en escolares, pero que ha sido utilizada para medir la producción oral de individuos con TDL (Nippold, Mansfield y Tomblin, 2009) y de individuos aprendiendo una segunda lengua (Larsen-Freeman, 1983). Este autor oponía su unidad-T a la división en oraciones, la cual percibía como un indicador pobre de desarrollo, especialmente si se demarcaba según la ortografía de cada escolar.

La unidad-T consiste en una cláusula principal con verbo conjugado acompañada de todas sus subordinadas. Hunt (1966) asocia su unidad-T con la por ese entonces llamada gramática transformacional – hoy mayormente conocida como generativa. Las cláusulas, explica, se pueden extender por “reduction and consolidation” (Hunt, 1966:738). Básicamente, reducir una cláusula independiente al estatus de una subordinada de otra, extendiendo ésta última; o si es consolidación, reducir el contenido de una cláusula a un componente de otra – e.g. “Era una ballena blanca. Era una ballena rara.” → “Era una ballena blanca rara”.

La selección de esta unidad recae no sólo en su claridad y coherencia con respecto a la segmentación misma en tanto nos permite pensar la complejidad semántica en escolares sin y con sospecha de TDL, sino también es una unidad muy extendida en el área de la psicolingüística por lo que los podemos comparar esta investigación con antecedentes. Además, también supera el problema de la ambigüedad de las oraciones [sentences], como bien aclara el National Council of Teachers of English en un reporte retomado en el texto de Hunt (1965) “His construction of the T-unit; for example, overcomes the problem of what is or is not a sentence and therefore establishes a base unit that can be systemated” (Hunt, 1965:05). Y también es compatible con herramientas de transcripción como CLAN cuyo acrónimo significa Computerized Language Analysis. Este programa nos permite estudiar la interacción conversacional, el aprendizaje de una lengua, los desórdenes del lenguaje a partir de un conjunto variado de corpora (MacWhinney; 2017:8) e incluye recursos para anotación gramatical y métricas lingüísticas.

Sin embargo, la unidad-T ha sido cuestionada. Si se origina del análisis de textos escritos, ¿podría ser problemático a la hora de trabajar con entrevistas realizadas de forma oral? Partidarios de una segmentación prosódica para estudios de pragmática como Cabedo Nebot (2009), indican la dificultad de establecer límites sonoros claros para generar unidades divisibles “Así pues, la duda reside en el hecho de que si equiparamos pausa a silencio hay que señalar que en el discurso oral no siempre marcamos concienzudamente silencios claros y precisos, es decir, que nuestra separación en grupos fónicos no siempre es precisa.” (2009:35). La objeción de Cabedo Nebot a la noción de oración como unidad de segmentación es aplicable parcialmente también a la unidad-T: si bien en los textos escritos acabados son infrecuentes, las reformulaciones e interrupciones abundan en la oralidad afectando la completitud de las cláusulas a medir. Frecuentemente, la unidad-T en textos orales se conformará de cláusulas interrumpidas y de réplicas escuetas que presuponen lo ya dicho por el interlocutor.

A su vez, otra crítica realizada a esta unidad refiere a su ambigüedad, como plantea Taina Mylläri (2020) “However, the relationship between clauses can sometimes be ambiguous, which makes it hard to determine whether a clause is coordinated or subordinated (Lieko, 1992, pp. 29–31; Quirk et al., 1972, pp. 795–796).” (Mylläri, 2020:15).

Sin embargo, a pesar de estos inconvenientes la autora también señala lo difícil de .





encontrar unidades de segmentación en las cuales los corpora encajen sin excepciones:

“In other words, as Rimmer (2006, p. 508) points out, authentic language does not always fit “into neat pigeon holes”. It is therefore important to explicitly define the production units used and to make visible the exceptions allowed or the amount of data omitted. This information should always be included when reporting research findings.” (Mylläri, 2020:21).

En este sentido, que encontremos ciertas desventajas en la unidad-T no implica que segmentaciones alternativas no las posean también.

Otras unidades que se tuvieron en cuenta y permitieron la comparación con Unidad-t, son, por ejemplo, la propuesta por Cabedo Nebot (2009) que tienen en cuenta criterios prosódicos, la de Crookes (1990) que trabaja también con la noción de enunciado y la expuesta por Lintunen y Mäkilä (2014) que combinan criterios prosódicos y sintácticos.

En base a esta recapitulación, nosotros tomaremos la unidad-T para realizar comparaciones con una muestra reducida del corpus y ver los alcances de dicha definición, teniendo en cuenta si es funcional a las transcripciones que involucran casos de sospecha de TDL y a los objetivos investigativos a los que se orienta el corpus.

OBJETIVOS

- Analizar los distintos criterios de segmentación utilizados en diferentes investigaciones basadas en estudios de corpus.
- Observar los alcances de los criterios de segmentación y sus distintas unidades a la hora de trabajar con casos con sospecha de TDL o TDL diagnosticado.
- Sistematizar los criterios utilizados teniendo en cuenta una perspectiva generativista de la adquisición del lenguaje.

METODOLOGÍA

Se llevará a cabo un relevamiento bibliográfico de estudios realizados acerca de criterios de segmentación y trabajos donde dichas segmentaciones se hayan aplicado, a fin de poder realizar una sistematización de las categorías y conceptos. A su vez, también se realizará la comparación de distintas unidades de medida en términos de ventajas y desventajas, de adecuación, implementación y coherencia.

CONCLUSIONES

Como podemos observar, la segmentación de unidades responde a múltiples criterios, perspectivas, necesidades de investigación y definiciones de complejidad. En la adscripción procuraremos sistematizar algunos criterios y unidades para observar sus alcances desde una perspectiva generativista. En este sentido, la unidad-T presenta los criterios necesarios para trabajar con el corpus en tanto entrevistas orales transcritas, como así también nos permite ver sus alcances a la hora de trabajar con escolares con sospecha de TDL.

BIBLIOGRAFÍA BÁSICA

Cabedo Nebot, A. (2009). *Segmentación prosódica de la conversación coloquial: sobre el grupo entonativo como mecanismo demarcativo de unidades mínimas*. Universitat de València Servei de





Publicacions, Valencia.

Crookes, G. (1990) "The Utterance, and Other Basic Units for Second Language Discourse Analysis" en *Applied Linguistics*, Vol. 11, No. 2. Oxford University Press.

Dotti, H. M., Corrêa, L. M. S., Rivera, G., Benassi, J., Formichelli, M. C. (2018) "Una evaluación de la comprensión de estructuras sintácticas con alto costo de procesamiento en niños en edad escolar" en *Revista Argentina de Ciencias del Comportamiento*, vol. 10, núm. 2, 2018, Mayo-, pp. 37-57. Universidad Nacional de Córdoba.

Halliday, M. A. K. (1963) "The tones of English" en *Archivium Linguisticum* 15: 1-28.

Hunt, K. W. (1965). *Grammatical Structures Written at Three Grade Levels*. NCTE Report No. 3.

Lintunen, P. y Mäkilä, M. (2014). Measuring syntactic complexity in spoken and written learner language: Comparing the incomparable?. *Research in Language*, 12(4), 377-399.

MacWhinney, B. (2017). Tools for analyzing talk part 2: The CLAN program. *Talkbank. Org*, (2000).

Mylläri, T. (2020). Words, clauses, sentences, and T-units in learner language: Precise and objective units of measure?. *Journal of the European Second Language Association*, 4(1).

Nippold, M. A., Mansfield, T. C., Billow, J. L. (2009) "Syntactic Development in Adolescents with a History of Language Impairments: A Follow-Up Investigation" en *American Journal of Speech-Language Pathology*, Vol. 18, 241-251. American Speech-Language-Hearing Association.

Larsen-Freeman, D. (1983) "Assessing global second language proficiency" en H. W. Seliger y M. H. Long (eds.): *Classroom Oriented Research in Second Language Acquisition*. Rowley, Massachusetts. Newbury House.

