

Medidas de longitud en la investigación empírica de ASL

¿Medidas de complejidad sintáctica o de fluidez?¹

Irini Mavrou & María Cecilia Ainciburu

Universidad Antonio de Nebrija | Universidad de Siena y Universidad Antonio de Nebrija

El objetivo del presente estudio fue analizar el controvertido estatus de tres medidas lingüísticas ampliamente utilizadas en la investigación empírica de Adquisición de Segundas Lenguas, como son la longitud promedio de la unidad terminal, la longitud promedio de la cláusula y la longitud promedio de la oración. Sesenta y cinco aprendientes griegos de español como lengua extranjera realizaron dos tareas escritas y su actuación se evaluó con base en un total de 24 medidas de precisión, fluidez y complejidad sintáctica y léxica. Mediante la aplicación del Análisis Factorial Común se pudo constatar que las medidas de longitud (longitud promedio de la unidad terminal y de la cláusula) y de subordinación cargan en factores diferentes y, más que fluidez, dichas medidas evalúan diferentes facetas del constructo de complejidad sintáctica. En cuanto a la longitud promedio de la oración, los resultados sugieren que esta medida se ve influenciada por factores relacionados con su cómputo, con la extensión de los textos analizados, así como por las convenciones lingüísticas que rigen la lengua materna o extranjera.

Palabras clave: complejidad lingüística, fluidez, medidas de longitud, expresión escrita, Análisis Factorial Común

Abstract

The purpose of this study was to examine the controversy surrounding three linguistic measures widely used in the empirical research of Second Language Acquisition, e.g., mean length of T-Unit, mean length of clause, and mean length of utterance. The corpus consisted of two compositions written by a group of 65 Greek learners of Spanish as a foreign language, which were analysed on the basis of 24 linguistic measures of accuracy, fluency, and syntactic and lexical complexity. The results obtained using Common Factor Analysis revealed that measures of mean length (mean length of T-Unit and mean length of clause) and subordination measures load on different factors, and these measures, more than fluency, gauge different aspects of syntactic complexity. On the other hand, mean length of utterance is likely to be influenced by factors related to the method of measurement, the length of the texts being analysed, as well as the linguistic conventions of both one's native and foreign language.

Keywords: linguistic complexity, fluency, measures of mean length, written production, common factor analysis

¹ Accepted version: Mavrou, I., & Ainciburu, M. C. (2019). Medidas de la longitud en la investigación empírica de ASL: ¿Medidas de complejidad sintáctica o de fluidez? *Revista Española de Lingüística Aplicada / Spanish Journal of Applied Linguistics*, 32(1), 125-154. doi:10.1075/resla.15028.mav

This manuscript is under copyright. The publisher should be contacted for permission to re-use the material in any form.

1. Introducción

En el ámbito de Adquisición de Segundas Lenguas (ASL) los constructos de complejidad, precisión y fluidez (CAF; *Complexity, Accuracy, Fluency*) han sido ampliamente estudiados, sobre todo a partir de la década de los 90, y han llegado a convertirse en el principal foco de investigación en ASL (Housen y Kuiken, 2009, p.462). Ahora bien, no constituyen una teoría o un programa de investigación por sí solos (Palloti, 2009), tampoco existe un acuerdo unánime con respecto a su conceptualización, medición empírica e interdependencia (Housen y Kuiken, 2009; Housen, Kuiken, y Vedder, 2012). El interés del presente estudio se centra en la medición de estos constructos y, más precisamente, en esclarecer el controvertido estatus de tres medidas lingüísticas ampliamente utilizadas en la investigación empírica de ASL, a saber, longitud promedio de la unidad terminal o unidad t, longitud promedio de la cláusula y longitud promedio de la oración.

La necesidad de ahondar en la reflexión sobre dichas medidas se deriva de las inconsistencias, detectadas en la bibliografía especializada, relacionadas con el constructo que se pretende evaluar mediante ellas. Por ejemplo, en varios estudios la longitud promedio de la unidad terminal y la longitud promedio de la cláusula se emplearon como indicadores de la complejidad sintáctica (Kormos, 2011; Sotillo, 2000; Torres González, 1993; Torres López, 1999; Treffers-Daller, 2013; Vajjala y Meurers, 2012; Véliz, 1996, 1999), mientras que otros autores trataron estas medidas como propias de la fluidez (Larsen-Freeman, 2006; Mohammadzadeh Mohammadabadi, Dabaghi, y Tavakoli, 2013; Nariman-Jahan y Rahimpour, 2011; Navés, Torras, y Celaya, 2003; Rahimpour y Hosseini, 2010; Salimi y Fatollahnejad, 2012; Torras y Celaya, 2001; Wolfe-Quintero, Inagaki, y Kim, 1998; Yang y Sun, 2015). Una postura diferente mantiene el grupo de investigadores que trabaja dentro del marco del proyecto BAF [*Barcelona Age Factor*] (Muñoz, 2006) quienes sostienen que la longitud promedio de la cláusula y de la oración no deberían ser vistas ni como propias de la fluidez ni de la complejidad sintáctica, sino como características de otro constructo aún no determinado (Celaya, Pérez-Vidal, y Torras, 2000/2001; Navés, 2007).

Teniendo en cuenta las discrepancias señaladas, el presente estudio aborda empíricamente la problemática referida al constructo que evalúan las medidas de longitud a través del análisis de un corpus de producciones escritas de aprendientes

griegos de español como lengua extranjera (ELE). Es cierto que el debate acerca de estas medidas gira en torno a si su uso es más adecuado para operacionalizar la complejidad sintáctica o la fluidez. Ahora bien, la naturaleza multicomponencial de la actuación del aprendiente de lengua extranjera, así como las técnicas estadísticas utilizadas en la presente investigación, han hecho imprescindible el uso de varias medidas lingüísticas correspondientes a todos los constructos de CAF.

2. Conceptualización y medición de los constructos de CAF

Desde la perspectiva psicolingüística de ASL, los constructos de CAF emergen como los componentes principales de la actuación y del desarrollo interlingüístico en una segunda lengua (L2) o lengua extranjera (LE) (Housen y Kuiken, 2009; Housen et al., 2012; Larsen-Freeman, 2009; Wolfe-Quintero et al., 1998), de ahí la necesidad de considerarlos conjuntamente a la hora de evaluar las producciones lingüísticas de los aprendientes de L2/LE.

La complejidad sintáctica se concibe en términos de variedad de formas y estructuras lingüísticas empleadas por el aprendiente de LE en su producción oral o escrita (Ortega, 2003; Wolfe-Quintero et al., 1998) y su evaluación se hace, principalmente, a partir de dos tipos de medidas, las que captan la densidad de subordinación y las medidas de longitud (véase Tabla 1). Los principales componentes de estas medidas son la *unidad terminal o unidad t* y la *cláusula* que se definen y explican en detalle en el siguiente apartado.

La complejidad léxica hace referencia a la cantidad y variedad del vocabulario utilizado por el aprendiente de LE (McCarthy y Jarvis, 2007, 2010). Otros aspectos relacionados con el vocabulario son la sofisticación léxica que se refiere a la dificultad o singularidad de las palabras producidas por el aprendiente, y la densidad léxica que representa la proporción de ítems léxicos (palabras de contenido referencial y semántico) en relación con el número total de palabras o palabras funcionales (Johansson, 2008; Laufer y Nation, 1995; Lu, 2012; Šišková, 2012; Wolfe-Quintero et al., 1998). La Tabla 2 recoge algunas medidas utilizadas en la evaluación de estos aspectos.

Tabla 1. Medidas de complejidad sintáctica*

Medidas de longitud	Medidas de subordinación
Longitud promedio de la unidad terminal*	Cláusulas/Unidades terminales
Longitud promedio de la oración*	Cláusulas dependientes/Unidades terminales
Longitud promedio de la cláusula*	S-nodes/Unidades terminales **
Longitud promedio de la AS-unit***	S-nodes/AS-units

* Las medidas de longitud se marcan con asterisco (*) debido a la controversia suscitada con respecto al constructo que evalúan. Para un mayor esclarecimiento del constructo de complejidad sintáctica y de las prácticas utilizadas para su medición se remite a (1) la revisión de Bulté y Housen (2012) de las medidas de complejidad lingüística utilizadas en 40 estudios que se circunscriben al ámbito de la enseñanza mediante tareas, (2) el reporte técnico de Wolfe-Quintero et al. (1998) orientado a las medidas de complejidad, precisión y fluidez empleadas en 39 estudios dirigidos exclusivamente al discurso escrito, y (3) el estudio de metaanálisis de Ortega (2003; véase también Ortega, 2012). ** Según Kobayashi (2009): “S nodes (sentence-nodes) are ‘clauses, identified by a tensed verb’ (Wesche & Ready, 1985) or they can include infinitives and gerunds as well as tensed verbs (Brock, 1986)” (p.29). *** Foster, Tonkyn y Wigglesworth (2000) definen la AS-unit como “a single speaker’s utterance consisting of an independent clause, or sub-clausal unit, together with any subordinate clause(s) associated with either” (p.365).

Tabla 2. Medidas del léxico*

Constructo	Medidas
Densidad léxica	Ítems léxicos/Palabras Ítems léxicos/ Ítems funcionales
Sofisticación léxica	Advanced Guiraud=Formas avanzadas/ $\sqrt{\text{Palabras}}$ Medida Lambda calculada con el programa P_Lex (Meara y Miralpeix, 2007) Formas sofisticadas/Formas Ítems léxicos sofisticados/Ítems léxicos
Sofisticación verbal	Formas verbales sofisticadas/ Verbos Formas verbales sofisticadas/ $\sqrt{2}$ *Verbos Formas verbales sofisticadas ² / Verbos
Diversidad léxica	Type/Token ratio (TTR) ** Mean Segmental TTR TTR Corregida=Types/ $\sqrt{2}$ *Tokens Root TTR (Índice de Guiraud)=Types/ $\sqrt{\text{Tokens}}$ Índice de Uber= $\text{Log}^2\text{Types}/\text{Log}(\text{Tokens}/\text{Types})$ Medida D calculada con el programa VOCD del <i>Computerized Language Analysis</i> (MacWhinney, 2000) Measure of Textual Lexical Diversity (McCarthy y Jarvis, 2010)
Variación léxica	Formas léxicas/ Ítems léxicos
Variación verbal	Formas verbales/Verbos Formas verbales ² /Verbos Formas verbales/ $\sqrt{2}$ *Verbos

* Para la elaboración de la tabla se ha recurrido a los trabajos de Bulté y Housen (2012), Lu (2012), McCarthy y Jarvis (2010), Roberto, Martí y Salamó (2012), Šišková, (2012) y Wolfe-Quintero et al. (1998). Para otras variaciones de las medidas de densidad, sofisticación y variación léxicas véase Wolfe-Quintero et al. (1998). ** Por *token* se entienden aquellas palabras o ítems utilizados en el discurso oral o escrito. Por otro lado, el término *type* (o forma) alude a las formas únicas que contiene el discurso.

Por precisión lingüística se entiende la habilidad del aprendiente de LE de producir lengua carente de errores tomando como modelo la norma, que suele ser la de un hablante nativo ideal de la lengua objeto (Housen y Kuiken, 2009; Housen et al., 2012; Palloti, 2009; Skehan y Foster, 2001; Wolfe-Quintero et al., 1998). Según Wolfe-Quintero et al. (1998, p.33), la precisión depende de tres fuentes que interactúan: el grado de representación del conocimiento, la fuerza de representaciones que compiten entre sí y el grado de automatización en la producción del lenguaje. En cuanto a la evaluación del constructo, se suele optar por medidas que

toman en cuenta el número total de errores en relación con el número total de unidades terminales, cláusulas o palabras, o medidas que se basan en los conceptos de unidad terminal libre de errores y cláusula libre de errores (véase Tabla 3).

Tabla 3. Medidas de precisión lingüística*

Medidas basadas en unidades libres de errores	Cómputo
Unidades terminales libres de errores (UTLE)	UTLE/Unidades terminales UTLE/Oraciones UTLE/Palabras
Cláusulas libres de errores (CLLE)	CLLE/Cláusulas CLLE/Oraciones CLLE/Unidades terminales
Medidas basadas en la cuantificación de los errores	Cómputo
	Errores/Unidades terminales Errores/Cláusulas Errores específicos/Unidades terminales Errores específicos/Cláusulas Errores/Palabras Errores por cada 100 palabras Errores específicos/Palabras

* Para una descripción detallada de estas medidas véase Wolfe-Quintero et al. (1998).

En lo que concierne al último constructo de CAF, la fluidez, Towell y Dewaele (2005) señalan que en ella se implican variables como la cantidad de conocimiento y el grado en que dicho conocimiento ha sido procedimentado y automatizado de manera explícita. Ahora bien, es preciso señalar que la mayoría de estas conceptualizaciones se han elaborado teniendo como referencia la producción oral.² Por lo que respecta a la fluidez escrita, podría concebirse como:

la habilidad del aprendiente de LE de acceder y recuperar segmentos de su memoria y traducir de manera eficiente su pensamiento dentro de un intervalo de tiempo determinado, tanto en términos de cantidad como de velocidad, y de modo equiparable, aunque no necesariamente idéntico, al de un hablante o escritor competente, nativo o no. (Mavrou, 2016, p.62)

Partiendo de esta definición, el constructo de fluidez puede evaluarse utilizando la frecuencia absoluta de ciertas unidades de análisis del discurso, medidas que toman en cuenta la velocidad de producción y, según autores como Wolfe-Quintero et al. (1998), las medidas de longitud (véase Tabla 4). Son estas últimas las que se han utilizado de manera poco consistente en la investigación empírica de ASL, ya sea para medir la complejidad sintáctica o la fluidez (véase apartado 1), y las que constituyen el foco principal de la presente investigación.

² Schmidt (1992) y Kormos y Dénes (2004) recogen varias definiciones de este constructo.

Tabla 4. Algunas medidas de fluidez escrita propuestas por Wolfe-Quintero et al. (1998)

Medidas de frecuencia absoluta	Palabras Verbos Cláusulas Oraciones Unidades terminales
Medidas de velocidad de producción	Palabras/Minutos
Medidas de longitud	Longitud promedio de la unidad terminal* Longitud promedio de la oración* Longitud promedio de la cláusula*
Medidas de longitud+precisión	Longitud promedio de la unidad terminal libre de errores Longitud promedio de la cláusula libre de errores

* Las medidas de longitud se marcan con asterisco (*) debido a la controversia suscitada con respecto al constructo que evalúan.

3. Conceptualización, medición y problemas metodológicos relacionados con las unidades de análisis del discurso

Los principales constituyentes de las medidas de longitud –objeto del presente estudio–, así como también de la mayoría de las medidas lingüísticas presentadas en el apartado anterior, son la unidad terminal o unidad t (UT) y la cláusula (CL). Estas unidades de análisis del discurso se remontan a Hunt (1965) quien definió la UT como “one main clause with all the subordinate clauses attached to it” (p.36) y la CL como “a structure with a subject and a finite verb (a verb with a tense marker)” (p.31). En cuanto a la relación matemática entre la longitud promedio de la UT (LUT) y la longitud promedio de la CL (LCL), Hunt (1965) señala: “For any body of writing, the average clause length (expressed in words) multiplied by the average number of clauses per T-unit (expressed as a decimal ratio) will exactly equal the average T-unit length (expressed in words)” (p.49).

En la investigación en español como lengua materna (LM) o L2/LE, la UT se define de modo análogo como una CL principal a la que se le añaden otras subordinadas o proposiciones. Se trata de unidades dotadas de sentido que suelen delimitarse con un punto o un signo de interrogación al final, a excepción de las oraciones yuxtapuestas y coordinadas que tienen más de una UT (Checa García, 2005; Torres González, 1997; Torres López, 1999, entre otros).

A pesar de que la UT se considera una medida apropiada para el análisis del discurso escrito (Skehan, 2003), fue criticada por Bardovi-Harlig y Bofman (1988) debido a que, en su definición, contiene la subordinación pero no la coordinación, por lo que las autoras argumentaron a favor del uso de la oración que tiene “a certain degree of psychological reality in that it allows researchers to glimpse, through the

learners's own production, how the learner views the structure of the English sentence" (p.5). Según Bardovi-Harlig y Bofman (1988), un análisis basado en el concepto de oración presenta las siguientes ventajas:

First, a sentence-based analysis better characterizes learner knowledge; second, it facilitates comparison across learners and across stages of second language development; and, third, it encourages dialogue between classroom teachers and researchers by providing a common unit of analysis. (p.5)

En lo que concierne al concepto de CL, su definición entraña ciertas dificultades. Si se adopta la definición de Hunt (1965), se identifica como CL una construcción sintácticamente libre o dependiente con el verbo en forma personal, mientras que otros autores consideran en su definición el uso de las formas no personales del verbo salvo aquellas utilizadas en las perífrasis verbales (Bardovi-Harlig y Bofman, 1988; Checa García, 2005; Torres González, 1993, 1997, entre otros). En este caso, en lugar de CL, se suele utilizar el término S-nodes.

De la conceptualización de la CL dependen en gran medida los criterios utilizados en la segmentación de las muestras del discurso. Por ejemplo, en el estudio de Torres López (1999) las formas no personales que funcionaban como elementos verbales o llevaban sus propios complementos se trataron como una CL más. Por el contrario, si funcionaban como nombre, adjetivo y adverbio (en el caso del infinitivo, del participio y del gerundio respectivamente) se consideraron incorporadas en la oración o la CL a la que pertenecían (véase también Torres González, 1993, pp.54-55). En cambio, Martínez Arbeláiz (2004) no consideró como CL aquellos fragmentos en los que las palabras 'cuando' y 'mientras' aparecían sin verbo conjugado.

Las diferentes conceptualizaciones de las unidades de análisis del discurso no constituyen en primera estancia un problema metodológico, ya que el investigador puede optar por una de estas y justificar su decisión. Sin embargo, surgen ciertas inconsistencias a la hora de comparar los resultados obtenidos en diferentes estudios puesto que diferentes conceptualizaciones de las UT y CL arrojarán valores diferentes tanto de las LUT y LCL como del Índice de Subordinación ($IS=CL/UT$). A modo de ejemplo, se puede considerar la siguiente oración compuesta por 4 palabras (PAL): 'Pienso ir al teatro'. Si se sigue el criterio de Hunt, la LCL equivale a $LCL=4PAL/1CL=4$ y el $IS=1CL/1UT=1$. Por otro lado, si se toman en cuenta en la definición de la CL las formas no personales del verbo, $LCL=4PAL/2CL=2$,

$IS=2CL/1UT=2$. En ambos casos coincide solo el valor de la LUT:
 $LUT=PAL/UT=4/1=4$.

Bulté y Housen (2012, pp.39-40) sostienen que la definición de la CL proporcionada por Hunt es más operativa pero tiene un alcance limitado ya que, al calcularse la LCL, se ignora el uso de construcciones con formas no personales [*nonfinite constructions*] o de cláusulas con sujeto elíptico [*clauses with subject ellipsis*]. Sin embargo, se debe aclarar que la elipsis del sujeto no siempre provoca confusiones lingüísticas. Por ejemplo, en idiomas como el inglés la ausencia del sujeto puede dificultar la comprensión del mensaje; por el contrario, en otros idiomas como el español o el griego esa elipsis no solo es aceptada sino que, en ocasiones, resulta pragmáticamente más adecuada. En cuanto a las definiciones del concepto de CL que tienen en cuenta las formas no personales del verbo, Bulté y Housen (2012, p.39) las consideran como lingüísticamente más válidas porque respetan la integridad lingüística del discurso que se produce en los niveles no nativos más avanzados en los que se observa un aumento en el uso de proposiciones y participios (véase también Wolfe-Quintero et al., 1998, pp.73-74).

Asimismo, debe señalarse que las formas no personales abundan en el discurso de aprendientes que poseen un nivel de competencia lingüística bajo en la LE sea por desconocimiento de ciertas reglas gramaticales o de los mecanismos de vinculación sintáctica que rigen la lengua meta. De igual manera, un abuso de las formas no personales del verbo podría deberse a la transferencia del conocimiento que se dispone de otros idiomas. A modo de ejemplo, se reproducen a continuación dos fragmentos, ambos extraídos del corpus utilizado en el presente estudio:

- Ejemplo (1): ‘Dijo a su mujer sentirse* [=sentarse]’ (Informante 6)
- Ejemplo (2): ‘Le pidió que se sentara’ (Informante 25)

Si se toman en cuenta las formas no personales, ambos fragmentos contienen 2CL y 5PAL, por lo que la LCL de ambos equivale a 2.5. Sin embargo, el segundo fragmento es completamente correcto, adecuado y más elaborado sintácticamente. En cuanto al primer fragmento, podría reflejar el discurso de un aprendiente que desconoce el modo subjuntivo o que domina el inglés y transfiere este conocimiento al español, suposición esta que se justifica por la similitud que presentan los siguientes fragmentos:

- ‘Dijo a su mujer sentarse’ (español, incorrecto)
- ‘*He asked her wife to sit down*’ (inglés, correcto)

Una última cuestión a la que debe prestarse atención es la referente al uso selectivo de medidas lingüísticas en la mayoría de los estudios empíricos de ASL y a la redundancia metodológica cuando estas evalúan la misma faceta de un constructo determinado. Por ejemplo, para la complejidad sintáctica se suele optar por una o dos medidas y estas suelen ser medidas que evalúan la densidad de subordinación. Para entender mejor las implicaciones que dicha selección puede tener, es posible considerar el siguiente caso: al comparar dos grupos (por ejemplo, hablantes nativos y no nativos, aprendientes con diferente LM) respecto al IS y obtener valores de significación estadística mayores que .05, se llegaría a la conclusión de que los dos grupos provienen de la misma población. Sin embargo, aun si los dos grupos utilizan la misma cantidad de subordinación, probablemente uno de ellos utilice un mayor número de elementos dentro de las CL o UT, hecho que no se percibe satisfactoriamente con medidas como el IS. Por otro lado, si la LCL se mantiene constante, las posibles diferencias en cuanto al valor del IS podrían atribuirse a un mayor o menor uso de oraciones subordinadas. De ahí la necesidad de emplear tanto medidas de longitud como de subordinación y examinar sus valores conjuntamente teniendo en cuenta la relación matemática $LUT=IS \times LCL$ (Hunt, 1965).

4. Análisis Factorial Exploratorio: una herramienta útil al servicio del investigador de ASL

El Análisis Factorial Exploratorio (AFE) es una técnica estadística que permite explorar los constructos o variables latentes de las variables medidas u observadas por el investigador. A pesar de que el AFE se utiliza extensamente en Psicometría, en ASL ha tenido una menor difusión debido, probablemente, al tipo de variables y diseños metodológicos con los que se suele trabajar en este ámbito (Mavrou, 2015). Ahora bien, dicha técnica es especialmente valiosa para esclarecer cuestiones como las planteadas en la presente investigación, esto es, explorar qué constructo de CAF (variables latentes) representan mejor las medidas de longitud (variables medidas). Además, se trata de una técnica relativamente potente, ya que puede revelar patrones de agrupación de variables que no se detectan fácilmente mediante una simple correlación (por ejemplo, es posible que haya una correlación moderada entre dos variables o medidas lingüísticas sin que ambas representen necesariamente el mismo constructo).

Según Wolfe-Quintero et al. (1998), si bien el AFE había sido utilizado para el estudio de medidas de competencia lingüística, este no fue el caso para las medidas de CAF. A pesar de que han pasado casi dos décadas desde la publicación de su reporte técnico y aunque las medidas de CAF han sido herramientas de análisis del discurso utilizadas en innumerables investigaciones de ASL, apenas encontramos en ellas un tratamiento estadístico basado en los métodos del AFE.

Wolfe-Quintero et al. (1998) citan un estudio llevado a cabo por Ortega en 1995 y resumen sus resultados de la siguiente manera:

The association of length measures with fluency was also suggested by Ortega (1995) in her study of the speech of 32 learners. In a factor analysis, Ortega found that length (words per utterance) was strongly related to rate (syllables per second), and that both were part of one factor presumed to be fluency, but not related to factors associated with lexical complexity or grammatical accuracy. (Wolfe-Quintero et al., 1998, p.14)

Skehan y Foster (1997) utilizaron el Análisis de Componentes Principales (uno de los métodos del AFE) pero interpretaron sus resultados en términos de *trade-offs* entre los constructos de CAF.³ Además, la matriz de factores que aparece en su artículo no proporciona una imagen clara de cómo cargaron las medidas lingüísticas que utilizaron para elicitar los constructos de CAF en los tres factores obtenidos tras aplicar el AFE.

Celaya, Pérez-Vidal y Torras (2000/2001) optaron por el Análisis de Componentes Principales a fin de determinar qué medidas lingüísticas evalúan de manera más adecuada la competencia escrita de alumnos bilingües (catalán-castellano), aprendientes de inglés como LE. Para ello utilizaron un total de 17 medidas, 10 de complejidad, 5 de fluidez y 2 de corrección. Aunque las autoras describen los resultados de su análisis, la omisión de los valores numéricos (valores descriptivos de las medidas utilizadas, magnitud de las correlaciones y de las cargas factoriales, indicadores de adecuación muestral, etc.) no permiten apreciar de manera más exhausta sus resultados.

En un estudio más reciente, Vercellotti (2012) utilizó modelos lineales jerárquicos con el objetivo de examinar el desarrollo y los posibles *trade-offs* entre los constructos de CAF en la producción oral de aprendientes de inglés como L2. Aunque la investigadora proporciona las correlaciones entre las 9 medidas lingüísticas utilizadas en su estudio, la interpretación de dichas correlaciones resulta ardua debido

³ El término *trade-off* alude al mejor rendimiento con respecto a una dimensión lingüística concreta (complejidad, precisión y fluidez) en detrimento de otra(s).

a que correlaciones de una magnitud de .2 fueron estadísticamente significativas al nivel .01 al igual que las correlaciones de una magnitud de aproximadamente .8. Es por ello que el AFE se considera más potente que la correlación, ya que permite reducir la dimensionalidad de los datos en un conjunto menor de factores o componentes.

¿A qué se debe el escaso uso del AFE en los estudios en los que se han empleado las medidas de CAF? Existen al menos dos respuestas plausibles para esta pregunta. En primer lugar, la aplicación del AFE requiere un número elevado de medidas lingüísticas, hecho que, probablemente, acarrearía la redundancia en la medición. Además, la inclusión de varias medidas de un constructo determinado puede arrojar resultados confusos, difíciles de sintetizar o interpretar, e incluso contradictorios. En segundo lugar, parece que, en la mayoría de los estudios sobre CAF, la selección de las medidas lingüísticas se basa en una mera revisión bibliográfica, dando prioridad al uso de aquellas medidas propuestas por otros autores o las más frecuentemente utilizadas en la investigación empírica de ASL. Esto es inevitable, pero también necesario ya que facilita la comparabilidad entre diferentes estudios. Sin embargo, no permite esclarecer cuestiones importantes como la adecuación y validez de ciertas medidas para evaluar determinados constructos. Al adoptar una perspectiva exploratoria, el presente estudio quiere avanzar en esta dirección, centrándose específicamente en las medidas de longitud.

5. Metodología

En función de lo expuesto en el apartado anterior, la técnica multivariante del AFE se consideró como la opción más adecuada para abordar los objetivos del estudio. Es preciso aclarar que, a pesar del interés puesto en las medidas de longitud, fue imprescindible incluir un amplio número de medidas lingüísticas por dos razones: (a) la actuación del aprendiente de LE no es un constructo monolítico sino multifacético, que debe evaluarse considerando sus diferentes facetas de manera conjunta; (b) conseguir soluciones factoriales relativamente estables a la hora de aplicar el AFE requiere, entre otros, una ratio de variables por factor de al menos 4:1 (para una discusión más detallada véase Mavrou, 2015). Al utilizar 24 medidas lingüísticas, se consideró que dicho supuesto se cumpliría de manera satisfactoria.

5.1 Participantes

La muestra del estudio estuvo compuesta por 65 aprendientes griegos de ELE, inscritos en cursos del Instituto Cervantes de Atenas y del Centro de Idiomas Extranjeros de la Universidad Kapodistriaca de Atenas. Todos tenían un nivel intermedio de ELE (B2), según los criterios establecidos mediante los exámenes de nivelación de los alumnos. Todos eran hablantes nativos de griego y su edad oscilaba entre los 20 y los 38 años ($M=26.26$, $SD=4.94$).

5.2 Tareas escritas

Dos tareas narrativas basadas en material gráfico sirvieron de estímulo para la expresión escrita.⁴ Skehan (2009) apunta que las narrativas están más orientadas por el *input* y menos negociables, ya que requieren el uso de un vocabulario determinado. Ello hace que se consideren demandantes a nivel de codificación lingüística (Kormos, 2011; Kormos y Trebits, 2012), aunque no necesariamente a nivel de conceptualización del mensaje. Además, este tipo de tareas acarrea menos problemas de fiabilidad en cuanto a su evaluación, en comparación con las tareas abiertas que, con frecuencia, provocan que el aprendiente recurra a un vocabulario que ya domina (Weir, 2005).

Para los propósitos del estudio, se manipularon las condiciones de actuación de las dos tareas. En concreto, los informantes realizaron la primera (T1) bajo presión de tiempo (10 minutos y sin planificación previa) y la segunda (T2) bajo condiciones relativamente más favorables de actuación (20 minutos y sin planificación previa), si bien, en ambos casos, se les pidió a los informantes que utilizaran tiempos verbales del pasado.

5.3 Procedimiento de recogida y tratamiento de los datos

El procedimiento de recogida de datos tuvo lugar en los correspondientes centros de estudio de ELE de los informantes y la administración de las dos tareas se hizo utilizando la técnica de contrabalanceo entresujeto (la mitad de los informantes

⁴ La primera tarea se puede encontrar en Mavrou (2013, p.91) y la segunda en Gilabert (2005, p.399).

realizaron las tareas en el orden T1T2 y la otra mitad en el orden T2T1, con una pausa de aproximadamente 5 minutos entre tareas). Se proporcionaron instrucciones *a priori* en español y, en caso necesario, se ofrecieron aclaraciones también en griego para evitar posibles confusiones o malentendidos. Además, se informó a los participantes de que no podían consultar manuales o apuntes de clase, tampoco podían sobrepasar el tiempo establecido para la realización de las tareas.

Después de obtener el corpus de redacciones escritas, su análisis se realizó a partir de las 24 medidas lingüísticas recogidas en la Tabla 5. Para el tratamiento cuantitativo de los datos fue necesario tomar una serie de decisiones que se detallan en la continuación.

Del cálculo de las palabras se excluyeron los nombres propios de persona, nombres de ciudades, parques, programas televisivos, números, interjecciones y abreviaturas. Por lo que respecta a los ítems léxicos (palabras de contenido referencial y semántico), y siguiendo las pautas adoptadas en otros estudios de ASL que examinaron los constructos de CAF (Gilbert, 2005; Lu, 2012; Vajjala y Meurers, 2012) se incluyeron en el análisis: sustantivos, adjetivos, verbos independientemente de si estaban bien conjugados o no, adverbios con una base adjetival incluidos los que pueden funcionar tanto como un adjetivo como un adverbio (por ejemplo, ‘rápido’) y los que se forman uniendo el sufijo *-mente* a una raíz adjetival (por ejemplo, ‘concretamente’). Por otro lado, no se incluyeron los verbos modales, los verbos ser, estar y haber, las palabras de clase cerrada (preposiciones, interjecciones, pronombres, etc.), los adverbios que no terminan en *-mente*, y las palabras señor, señora, señorita. Las palabras que diferían en cuanto a su morfología flexiva se contaron como *tokens* del mismo *type*, mientras que las que diferían con respecto a su morfología derivativa se consideraron como *tokens* de diferente *type*. Para el cálculo de *types* y *tokens* y la lematización del corpus se utilizaron las herramientas V_Words v2.0 (Meara y Miralpeix, 2016) y Grampal (Moreno, 1991; Moreno y Goñi, 1995).

Tabla 5. Medidas lingüísticas utilizadas para evaluar la actuación de los informantes en las dos tareas*

Constructos	Medidas lingüísticas
Complejidad sintáctica	1. Índice de Subordinación 1: $IS1=CL/UT$
	2. Índice de Subordinación 2: $IS2=S-nodes/UT$
	3. Índice de Subordinación 3: $IS3=CL/OR$
	4. Complejidad sintáctica 1: $CS1=CLdep/UT$
	5. Complejidad sintáctica 2: $CS2=CLdep/CL$
Complejidad léxica	6. TTR Corregida: $TTRCorr=Types/\sqrt{2*PAL}$
Diversidad léxica	

Variación léxica	8. Variación léxica 1: $VL1 = \text{Typeslex} / \text{PALlex}$
	9. Variación léxica 2: $VL2 = \text{Typeslex} / \sqrt{2} * \text{PALlex}$
Densidad léxica	10. Densidad léxica: $DL = \text{PALlex} / \text{PAL}$
	11. $UTLE.UT = UTLE / UT$
	12. $CLLE.CL = CLLE / CL$
	13. Porcentaje de precisión lingüística: PPL
Precisión lingüística	14. Precisión lingüística: $PL = \text{Errores} / UT$
	15. UTLE
	16. CLLE
	17. PAL
	18. UT
Fluidez	19. CL
	20. S-nodes
	21. OR
	22. Longitud promedio de la UT: $LUT = \text{PAL} / UT$
Medidas de longitud	23. Longitud promedio de la CL: $LCL = \text{PAL} / CL$
	24. Longitud promedio de la OR: $LOR = \text{PAL} / OR$

* UT=Unidades terminales; CL=Cláusulas; OR=Oraciones; CLdep=Cláusulas dependientes; PAL=Palabras; lex=léxicos; UTLE=UT libres de errores; CLLE=CL libres de errores.

En cuanto a los errores, se contabilizaron las incorrecciones léxicas, los errores debidos a la interferencia de la LM u otras LE, las expresiones inventadas, los errores que afectaban a las estructuras gramaticales, la elección equivocada del tiempo verbal, los errores de concordancia de género y número, los problemas de concordancia entre determinantes y adjetivos y el correspondiente sustantivo, las confusiones entre verbos reflexivos y no reflexivos, así como la ausencia de tildes en las desinencias verbales de los verbos en pasado. En cambio, no se contaron como errores las alteraciones en el orden de las palabras.

La codificación de los errores se llevó a cabo por dos personas, la primera investigadora y una profesora nativa de español con experiencia en la enseñanza de ELE. La confiabilidad interevaluador se obtuvo mediante el Coeficiente de Correlación Intraclase (CCI) (modelo matemático de efectos mixtos, acuerdo absoluto, medidas individuales) y fue de $CCI=.850$, 95%IC [.577, .932] para la T1 y $CCI=.897$ 95%IC [.588, .960] para la T2.

Por lo que se refiere a las medidas que se basan en la frecuencia absoluta de ciertas unidades de análisis (medidas 15-21), Wolfe-Quintero et al. (1998) sostienen que su uso es deseable siempre que se imponga un límite de tiempo para el desempeño de la tarea o conceptual por el tipo y las características de esta. Ambos criterios se cumplen en el contexto de la presente investigación. Asimismo, a pesar de que se prescindió de medidas de velocidad de producción, en líneas generales, se constató que los informantes utilizaron todo el tiempo que se les había concedido para llevar a cabo las tareas.

6. Resultados

Las Tablas 6 y 7 presentan las medidas descriptivas de los indicadores utilizados para evaluar la actuación de los informantes en las T1 y T2 respectivamente.

Tabla 6. Descripción de las medidas lingüísticas utilizadas en la T1

	M	SD	Asimetría	Curtosis
T1IS1	1.39	0.22	1.29	2.14
T1IS2	1.67	0.32	1.31	2.44
T1IS3	2.25	0.51	0.65	1.28
T1CS1	0.40	0.22	1.18	2.07
T1CS2	0.27	0.10	0.15	0.23
T1TTRCorr	4.06	0.43	0.76	0.63
T1VL1	0.77	0.08	-0.18	-0.32
T1VL2	3.63	0.48	0.43	0.30
T1IU	16.21	2.52	1.22	2.23
T1DL	0.42	0.03	0.42	0.98
T1PPL	85.93	5.75	-0.23	-0.93
T1PL	1.28	0.61	0.91	1.40
T1UTLE.UT	0.33	0.17	0.48	0.08
T1CLLE.CL	0.41	0.18	0.17	-0.60
T1UTLE	4.38	3.01	1.15	1.14
T1CLLE	7.40	4.55	0.81	0.17
T1PAL	111.71	27.65	1.11	1.26
T1UT	12.71	3.62	0.67	0.44
T1CL	17.37	4.67	0.63	0.08
T1S-nodes	20.85	5.52	0.45	-0.26
T1OR	8.03	2.66	1.24	1.24
T1LUT	9.11	2.05	0.94	1.29
T1LCL	6.56	1.09	0.82	0.99
T1LOR	14.59	3.33	0.15	-0.40

En primer lugar, se examinaron las medidas de asimetría y de curtosis. Como se puede apreciar en las Tablas 6 y 7, de los 96 valores obtenidos (24 de asimetría y 24 de curtosis por tarea) 67 estaban dentro del intervalo de $|-1 +1|$, hecho indicativo del cumplimiento del supuesto de normalidad. En cuanto a los demás valores, 22 se encontraban dentro del intervalo de $|-2 +2|$ y 7 fueron mayores que 2 (solo curtosis).

Tabla 7. Descripción de las medidas lingüísticas utilizadas en la T2

	M	SD	Asimetría	Curtosis
T2IS1	1.61	0.27	0.52	0.27
T2IS2	1.89	0.42	1.14	2.15
T2IS3	2.53	0.54	0.60	-0.21
T2CS1	0.61	0.25	0.46	0.30
T2CS2	0.36	0.10	-0.45	0.23
T2TTRCorr	3.97	0.38	0.31	-0.30
T2VL1	0.71	0.06	-0.04	0.06
T2VL2	3.70	0.48	0.73	0.36
T2IU	14.41	1.44	0.25	-0.45
T2DL	0.40	0.03	0.73	1.79
T2PPL	86.14	5.15	-0.69	0.78
T2PL	1.52	0.56	0.61	-0.16

T2UTLE.UT	0.23	0.16	0.48	-0.36
T2CLLE.CL	0.38	0.17	-0.08	-0.40
T2UTLE	3.06	2.42	1.08	1.25
T2CLLE	7.94	4.48	0.72	0.29
T2PAL	139.23	36.93	1.30	2.37
T2UT	12.82	3.71	0.67	0.76
T2CL	20.34	5.86	0.60	0.69
T2S-nodes	23.78	6.84	0.57	1.20
T2OR	8.29	2.71	1.13	2.41
T2LUT	11.23	2.57	1.03	0.99
T2LCL	7.01	1.23	0.57	0.29
T2LOR	17.60	4.34	0.99	1.02

Con el objetivo de valorar si estas desviaciones del supuesto de normalidad eran severas, los valores de asimetría y de curtosis se dividieron entre su error típico, y el resultante valor tipificado [*z-score*] fue utilizado para evaluar la significación estadística (Fidell y Tabachnick, 2003; Field, 2009). Con este procedimiento se constató que 13 valores tipificados de asimetría (medidas: T1IS1, T1IS2, T1CS1, T1IU, T1PL, T1UTLE, T1PAL, T1OR, T2IS2, T2UTLE, T2PAL, T2OR, T2LUT) y 7 valores tipificados de curtosis (medidas: T1IS1, T1IS2, T1CS1, T1IU, T2IS2, T2PAL, T2OR) eran mayores que 3.29. No obstante, dado que el número total de estos valores fue pequeño, no se consideró necesario proceder a transformaciones adicionales [*data normalization*].

A continuación, se aplicó la técnica multivariante del AFE.⁵ Se efectuaron dos modelos, uno por tarea, utilizando el método del Análisis Factorial Común (AFC) debido a que proporciona valores menos inflados en comparación con el método del Análisis de Componentes Principales (Conway y Huffcutt, 2003; Fabrigar, Wegener, MacCallum, y Strahan, 1999). La extracción de los factores se realizó mediante la Factorización de Ejes Principales por ser un método del AFC relativamente blindado a violaciones del supuesto de normalidad (Fabrigar et al., 1999, p.277), y se optó por la rotación oblicua (*oblimin directo*) ya que permite obtener soluciones precisas, simples, reproducibles y más realistas (Conway y Huffcutt, 2003; Costello y Osborne, 2005; Fabrigar et al., 1999). En cuanto a la determinación del número de los factores que iban a ser retenidos, se llevó a cabo utilizando el gráfico de sedimentación y el porcentaje de la varianza total explicada, estableciendo un límite del 80% (Hair, Black, Babin, y Anderson, 2010; Rietveld y van Hout, 1993).

⁵ Es necesario aclarar que, si bien el uso del AFE es recomendable con tamaños muestrales relativamente grandes, de Winter, Dodou y Wieringa (2009) señalan que este no debe ser el único criterio para utilizar o descartar dicha técnica, y que su aplicación con tamaños muestrales moderados y en una fase de investigación exploratoria podría revelar patrones latentes valiosos.

Con el fin de llegar a soluciones factoriales interpretables, fue necesario eliminar la medida T1IS2 (solo primer modelo), así como las medidas de densidad léxica (ambos modelos), puesto que las correlaciones entre las T1DL y T2DL, por un lado, y el resto de las medidas de complejidad léxica, por otro, fueron débiles o insignificantes.⁶ En lo que se refiere a este último punto, es oportuno mencionar que en dos AFE que se efectuaron (uno por tarea), utilizando solo los indicadores que evaluaban el léxico, las medidas de densidad léxica cargaron en un factor distinto al de las demás medidas de complejidad léxica. Por tanto, se decidió prescindir de las T1DL y T2DL ya que su inclusión podría haber comprometido la plausibilidad de los modelos del AFE. Sin embargo, se incluyeron las medidas LUT y LCL –a pesar de que, como se verá más adelante, la primera acabó siendo una variable compleja– con el objetivo de profundizar en el constructo que estas representan al ser, precisamente, el objetivo del presente estudio.

Las Tablas 8 y 9 presentan los modelos del AFC efectuados para las T1 y T2 respectivamente. De los resultados obtenidos en relación con el primer modelo se destaca que el índice Kaiser-Meyer-Olkin obtuvo un valor mayor que .50 ($KMO=.595$) y la prueba de esfericidad de Bartlett fue significativa ($\chi^2(231)=2982.07$, $p < .001$), hecho que sugiere una adecuación muestral aceptable y buena factorabilidad. El Factor 1 se formó por las medidas PPL, PL, UTLE.UT y CLLE.UT, medidas similares del mismo constructo subyacente, la precisión lingüística. En este mismo factor cargaron también las medidas UTLE y CLLE, aunque sus cargas factoriales fueron ligeramente más bajas en comparación con las cargas de las demás medidas de precisión, probablemente, porque las UTLE y CLLE son medidas de frecuencia absoluta y no ratios. En el Factor 2 saturaron las medidas de complejidad sintáctica, siendo el IS1 la medida que presentó la saturación factorial más alta ($\lambda=.981$). El Factor 3 se formó por las medidas de fluidez PAL, S-nodes, UT y CL, presentando las dos primeras cargas más elevadas en comparación con las dos restantes. Las medidas IU, VL1, V2, TTRCorr formaron el Factor 4 de complejidad léxica. El IU parece ser la medida más representativa del constructo en cuestión, al menos en relación con los demás indicadores de complejidad léxica utilizados. Dicha

⁶ Las correlaciones entre las medidas de complejidad léxica en la T1 fueron las siguientes: T1DL-T1TTRCorr: $r=.115$, $p=.363$; T1DL-T1VL1: $r=-.043$, $p=.732$; T1DL-T1VL2: $r=.176$, $p=.162$; T1DL-T1IU: $r=.254$, $p=.041$. En cuanto a la T2, se obtuvieron las siguientes correlaciones: T2DL-T2TTRCorr: $r=-.062$, $p=.624$; T2DL-T2VL1: $r=-.270$, $p=.030$; T2DL-T2VL2: $r=.015$, $p=.906$; T2DL-T2IU: $r=.009$, $p=.941$.

medida saturó solo en el factor de complejidad léxica y de manera fuerte ($\lambda=.925$). Por otro lado, las medidas VL1, VL2 y TTRCorr, además del factor de complejidad léxica (Factor 4), saturaron también en el factor de fluidez (Factor 3) y sus cargas factoriales en este último factor fueron moderadas ($.40 < \lambda < .50$).

Tabla 8. Análisis Factorial Común para las medidas lingüísticas utilizadas en la T1*

Medidas	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	h^2 final
T1PPL	.977						.875
T1CLLE.CL	.966						.937
T1UTLE.UT	.964						.905
T1PL	-.876						.863
T1CLLE	.754						.950
T1UTLE	.741						.897
T1IS1		.981					.948
T1CS1		.977					.998
T1CS2		.884					.902
T1PAL			.926				.977
T1S-nodes			.921				.928
T1CL			.872				.992
T1UT			.825				.989
T1OR			.676			-.600	.946
T1IU				.925			.864
T1TTRCorr			.410	.834			.965
T1VL1			-.414	.805			.768
T1VL2			.478	.736			.912
T1LCL					.965		.986
T1LUT		.625			.671		.987
T1IS3						.900	.991
T1LOR						.889	.990
% Varianza explicada	35.27	20.56	13.03	11.55	7.28	5.82	
Autovalor	7.76	4.52	2.87	2.54	1.60	1.28	
% Varianza acumulada			94.64 %				
Satur. al cuadrado de la rotación	6.13	4.12	5.57	3.17	2.68	2.86	
Kaiser-Meyer-Olkin			.595				
Test de Bartlett		$\chi^2=2982.07$	$gl=231$	$p < .001$			

* En negrita se marcan las cargas factoriales mayores que .60.

La LUT y la LCL se situaron en un factor distinto (Factor 5; medidas de longitud), aunque la LUT resultó una variable compleja al presentar cargas elevadas y casi idénticas en cuanto a su magnitud ($.60 < \lambda < .70$) tanto en el Factor 2 de complejidad sintáctica como en el Factor 5 de medidas de longitud. En cambio, la LCL cargó fuertemente solo en el Factor 5 ($\lambda=.965$). Finalmente, el Factor 6 se formó por dos medidas, la longitud promedio de la oración (T1LOR) y el número total de CL dividido entre el número total de oraciones (T1IS3). En este mismo factor saturó también el número total de oraciones (T1OR). Sin embargo, al igual que la LUT, la LOR resultó una variable compleja, puesto que cargó también en el factor de fluidez

(Factor 3), y la magnitud de las cargas factoriales en ambos casos fue muy similar ($.60 < \lambda < .70$). Dado que el denominador común de las medidas que cargaron en el Factor 6 fue el concepto de oración, se decidió denominarlo ‘medidas basadas en oración’.

Tabla 9. Análisis Factorial Común para las medidas lingüísticas utilizadas en la T2*

Medidas	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	h^2 final
T2PAL	.994						.875
T2CL	.890						.937
T2UT	.884						.905
T2S-nodes	.869						.863
T2OR	.792					-.473	.950
T2VL2	.712				-.509		.897
T2IS1		.969					.948
T2CS1		.959					.998
T2IS2		.890					.902
T2CS2		.884					.977
T2UTLE.UT			-.932				.928
T2CLLE.CL			-.908				.992
T2PL			.907				.989
T2PPL			-.906				.946
T2UTLE			-.816				.864
T2CLLE			-.735				.965
T2LCL				.971			.768
T2LUT		.610		.691			.912
T2IU					-.900		.986
T2VL1					-.766		.987
T2TTRCorr	.595				-.677		.991
T2IS3						.932	.990
T2LOR				.425		.844	
% Varianza explicada	31.34	22.96	16.36	9.13	6.95	5.65	
Autovalor	7.21	5.28	3.76	2.10	1.60	1.30	
% Varianza acumulada			94.64 %				
Satur. al cuadrado de la rotación	5.98	5.01	5.13	2.63	2.73	2.80	
Kaiser-Meyer-Olkin			.594				
Test de Bartlett		$\chi^2=3267.34$	$gl=253$	$p < .001$			

* En negrita se marcan las cargas factoriales mayores que .60.

Los resultados arrojados por el segundo modelo del AFC (Tabla 9) fueron casi idénticos. Además, puede observarse que el número total de oraciones (T2OR) cargó fuertemente en el Factor 1 de fluidez ($\lambda=.792$) y débilmente en el Factor 6 de ‘medidas basadas en oración’ ($\lambda=.473$), resultado que parece sugerir que dicha medida más bien capta la fluidez del discurso. Las diferencias en la magnitud de las cargas factoriales podrían atribuirse a las condiciones de implementación de las tareas (T1 realizada bajo presión de tiempo y T2 bajo condiciones relativamente más óptimas de actuación y, tal vez, más realistas al menos cuando se trata de tareas escritas en LE).

En cuanto a la medida T2VL2, cargó en el factor de fluidez (Factor 1) de manera más fuerte ($\lambda=.712$) que en el Factor 5 de complejidad léxica ($\lambda=.509$), resultado que podría explicarse por el hecho de que el denominador de la ratio de la medida VL2 considera el doble de palabras ($VL2=Typeslex/\sqrt{2}*PALlex$). Por último, la LUT de nuevo saturó en dos factores, el Factor 2 de complejidad sintáctica ($\lambda=.610$) y el Factor 4 de medidas de longitud ($\lambda=.691$), en el que solo cargó fuertemente la LCL ($\lambda=.971$).

7. Discusión

Las inconsistencias detectadas en la investigación empírica de ASL en cuanto al constructo que representan las medidas LUT, LCL y LOR, y su empleo en diferentes estudios para medir diferentes constructos (complejidad sintáctica o fluidez) constituyeron el punto de partida del presente estudio.

Los resultados obtenidos mostraron que la LCL tiene un comportamiento diferente al de las demás medidas de complejidad sintáctica analizadas. Dos interpretaciones pueden considerarse plausibles. Por un lado y de acuerdo con Navés (2007), la LCL representa probablemente un constructo diferente al de la complejidad sintáctica o al de la fluidez. Por otro lado, es razonable asumir que la LCL evalúa una faceta muy específica del constructo de complejidad sintáctica (véase taxonomía de Bulté y Housen, 2012) y, por eso, cargó en un factor distinto. Esta afirmación se refuerza por el hecho de que la mayoría de las medidas de complejidad sintáctica utilizadas en el presente estudio evaluaban la densidad de subordinación (IS1, IS2, CS1, CS2), por lo que su agrupación en el mismo factor era esperable.

No obstante, es oportuno hacer una observación adicional. La longitud promedio de la CL en las T1 y T2 fue $M=5.8$ ($Rango=4.0-9.5$) y $M=6.4$ ($Rango=4.6-10.4$) palabras respectivamente. El punto interesante en relación con estos valores es que la LCL mínima fue alrededor de 4 segmentos, número que coincide con la cantidad de ítems que pueden mantenerse en el foco de atención de la memoria operativa (véase Cowan, 2001; Unsworth y Engle, 2006, 2007).

El estudio de Robinson (1997, citado por DeKeyser, 2016) mostró que la adquisición de estructuras complejas –que requieren la búsqueda y uso de reglas gramaticales determinadas en LE– se predijo por la capacidad de memoria. Towell y

Dewaele (2005) remiten a Temple quien sugiere que, durante los procesos de producción en LE, la memoria operativa se ocupa de almacenar y coordinar los fragmentos que se van a producir y que, dadas las limitaciones de capacidad de este mecanismo, “parallel processing breaks down and is replaced by serial processing with word-by-word or phrasal type of production (Temple, 1997, 2000)” (Towell y Dewaele, 2005, p.217). Hasta qué punto la LCL o, lo que es lo mismo, el factor que fue representado por esta medida, se podría definir atendiendo solo a factores lingüísticos inherentes a la estructura de la lengua meta o considerando también las capacidades cognitivas del aprendiente, como la capacidad de la memoria operativa, es una cuestión que la investigación futura debería examinar en mayor profundidad.

En cuanto a la LUT, a diferencia de Wolfe-Quintero et al. (1998) y Larsen-Freeman (2006) quienes trataron la LUT como medida de fluidez, los resultados del presente estudio sugieren que dicha medida evalúa una faceta del constructo de complejidad sintáctica, aunque probablemente no constituye una medida ‘pura’ de este. Es posible que la LUT refleje algo más general, como el nivel de competencia lingüística global del aprendiente de LE (véase Bulté y Housen, 2012). Por otro lado, la pregunta que surge es qué comparten las medidas LUT y LCL, puesto que ambas cargaron en el mismo factor.

Nuevamente, se puede afirmar que el número de segmentos que se incluyen en un fragmento pequeño (CL) o más largo (UT) depende de la capacidad residual de la memoria operativa para llevar a cabo procesos simultáneos de almacenamiento, repaso mental y manipulación de estos segmentos. Alternativamente, se puede especular que la longitud de los fragmentos utilizados en un texto depende de factores externos o individuales. Por ejemplo, algunas personas escribirán fragmentos más largos sea por preferencia estilística (Holmes, 1985), porque están dispuestas a asumir más riesgos al menos cuanto se trata de una LE (Skehan y Foster, 2001), o simplemente porque las oraciones largas son aceptables según las convenciones que rigen el discurso escrito en diferentes idiomas (por ejemplo, en comparación con el inglés, el español académico admite oraciones más largas).

Otro resultado interesante del presente estudio concierne a las medidas que se basan en el concepto de oración, en concreto, la LOR y el IS3. Ambas medidas cargaron en el mismo factor y este patrón de resultados se obtuvo en los dos modelos del AFC efectuados. Además, solo en este caso una medida de longitud de una unidad de análisis del discurso concreta, la oración, cargó en el mismo factor con una medida

de subordinación calculada a partir de la misma unidad de análisis ($IS3=CL/OR$). Por tanto, la saturación de estas medidas bajo el mismo factor podría deberse a factores relacionados con su cómputo o, tal vez, a factores ajenos como las convenciones que rigen la escritura en la LM o la LE y la transferencia de los patrones oracionales de una lengua a la otra. Asimismo, coincidimos con Larsen-Freeman (2009) que cuestionó la utilidad y aplicabilidad de la LOR en la investigación empírica de ASL.

While the MLU [Mean Length Utterance] is admittedly problematic in certain respects, even this limited way of calibrating development does not work well in the case of the acquisition of second or foreign languages, where learners are more cognitively mature and where they rely a great deal on formulaic utterances. (Larsen-Freeman, 2009, p.579)

Por otro lado, no se debe obviar una limitación importante del presente estudio, el hecho de que la extensión de los textos analizados fue bastante reducida. Según Ortega (2012), “mean length of utterance is a good index of grammatical development in a first language only at a gross level of interpretation and when calculated on samples containing a minimum length of 50 utterances” (p.141). El umbral mínimo de 50 oraciones constituye, probablemente, un requisito imprescindible al examinar la producción lingüística también en una LE.

Aunque el principal foco del estudio fueron las medidas LUT, LCL y LOR, también se obtuvieron algunos resultados interesantes con respecto a las medidas de complejidad léxica. En concreto, en el presente estudio las transformaciones algebraicas de la ratio de *type/token* (TTRCorr) y de variación léxica (VL2) presentaron cargas elevadas tanto en el factor de complejidad léxica como en el factor de fluidez. Además, la medida VL2 calculada a partir de la T2 cargó solo en el factor de fluidez. Estos resultados corroboran lo sostenido por Wolfe-Quintero et al. (1998), que estas medidas no se pueden considerar como ‘puras’ debido a la transformación aplicada al número total de *tokens*. Asimismo, refuerzan la opinión de varios autores sobre la sensibilidad de ciertas medidas a la longitud de los textos analizados (Dewaele y Pavlenko, 2003; McCarthy y Jarvis, 2007, 2010; Richards, 1987; Richards y Malvern, 2000; Wolfe-Quintero et al., 1998) y parece que esto no se puede evitar incluso cuando se trata de corpus lematizados.

Por otro lado, se podría afirmar con cierto grado de seguridad que el Índice de Uber constituye una medida fiable de la complejidad léxica de textos breves, como los que produjeron los informantes del presente estudio, conclusión a la que llegó

también Jarvis (2001, en Dewaele y Pavlenko, 2003, pp.129-130), mientras que la VL1 podría ser una buena alternativa para evaluar la variación léxica de estos textos.

Por último, es necesario remarcar algunas cuestiones metodológicas relacionadas con la técnica multivariante del AFE. Aunque el tamaño muestral del presente estudio fue moderado, las comunalidades de las variables fueron altas. Además, se obtuvieron cargas factoriales elevadas ($\lambda > .60-.70$) y las saturaciones de los ítems en los factores estudiados fueron fuertes. Sin embargo, es también cierto que el número de variables de dos factores fue insuficiente (menos de tres ítems) (Costello y Osborne 2005; Fabrigar et al., 1999). En cuanto a las variables complejas que se detectaron, así como algunas variables redundantes, se recomienda en futuros estudios realizar los mismos modelos del AFE excluyendo las primeras y creando puntajes compuestos para las segundas.

8. Conclusiones

La evidencia recogida en este estudio parece confirmar la idoneidad de ciertas medidas lingüísticas frente a otras a la hora de operacionalizar determinados constructos. Sin embargo, no logra resolver definitivamente la problemática acerca del constructo que evalúan las medidas LUT y LCL. Además, el análisis realizado suscita nuevos interrogantes en lo referente al constructo que representan aquellas medidas que se basan en el concepto de oración.

Efectivamente, una serie de limitaciones obliga a considerar los resultados del presente estudio con precaución y establece barreras a la posibilidad de generalización de estos. En concreto, la segmentación de los textos en UT, CL y OR y el cálculo de los errores se hicieron en función de los textos producidos y teniendo en cuenta los problemas lingüísticos encontrados en estos. Por tanto, la extrapolación de estos procedimientos a otros géneros discursivos o a informantes de una LM diferente, e incluso la implementación de otro tipo de tareas bajo condiciones de actuación distintas, deberían suponer un replanteamiento y modificación del método de análisis y tratamiento cuantitativo.

Asimismo, dado que en el presente estudio se consideraron principalmente las medidas genéricas de los constructos de complejidad, precisión y fluidez, sería de especial utilidad incluir en nuevas investigaciones un mayor número de medidas específicas de estos constructos. Por ejemplo, además de la complejidad sintáctica, se

pueden considerar medidas que evalúen la complejidad de ideas y de contenido, así como medidas que capten la gravedad y el tipo de los errores. Del mismo modo, en futuros estudios sería recomendable el análisis de textos más extensos, procedentes de informantes con un nivel de competencia lingüística mayor o menor, o bien una réplica del mismo estudio con informantes de una LM diferente.

La consideración conjunta de los parámetros mencionados sería de gran valor para superar los problemas relacionados con el nivel de determinación de los factores y la presencia de variables complejas, al menos cuando el tratamiento estadístico incluye técnicas como el AFE, así como para llegar a conclusiones más firmes sobre el constructo que evalúan los indicadores basados en la longitud promedio de las unidades de análisis UT, CL y OR.

Referencias bibliográficas

- Bardovi-Harlig, K., y Bofman, T. (1988, March). *A second look at T-unit analysis*. Comunicación presentada en 22nd Annual Meeting of the Teachers of English to Speakers of Other Languages, Chicago, IL. Disponible en: <http://files.eric.ed.gov/fulltext/ED295478.pdf>
- Bulté, B., y Housen, A. (2012). Defining and operationalising L2 complexity. En A. Housen, F. Kuiken, y I. Vedder (Eds.), *Dimensions of L2 performance and proficiency. Complexity, accuracy and fluency in SLA* (pp. 21–46). Amsterdam, The Netherlands: John Benjamins. <https://doi.org/10.1075/llt.32.02bul>
- Celaya, M. L., Pérez-Vidal, C., y Torras, M. R. (2000/2001). Matriz de criterios de medición para la determinación del perfil de competencia lingüística escrita en inglés (LE). *RESLA*, 14, 87–98.
- Checa García, I. (2005). Medidas de madurez sintáctica aplicadas a lecturas de ELE. *Interlingüística*, 16(1), 273–285.
- Conway, J. M., y Huffcutt, A. I. (2003). A review and evaluation of exploratory factor analysis practices in organizational research. *Organizational Research Methods*, 6(2), 147–168. <https://doi.org/10.1177/1094428103251541>
- Costello, A. B., y Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, 10(7). Disponible en: <http://pareonline.net/pdf/v10n7.pdf>
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1), 87–115. <https://doi.org/10.1017/S0140525X01003922>
- DeKeyser, R. (2016). Of moving targets and chameleons: Why the concept of difficulty is so hard to pin down. *Studies in Second Language Acquisition*, 38(2), 353–363. <https://doi.org/10.1017/S0272263116000024>
- Dewaele, J. M., y Pavlenko, A. (2003). Productivity and lexical diversity in native and non-native speech: A study of cross-cultural effects. En V. Cook (Ed.), *Effects of the second language on the first* (pp. 120–141). Clevedon, UK: Multilingual Matters.

- de Winter, J. C. F., Dodou, D., y Wieringa, P. A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research*, 44, 147–181. <https://doi.org/10.1080/00273170902794206>
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272–299. <https://doi.org/10.1037/1082-989X.4.3.272>
- Fidell, L. S., y Tabachnick, B. G. (2003). Preparatory data analysis. En I. B. Weiner (Ed. -in-Chief), J. A. Schinka, y W. F. Velicer (Vol. Eds.), *Handbook of psychology: Research methods in psychology* (Vol. 2, pp. 115–141). New York: Wiley.
- Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London, UK: SAGE Publications.
- Foster, P., Tonkyn, A., y Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21(3), 354–375. <https://doi.org/10.1093/applin/21.3.354>.
- Gilabert, R. (2005). *Task complexity and L2 narrative oral production* (Tesis doctoral inédita, Universitat de Barcelona). Disponible en: <http://hdl.handle.net/2445/35007>
- Hair, J. F. Jr., Black, W. C., Babin, B. J., y Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Upper Saddle River, NJ: Prentice Hall.
- Holmes, D. I. (1985). The analysis of literary style: A review. *Journal of the Royal Statistical Society, Series A (General)*, 148(4), 328–341. <https://doi.org/10.2307/2981893>
- Housen, A., y Kuiken, F. (2009). Complexity, accuracy, and fluency in Second Language Acquisition. *Applied Linguistics*, 30(4), 461–473. <https://doi.org/10.1093/applin/amp048>
- Housen, A., Kuiken, F., y Vedder, I. (2012). Complexity, accuracy and fluency: Definitions, measurement and research. En A. Housen, F. Kuiken, y I. Vedder (Eds.), *Dimensions of L2 performance and proficiency. Complexity, accuracy and fluency in SLA* (pp. 1–20). Amsterdam, The Netherlands: John Benjamins. <https://doi.org/10.1075/llt.32.01hou>
- Hunt, K. W. (1965). *Grammatical structures written at three grade levels*. NCTE Research Report No. 3. Champaign, IL: National Council of Teachers of English.
- Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: A developmental perspective. *Lund Working Papers in Linguistics*, 53, 61–79.
- Kobayashi, M. (2009). *Hitting the mark: How can text organisation and response format affect reading test performance?* New York: Peter Lang.
- Kormos, J. (2011). Task complexity and linguistic and discourse features of narrative writing performance. *Journal of Second Language Writing*, 20, 148–161. <https://doi.org/10.1016/j.jslw.2011.02.001>
- Kormos, J., y Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32(2), 145–164. <https://doi.org/10.1016/j.system.2004.01.001>
- Kormos, J., y Trebits, A. (2012). The role of task complexity, modality, and aptitude in narrative task performance. *Language Learning*, 62(2), 439–472. <https://doi.org/10.1111/j.1467-9922.2012.00695.x>

- Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics*, 27(4), 590–619. <https://doi.org/10.1093/applin/aml029>
- Larsen-Freeman, D. (2009). Adjusting expectations: The study of complexity, accuracy, and fluency in Second Language Acquisition. *Applied Linguistics*, 30(4), 579–589. <https://doi.org/10.1093/applin/amp043>
- Laufer, B., y Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307–322 <https://doi.org/10.1093/applin/16.3.307>
- Lu, X. (2012). The Relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal*, 96(2), 190–208. <https://doi.org/10.1111/j.1540-4781.2011.01232.x>
- MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Martínez Arbeláiz, A. (2004). Índices de progreso en la producción escrita de estudiantes de español en situación de inmersión. *Revista Electrónica de Lingüística Aplicada*, 3, 115–145.
- Mavrou, I. (2013). Precisión lingüística y complejidad sintáctica: ¿hasta qué punto compiten entre sí por los recursos limitados del aprendiz? *Revista Nebrija de Lingüística Aplicada a la Enseñanza de Lenguas*, 14, 75–91.
- Mavrou, I. (2015). Análisis factorial exploratorio: Cuestiones conceptuales y metodológicas. *Revista Nebrija de Lingüística Aplicada a la Enseñanza de Lenguas*, 19, 71–80. <https://doi.org/10.26378/rnlael019283>
- Mavrou, I. (2016). Complejidad, precisión, fluidez y léxico: Una revisión. *Moderna språk*, 110(1), 49–69.
- McCarthy, P. M., y Jarvis, S. (2007). Vocd: A theoretical and empirical evaluation. *Language Teaching*, 24(4), 459–488. <https://doi.org/10.1177/0265532207080767>
- McCarthy, P. M., y Jarvis, S. (2010). MTL-D, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42(2), 381–392. <https://doi.org/10.3758/BRM.42.2.381>
- Meara, P. M., y Miralpeix, I. (2007). *Vocabulary size estimator*. Swansea: Lognostics.
- Meara, P. M., y Miralpeix, I. (2016). *Tools for vocabulary research*. Bristol, UK: Multilingual Matters.
- Mohammadzadeh Mohammadabadi, A., Dabaghi, A., & Tavakoli, M. (2013). The effects of simultaneous use of pre-planning along +/-Here-and-Now dimension on fluency, complexity, and accuracy of Iranian EFL learners' written performance. *International Journal of Research Studies in Language Learning*, 2(3), 49–65. <https://doi.org/10.5861/ijrsl.2012.168>
- Moreno, A. (1991). *Un modelo computacional para el análisis y generación de la morfología del español* (Tesis doctoral inédita, Universidad Autónoma de Madrid). Disponible en: <https://repositorio.uam.es/handle/10486/12294>
- Moreno, A., y Goñi, J. M. (1995). GRAMPAL: a morphological processor of Spanish implemented in Prolog. En M. Alpuente y M. I. Sessa (Eds.), *Proceedings of Joint Conference on Declarative Programming (GULP-PRODE 95)* (pp. 321–331). Salerno, Italy: Palladio Editrice.
- Muñoz, C. (2006). The effects of age on foreign language learning: The BAF Project. En C. Muñoz (Ed.), *Age and the rate of foreign language learning* (pp. 1–40). Clevedon, UK: Multilingual Matters. <https://doi.org/10.21832/9781853598937-003>

- Nariman-Jahan, R., y Rahimpour, M. (2011). The effects of planning and proficiency on language production of writing task performance. *Educational Research*, 2(9), 1528–1537.
- Navés, T. (2007, Abril). *Analytical measures of learners' written interlanguage*. Comunicación presentada en el XXV Congreso Internacional de AESLA, Universidad de Murcia, España.
- Navés, T., Torras, M. R., y Celaya, M. L. (2003). Long-term effects of an earlier start. En S. Foster-Cohen y S. Pekarek (Eds.), *EUROSLA-Yearbook. Annual Conference of the European Second Language Association* (pp. 103–130). Amsterdam, The Netherlands: John Benjamins.
- Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24(4), 492–518. <https://doi.org/10.1093/applin/24.4.492>
- Ortega, L. (2012). Interlanguage complexity: A construct in search of theoretical renewal. En B. Szmrecsanyi y B. Kortmann (Eds.), *Linguistic complexity: Second language acquisition, indigenization, contact* (pp. 127–155). Berlin: de Gruyter. <https://doi.org/10.1515/9783110229226.127>
- Palloti, G. (2009). CAF: defining, refining and differentiating constructs. *Applied Linguistics*, 30(4), 590–601. <https://doi.org/10.1093/applin/amp045>
- Rahimpour, M., y Hosseini, P. (2010). The impact of task complexity on L2 learners' written narratives. *English Language Teaching*, 3(3), 198–205. <https://doi.org/10.5539/elt.v3n3p198>
- Richards, B. (1987). Type/token ratios: what do they really tell us? *Journal of Child Language*, 14(2), 201–209. <https://doi.org/10.1017/S0305000900012885>
- Richards, B., y Malvern, D. (2000, September). *Measuring vocabulary richness in teenage learners of French*. Comunicación presentada en British Educational Research Association Conference, Cardiff University.
- Rietveld, T., y van Hout, R. (1993). *Statistical techniques for the study of language and language behaviour*. Berlin, Germany: De Gruyter Mouton.
- Roberto, J. A., Martí, M. A., y Salamó, M. (2012). Análisis de la riqueza léxica en el contexto de la clasificación de atributos demográficos latentes. *Procesamiento de Lenguaje Natural*, 48, 97–104.
- Salimi, A., y Fatollahnejad, S. (2012). The effects of strategic planning and topic familiarity on Iranian intermediate EFL learners' written performance in TBLT. *Theory and Practice in Language Studies*, 2(11), 2308–2315. <https://doi.org/10.4304/tpls.2.11.2308-2315>
- Schmidt, R. W. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition*, 14(4), 357–385. <https://doi.org/10.1017/s0272263100011189>
- Šišková, Z. (2012). Lexical richness in EFL students' narratives. *University of Reading Language Studies Working Papers*, 4, 26–36.
- Skehan, P. (2003). Task-based instruction. *Language Teaching*, 36(1), 1–14. <https://doi.org/10.1017/S026144480200188X>
- Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30(4), 510–532. <https://doi.org/10.1093/applin/amp047>
- Skehan, P., y Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research*, 1(3), 185–211. <https://doi.org/10.1177/136216889700100302>

- Skehan, P., y Foster, P. (2001). Cognition and tasks. En P. Robinson (Ed.), *Cognition and second language instruction* (pp. 181–205). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139524780.009>
- Sotillo, S. M. (2000). Discourse functions and syntactic complexity in synchronous and asynchronous communication. *Language Learning & Technology*, 4(1), 82–119.
- Torras, M. R., y Celaya, M. L. (2001). Age-related differences in the development of written production. An empirical study of EFL school learners. *International Journal of English Studies*, 1(2), 130–126.
- Torres González, A. N. (1993). *Madurez sintáctica en estudiantes no universitarios de la zona metropolitana de Tenerife* (Tesis doctoral, Universidad de La Laguna, Tenerife). Disponible en: <http://dialnet.unirioja.es/servlet/tesis?codigo=463>
- Torres González, A. N. (1997). Complejidad sintáctica en escritores de las islas canarias. *REALE*, 8, 105–125
- Torres López, M. C. (1999). Índices primarios de madurez sintáctica: análisis comparativo en alumnos de COU. *REALE*, 12, 93–118.
- Towell, R., y Dewaele, J. M. (2005). The role of psycholinguistic factors in the development of fluency amongst advanced learners of French. En J. M. Dewaele (Ed.), *Focus on French as a foreign language: Multidisciplinary approaches* (pp. 210–239). Clevedon, UK: Multilingual Matters.
- Treffers-Daller, J. (2013). Measuring lexical diversity among L2 learners of French: an exploration of the validity of D, MTL D and HD-D as measures of language ability. En S. Jarvis y M. Daller (Eds.), *Vocabulary knowledge: Human ratings and automated measures* (pp. 79–104). Amsterdam, The Netherlands: John Benjamins. <https://doi.org/10.1075/sibil.47.05ch3>
- Unsworth, N., y Engle, R. W. (2006). Simple and complex memory spans and their relation to fluid abilities: Evidence from list-length effects. *Journal of Memory and Language*, 54(1), 68–80. <https://doi.org/10.1016/j.jml.2005.06.003>
- Unsworth, N., y Engle, R. W. (2007). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, 133(6), 1038–1066. <http://dx.doi.org/10.1037/0033-2909.133.6.1038>
- Vajjala, S., y Meurers, D. (2012, June). *On improving the accuracy of readability classification using insights from second language acquisition*. Comunicación presentada en 7th Workshop on Innovative Use of NLP for Building Educational Applications, The Association for Computational Linguistics, Montréal, Canada.
- Véliz, M. (1996). Recurrencia textual y competencia narrativa en escolares de educación básica y media. *REALE*, 5, 107–126.
- Véliz, M. (1999). Complejidad sintáctica y modo del discurso. *Estudios Filológicos*, 34, 181–192.
- Vercellotti, M. L. (2012). *Complexity, accuracy, and fluency as properties of language performance: The development of the multiple subsystems over time and in relation to each other* (Tesis doctoral inédita, University of Pittsburgh). Disponible en: http://d-scholarship.pitt.edu/12071/1/Vercellotti_CAF_v3.pdf
- Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. Basingstoke, Hampshire, UK: Palgrave Macmillan.
- Wolfe-Quintero, K., Inagaki, S., y Kim, H. Y. (1998). *Second language development in writing: Measures of fluency, accuracy & complexity (Technical Report 17)*.

Honolulu, HI: University of Hawai'i, Second Language Teaching and Curriculum Center.

Yang, W., y Sun, Y. (2015). Dynamic development of complexity, accuracy and fluency in multilingual learners' L1, L2 and L3 writing. *Theory and Practice in Language Studies*, 5(2), 298–308. <https://doi.org/10.17507/tpls.0502.09>