

ANÁLISIS INFORMATIZADO DE ERRORES: ENWIL

CARMEN VALERO
GUZMÁN MANCHO
CARMEN FLYS
ESPERANZA CERDÁ

mcarmen.valero@uah.es
guzman.mancho@uah.es
carmen.flys@uah.es
esperanza.cerda@uah.es

Universidad de Alcalá

Resum. **Anàlisi informatitzada d'errors: ENWIL.** La comunicació en la segona llengua evidencia l'estat d'interllengua de l'aprenent mitjançant els errors produïts. L'aplicació de la informàtica a l'anàlisi dels errors venç les limitacions pròpies dels plantejaments clàssics d'aquesta eina descriptiva d'ASL. Tant és així, que les anàlisis informatitzades d'errors (AIE) faciliten la descripció tan comprensiva com es vulgui dels estadis de la interllengua de la població examinada. Aquest és el propòsit de l'AIE anomenat *English Written Interlanguage* (ENWIL) que presentem en aquest assaig: el tractament computeritzat dels errors comesos als nivells lingüístics i derivats de l'acció d'escriure. Un estudi preliminar compara dos estadis de la interllengua de la nostra població, arran del seu grau d'efectivitat comunicativa i precisió gramatical i com aquestes s'han desenvolupat. Per últim, s'assenyala que ENWIL s'integra dins de la línia d'investigació cada vegada més activa dels AIE; ho demostren projectes internacionals tals com ICLE, COALA, etc.

Paraules clau: interllengua, adquisició de segones llengües, anàlisi d'errors.

Abstract. **Computerised error analysis: ENWIL.** The analysis of errors committed in second language communication can illustrate the stage of the learner's interlanguage. The application of computer technology to error analysis is useful in overcoming some of the inherent limitations of the classical approach to this SLA descriptive tool. In fact, Computerized Error Analysis (CEA) can facilitate a comprehensive description of each stage of the subjects' interlanguage development. This is precisely the purpose of the *English Written Interlanguage* project (ENWIL): the computerization of different types of errors committed while writing. This preliminary study compares two interlanguage stages of our subjects on the basis of their degree of communicative efficiency and grammatical accuracy and how these develop. This ENWIL project, together with others such as ICLE or COALA, illustrate that CEA is an active and fast developing area of research.

Key words: interlanguage, second language acquisition, error analysis.

1. Introducción: Análisis de errores y adquisición de segundas lenguas

La Adquisición de Segundas Lenguas (ASL *Second Language Acquisition* (SLA)) es en nuestros días un área de investigación de interés creciente y necesaria para la enseñanza de lenguas extranjeras. Podríamos definirla a grandes rasgos como una propuesta científica cuyo objetivo principal es llegar a conocer los mecanismos según los cuales los aprendices adquieren una segunda lengua (SL) y la posterior aplicación de los resultados a la enseñanza y/o aprendizaje de las lenguas. Desde esta perspectiva el Análisis de Errores (AE), como una subárea de ASL, se convierte en una propuesta metodológica interesante para describir el proceso de esta adquisición lingüística.

La descripción de dicho proceso nos lleva al estudio del sistema lingüístico del alumno o su *interlingua* (IL). Ésta puede definirse como una gramática mental en la que el aprendiz internaliza el conjunto de reglas de uso de la SL y se caracteriza, según Ellis [3, p. 33–34], por ser de naturaleza individual; construirse a partir de la comprensión y producción en la SL; ser permeable, es decir, sobre ella se ejercen tanto influencias externas (de *input*) como internas (estrategias de aprendizaje); ser transitoria, es decir, se crean, añaden u omiten nuevas reglas a lo largo del proceso de aprendizaje; y ser también fosilizable ya que hay determinadas estructuras, vocablos, sonidos o reglas que se fijan en la mente del aprendiz y son difíciles de modificar aún cuando no se ajusten al sistema de la SL.

En definitiva la *interlingua* se articula en la mente de cada aprendiz según la naturaleza variable de las reglas de uso de la SL y se manifiesta en la producción en dicha lengua. Las desviaciones o errores cometidos en su uso muestran el grado de diferenciación entre la IL y el sistema de la SL, ya sea de un solo sujeto o de un colectivo y a lo largo del proceso de aprendizaje de la SL. De este modo, el análisis de errores constituye una herramienta valiosa para estudiar el proceso de adquisición de la SL, a la vez que constituyen una prueba del carácter variable de la IL de los aprendices.

2. Análisis informatizado de errores (AIE)

El Análisis de Errores (AE) como metodología de estudio sigue siendo, sin embargo, polémico dado que existen posturas distintas. Si bien la comunidad científica está de acuerdo de manera general en que los errores cometidos en la SL aportan datos importantes en cuanto a ASL, existen voces detractoras a la hora de utilizar AE de manera discriminada en la investigación o en

el aula. Sus críticas se fundamentan en ciertas limitaciones, por un lado de orden metodológico y, por otro lado, de orden temático. Dagneaux, Dennes, & Granger [1, p. 164] las resumen del siguiente modo:

Limitaciones de orden metodológico: AE está basado en datos de aprendices heterogéneos y las unidades analíticas de los errores no están muy bien delimitadas.

Limitaciones de orden temático: AE no puede considerar el fenómeno de AVOIDANCE (es decir, AE no sistematiza las unidades analíticas que el aprendiz evita utilizar); AE está restringido a lo que el aprendiz NO puede realizar (es decir, no considera lo que SÍ puede realizar); y AE ofrece una visión estática del proceso de aprendizaje de SL.

A pesar de estas limitaciones, AE sigue constituyendo una línea de investigación importante en ASL y recientemente, con la incorporación de la informática a la metodología tradicional del AE, han surgido nuevas propuestas tales como el *Análisis Informatizado de Errores* (AIE), área a la que pertenece el presente trabajo.

AIE se fundamenta en la conjunción de la tecnología informática y la metodología tradicional de AE. Su carácter innovador consiste en la posibilidad de poder disponer y procesar gran cantidad de datos o errores en muy poco tiempo, a partir de los cuales pueden extraerse conclusiones para la elaboración de materiales u otros fines. Esta asociación en realidad se fundamenta en la unión de la Lingüística Computacional, disciplina que basa su análisis en corpora de lenguaje natural computerizado, y en la enseñanza de segundas lenguas, asociación que ha dado ya sus primeros frutos con la elaboración de gramáticas descriptivas y de programas automáticos de corrección. Un buen ejemplo de las primeras es *Collins Cobuild English Grammar* (1996), elaborada a partir del tratamiento informatizado de ejemplos reales de la lengua inglesa y, en el caso de los segundos, los programas de corrección incorporados en los tratamientos de texto Microsoft Word para Windows.

No hay que olvidar, sin embargo, algunos problemas que plantea el uso de este material didáctico por parte de los aprendices de una SL (inglés en este caso). Las gramáticas descriptivas hacen generalmente hincapié en las reglas de uso de la L2, pero la selección y el orden de la información gramatical y discursiva no suelen estar organizados de acuerdo con la interlengua de nuestros aprendices o alumnos.

En cuanto a los programas de corrección, tal y como Granger y Meunier opinan [5], éstos son de escaso uso para el aprendiz de la L2 porque los errores que los aprendices producen difieren ampliamente de los errores que cometen los hablantes nativos, y no se suelen incluir en los programas de corrección.

En definitiva, estos materiales van destinados a un público que utiliza la L2 como un hablante nativo. De ello se desprende la necesidad de elaborar materiales didácticos, y/o programas automáticos de corrección destinados a los aprendices de una SL y adaptados a su interlingua.

Desde nuestro punto de vista el AIE constituye el primer paso para llevar a cabo esta tarea, ya que este tipo de análisis permite, en primer lugar, la catalogación y computerización de información extensiva de producción real del aprendiz de SL. En segundo lugar, el procesamiento de dicha información favorece la sistematización y conocimiento de la evolución de la interlingua de los aprendices objeto de estudio. Y, en tercer lugar, los resultados obtenidos darán cuenta de las áreas o aspectos de la lengua en la que los alumnos tienen dificultades o necesitan un refuerzo especial. Y, finalmente, con estos datos se pasa a la elaboración de material didáctico adecuado para el tipo de aprendiz de esta SL.

Siguiendo esta línea innovadora hemos desarrollado ENWIL.

3. English Written Interlanguage (ENWIL)

3.1. OBJETIVOS

English Written Interlanguage (ENWIL) es un proyecto de investigación que se viene desarrollando en el Departamento de Filología Moderna de la Universidad de Alcalá desde 1997. Su objetivo principal es la descripción de la interlingua escrita de estudiantes de la carrera de Filología Inglesa de dicha universidad de la promoción 1996–2000, con el propósito de diseñar materiales didácticos orientados a este tipo de aprendices.

Nuestro interés en el registro escrito en la SL en el marco de estudios superiores se basa fundamentalmente en la siguiente consideración: esta actividad no sigue un modelo único sino que está determinada por el sistema de la lengua que se usa y por unas convenciones culturales específicas. Es decir, la presencia o ausencia de determinadas convenciones gramaticales y comunicativas en la actividad de escribir contribuyen en mayor o menor medida a conferir respectivamente precisión gramatical y fluidez comunicativa al texto escrito. Por tanto, dado que uno de los objetivos de la enseñanza de lenguas extranjeras es formar comunicadores competentes, consideramos importante el análisis de textos escritos en inglés de los que hoy son nuestros alumnos, pero que mañana serán profesionales, con el fin de conocer las áreas de conflicto de su IL, y, en definitiva, concienciarles del uso de una serie de variables tales como el uso correcto de las normas gramaticales, la elección adecuada de

vocablos, la importancia del contexto, el desarrollo lógico del texto, la consideración de ciertas convenciones de puntuación, así como de las expectativas culturales del público receptor. Con el fin de que la enseñanza sea más eficaz habrá que incidir en aquellos aspectos en los que el alumno tiene dificultades. Un modo de conocer estas dificultades es a través del análisis de sus textos escritos, aspecto que pretendemos cubrir con ENWIL.

3.2. METODOLOGÍA SEGUIDA POR ENWIL

La metodología científica de ENWIL se fundamenta en las líneas de investigación de la Lingüística Computacional, si bien nuestro corpus no consiste en lenguaje natural computerizado, sino que nos basamos en lenguaje real. Para ello hemos diseñado un programa informatizado según el *software ACCESS* que parte del tratamiento de textos *Microsoft Word*: mientras éste actúa de soporte para editar el texto, aquél hace funciones de administrador y gestor de los errores marcados.

El método de trabajo de ENWIL es el siguiente: selección e introducción de los textos en soporte informático; identificación y localización de los errores cometidos en el texto; y generación de informes. Tal labor se facilita con el uso de la Barra de Herramientas, creada específicamente para ENWIL. Los pasos seguidos son:

En primer lugar, tras la *selección* al azar de una serie de textos producidos por los alumnos como tarea, procedemos a su *introducción* en soporte informático mediante el tratamiento de Textos Word.

Una vez introducido el texto, el siguiente paso es la *identificación* de los errores cometidos. Para esta tarea, ENWIL facilita la selección de un texto preparado para la identificación de sus errores ya que la pantalla base muestra una barra de herramientas donde se encuentran las cinco taxonomías de errores derivadas de los siguientes niveles descriptivos: errores en el nivel morfosintáctico (MS), errores en el nivel léxico-semántico (L), errores en el nivel discursivo (D), errores en el nivel ortográfico (SP) y errores en el nivel de puntuación (P).

En la elaboración de nuestra clasificación hemos tenido en cuenta las numerosas clasificaciones de errores morfológicos y sintácticos ya existentes (ver resumen en Dulay *et al.* [2]; Fernández [4]), así como otras clasificaciones más recientes en las que se tienen en cuenta aspectos que van más allá de la oración, es decir, aspectos discursivos (ver Haden [6], Fernández [4]). Y sobre todo hemos considerado aquellos errores cometidos por nuestros alumnos a lo largo de los diferentes cursos (Valero *et al.* [7]). Cabe indicar que se trata

de una clasificación abierta que nos permite incorporar nuevos errores en cualquier momento, ya que, dado lo particular de la investigación (estudio de la IL durante 4 años), creemos necesaria la continua adaptación de ésta a los distintos estadios de adquisición de la SL.

En cuanto a las categorías de errores consideradas, los parámetros que definen los errores en el nivel morfo-sintáctico se basan en: tipo de oración, función sintáctica, clase de palabra y categoría morfológica. Por ejemplo:

MNPL: Forma errónea en la categoría morfológica de pluralización en la clase léxica *nombre*.

SNPD: uso erróneo del determinante en el sintagma nominal.

SSSS: error en la oración subordinada en cuanto al uso de la partícula subordinante.

Los errores léxico-semánticos se organizan de acuerdo con el origen del error cometido, es decir, influencia de la L1 o de otras lenguas conocidas, desconocimiento de la L2, o *avoidance*. Por ejemplo:

LT: Traducción literal de vocablos, conceptos, frases hechas, etc., calcos, anglicismos.

LFC: Falsos amigos.

LWI: Elección equivocada de un vocablo dentro del mismo campo semántico.

Los errores discursivos se dividen en cuatro grandes grupos: organización y estructura del texto; elementos de cohesión y coherencia; elementos de registro, estilo y adecuación; y estructura del texto en general. Algunos ejemplos son:

D1TOP: Dentro del apartado dedicado a la estructura y organización lógica del texto, falta de enunciación clara del tema (*topic sentence*);

D22REFØ: Dentro del grupo de los factores que afectan a la cohesión del texto, falta de referencia pronominal en particular.

D24AUD: Dentro del apartado dedicado a la estructura textual general del texto, falta en cuanto a tener en cuenta a la audiencia lectora.

Los errores de ortografía se refieren a errores en cuanto a una ortografía incorrecta, inconsistente o errónea por influencia de la L1 u otra lengua. Por ejemplo:

SPS: Ortografía incorrecta.

SPINCONS: Ortografía inconsistente.

SPST: Error de ortografía por influencia de la L1 u otra lengua conocida.

Por último, en cuanto a los errores de puntuación, existen diferentes categorías referidas a un uso incorrecto, la ausencia de un signo o el uso excesivo

del mismo. Por ejemplo: P; P; o P"Ø. Los dos primeros tratan del uso erróneo de la coma y el punto y coma, mientras que el tercero se refiere a la falta (u omisión) de las comillas.

Una vez llevada a cabo la *identificación de errores* de acuerdo con la clasificación elaborada y que cuenta con aproximadamente 230 categorías diferentes, repartidas en cinco apartados distintos como acabamos de exponer, se pasa a la LOCALIZACIÓN de los errores en el texto.

Para la localización de los errores dentro de cada texto, ENWIL distingue entre los siguientes tipos de errores: uso (o bien presencia del elemento cuando éste no debería ser utilizado), mal uso (o bien, selección paradigmática errónea) u omisión (o bien, ausencia del elemento, cuando éste sí debería ser utilizado): en los dos primeros casos, el error se localiza tras el elemento usado; en el tercer caso, el error se localiza tras el espacio que debiera ocupar el elemento omitido.

Los errores discursivos, por su parte, pueden aparecer por encima del nivel oracional, y de hecho, un texto completo puede ser objeto de algún tipo de error discursivo. ENWIL describe los errores en el nivel discursivo en tres posiciones diferentes textuales. En primer lugar, los elementos discursivos cuyo uso es erróneo u omitido se localizan tras el elemento en cuestión o en el espacio de la palabra omitida, respectivamente. En segundo lugar, las diferentes partes lógicas de un párrafo pueden ser fuente, asimismo, de error; su localización respectiva, por tanto, se realiza tras el signo de puntuación, o al final del párrafo. En tercer lugar, los errores discursivos que se refieren a todo el texto se localizan al final del texto.

ENWIL ofrece también la opción de introducir *comentarios* referidos al uso erróneo o aportando una solución al problema de modo que luego facilite su corrección o sea útil para la elaboración de materiales. Tales comentarios se introducen en el texto con solo pinchar en la barra de herramientas en el icono correspondiente y permanecen ocultos a menos que queramos visualizarlos.

Una vez localizados los errores, pasamos a la *generación de informes*. ENWIL permite esta generación sirviéndonos del *software* ACCESS. Estos informes tienen gran validez para la descripción de la IL de la población objeto de estudio. Y, además, dado que los errores se describen a partir de los niveles descriptivos morfosintácticos, semánticos, discursivos, ortográficos y de puntuación, la información nos permite identificar en qué nivel concreto los alumnos tienen problemas.

3.3. EJEMPLO DE INFORMACIÓN CONSEGUIDA A TRAVÉS DE ENWIL

ENWIL ha resultado un eficaz gestor y administrador de datos como pone en manifiesto la generación de información de orden cuantitativo. Los datos obtenidos son el resultado de la investigación preliminar de dos series de textos escritos por alumnos de Filología Inglesa en la Universidad de Alcalá, correspondientes a la promoción 1996–2000. Llamaremos a la primera serie, *serie A*, y corresponde a los textos escritos por alumnos con unos conocimientos de inglés de carácter pre-universitario; la segunda serie, o *serie B*, corresponde a los textos producidos por alumnos que han recibido cierta instrucción sobre la composición en inglés durante su primer año en la Facultad.

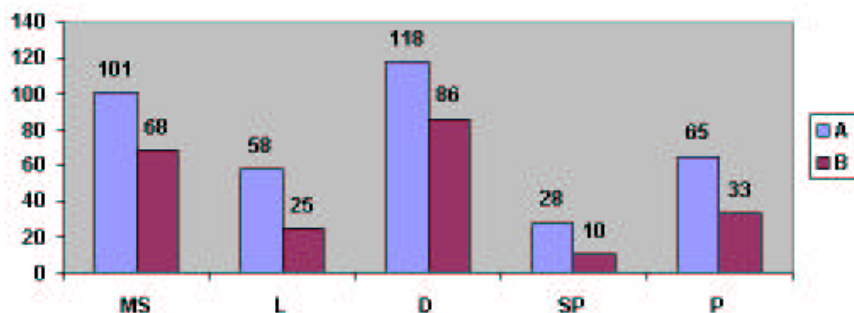


Figura 1: Número de errores cometidos en series A y B

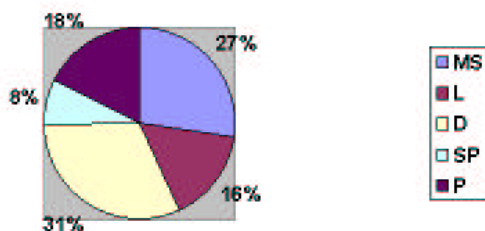


Figura 2: Distribución porcentual de errores cometidos en serie A

Como observamos en la figura 1, la serie A muestra un mayor número

de errores que la serie B. En cuanto a la serie A, según vemos en la figura 2, se observa un mayor porcentaje de errores en los niveles discursivo (D) y morfo-sintáctico (MS). Los errores D se relacionan con una falta de fluidez comunicativa en cuanto a la organización textual en la lengua inglesa y los errores MS tienen que ver con la falta de precisión gramatical. Dado el estadio del proceso de aprendizaje de la lengua inglesa de los alumnos objeto de estudio, interpretamos que la falta de fluidez comunicativa y de precisión gramatical son síntomas del desconocimiento por parte del alumno de las reglas de uso y de forma del inglés, y posiblemente, de la ausencia en las programaciones de la asignatura en los institutos o centros de enseñanza medias de contenidos referidos a la composición en inglés.

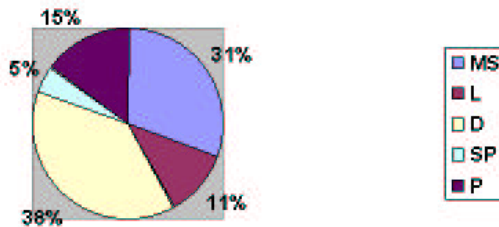


Figura 3: Distribución porcentual de errores cometidos en serie B

En lo que concierne a la serie B, la figura 1 pone de manifiesto que los mismos alumnos siguen cometiendo más errores en los niveles discursivo y morfo-sintáctico, como también se observa en la distribución porcentual de los errores en la serie B (fig. 3). De nuevo, cabe la misma interpretación a estos datos acerca de la falta de fluidez comunicativa y de precisión gramatical. No obstante, resulta importante comparar las series A y B (fig. 1): así se observa que la serie B desciende en el número de errores cometidos en cada uno de nuestros niveles analíticos. Esta evolución supone, a nuestro juicio, una toma de conciencia (*awareness*) de los aspectos discursivos y gramaticales tras el proceso de enseñanza/aprendizaje de éstos durante el primer curso.

En definitiva, ENWIL constituye un ejemplo de AIE que posibilita el estudio de los estadios de la IL de un conjunto de estudiantes de inglés como SL. Hay que reseñar que dicho programa puede adaptarse a la adquisición de otras segundas lenguas ya que permite el cambio de variables o la introducción

de otras nuevas. Es decir, una simple adaptación de las clases de errores tipificadas para el estudio de inglés como SL, permitiría utilizar el programa para el estudio, por ejemplo, de español como SL o francés, o alemán, incluso cuando la L1 sea diferente.

4. El futuro del análisis informatizado de textos

ENWIL no es, sin embargo, un proyecto aislado de AIE. Tal línea de investigación ha despertado el interés de otros grupos internacionales que han desarrollado, o están desarrollando, proyectos similares. Podríamos citar el *International Corpus of Learner English* (ICLE) y *Computer-Aided Error Analysis* (CAE) ambos desarrollados en la Universidad de Lovaina, Bélgica, o bien a COALA (*Computer Aided Linguistic Analysis*), desarrollado en Australia.

Los dos primeros son los que guardan una relación más afín con ENWIL. En cuanto al primero, el interés de ICLE reside en la compilación de un corpus informatizado de errores producido por estudiantes de inglés pero con diferentes L1 (Granger, [5, p. 57]). ICLE parte de las siguientes variables: (1) Tipo de aprendiz: “inglés como lengua extranjera”; (2) Edad: adultos; (3) Estadio de interlenguaje: avanzado; (4) Tarea: producción de un ensayo (Granger, [5, p. 26]).

En este proyecto, el tratamiento informatizado se centra en los errores morfo-sintácticos. Los errores de tipos discursivo y léxico-semántico se investigan con la ayuda del *software* de concordancias TACT, limitando su estudio a la generación de información sobre la frecuencia de uso y distribución léxica de los elementos bajo análisis.

CAE sigue un procedimiento similar pero su objeto de estudio son los aprendices de SL con una misma L1, el francés.

5. Conclusión

El *Análisis Informatizado de Errores* (AIE) supone un nuevo estadio en el Análisis de Errores como metodología de estudio en ASL para conocer el proceso de adquisición de SL.

EL programa ENWIL (*English Written Interlaguage*), desarrollado en la Universidad de Alcalá y aplicado a una promoción de alumnos de Filología Inglesa ha permitido el tratamiento y procesamiento computerizado de gran cantidad de datos sobre su producción en lengua inglesa. Una vez analizados

los textos producidos por los alumnos a lo largo de los cuatro años de la carrera, ENWIL estará en disposición de poder extraer conclusiones acerca de la descripción completa e íntegra de la *interlingua* de los aprendices objeto de estudio. Con estos datos en la mano, las aplicaciones pedagógicas son valiosas dado que al conocer *a priori* qué áreas lingüísticas ofrecen más o menos conflicto, se pueden elaborar material didáctico, gramáticas o bien correctores de estilo destinados a estos aprendices.

Los resultados obtenidos hasta este momento tras el análisis de dos series de textos producidos en distintos momentos de sus estudios muestran un mayor porcentaje de errores en los niveles morfosintácticos y discursivos pero también una modificación en la distribución porcentual de los mismos. El análisis de datos obtenidos a lo largo de su carrera nos permitirá un estudio de la evolución de su interlingua.

Referencias

- [1] DAGNEAUX, E. & DENNES, M. & GRANGER, S. (1998). Computer-Aided Error Analysis. *System*, 26: 163-174.
- [2] DULAY, H. & BURT, M. & KRASHEN, S. (1982). *Language Two*. Oxford: Oxford University Press.
- [3] ELLIS, R. (1997). *Second Language Acquisition*. Oxford: Oxford University Press.
- [4] FERNÁNDEZ, S. (1997). *Interlingua y análisis de errores en el aprendizaje del español como lengua extranjera*. Madrid: Edelsa.
- [5] GRANGER, S. (1994). The Learner Corpus: A Revolution in Applied Linguistics. *English Today*, 39: 25-29.
- [6] HADEN, R. (1987). Discourse Error Analysis. J. Mohaghan, ed., *Grammar in the Construction of Texts*. London: Pinter. 33-40.
- [7] VALERO GARCÉS, C. & MANCHO, G. & FLYS, C. & CERDÁ, E. (en prensa). Evolución de la Interlingua y Análisis de Textos: ENWIL y el Análisis de Errores en la Expresión Escrita en EFL. *Perspectivas Aplicadas de la Organización del Conocimiento y del Procesamiento y Uso del Lenguaje*. Logroño: Universidad de la Rioja.