

Tipología de errores gramaticales para un corrector automático

Ana M^a Díaz Villa

Dpto. de Lingüística Computacional. Real Academia Española
anama@rae.es

Resumen: En este artículo se presenta una propuesta de tipología de errores gramaticales, orientada al desarrollo de un corrector gramatical y de estilo automático que centra la atención en el tratamiento de los errores de motivación cognitiva. Concretamente, en el ámbito de la corrección gramatical, se entiende además que la granularidad de la tipología condiciona en gran medida la adecuación de los mensajes de error proporcionados al usuario en la fase de diagnóstico, así como la adaptación del sistema desarrollado para desempeñar otras tareas (CALL); de este modo, la función de la tipología de errores resulta ser mucho más importante en la creación de un corrector gramatical y de estilo automático de lo que se ha dicho hasta la fecha en otros estudios.

Palabras clave: corrección gramatical, tipología de errores, errores de motivación cognitiva

Abstract: This paper presents a proposal for a grammar error typology oriented towards the development of an automatic grammar and style checker. Cognitively-motivated errors have been considered the first and foremost objective. In particular, in the area of grammar checking, the granularity of the error typology greatly influences the adequacy of the error messages shown to the final user during the diagnosis phase, and, besides, allows an easy adaptation of the system to other language tasks (CALL). Thus, the function of an error typology turns to be more important in the development of a grammar and style checker than studies in the field have acknowledged to date.

Keywords: grammar checking, error typology, cognitively-motivated language errors

1. *Introducción*

El desarrollo de las numerosas tipologías de errores ortográficos, gramaticales y/o de estilo que se pueden consultar en la actualidad ha estado siempre ligado a la creación de sistemas de corrección automática de textos. La tipología aquí presentada atiende al desarrollo de un corrector gramatical y de estilo automático, restringido a textos con un grado medio-alto de revisión, escritos en español peninsular por hablantes nativos. En este artículo se tratarán todas las cuestiones relacionadas con la tipología de errores gramaticales, dejando la clasificación de los errores de estilo como el objetivo principal de trabajos posteriores.

Aún hoy día no existe —ni parece posible que exista en un futuro— una tipología de errores universal, disponible para cualquier sistema que se desee construir. Esto es debido a que en diferentes lenguas y en diversos estudios pueden variar considerablemente los

tipos de error identificados, así como los ejes en torno a los cuales se define la clasificación propuesta. Así pues, la elaboración de una tipología de errores nunca puede ser llevada a cabo independientemente del objetivo para el cual haya sido concebida.

En la gran mayoría de los sistemas de corrección automática existentes hasta la fecha —ya se trate de corrección ortográfica o de corrección gramatical y de estilo—, la motivación fundamental para la elaboración de una tipología de errores parece ser la estructuración y organización de todos los casos de que pretende dar cuenta el sistema, tanto en la fase de diseño del mismo como en la presentación de la aplicación final desarrollada. Sin embargo, en este artículo se considera que el papel que desempeña una tipología de errores en el desarrollo de un corrector automático —al menos, en el caso de la corrección gramatical— va más allá de la mera organización de los problemas tratados.

2. Tipologías. Orientación

Desde los primeros estudios acerca de la clasificación de errores gramaticales, existe una discusión abierta en cuanto a la orientación que debe seguir una tipología de errores:

- Veronis (1988) establece una distinción principal entre errores de competencia y errores de actuación; esto es, entre errores debidos a ciertas carencias cognitivas de los hablantes respecto de la lengua utilizada y errores meramente fortuitos.
- Ramírez Bustamante *et al.* (1994) presentan una tipología determinada por los efectos lingüísticos que producen los errores, en lugar de atender a las causas que los provocan; esta decisión es contraria a lo expuesto en los estudios de Veronis (1988), Corder (1992) y Vosse (1992), donde, según se dijo más arriba, se trata de relacionar cada tipo de error con un problema de competencia o bien con un fallo de actuación (lo que para Ramírez Bustamante *et al.* supone la causa del error).
- Por último, Oliva (1997), aparte de clasificar los errores según se trate de problemas de competencia o de actuación, entiende que considerar además la frecuencia con que se produce cada tipo de error resulta muy útil para controlar la cobertura del corrector automático en desarrollo; y, del mismo modo, presenta una nueva distinción, esta vez en función de la complejidad del aparato formal necesitado para la detección de cada tipo de error propuesto.

En el caso de esta tipología —y tomando como referencia las clasificaciones expuestas— se ha optado por presentar la siguiente orientación:

- En primer lugar, se ha establecido una distinción entre errores de motivación cognitiva —o de competencia— y errores fortuitos —o de actuación—. A diferencia de lo considerado en Ramírez Bustamante *et al.* (1994), no se entiende en este caso que hablar de errores cognitivos y de errores fortuitos implique establecer la tipología en función de las causas de los errores, en lugar de atender a los efectos. Antes bien, en el caso de los errores cognitivos la clasificac-

ión tiene lugar atendiendo a las causas que provocan tales errores; suelen ser, de hecho, los tipos con *nombre propio*, por decirlo de algún modo: se trata de los casos mencionados en obras de referencia y, en la mayoría de las ocasiones, son fácilmente predecibles en función del contexto (independientemente de que su detección implique unos u otros niveles de procesamiento lingüístico). Además, una clasificación de los errores de motivación cognitiva basada en las causas que los provocan es mucho más informativa para la corrección gramatical, por cuanto permite proporcionar diagnósticos mucho más explicativos y útiles a los usuarios del sistema (esta cuestión se tratará más adelante).

Los errores fortuitos, en cambio, no resultan fáciles de diagnosticar, dado que tampoco son fácilmente predecibles; sin embargo, sí son, por lo general, solucionables de manera algorítmica. La inclusión de este tipo de errores dentro de la tipología no parece, pues, favorecer el ciclo detección–diagnóstico–corrección, por lo que contar con una clasificación de los mismos se ha considerado menos relevante en el desarrollo de un corrector automático. Con todo, sí se mencionan algunos casos de este tipo, atendiendo en esta ocasión a los efectos que producen en el texto.

- En segundo lugar, se ha procurado agrupar los errores atendiendo al nivel de análisis que se ve afectado en cada caso. Sin embargo, como comentan Ramírez Bustamante *et al.* (2000), el nivel de descripción lingüística al que pertenece un error determinado no siempre se corresponde con el nivel de procesamiento lingüístico en que puede ser tratado ese mismo error: por ejemplo, numerosos errores relacionados con cuestiones sintácticas y semánticas que se producen en contextos muy concretos pueden ser detectados sin llegar siquiera a un nivel de procesamiento sintáctico.

Esta segunda agrupación atiende, pues, a un criterio mixto entre niveles de análisis afectados y niveles de procesamiento del texto; así, los tipos de errores propuestos en este caso presentan una distinción entre errores léxicos, sintácticos,

semánticos, de concordancia y de puntuación (aparte de otros errores fortuitos, no clasificables como pertenecientes a ninguno de estos tipos).

3. Función

Tal como se comenta en numerosos estudios que abordan este tema —Ramírez Bustamante *et al.* (1994), Wedbjer Rambell (1999) y Becker *et al.* (1999)—, contar con una tipología de errores en el diseño de un sistema de corrección automática de textos permite definir exactamente, por una parte, la clase de errores que se pretende detectar y tratar; por otra parte, ayuda a evaluar el funcionamiento de dicho sistema en cuanto a la cobertura de casos obtenida.

Concretamente en el ámbito de la corrección gramatical, la tipología de errores cobra una importancia mayor aún:

- En primer lugar, los errores detectados pueden ser de muy diversa naturaleza y afectar a una sola palabra, a un sintagma o a una oración completa (esto hace aún más necesaria una clasificación apropiada de los errores).
- En segundo lugar, principalmente cuando se trata de errores de motivación cognitiva, además de la mera detección del error es necesario, al menos, proporcionar un diagnóstico adecuado para cada situación y, si es posible, una corrección apropiada en función del diagnóstico presentado —a diferencia de lo que sucede en un corrector ortográfico, donde tan solo se detectan los errores que dan lugar a una secuencia anómala de caracteres y el diagnóstico puede ser idéntico en todos los casos (la aparición de una palabra desconocida por el sistema), independientemente de la corrección proporcionada—. En ese caso, la simple adscripción de cada error a un tipo determinado dentro de la tipología (siempre que esta no sea excesivamente burda) puede suponer un diagnóstico adecuado que resulte, en forma de mensaje de error, lo suficientemente informativo para el usuario de la aplicación. De nuevo en el caso de los errores de motivación cognitiva, si la tipología ha sido elaborada en función de las causas, puede considerarse incluso la adaptación posterior del sistema de corrección para la

enseñanza de una lengua a hablantes no nativos (CALL, mediante la especialización de los mensajes de error, por ejemplo).

- En tercer lugar, una tipología estructurada de manera adecuada permite llevar a cabo estudios acerca de los errores que se cometen con mayor frecuencia en una lengua (una vez que se dispone del corrector que implementa dicha tipología).

La creación de la tipología de errores debe ser, así, una tarea inicial en el desarrollo de un corrector gramatical, pero en continuo uso y revisión durante todo el proceso de creación del sistema.

En la fase inicial de desarrollo del corrector gramatical, la tipología aquí descrita cumple dos funciones esenciales:

1. Anotación del corpus de errores. Este corpus supone un referente real muy útil en la construcción del corrector, ya que de él se pueden obtener los casos de error más frecuentes y los contextos en que estos se producen, con el fin de implementar gramáticas u otro tipo de recursos que ayuden a la detección y propuesta de corrección de aquellos. En este caso se ha elaborado mediante la recogida de oraciones pertenecientes a textos con un grado medio-alto de revisión, escritos en español peninsular por hablantes nativos. Para su elaboración se ha procedido a la lectura y revisión de textos de prensa, libros de ficción y obras de divulgación general. Hasta la fecha se han codificado alrededor de 650 errores. Además de este corpus de errores, existe otro corpus complementario, formado a partir de todas las oraciones del *Corpus de referencia del español actual (CREA)* en las que aparece codificado al menos un error, y que puede ser utilizado del mismo modo que el corpus anterior, con un total de 44 867 casos.
2. Organización de la información contenida en los recursos desarrollados. La estructuración de todos los componentes de que se sirve el sistema en función de los tipos de errores estudiados permite observar en qué medida se resuelven los problemas de tratamiento e implementación para cada tipo propuesto en la tipología.

4. Cuestiones de diseño

4.1. Fuentes

Existen numerosas obras de referencia que tratan el problema de la corrección gramatical en español, tales como gramáticas, diccionarios normativos, libros de estilo, etc. La información contenida en ellas —la explicación de determinados problemas, acompañada de una serie de ejemplos ilustrativos— es muy útil en la elaboración de una tipología de errores.

Pese a ello, en ninguna de las obras de ese tipo se presentan datos cuantitativos acerca de los errores más comunes cometidos en textos reales y, en ocasiones, simplemente reflejan la interpretación personal de cada autor acerca de cada problema lingüístico tratado.

Según apunta Wedbjer Rambell (1999), resulta de enorme importancia para el desarrollo de un sistema de corrección gramatical y de estilo conocer qué tipos de error se producen realmente en los textos con mayor frecuencia, sistematizados de la forma adecuada. Esto permite, además, dirigir los esfuerzos a intentar resolver los problemas más comunes.

Por esta razón, la elaboración de la tipología de errores se ha basado aquí, en primer lugar, en la observación de textos de diverso tipo y de los errores que se producen en ellos. Dado que sería necesario un corpus de enormes dimensiones para alcanzar una representatividad máxima de los errores que se cometen habitualmente, se ha recurrido además —y de forma paralela a la recogida de casos reales— al vaciado de algunas obras de referencia de prestigio —por ejemplo, el *Diccionario de dudas y dificultades de la lengua española* de Seco (2004) y una versión preliminar del *Diccionario panhispánico de dudas*, de la Real Academia Española—.

4.2. Límites

Antes de tratar algunos aspectos concretos del diseño de esta tipología, es necesario especificar qué se ha considerado aquí un error gramatical.

En este caso, se considera que un error es de tipo gramatical cuando afecta a algún aspecto en uno de los niveles de análisis gramatical —como son el morfológico, el sintáctico o incluso el léxico-semántico—. Esta concepción de error gramatical se diferencia así de la de Wedbjer Rambell (1999), donde la distinción entre errores ortográficos (*spelling*

errors, fuera del ámbito de estudio de esta tipología) y gramaticales depende del contexto necesario para la detección de cada error en cada caso: la identificación de errores ortográficos no tiene en cuenta el contexto, mientras que en el caso de los errores gramaticales se considera que el contexto máximo para su detección es la oración.

Aquí, sin embargo, se considera que un error como el que se produce en **andé*, si bien podría ser detectado sin siquiera contar con un contexto mínimo, es de naturaleza gramatical, por cuanto supone una anomalía en el paradigma flexivo del verbo *andar* (en estos casos, además, la distancia que existe entre el error y la corrección suele ser demasiado grande como para que se pueda dar cuenta de ellos con la ayuda de un corrector ortográfico convencional).

En cuanto al diseño de la tipología de errores, no se debe olvidar que este se ve condicionado por la utilidad que se pretenda dar a este recurso. En este caso, hay dos aspectos que se ha considerado necesario fijar, presentados a continuación.

- **Granularidad.** En la cualificación de los errores gramaticales que forman parte de esta tipología, cada tipo debería estar perfectamente definido y separado de los demás. Esto se puede conseguir atribuyendo a cada caso unos rasgos determinados y un lugar concreto dentro de una jerarquía previamente fijada. A medida que se van incluyendo más niveles en ella, mayor es la posibilidad de que se produzcan solapamientos entre tipos diferentes; por otra parte, si no se incluye ninguna subclasificación, la tipología podría resultar excesivamente burda y poco funcional. Por estas razones, esta tipología no presenta una granularidad excesiva —a diferencia de lo observado en otros proyectos (Wedbjer Rambell, 1999)—: se ha optado por proponer una jerarquía que establece dos grupos iniciales, en los que se incluyen varios tipos diferentes, con una posibilidad de subespecificación máxima de dos niveles, según se muestra en la sección dedicada a la exposición de cada tipo de error.
- **Grado de severidad del error.** Aparte de la asignación de cada tipo a una posición concreta dentro de la jerarquía, para los errores gramaticales se ha esta-

blecido una nueva diferenciación: dado que no siempre el error cometido supone una violación severa de algún aspecto en los niveles de análisis lingüístico tratados, este puede ser clasificado como error severo, menos severo, preferencia de uso o recomendación de alguna obra de referencia consultada, forma preferida por la norma culta o forma vulgar o coloquial.

5. Tipología de errores

Esta se estructura en torno a errores de motivación cognitiva y errores fortuitos.

5.1. Errores de motivación cognitiva

En este apartado se recogen cinco tipos principales de error gramatical:

1. Errores de puntuación. Gran parte de los casos tratados aquí parece obedecer a meros errores fortuitos; sin embargo, algunos de ellos podrían atender también a ciertos problemas cognitivos de los hablantes. Resulta, pues, difícil establecer un límite claro entre los casos provocados por una u otra causa. Dado, además, que los problemas encontrados en el uso de los signos de puntuación son bastante predecibles, se ha optado por tratar este tipo dentro de los errores de motivación cognitiva, con una división en tres subtipos diferentes:
 - a) Problemas relacionados con la colocación de espacios entre signos de puntuación:

Con las víctimas de hoy son ya 4.133(3.138 palestinos y 923 israelíes) las personas muertas (...)
 - b) Errores en el uso de la caja tipográfica:

¿Ya nos vamos? pero si acabamos de llegar...
 - c) Casos en que no se produce un emparejamiento correcto entre signos de interrogación, paréntesis, comillas, etc. de apertura y de cierre:

Le falta energía? Pues tiene centenares de botes donde elegir.
 - d) Problemas de colocación de signos de puntuación en general, por omisión, adición innecesaria o incorrección en el orden seguido:

En Islandia, un país modélico en la gestión de los recursos marinos a, largo plazo, los proveedores del Grupo Carrefour pescan (...)

2. Errores léxicos. Dentro de este tipo se incluyen los errores que pueden ser registrados en las piezas léxicas (simples o complejas) que forman parte de una oración.

Los subtipos de error considerados en este apartado son los siguientes:

- a) Espaciado (omisión o adición de un espacio, dando lugar a una palabra o secuencia incorrecta):

No había si quiera un consenso sobre la fecha ni sobre los objetivos.
- b) Acentuación (omisión o adición innecesaria de una tilde, dando de nuevo lugar a una palabra incorrecta o diferente de la esperada en un contexto determinado):

(...) una patrulla española se vió ayer inmersa en una emboscada(...)
- c) Correcciones en la escritura de nombres propios, acrónimos y abreviaturas:

Familias de británicos muertos en Irak buscan el apoyo de Moore (frente a Iraq, la única forma citada en el Diccionario panhispánico de dudas);
Se encontrará ud. en esta página (...) (en lugar de Ud.).
- d) Uso de una variante gráfica no correcta:

Los shiís respetaban la Suna (en lugar de chiés).
- e) Errores en alguna forma del paradigma flexivo de un sustantivo, un adjetivo o un verbo:

Era costumbre que se bendeciera aquel día.
- f) Errores producidos en construcciones complejas que superan el nivel de la palabra:

En base a este perfil, el desarrollo previsible de las actuaciones podría derivar hacia una sucesión de ataques (...) (en lugar de con base en).
- g) Problemas de paronimia, o de los llamados “pares de confusión”:

- Paronimia sintáctica; se trata de uno de los tipos de error más frecuentes observados en el corpus de errores:
*Pero **si** que creo que hay una gran diferencia en eso.*
 - Paronimia semántica:
*La actuación del masajista pareció **surgir** efecto (en lugar de **surtir**).*
 - h) Mala selección léxica en la coaparición de palabras, en cuanto a las propiedades morfológicas de las mismas:
*Eres un **malo** amigo (en lugar de **mal**).*
 - i) Problemas en la escritura de extranjerismos:
*Se puede decir **a grosso modo** que hay tres tipos principales de lesiones (en lugar de **grosso modo**).*
 - j) Americanismos léxicos; no se trata de errores, sino más bien de formas condicionadas por un uso marcado en español peninsular:
*La dureza de los **fluorescentes opacaba** el brillo de las baldosas, y olía a lejía.*
- 3. Errores sintácticos. Todos los casos tratados aquí sobrepasan el nivel de la palabra. Los subtipos desarrollados son los siguientes:
 - a) Mala categorización de algún elemento dentro de la oración:
*eran **los peores pagados** de la empresa.*
 - b) Construcciones sintácticas incorrectas:
*Ella vive muy **cerca mío** (por de mí).*
 - c) Problemas de rección sintáctica; aquí se incluyen numerosos casos:
 - Errores de rección preposicional:
*Esta organización **apuesta al** desarrollo de un amplio trabajo en redes (en lugar de **apuesta por**).*
 - Uso incorrecto de la preposición en el complemento directo:
*Bruselas **teme a la corrupción**.*
 - Mala rección de tiempo verbal (problemas relacionados con la *consecutio temporum*):
*Ayer por la tarde el grupo terrorista **ha publicado** en un sitio web un vídeo (...)*
- Mala rección de modo verbal:
*Hay una pequeña probabilidad de que **se recupera** espontáneamente.*
 - Loísmo, leísmo y laísmo:
*Al guardarlos debe **ponerlos** un nombre a cada uno (...)*
 - Queísmo y dequeísmo:
*Muchos **se han olvidado que** en Zaragoza hay, desde hace 25 años, un instituto que lleva el nombre de los hermanos Marx.*
 - d) Expresión de un complemento agente en pasivas reflejas:
*La fuga se detectó **por un trabajador de la fábrica**.*
 - e) Incorrección en el orden de palabras:
*Sabe, desde el pasado verano, (...) que sus peroratas contra la ocupación se vigilan **de muy cerca**.*
 - f) Americanismos sintácticos; tampoco aquí se trata de un error propiamente dicho, sino más bien de un uso marcado (por ser muy poco frecuente en el español peninsular) de ciertas estructuras:
*Las niñas **jugaban béisbol**.*
- 4. Errores de concordancia. Si bien estos errores podrían ser tratados como un subtipo de error sintáctico, dada la complejidad de los mismos, se ha decidido que deben ser clasificados como un tipo independiente. En esta ocasión se proponen cuatro subtipos principales:
 - a) Concordancia intrasintagmática:
*Los tres palestinos fueron víctimas de un ataque con dos misiles realizado por parte **del helicópteros israelíes** (...)*
 - b) Concordancia intersintagmática:
*El ramadán es el noveno mes del calendario lunar y es **el mes en lo que** comenzó la revelación del Corán.*
Tal vez sea necesario especificar que en los casos de oraciones explicativas entre comas se ha entendido que la concordancia es intersintagmática, así como en la relación que mantienen un pronombre o adverbio relativo y su antecedente, o los problemas encontrados en sujetos u otro tipo de argumentos coordinados.

- c) Concordancia con femeninos débiles. Este tipo de concordancia es el que se produce con los sustantivos femeninos que comienzan con una *a* tónica y los determinantes y adjetivos que los modifican: con estos sustantivos algunos determinantes escogen un alomorfo femenino que es homógrafo del masculino. Tal particularidad de los femeninos débiles provoca cierta confusión en los hablantes en cuanto al género real de ese tipo de sustantivos, que llegan a ser considerados como masculinos a efectos de concordancia (ya sea esta intersintagmática o intrasintagmática):

(...) *y el PSOE se comprometió a aumentar cada año en un 25 % el esfuerzo económico en **este** área.*

*Se trata de un aula muy **pequeño**.*

- d) Concordancia *ad sensum*. Este último tipo es en realidad un caso especial de concordancia intersintagmática; la razón por la que se ha decidido considerar como un tercer subtipo es que no puede decirse que se trate realmente de un error:

***Un grupo** de investigadores españoles **han demostrado** la existencia de «autopistas» de viento (...)*

En realidad, algunos de los casos tratados como errores de concordancia no responden a un tipo de error cognitivo; sin embargo —al igual que sucede en los errores de puntuación—, se ha considerado más acertado incluir todos los tipos dentro de este apartado, en lugar de separarlos según se deban a un error fortuito o a una falta de competencia del autor.

5. Errores semánticos. Aquí se incluyen los casos relacionados con un mal uso de ciertas piezas léxicas o de ciertas construcciones en cuanto al significado de las mismas (se debe mencionar que el único caso no incluido aquí es el de los parónimos semánticos, por haber considerado más interesante un tratamiento de los mismos paralelo al de los parónimos sintácticos).

Se pueden incluir aquí cuatro subtipos diferentes:

- a) Errores relacionados con una confusión en cuanto al significado de ciertas piezas léxicas, sin que se trate de parónimos semánticos:

*Despacio, Silas **resiguió** el camino que recorría aquella marca (...)*

- b) Mala selección léxica en la coaparición de palabras, en cuanto a las propiedades semánticas de las mismas:

*Este personaje se volvió **muy inmortal**.*

Dentro de este subtipo se podrían incluir, además, los errores de concordancia semántica.

- c) Problemas de duplicación de la negación:

*Quiso **evitar** que **no** se cayera.*

- d) Redundancia semántica. Los errores de este tipo pueden entenderse como errores gramaticales por cuanto afectan al nivel semántico, en especial en ciertas construcciones como *subir arriba, bajar abajo*, etc.

5.2. Errores fortuitos

Los errores incluidos en este grupo son, según se ha mencionado anteriormente, aquellos que podrían ser detectados por cualquier hablante nativo de nuestra lengua y que no tienen una motivación cognitiva; antes bien, parecen ser problemas originados simplemente por una falta de atención en la escritura. En este grupo quedan los casos que no han podido ser asimilados a ninguno de los tipos anteriores, clasificados de la siguiente manera:

1. Omisión o inclusión de un espacio, sin que se pueda considerar que se trata de errores de puntuación o de errores léxicos:

***U no** de los principales problemas es la falta de preparación del personal.*

2. Omisión de elementos necesarios dentro de una oración:

*Los marines dispararon contra el camión y luego contra una furgoneta y un vehículo agrícola, **en los** viajaban las otras seis víctimas mortales.*

3. Repetición innecesaria de palabras o de categorías dentro de una oración:

Se trata de un programa gratuito que que se puede conseguir e instalar fácilmente.

*Alía **anunció ha asegurado** que el error se debe a la “gran cantidad de facturas” (...).*

Como se puede observar, no se ha incluido en esta clasificación ningún error de tipo pragmático, ya que todavía no existe una estrategia de detección adecuada y fiable para ese nivel de análisis lingüístico —debido, realmente, a su escaso desarrollo desde el punto de vista computacional—. Pese a que pueda parecer que sucede lo mismo en el caso de los errores semánticos, tal como apuntan Bustamante *et al.* (1994) —y a diferencia de lo considerado por Vosse (1992)—, hay en la semántica una serie de rasgos que muestran una correlación en la sintaxis. Este hecho posibilita la detección de ese tipo de errores con técnicas de procesamiento de bajo nivel, si bien se requiere además la clasificación de las piezas léxicas en función de ciertos rasgos sintáctico-semánticos (posibilidad de gradación de los adjetivos o predicación de estos respecto de seres animados, entre otros) y un estudio más exhaustivo del contexto, teniendo en cuenta las colocaciones más comunes y las posibles restricciones léxicas de tales piezas.

6. Conclusiones

En este estudio se ha considerado que la elaboración de una tipología de errores de motivación cognitiva lo suficientemente granular resulta imprescindible en el desarrollo de un corrector automático, dado que desempeña un papel fundamental en la determinación del diagnóstico producido para cada error (dependiendo de esta la adecuación de los mensajes de error presentados al usuario) y, además, permite llevar a cabo adaptaciones posteriores de la aplicación desarrollada (por ejemplo, para la enseñanza de segundas lenguas). La clasificación de los errores fortuitos, por otra parte, ocupa un lugar secundario —si cabe— dentro de esta tipología, dado que se trata de casos difíciles de predecir y diagnosticar (y no son, por tanto, el objetivo principal en una clasificación de los errores de que debe dar cuenta un corrector automático).

En cuanto a la distinción entre errores ortográficos y gramaticales, se ha considerado que no es el contexto necesario para la detección de los mismos lo que lleva a incluirlos en

un grupo o en otro, sino más bien el hecho de que se vea afectado algún nivel de análisis o la distancia a la que se encuentra la corrección respecto del error, en cada caso.

Bibliografía

- M. Becker, A. Bredenkamp, B. Crysman, y J. Klein. 1999. Annotation of error types for German newsgroup corpus. En *Proceedings of the ATALA workshop on Treebanks*, pages 89–100, Paris.
- S. P. Corder. 1992. La importancia de los errores del que aprende una segunda lengua. En J. Muñoz Licerias, editor, *La adquisición de las lenguas extranjeras*, Madrid (citado por Bustamante *et al.*, 1994).
- K. Oliva. 1997. Techniques for accelerating a grammar-checker. En *Proceedings of the 5th Conference on Applied Natural Language Processing*, Washington D.C.
- F. Ramírez Bustamante, Y. Rodríguez Sellés, y F. Sánchez León. 1994. Tipología de errores gramaticales del español para un sistema automático de corrección de textos. En *Actas del X Congreso de Lenguajes Naturales y Lenguajes Formales*, Sevilla.
- F. Ramírez Bustamante, T. Declerck, y F. Sánchez León. 2000. Towards a theory of textual errors. En *Proceedings of the 3rd International Workshop on Controlled Language Applications (CLAW'00)*, Seattle.
- M. Seco Reymundo. 2004. *Diccionario de dudas y dificultades de la lengua española*. Espasa Calpe, Madrid, 10^a edición. 5^a reimpresión.
- J. Veronis. 1988. Morphosyntactic correction in natural language interfaces. En *Proceedings of the 13th International Conference on Computational Linguistics (COLING'88)*, pages 708–713.
- T. Vosse. 1992. Detecting and correcting morpho-syntactic errors in real texts. En *Proceedings of the Third Conference on Applied Natural Language Processing*, pages 111–118, Trento.
- O. Wedbjer Rambell. 1999. Error typology for automatic proof-reading purposes. En A. Sâgvall Hein, editor, *Reports from the SCARRIE project*, Uppsala.