

Recursos informáticos para la detección del plagio académico

Computer resources for academic plagiarism detection

Adrián Cabedo Nebot

Grupo *Val.Es.Co*

Universitat de València

Recibido el 15 de enero de 2010

Aprobado el 26 de febrero de 2010

Resumen: En el ámbito académico es posible que, en determinadas circunstancias y por diferente finalidad, se requiera la elaboración de un informe descriptivo sobre un caso de plagio o de atribución de autoría. Esta tarea debe realizarla un lingüista perfectamente formado y que, en el momento de redactar su análisis, posea una metodología compacta y sistemática. Este artículo, en tal sentido, esboza un protocolo de actuación, mediante el uso de algunas herramientas informáticas que facilitan la detección del plagio en el marco académico.

Palabras clave: Plagio académico. Autoría. Informe. Copia.

Summary: In an academic environment it is possible that, under certain circumstances and for different purposes, some kind of descriptive report on a case of plagiarism or attribution will be required. This task must be performed by a fully trained linguist and, at the time of writing his analysis, this one should apply a compact and a systematic methodology. This article, in this sense, outlines an action protocol, using some tools that facilitate the detection of plagiarism in an academic framework.

Keywords: Academic plagiarism. Authorship. Report. Copy.

1. Introducción.

Dentro del ámbito de la lingüística aplicada, existe una disciplina profesional de especial auge en los últimos tiempos: la lingüística forense. Se trata de una disciplina poliédrica, dado que sus objetivos de estudio son diversos: la determinación de plagio entre textos, la pertenencia de un texto a un autor determinado, la identificación de voces...

Las aproximaciones heurísticas a estos objetivos dependen de cada una de las categorías señaladas. Así, el lingüista, además de tener conocimientos sobre el lenguaje, deberá poseer también un amplio bagaje en el dominio de instrumentos informáticos de diferente tipología.

En la fonética forense, por razones de análisis del material empleado (archivos de audio), el dominio de la informática y de conocimientos sobre acústica del sonido tendrán mayor importancia, es decir, la objetivación de los datos es especialmente relevante; por otro lado, en la determinación de plagio o autoría es importante tanto la sistematización cuantitativa de datos como la interpretación lingüística de estos (BRIZ 2006).

Así, cuantificar datos en fonética forense (análisis de formantes, de la F0, de la duración de las vocales...) es más concluyente que la mera cuantificación de elementos (cantidad de categorías gramaticales, repeticiones léxicas, etc.) en la determinación de plagio/autoría. Ello no significa que la interpretación subjetiva del investigador no sea necesaria en ambas categorías, sino que el análisis textual, siempre que la copia textual no sea *verbatim*, depende del criterio personal del lingüista, de su conocimiento del campo y de sus recursos lingüísticos. Sobre ese conocimiento previo, cuantificar datos es una herramienta posterior de gran importancia para corroborar las hipótesis cualitativas previas del investigador.

Últimamente, el plagio académico se ha convertido en una constante y, en ese sentido, la sociedad cuestiona los valores de honestidad y esfuerzo intelectual que se realizan en las escuelas en la actualidad (CARROLL and APPLETON, 2001; BAMFORD and SERGIU, 2005). Así, será labor de los docentes erradicar esa tendencia al plagio en los estudiantes (GOTTERBARN and MILLER, 2006).

La intención de este artículo es proponer un protocolo para el análisis textual en casos de determinación de plagio/autoría. Como protocolo, se ofrece una metodología general y orientativa basada en el análisis de un caso real. Así, el propósito es doble: presentar una guía para la elaboración de informes de autoría y, por otro, exponer el

informe de un caso de atribución de autoría¹. Ambos objetivos se inscriben en el ámbito académico.

En general, son muchos los factores que condicionan el plagio (PARK, 2003):

1. Género. Es más común el plagio entre hombres que entre mujeres.
2. Edad y madurez. Los jóvenes suelen plagiar más que los mayores.
3. Cociente intelectual. La gente con cociente intelectual menor suele copiar más, aunque también se da el proceso inverso: alumnos con un nivel elevado de cociente intelectual copian también en gran medida.
4. Vida social. Alumnos con una vida social activa suelen copiar más porque el tiempo que dedican a la parcela social les impide dedicarse más intensamente a la preparación de trabajos y exámenes.
5. Factores de la personalidad del estudiante. Los estudiantes copian si tienen un comportamiento agresivo, pero también si quieren homogeneizarse con sus iguales.
6. Actitud del estudiante hacia la asignatura. En este caso, la motivación del alumnado influye en una mayor o menor copia académica.

En tal sentido, cabe destacar que este guión o estas pautas deben adaptarse en cada caso, ya que todos los textos disponen de una estructura característica y, además, en cada texto el investigador dará prioridad a unas cuestiones lingüísticas frente a otras.

2. Protocolo y metodología.

Son múltiples las situaciones en las que un investigador especializado puede recibir una petición de determinación de plagio/autoría. Por ejemplo, el marco de la lingüística forense suele circunscribirse a casos de posible repercusión jurídica (ALCARAZ, 2005); esto es, ante la posibilidad de una demanda penal o civil, los interesados pueden solicitar, en determinadas circunstancias, un análisis de autoría o plagio de un texto.

Aún así, también es frecuente encontrar estas solicitudes en el marco de la dinámica docente, puesto que en la actualidad muchos estudiantes recurren al plagio o a la copia indiscriminada de textos procedentes de páginas de Internet.

En general, los requerimientos de determinación de autoría son debidos a causas de diferente índole. Lo importante, por ello, es establecer y aplicar un protocolo constante en la elaboración de este tipo de informes valorativos. Al fin y al cabo, un

¹ Los datos que aparecen en este artículo no ponen en peligro la identidad de sus autores ni las circunstancias en las que se solicitó la determinación de autoría. En ningún caso este informe tuvo repercusión judicial ni supuso sanción administrativa para los involucrados.

lingüista no es un juez y, por tanto, sus conclusiones serán simplemente descriptivas y no condenatorias. En ese sentido, un lingüista determina el nivel de semejanza textual entre dos textos para establecer el grado de plagio o la atribución de autoría.

Los informes deben mantener la objetividad y la discreción en las afirmaciones realizadas. Por tanto, deberá prescindirse de comentarios valorativos que no se sustenten en datos empíricos y es, en este caso, en el que el apartado cuantitativo y el cualitativo se fusionan en un constructo de mayor repercusión explicativa (TURELL, 2005).

Por ello, un informe de atribución de autoría deberá basarse, necesariamente, en la cuantificación de los datos y en el análisis valorativo de estos. Así, si los textos analizados son muy extensos y los datos recogidos ingentes, se deberá recurrir a la aplicación de pruebas estadísticas convenientes para fortalecer el dictamen efectuado (SEAWARD and MATWIN, 2009; BARRÓN, ROSSO, PINTO and JUAN, 2008).

Imagínese que se reciben varios textos sobre los cuales quiere establecerse el grado de cercanía y, en tal caso, concluir un posible caso de homogeneidad textual. La primera tarea del investigador es leer los textos detenidamente y formular hipótesis previas. Esta lectura inicial estará más o menos condicionada según la información que se haya facilitado por parte de los solicitantes; por ejemplo, si se quiere saber si un texto B es una copia de A, se tenderá a observar esa circunstancia en esa dirección interpretativa, pero no en aquella en la que A es copia de B. Esto es un error, ya que el investigador debe partir siempre de la imparcialidad y mantenerse en una posición neutra en todo momento.

Una vez elaboradas las hipótesis descriptivas previas, el análisis debe incidir en los siguientes aspectos para cada texto²:

² Como se observa en estos ítems, la determinación de la autoría o del plagio supone introducir el comentario lingüístico de textos en el ámbito explicativo. Por ello, la persona encargada de realizar informes evaluativos debe ser un lingüista experto y conocer aspectos lingüísticos característicos de cualquier tipo de texto, es decir, se trata de aplicar la metodología tradicional del análisis de textos a la determinación de autoría o plagio.

Cantidad de páginas, párrafos, líneas, palabras y caracteres con y sin espacio	
Información tipográfica:	Subrayados, negrita, cursiva, mayúsculas, versales...
Elementos paratextuales:	Títulos, subtítulos, apéndices, anexos...
Tipología textual	<ol style="list-style-type: none"> 1. Descriptivo, argumentativo, prescriptivo... 2. Registro (formal/informal)
Disposición estructural	<ol style="list-style-type: none"> 1. Ausencia o presencia de marcas de división estructural. P.ej.: <i>1. Introducción/ 2. Estado de la cuestión...</i> 2. Orden de la información introducida.
Contenido del texto	<ol style="list-style-type: none"> 1. Tema 2. Argumento 3. Perspectiva narrativa. P.ej.: un texto argumentativo puede defender o atacar una postura.
Plano morfosintáctico:	<ol style="list-style-type: none"> 1. Categorías gramaticales: <ul style="list-style-type: none"> – Cantidad de sustantivos, verbos, adjetivos y categoría predominante 2. Modo indicativo/subjuntivos de los verbos Orden de palabras. 3. Construcciones anómalas y errores (*me se, *la digo...). 4. <i>Etc.</i>
Plano léxicosemántico	<ol style="list-style-type: none"> 1. Semántica de las palabras: <ul style="list-style-type: none"> – Sustantivos concretos/abstractos – Adjetivos – <i>Etc.</i>
Plano pragmático	<ol style="list-style-type: none"> 1. Componente directo/indirecto del nivel oracional: <ul style="list-style-type: none"> – Ironías. – Humor. 2. Introducción de elementos coloquiales: <ul style="list-style-type: none"> – Marcadores del discurso (bueno, vamos, venga...). – Realzadores del emisor: <i>yo</i>. – Uso de la conjunción <i>que</i> para todo valor subordinado (incluido el adverbial). – <i>Etc.</i>

Adviértase que en todos los niveles de análisis propuestos hay un *etc.* final, puesto que, dentro del ámbito lingüístico, son innumerables los aspectos en los que un investigador podrá incidir para determinar la cercanía o no entre dos textos. Habitualmente, en la lectura previa, deberán señalarse cuáles serán los elementos textuales más relevantes para establecer la homogeneidad o heterogeneidad textual.

Por ejemplo, en textos en los que se utiliza generalmente un registro coloquial (correos electrónicos), serán frecuentes elementos del español coloquial (presencia de *yo*, de marcadores fáticos [*oye, venga...*], etc.) y, por tanto, serán tanto más esclarecedores los elementos no esperables (quizá elementos de un registro más formal).

Por su parte, otros autores explicitan *ítems* similares (BARRÓN, 2008):

1. Vocabulario utilizado.
2. Cambios de vocabulario.
3. Texto incoherente.
4. Puntuación.
5. Cantidad de texto común entre documentos.
6. Errores en común.
7. Distribución de las palabras.
8. Estructura sintáctica del texto.
9. Largas secuencias de texto en común.
10. Orden de similitud entre textos.
11. Dependencia entre ciertas palabras y frases
12. Frecuencia de palabras.
13. Preferencia por el uso de sentencias cortas o largas.
14. Legibilidad del texto.
15. Referencias incongruentes.

Así mismo, los puntos establecidos anteriormente no deben cumplimentarse en su totalidad en todo análisis de textos, dado que determinados factores serán más explicativos que otros según las circunstancias. Por tanto, son objetos explicativos que se activan alternativamente según el texto y sus peculiaridades.

Por otro lado, huelga decir que la tarea final de un informe forense sobre autoría o plagio es cercana a la de un comentario de textos y, en ese sentido, es importante que el investigador domine ampliamente los conocimientos lingüísticos necesarios para el análisis textual, y para aplicarlos según las necesidades del encargo y la idiosincrasia del texto analizado.

3. Cuantificación de datos y herramientas informáticas.

En la actualidad, el plagio académico se ha convertido en una constante. El problema asociado no es únicamente el plagio directo o intencionado de fragmentos textuales (bien procedan de internet, de libros, de apuntes de otros compañeros, etc.), sino la conciencia que, sobre ese plagio, tienen los plagiarios. En muchas ocasiones no se relaciona la apropiación de textos ajenos, más o menos extensos, con nociones como infracción o conducta indebida (BAMFORD and SERGIOU, 2005; PARK, 2003); esto debe despertar una gran preocupación en los docentes, ya que los alumnos no entienden que sus actuaciones sean recriminatorias.

El profesor universitario imparte conocimientos a una ingente cantidad de alumnos, con una media de 60 a 100 alumnos por clase. En esa circunstancia, puede optarse por dos vías académicas:

- Ofrecer la posibilidad al estudiante de que pueda realizar trabajos académicos optativos u obligatorios.
- Realizar un único examen final de la asignatura sin ningún tipo de trabajo o actividad extra.

En el primer caso, se corre el peligro de que el alumno acuda a internet y copie, en mayor o menor medida, diferentes informaciones. En ese sentido, es interesante manifestar la tendencia generalizada del estudiante de visitar y plagiar páginas de contenido enciclopédico (*es.wikipedia.org*, sobre todo). Ello evidencia diversas cuestiones, aunque quizá la más preocupante es la falta de recursos en cuanto a la búsqueda de información, dado que son muchas las fuentes bibliográficas (en general, todo artículo publicado en la red [en *pdf* o *word*] o cualquier otra página *html*).

Así, el profesor se encuentra con varios problemas: en primer lugar, la gran cantidad de trabajos optativos que puede recibir y, en segundo lugar, la dificultad de poder analizar concienzudamente la propiedad intelectual de cada texto recibido.

De alguna manera, el docente debe recurrir a herramientas de sistematización de tantos documentos, para poder agilizar las labores de persecución y detección de los fragmentos plagiados de forma evidente (THOMAS and COLLEGE, 2008; GRAY, SALLIS and MACDONELL, 1997; STEIN, KOPPEL and STAMATOS, 2007). En la disciplina humanística, es habitual incorporar fragmentos intertextuales en el texto propio, si bien no se considera plagio en el caso de que dichos fragmentos hayan sido correctamente introducidos y citados. Por ejemplo, piénsese en un fragmento como el siguiente:

La lingüística de perspectiva relevantista es actualmente una disciplina muy trabajada y explotada en muchos círculos académicos de Europa. Como decían SPERBER y WILSON (1994: 37): “la comunicación tiene éxito no cuando los oyentes reconocen el

significado lingüístico del enunciado, sino cuando infieren el 'significado' que el hablante le atribuye”.

En un caso como el anterior, no hay plagio, sino inclusión de un fragmento de texto debidamente introducido y citado. En la redacción académica esto se considera una actuación destacable, dado que permite visualizar la relación entre la interpretación del alumno junto al texto que, en cierto modo, ha despertado tal comentario. Una de las finalidades de la docencia universitaria es que el estudiante mejore su capacidad de análisis y reflexione sobre las palabras de otros, siempre desde un punto de vista metódico y científico, con pruebas y adecuadas argumentaciones de lo expresado.

La dificultad estriba en discernir qué se entiende por plagio en el ámbito de la asignatura impartida, y hasta qué punto se permite que el alumno incorpore citas externas sin parafrasearlas. Esta delimitación es crucial, dado que el profesor debe anticipar, previamente a la recepción de los trabajos, cuál es el límite de copia directa permitida. En tal sentido, se diferencian varios tipos de plagios (CLOUGH, 2003):

1. *Plagio verbatim* o copia literal.
2. *Parafraseo*: se cambia el léxico y el orden sintáctico.
3. *Plagio de fuentes secundarias*. Cuando se lee un texto que tiene una referencia a otro texto y se copia este Segundo fragmento. Habitualmente, se suele marcar en el texto fuente con un *apud*.
4. *Plagio de la forma*: se copia la tipografía o la estructura.
5. *Plagio de ideas*: es el más difícil de detectar.
6. *Plagio de autoría*: usurpar un texto y hacer pasar como propio un texto ajeno.

Cabe pensar, en tal sentido, que la copia directa de los textos, sobre todo en trabajos de carácter recopilatorio o descriptivo, es una acción habitual en el sector estudiantil. Ahora bien, esa apropiación directa puede ser admisible si se incluye bibliografía adicional.

Para poder analizar correctamente un amplio número de trabajos o documentos, el docente dispone de herramientas informáticas que han sido diseñadas para detectar cuestiones de plagio u autoría. De entre ellas, las más conocidas son *copycatch*, *wcopyfind*, *turnitin*, *viper*, etc.

Estos programas tienen como finalidad principal cuantificar el número de formas lingüísticas coincidentes entre varios textos. En realidad, son programas de coincidencia formal, en los que se busca cuántos caracteres juntos situados entre espacios en blanco coinciden linealmente en diferentes documentos (DREHER, 2007; BARRÓN, 2008; THOMAS and COLLEGE, 2008).

En otras palabras, si en un texto A aparece la palabra <corazón> en la oración *Juan no tiene un corazón bondadoso*, la secuencia “c+o+r+a+z+ó+n” (siete caracteres unidos)

se encuentra entre dos espacios en blanco, que la separan de “u+n” y de “b+o+n+d+a+d+o+s+o” respectivamente. Por supuesto, estos programas no buscan únicamente coincidencias formales entre espacios en blanco, sino que continúan cuantificando las coincidencias hasta que el carácter siguiente cambia durante un número “x” de caracteres.

Por ejemplo, piénsese en dos textos, A y B, con las siguientes secuencias:

Texto A	<i>Juan no tiene un corazón bondadoso para ser profesor de educación secundaria ni para nada más</i>
Texto B	<i>Juan no tiene un corazón bondadoso para ser profesor de educación universitaria ni para realizar acciones generosas</i>

Programas como los mencionados pueden realizar acciones según los intereses del investigador. Podrían analizar la secuencia y encontrar dos secuencias del texto B que corresponden al texto A (*Juan no tiene... y ni para...*); la primera se encuentra porque el carácter “u” de la palabra “universitaria” actúa como límite y rompe la coincidencia entre ambas secuencias.

Ahora bien, el programa podría continuar con la búsqueda de coincidencias si ese carácter límite se acompaña de un número específico de caracteres. Por ejemplo, si “universitaria” tiene 13 caracteres, hay 12 que siguen al carácter límite “u”. Una vez terminan esos caracteres, el programa puede seguir buscando coincidencias como si fuera un bloque de copia completo si se especifica que así debe hacerlo a partir de un número de caracteres establecido. Así, visualmente no se tendría dos secuencias diferenciadas, sino una sola en la que se incrustaría un número de caracteres distinto o inesperado.

Por ejemplo, si se indica al programa que busque más allá del carácter límite más 15 caracteres, entenderá como secuencia homogénea los ejemplos anteriormente expuestos, mientras que si la oración de un texto C fuera “*Juan no tiene un corazón bondadoso para ser profesor de educación universitaria de reprobada calidad ni para realizar acciones generosas*”, se entendería que hay dos secuencias homogéneas “Juan no tiene (...) educación” y “ni para”, puesto que el carácter límite “u” de *universitaria* se acompaña de más de quince caracteres: “(u)*universitaria de reprobada calidad*”.

Mayoritariamente, programas como los mencionados no reconocen la categoría de las formas léxicas que comparan, puesto que, como se ha señalado, buscan correlaciones de caracteres físicos, es decir, de grafías conjuntas, pero no entienden la semántica de esas formas. Esa labor interpretativa debe ser realizada posteriormente por el analista. Ello puede poner de manifiesto los recursos estilísticos del autor, según si utiliza más o menos léxico abstracto, más o menos formas verbales o nominales, etc.

Así pues, los programas de detección de plagio son herramientas de análisis cuantitativo. En principio, las copias directas de fragmentos amplios de texto son

visualmente evidentes gracias a los resultados obtenidos, ya que estos son mostrados con distintos colores (habitualmente, con tonalidades fuertes).

De los programas que se pueden encontrar *online*, se han testado dos: *wcopyfind* y *viper*³. El primero permite comparar las correspondencias formales entre dos o más textos introducidos por el investigador, mientras que el segundo permite contrastar los textos analizados con páginas de internet. Ambos proporcionan un porcentaje final de coincidencia textual, bien sea de los textos entre sí o de estos con las informaciones disponibles *online*.

4. Desarrollo de un análisis en un contexto real.

En un curso académico, se pidió a los alumnos de una asignatura de lengua española que realizaran tres trabajos:

1. Resumen del capítulo 1 de la *Teoría de la relevancia* (SPERBER y WILSON, 1994).
2. Resumen de dos artículos sobre la comunicación y sobre la perspectiva que sobre ella tienen los seguidores de la perspectiva relevantista:
 - TRUJILLO, F. (2001): “La teoría de la relevancia como base para una nueva interpretación de la comunicación”.
 - MOYA, C. (2006): “Relevancia e Inferencia: Procesos cognitivos propios de la comunicación humana”.
3. Comentario personal sobre el libro *El curioso incidente del perro a medianoche* (HADDON, 2004).

Hacer un resumen supone elaborar un texto descriptivo, por un lado, pero también sintético. Esta tarea puede incorporar, si procede, comentarios valorativos del alumnado, aunque no sea obligatorio. Por otro lado, este resumen no consiste en extraer fragmentos de los textos requeridos y, de algún modo, reducir su extensión, sino que depende de la lectura y de la jerarquización de contenidos que el lector (alumno o estudiante en este caso) decide. Así, un resumen debe contener las ideas básicas de un texto, si bien estas deben estar expresadas con palabras del lector y no debe componerse de citas u oraciones extraídas al azar del texto general.

Por su parte, un comentario personal sobre una novela supone leer el texto y, posteriormente, con bibliografía o no según las especificaciones del profesor, elaborar un texto valorativo o crítico, con aportaciones y comentarios personales. Por tanto, no es permisible copiar *verbatim*, opiniones de otros, a no ser para corroborar o relativizar las afirmaciones personales.

³ Ambos son gratuitos. Cabe especificar que estos programas han sido operativos durante la realización de este trabajo y que la última fecha en la que se consultó su operatividad fue el 07 de enero de 2009.

Lo más sencillo es cometer la última acción, es decir, mientras se realiza la lectura, se subrayan y se copian los fragmentos más destacables. Este comportamiento es reprobable, aunque lo es todavía más extraer citas sin atender a su relación con el todo, únicamente con la finalidad espuria de reducir la extensión de un texto más amplio.

En los casos aducidos anteriormente, se daban dos circunstancias. En el primer caso, el capítulo de la *Teoría de la relevancia* no se puede encontrar traducido al castellano por internet (sí en inglés), mientras que los textos del segundo trabajo sí estaban disponibles *online*. En ambos casos cabía cerciorarse de dos aspectos: cuánto se había plagiado de internet y cuánto se había plagiado de los textos originales.

En el primer caso existía la dificultad de que no se podía comparar con un texto fuente informatizado, puesto que la versión castellana de ese capítulo no estaba disponible. Por ello, para someter los trabajos recibidos a examen, se utilizaron dos vías:

- Se compararon los trabajos con las informaciones disponibles *online*.
- Se compararon los trabajos entre sí.

Esta última opción, la de comparar entre sí los trabajos, no tiene como finalidad establecer si los alumnos se copiaron entre ellos, sino observar cuántas coincidencias comunes respecto a un texto fuente aparecían. Esto es, puede reconstruirse el capítulo 1 de la *Teoría de la relevancia* según la copia que de él hayan realizado los alumnos.

En cualquier caso, no todos los alumnos copian deliberadamente en sus trabajos. La tendencia a la recopilación bibliográfica es apreciable en estudiantes de primer curso y, en general, muchos desarrollan altas capacidades de análisis y síntesis. En casos de plagio académico, suelen concurrir diferentes circunstancias: desidia, prisa, falta de estímulos, o, incluso, ignorancia en relación con la infracción (quizá porque el alumno desconozca que el plagio directo supone una conducta errónea).

4.1. Resultados.

En total, se recibieron 15 trabajos de la primera opción (trabajo sobre el capítulo 1 de la relevancia), 10 sobre los artículos y 24 sobre la novela de *El curioso incidente del perro a medianoche*. Sobre ellos, se investigó el posible plagio de estos textos no solo respecto a los textos originales, sino también de la copia en internet.

En principio, era esperable que ante un texto de temática específica y con unos conceptos y terminología de compleja comprensión, la copia fuera mayor; por su parte, en cuanto al comentario valorativo de la novela, dado que se trata de un texto narrativo con finalidad lúdica, parecía que el plagio directo fuera menos esperable.

Como ejemplo, se expondrán aquellos casos en los que se apreció un plagio notable y sistemático. El procedimiento que se empleó, como se ha comentado con anterioridad, fue realizar una comparativa con la información disponible en internet y, en segundo lugar, se compararon los documentos entre sí.

Los documentos se han anonimizado y renombrado con letras del alfabeto. A modo de muestra, el texto catalogado como B (en la opción del resumen de dos artículos sobre la teoría de la relevancia) presentó un 41% de plagio de páginas de internet, que sería la primera fase del análisis. Los resultados hallados por el programa *Viper* se distribuyen de la siguiente manera:

1. Porcentaje de caracteres continuos coincidentes entre un texto y las páginas encontradas en internet.
2. Listado de páginas plagiadas (cada una recibe un color distinto). En ellas también aparece un porcentaje que, en dicho caso, indica el grado de copia del total de información o texto de esa página. Por ejemplo, si aparece un 15% significa que del total de información de la página X se ha copiado ese tanto por ciento.
3. Texto original con las oraciones sombreadas según el color de la página plagiada que se incluía en el segundo punto. Si alguna oración no está sombreada con ningún color, quiere decir que es la parte original del alumno.

En el caso expuesto del texto B, los resultados fueron los siguientes:

41% de contenido coincidente

Lista de páginas plagiadas:

40%	http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-338X2006000100002&lng=en&nrm=iso
3%	http://html.rincondelvago.com/introduccion-a-la-pragmatica_maria-victoria-escandell.html
2%	http://html.rincondelvago.com/semiotica-visual-y-teoria-del-cine.html
2%	http://www.monografias.com/trabajos42/semantica-pragmatica/semantica-pragmatica3.shtml
2%	http://www.lapaginadelguion.org/contenido-pragma.htm
1%	http://www.iberletras.com/06.htm

Ejemplo del texto⁴:

*El modelo del código y el modelo inferencial no son incompatibles, existen varias maneras **se combinarlos.***

Por lo tanto Sperber y Wilson nos dicen:

“...la comunicación puede conseguirse codificando y descodificando mensajes, y también produciendo pruebas que originen la inferencia deseada... cada uno de los dos modelos resulta adecuado para un determinado tipo de comunicación: de ahí que elevar a cualquiera de ellos a la condición de teoría general de la comunicación sea un error”.

*En la comunicación participan a la vez mecanismos de codificación y mecanismos inferenciales y que pueden interactuar, es importante **decir que son específicamente independientes el uno del otro** y que la comunicación **es independiente el uno del otro.***

Durante el caso de la comunicación verbal humana Sperber y Wilson proponen un modelo en que la fuente y el destino son procesos de pensamiento, el codificador y el descodificador son capacidades lingüísticas, el mensaje es un pensamiento y el canal es el aire que transporta en la señal acústica.

Sin embargo, existe un vacío entre la representación semántica de las oraciones y los pensamientos comunicados por los enunciados. Este vacío no se llena con más codificación sino con inferencia.

En referencia los modelos inferenciales se diferencian bastante de los procesos de descodificación en que los primeros parten de un conjunto de ideas y llegan a una serie de conclusiones que derivan de forma lógica a las ideas, o por lo menos, están garantizados por las mismas, mientras que un proceso de descodificación parte de una señal y llegan a la recuperación de un mensaje que es asociado a la señal por una letra...

⁴ Se ha sustituido el sombreado en color azulado que aparecía en el informe original por una tipografía cursiva para indicar el plagio, mientras que en negrita se ha resaltado el texto original del alumno. Así mismo, esta es solo una muestra del texto completo del estudiante.

Como se observará, el plagio es prácticamente absoluto en este fragmento. Sin embargo, hay una cita explícita (“Sperger y Wilson nos dicen:”) que pone de manifiesto diferentes aspectos, como la equivocación en el nombre de uno de los autores (Sperber) o la incorrecta citación bibliográfica (no se indica el año ni el número de las páginas del texto de donde procede la cita).

En general, si algún estudiante copia fragmentos de páginas de internet, el profesor puede recurrir a fichas como la mostrada anteriormente, en las que aparece el grado de copia, las páginas copiadas y los fragmentos plagiados. Ello es sumamente ágil y sistemático mediante programas como *Viper*.

4.2. Segunda situación.

Imagínese que uno de los trabajos presentara un carácter descriptivo o que, en general, requiriera de la búsqueda de información bibliográfica para realizar un análisis contrastivo o valorativo. En el contexto académico, los trabajos de carácter sintético suponen la lectura y sistematización de datos externos, normalmente en forma de lectura crítica.

En estos trabajos de recopilación bibliográfica, el alumno, habitualmente, copia directamente los textos que encuentra en sus lecturas, pero no los parafrasea y eso convierte su actividad en un hecho reprochable. Ahora bien, el plagio directo es un procedimiento común entre el alumnado, consistente en adquirir diferentes textos (artículos o libros) y en copiarlos deliberadamente.

Ante esta circunstancia, el docente dispone de herramientas de contraste formal entre textos previamente introducidos. Piénsese que, como era el caso anteriormente expuesto, se solicitara al estudiante que realizara un resumen de un texto A que, por su parte, no está informatizado o que no está disponible *online*. En tal caso, el alumno podría optar por dos vías: una, la copia directa de fragmentos del texto A, con lo que la interiorización de conceptos se vería diluida; dos, la paráfrasis asimilativa de términos o ideas del texto original.

Lo ideal en estas situaciones es que el alumno leyera, entendiera y luego expusiera, en palabras propias, una síntesis del texto leído; no obstante, esta tendencia no es la más corriente. En el marco académico actual, por prisas o por otros factores extraacadémicos, el estudiante copia directamente para ahorrarse la segunda y la tercera fase del análisis, esto es, el entendimiento y la exposición propia.

Por tanto, para poder utilizar herramientas informáticas que estipulen el grado de coincidencia textual, el profesor necesita de varios elementos:

1. Los textos de los alumnos
2. El texto o textos originales sobre los que se pide un trabajo determinado.

El profesor no siempre dispondrá del texto informatizado, puesto que algunas lecturas solicitadas pueden proceder de libros o de artículos únicamente disponibles en papel. En ese caso, podría pensarse que no puede determinarse el plagio directo del texto original, pero no es realmente así. Ante un conjunto de textos del alumnado, el grado de copia respecto a un texto original determinado será más o menos similar o, como mínimo, será aproximado. Por ello, es interesante observar en qué medida se parecen los textos de los alumnos entre sí, dado que en ese caso se determinará el plagio o copia de un texto determinado.

En líneas generales, en el caso que se presenta, se detectó una copia directa (*verbatim*) de muchos fragmentos del capítulo. Mediante el uso del programa *wcopyfind* se establecieron plagios de entre el 20% al 40%, es decir, que, respecto a un texto fuente o texto original, se copió directamente ese porcentaje. Uno de los aspectos más positivos del programa es que los resultados se diversifican en dos pantallas en las que aparecen en rojo y subrayados los fragmentos textuales plagiados. Véase, al respecto, un ejemplo de los resultados hallados por el programa⁵:

Texto 1
<p><u>El entorno cognitivo de un individuo es un conjunto de supuestos a los que dicho individuo tiene acceso. Los seres humanos son dispositivos de procesamiento de información muy eficaces. La mayoría de discusiones sobre procesamiento de información, ya sea en el campo de la psicología experimental o de la inteligencia artificial, se han preocupado de la realización de objetivos absolutos. Los recursos generales para el procesamiento de información que posee un individuo no son muy flexibles. Por ellos la eficacia cognitiva a largo plazo consiste en mejorar nuestros conocimientos sobre el mundo como sea posible utilizando los recursos que poseemos. El problema para un eficaz procesamiento de la información a corto plazo es conseguir una asignación óptima de los recursos de procesamiento central. Una parte de la información es vieja y otra parte de la información no sólo es nueva, sino que no está conectada con nada en la representación del mundo que tiene el individuo. Cuando estos nuevos y viejos elementos de información interconectados se utilizan conjuntamente dentro de un proceso inferencial, es posible deducir información que no habría podido inferirse sin esta combinación de informaciones. Cuando el procesamiento de información nueva suscita esta clase de efecto de multiplicación decimos que relevante. Cuanto mayor es el efecto de multiplicación, mayor es la relevancia.</u></p>
Texto 2
<p><u>El entorno cognitivo de un individuo es un conjunto de supuestos a los que dicho individuo tiene acceso. La cuestión es cuáles supuestos concretos es más probable que</u></p>

⁵ Si bien en el informe final presentado por el programa *wcopyfind* los resultados coincidentes se presentan en rojo, en estos casos se ha sustituido el rojo por un subrayado.

ese individuo vaya a construir y procesar. Los seres humanos son dispositivos de procesamiento de información muy eficaces. La eficacia sólo puede definirse con relación a un objetivo. Algunos objetivos son absolutos, consisten en crear un estado de cosas concreto que puede existir o no en un momento dado. Otros objetivos, son relativos, consisten en elevar el valor de una determinada variable.

Así pues, la eficacia cognitiva a largo plazo consiste en mejorar nuestro conocimiento del mundo tanto como sea posible con los recursos que disponemos. Los recursos deben asignarse al procesamiento de información que probablemente vaya a suponer una contribución mayor a los objetivos cognitivos generales de la mente con el mínimo coste de procesamiento. Una parte de la información es vieja, otra parte de la información es nueva y no esta conectada con nada y otra parte es nueva y esta conectada con la información vieja. Cuando estos nuevos y viejos elementos de información interconectados se utilizan conjuntamente en un proceso inferencial es posible deducir información nueva. Cuando el procesamiento de información nueva suscita esta clase de efecto de multiplicación decimos que es relevante, y cuanto mayor es el efecto de multiplicación mayor es la relevancia.

En los ejemplos anteriores, el texto 1 y el texto 2 coinciden mayoritariamente en sus informaciones y en la manera de presentarlas. De este modo, es muy sencillo detectar la copia entre el alumnado. Ahora bien, en el caso anterior la copia no se ha realizado entre los autores de ambos textos, sino que estos han copiado por separado un texto base o texto original.

Si se observan los textos anteriores, se detectará fácilmente que la copia del autor del texto 1 respecto al texto fuente es mayor que la del texto 2. Adviértase al respecto que el procedimiento que realiza el autor del segundo texto es ampliar oraciones sueltas copiadas *verbatim* del texto fuente, con lo que, de algún modo, sí se produce una cierta tarea de redacción crítica. Por su parte, en el texto 1, la presencia de texto copiado *verbatim* es mayor y, en ese sentido, no hay tantas palabras propias de su autor (redacción personal valorativa) como en las del texto 2. Esa copia, la del texto 1, debe ser sancionada en mayor medida que la del texto 2, si bien debe reconocerse que ambos casos son incorrectos dentro de la legalidad y honradez académicas.

5. Conclusiones.

En general, a partir de lo expuesto anteriormente, se percibe que el plagio o copia directa es una actividad frecuente en la cotidianidad académica. El principal problema es que los propios alumnos entienden el plagio como un acto habitual y no sancionable. Actualmente, acceder a infinidad de información *online* es un arma de doble filo, dado que por una parte permite documentar a los estudiantes y, por otra, estos son seducidos por la posibilidad de realizar un trabajo de manera fácil y rápida, mediante el plagio directo.

Por otro lado, el profesor se encuentra ante situaciones complicadas, dado que si permite la realización de trabajos optativos se expone a la copia de textos (desde Internet

o desde textos escritos [libros, artículos de investigación, etc.]). Por eso, como pautas orientativas, este artículo ha presentado directrices básicas para el análisis de copia en los textos del alumnado.

En primer lugar, se ha observado que la metodología clásica del comentario de textos es importantísima en el informe cualitativo y valorativo de los textos. En segundo lugar, a partir de textos reales, se ha procedido al análisis directo mediante el uso de dos herramientas informáticas: *wcopyfind* y *viper*. El primero permite observar el porcentaje de copia textual entre varios textos. Ello puede ser útil en dos vías: observar la copia de los textos entre sí u observar el grado de coincidencia entre esos textos y un texto fuente originario. Finalmente, el programa *viper* realiza la misma operación, aunque incrementa las posibilidades analíticas, ya que compara los textos con toda la información que puede encontrarse *online*. Esto es útil porque permite cuantificar el grado de plagio de páginas de Internet y, además, se ofrece información detallada sobre esas páginas plagiadas (*urls*, *nombre de la página*, *grado porcentual de plagio*, etc.).

6. Bibliografía.

Alcaraz, E. “La lingüística legal: el uso, el abuso y la manipulación del lenguaje jurídico”, *Lingüística forense, lengua y derecho: conceptos, métodos y aplicaciones*, M. T. Turell (Ed.), Institut Universitari de Lingüística Aplicada, UPF, Documenta Universitaria, Barcelona, 2005.

Bamford, J.; Sergiou, K. “International Students and Plagiarism: an analysis of the reasons for plagiarism among international foundation students”, *Investigations in university teaching and learning*, vol. 2 (2), 2005, págs. 17-22.

Barrón-Cedeño, Alberto. *Detección automática de plagio en texto*, publicación electrónica. www.mavir.net/docs/Barron-DeteccionPlagioTexto.pdf [07-01-2010].

Barrón-Cedeño, A.; Rosso, P.; Pinto, D.; Juan, A. “On Cross-lingual Plagiarism Analysis using a Statistical Model”, *Proceedings of the ECAI'08 PAN Workshop: Uncovering Plagiarism, Authorship and Social Software Misuse*, 2008, págs. 9-13.

Briz, A. “La peritación lingüística: la identificación del plagio”, *Palabras, norma, discurso. En memoria de Fernando Lázaro Carreter*, Luis Santos Río (Ed.), Ediciones Universidad, Salamanca, 2006, págs. 167-183.

Carroll, Jude; Appleton, Jon. *Plagiarism a Good Practice Guide*, Oxford Brookes University, Oxford, 2001.

Clough, P. “Old and new challenges in automatic plagiarism detection”, *National UK Plagiarism Advisory Service*, JISC, publicación electrónica. http://ir.shef.ac.uk/cloughie/papers/pas_plagiarism.pdf [07-01-2010].

Dreher, Heinz. “Automatic Conceptual Analysis for Plagiarism Detection”, *Issues in Informing Science and Information Technology*, Vol. 4, 2007.

<http://proceedings.informingscience.org/InSITE2007/IISITv4p601-614Dreh383.pdf> [07-01-2010].

Gotterbarn, D.; Miller, K.; Impagliazzo, J. “Plagiarism and Scholarly Publications: An Ethical Analysis”, *36th ASEE/IEEE Frontiers in Education Conference*, 2006. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4116889> [07-01-2010].

Gray, A., Sallis, P. and MacDonell, S. “Software Forensics: Extending Authorship Analysis Techniques to Computer Programs”, *Proceedings of the 3rd Biannual Conference of the International Association of Forensic Linguists (IAFL)*, 1997, págs. 1-8.

Haddon, Mark. *El curioso incidente del perro a medianoche*, Salamandra, Barcelona, 2004.

Maurer, H.; Kappe, F.; Zaka, B. “Plagiarism - A survey”, *Journal of Universal Computer Science*, vol. 12, n. 8, 2006, págs. 1050-1084.

Moya, C. “Relevancia e Inferencia: Procesos cognitivos propios de la comunicación humana”, 2006. <http://www.scielo.org.co/pdf/fyf/n19/n19a02.pdf> [07-01-2010].

Park, C. “In Other (People’s) Words: plagiarism by university students— literature and lessons”, *Assessment & Evaluation in Higher Education*, Vol. 28, n. 5, 2003, págs. 471-488.

Putniņš, T. J.; Signoriello, D.; Jain, S.; Berryman, M. J.; Abbott, D. “Advanced text authorship detection methods and their application to biblical texts”, *Complex Systems : Microelectronics, MEMS and Nanotechnology*, 2005, págs. 1-13.

Roig, Miguel. “Avoiding plagiarism, self-plagiarism, and other questionable writing practices: A guide to ethical writing”, publicación electrónica. <<http://facpub.stjohns.edu/~roigm/plagiarism>> [07-01-2010].

Seaward, L.; Matwin, S. “Intrinsic Plagiarism Detection using Complexity Analysis”, *PAN’09*, 2009, págs. 56-61.

Sperber, Dan; Wilson, Deirdre. *Teoría de la relevancia*, Visor, Madrid, 1994.

Stein, B.; Koppel, M.; Stamatatos, E. “Plagiarism Analysis, Authorship Identification and Near-Duplicate Detection”, *ACM SIGIR Forum*, Vol. 41, n. 2, 2007, págs. 29-35.

Thomas, M.; College, I. “Plagiarism detection software”, *E-Leader Bangkok*, publicación electrónica. <http://www.g-casa.com/PDF/Bangkok%202008/Thomas,%20Bangkok,%202008.pdf> [07-01-2010].

Trujillo, F. “La teoría de la relevancia como base para una nueva interpretación de la comunicación”, 2001. <http://dialnet.unirioja.es/servlet/articulo?codigo=1183088> [07-01-2010].

Turell, M. T. “El plagio en la traducción literaria”, *Lingüística forense, lengua y derecho: conceptos, métodos y aplicaciones*. M. T. Turell (Ed.), Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, Documenta Universitaria, Barcelona, 2005.