

Constatación de la validez de los resúmenes adjuntos a artículos médicos de investigación de cara a la evaluación de resúmenes automáticos

IRIA DA CUNHA; JAUME LLOPIS

Instituto Universitario de lingüística aplicada

Universidad Pompeu Fabra

La Rambla, 30-32 08082 Barcelona

iria.dacunha@upf.edu; jllolis@ub.edu

RESUMEN

El resumen automático es un tema complejo sobre el que se está trabajando desde diversas perspectivas. Nuestro trabajo quiere llegar a un modelo lingüístico válido de resumen automático y, para ello, integramos varias perspectivas: textual, léxica, discursiva, sintáctica y comunicativa. En concreto, nos centraremos en artículos médicos en español. Estos artículos se publican obligatoriamente acompañados de sus respectivos resúmenes redactados por el mismo autor. De cara a la evaluación de la calidad de los resúmenes de nuestro sistema aquellos nos serán útiles para poder compararlos. Para corroborar que los abstracts de los autores son adecuados para este tipo de textos, hemos llevado a cabo un experimento estadístico con una técnica denominada Multidimensional Scaling, en el que han participado tres lingüistas y tres médicos.

Palabras clave: resumen automático, Multidimensional Scaling, artículo médico, abstract.

ABSTRACT

Automatic summarization is a complex subject which is being worked on from different perspectives. Our work aims at getting a valid linguistic model of automatic summarization and, for that, we integrate several perspectives: textual, lexical, discursive, syntactic and communicative. These articles are published with a compulsory summary written by the same author. These summaries will be useful for the evaluation of the quality of the summaries generated by our system because we will compare them. To corroborate that the authors' abstracts are suitable for this type of texts, we have carried out a statistical experiment using a technique called Multidimensional Scaling, in which three linguists and three physicians have participate.

Keywords: automatic summarization, Multidimensional Scaling, medical article, abstract.

1. INTRODUCCIÓN

La investigación sobre resumen automático está en auge hoy en día debido a que se trata de un recurso efectivo que nos ayuda a combatir la "sobreinformación" a la que estamos expuestos en la actualidad, debida entre otros motivos, a la aparición de

Internet. Enormes cantidades de datos llegan a nuestras manos por diversos medios y con diferentes finalidades, y resulta difícil asimilar toda esta información. Debemos seleccionar la que mejor se adapte a nuestros intereses para procesarla en el menor tiempo posible. En este sentido nace el resumen automático, como un recurso que permite economizar nuestro tiempo.

Los estudios sobre resumen automático se están realizando desde los años sesenta y se ha trabajado sobre ello desde diferentes perspectivas. Las técnicas más habituales utilizan un solo tipo de criterio, ya sea estadístico o lingüístico, y podrían clasificarse teniendo en cuenta si utilizan métodos superficiales (basados en frecuencias de palabras, títulos, posición de determinados fragmentos...), métodos de nivel medio (basados en cadenas léxicas, correferencia...) o métodos profundos (basados en la estructura del discurso, por ejemplo). Obras de referencia en este sentido son Mani & Maybury (1999) y Mani (2001).

Nuestro trabajo se centra en el resumen automático de artículos médicos en español y, en concreto, consideramos que un solo tipo de criterio es insuficiente para llegar a un resumen automático válido de este tipo de textos. En nuestra investigación integramos diversas perspectivas lingüísticas: textual, léxica, discursiva, sintáctica y comunicativa. Para una explicación más detallada del modo de integración de los cinco criterios utilizados puede acudirse a da Cunha & Wanner (2005).

Para validar un modelo tan complejo como éste, necesitamos material de referencia. Hay diferentes métodos para obtener este material, como que el investigador redacte sus propios resúmenes y los compare con los de su resumidor, o tomar algún sistema de resumen automático como referencia y utilizar sus resúmenes como modelo de contraste. Nosotros creemos que el mejor material de contraste es el resumen que escribe el propio autor de un artículo médico. Al ser esta hipótesis una de las bases de nuestro trabajo debemos validarla, lo cual hacemos en este artículo. Utilizamos la técnica *Multidimensional Scaling* para realizar el experimento, obteniendo resultados positivos.

Así, nuestro estudio sobre resumen de artículos médicos toma como punto de partida el resumen del autor, ya que consideramos que éste es el adecuado para este tipo de textos. Estos resúmenes nos servirán como modelo con el que comparar los ofrecidos por nuestro sistema y, posteriormente, validarlos o no. Como ya hemos comentado, para confirmar que el resumen del autor es el adecuado en este tipo de textos, además de

esbozar una serie de argumentos teóricos, llevamos a cabo una prueba estadística que en principio los validará.

La estructura del artículo comienza con la introducción que hemos realizado hasta ahora, continúa con un segundo apartado en el que se ofrece la motivación sobre la validez del resumen del autor, y sigue con una tercera sección donde se muestra la validación empírica de dicha motivación (explicación, metodología, resultados y análisis del experimento). Finalmente el cuarto apartado se dedica a las conclusiones del trabajo, el quinto a los agradecimientos a los participantes en el experimento y el sexto a la bibliografía consultada.

2. MOTIVACIÓN DE LA VALIDEZ DEL RESUMEN DEL AUTOR

Creemos que el resumen que un autor redacta a partir de un artículo que ha escrito previamente es válido por varios motivos.

El primero de ellos es que el resumen del autor está orientado a especialistas en la materia, y al ser él mismo uno de ellos, sabrá con exactitud cuáles son los contenidos que debe incluir en el resumen, es decir, los más relevantes, y de los que debe prescindir por tratarse de información innecesaria. Estas consideraciones remiten a Cabré (2002:89):

Los textos devienen así producto de operaciones lingüístico-cognitivas realizadas en unas determinadas circunstancias discursivas. Estas circunstancias implican el emisor y el receptor (tipo de emisor y receptor, intenciones y nivel de conocimiento que ambos poseen sobre el tema), la situación (el medio en que se produce la comunicación y el sistema de transferencia utilizado), el propósito y las expectativas del emisor y receptor con relación a su interacción. Cada uno de estos factores posee un determinado valor en cada acto comunicativo y su conjunto explica la configuración de un tipo de texto que pretende ser adecuado a las circunstancias en las que se produce.

En segundo lugar, es el mismo autor quien escribe tanto el artículo como el resumen, con lo cual estos estarán estrechamente ligados y se corresponderán en cuanto a estructura y contenido.

En tercer lugar, la propia revista en donde se publicará el artículo (en nuestro estudio se trata de *Medicina Clínica*) facilita a los autores una guía de estilo para publicaciones médicas (<http://www.doyma.es/revista/info/pdf/002Normas.pdf>).

Como cuarto motivo tenemos en cuenta la obligación de los autores de seguir una estructura consensuada basada en cuatro apartados fijados: Introducción, Pacientes y métodos, Resultados y Discusión (estructura IMRD). Esta idea enlaza con van Dijk (1989:165): "[...] la aceptabilidad de la publicación depende de una serie de criterios que exigen métodos e informes adecuados".

Finalmente, debemos ser conscientes de que el artículo se publicará *con* el resumen, es decir, que el redactado de ambos debe ir al unísono para que tenga coherencia. Caso distinto sería el de los resúmenes que se envían a congresos, ya que posteriormente el artículo podría variar y no seguir exactamente las ideas del resumen inicial.

3. EXPERIMENTO ESTADÍSTICO PARA VALIDAR EL RESUMEN DEL AUTOR COMO PUNTO DE PARTIDA

Para corroborar la hipótesis de que el resumen del autor es adecuado para este tipo de artículos médicos, hemos realizado un experimento en el cual han colaborado tres médicos y tres lingüistas. A continuación se explica en qué consiste el *Multidimensional Scaling* (la técnica empleada), se explica la metodología que hemos seguido para el experimento, se ofrecen los resultados y se realizan algunas observaciones lingüísticas sobre ellos.

3.1. *MULTIDIMENSIONAL SCALING*

El *Multidimensional Scaling* es una técnica de reducción de dimensiones. Tiene como objetivo buscar una representación en dimensión 2 (en un plano) que sea lo más parecida a la representación de la nube de puntos original en dimensión d (siendo d el número de variables con las que se ha trabajado; en nuestro caso la d es igual a 96). Se trata, por lo tanto, de transportar un espacio d -dimensional a uno de dimensión 2. El procedimiento usado consiste, en realidad, en la representación de los individuos estudiados en las dos primeras componentes principales de un Análisis de Componentes

Principales. Estas dos componentes principales captan un máximo de variabilidad (de información) y, gracias a ello, la reducción de dimensiones se realiza minimizando la información perdida. En nuestro análisis las dos primeras componentes explican el 68,4% de la información. Teniendo en cuenta que estamos trabajando con 96 variables, significa esto que existe una fuerte correlación entre las variables y que la representación que nos hace el *Multidimensional Scaling* es realmente informativa.

3.2. METODOLOGÍA DEL EXPERIMENTO

El experimento ha consistido en ofrecer por separado a estas seis personas (tres médicos y tres lingüistas) cinco artículos médicos (sin los resúmenes de los autores) y pedirles que seleccionasen en cada texto los contenidos que considerasen indispensables para construir un buen resumen del mismo, ofreciéndoles unos parámetros de longitud máxima. Las instrucciones que se ofrecieron a los seis informantes fueron las siguientes:

- **Subraye** en cada texto los fragmentos que considere *indispensables* para construir un buen resumen del mismo (excepto títulos y subtítulos):
 - Originales = máximo 20 líneas subrayadas (aprox.)
 - Originales Breves = máximo 15 líneas subrayadas (aprox.)

En la Tabla 1 puede observarse una muestra de los contenidos seleccionados por el autor, los médicos y los lingüistas. En la primera columna se incluyen los diferentes contenidos de los textos y, en la segunda, el nombre asignado a cada uno de ellos, es decir, a las variables (X12, X13, etc). En la tercera columna observamos los contenidos seleccionados para el resumen por el autor y, en las siguientes, los contenidos escogidos por médicos y lingüistas. La selección o no de las variables se refleja en la tabla mediante ceros y unos, respectivamente.

contenido	variable	autor	med 1	med 2	med 3	ling 1	ling 2	ling 3
La mediana de edad a la seroconversión fue de 25 años,	X12	1	1	1	1	0	0	0
la mediana del año de seroconversión fue 1991,	X13	1	0	0	0	0	0	0
y la mediana de seguimiento fue 4 años (rango, 0-14,59).	X14	1	0	0	0	0	0	0
El período de incubación del sida hasta diciembre de 1996 fue de 10 años.	X15	1	1	1	1	0	0	1
Sólo la edad (riesgo relativo [RR] 1,026; intervalo de confianza [IC] del 95%, 1,005-1,047)	X16	1	1	1	1	0	0	1
y el sexo (RR, 0,58; IC del 95%, 0,35-0,99) se asociaron de forma estadísticamente significativa a la progresión del sida	X17	1	1	1	1	0	0	0
sin observarse un efecto del año de calendario	X18	1	1	1	1	0	0	0

Tabla 1. Muestra de los contenidos seleccionados por el autor, los médicos y los lingüistas.

En este experimento partimos de la idea de que los médicos tienen la base teórica para seleccionar los contenidos más adecuados, y de que los lingüistas tienen el conocimiento "lingüístico" para seguir la estructura del texto y deducir qué informaciones escoger. Como ya hemos mencionado, consideramos que el resumen del autor es adecuado, y la semejanza de los resúmenes de los autores con los de los médicos constataría esta afirmación.

Así, hemos comparado los contenidos del resumen del autor con los contenidos seleccionados por estos tres médicos y tres lingüistas y, como veremos, se ha comprobado que, efectivamente, el autor coincide mayoritariamente con los tres primeros. Para cuantificar los resultados hemos utilizado el programa estadístico *Statgraphics* mediante el cual se ha realizado, como ya se ha comentado, un *Multidimensional Scaling* para comprobar el grado de coincidencia entre los resúmenes del autor, de los médicos y de los lingüistas.

3.3. RESULTADOS

Los resultados del experimento se muestran en el Gráfico 1. La primera componente principal aparece reflejada como Col_100 y la segunda componente principal como Col_101. Se observa que los tres médicos y los autores de los artículos

seleccionan contenidos muy similares a la hora de redactar los resúmenes de los artículos, mientras que los lingüistas no se asemejan ni siquiera entre ellos. Mediante esta validación estadística tenemos una primera prueba empírica de que el resumen del autor es adecuado, ya que otros tres especialistas en medicina así lo constatan.

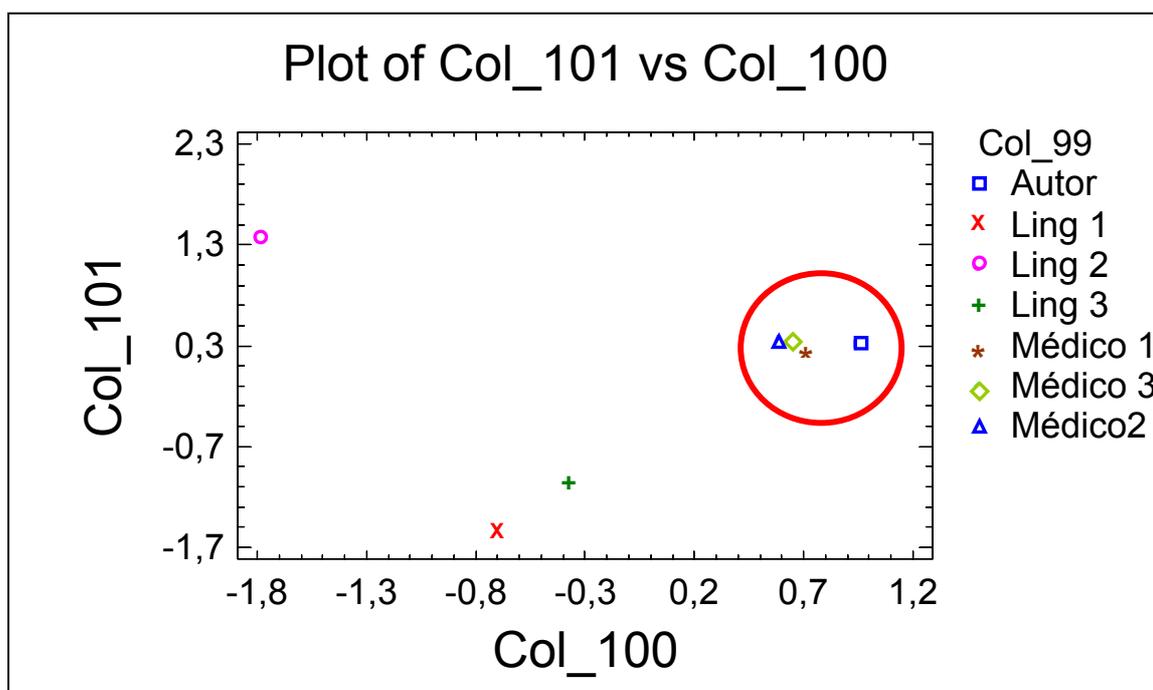


Gráfico 1. *Multidimensional Scaling* a partir de los contenidos escogidos para los resúmenes.

3.4. ANÁLISIS LINGÜÍSTICO DE LOS RESULTADOS

Después de realizar este experimento y de constatar la coincidencia entre los resúmenes de los médicos y la gran diferencia entre los fragmentos seleccionados por estos dos colectivos de profesionales (médicos y lingüistas), se ha observado qué tipo de contenidos ha escogido cada uno de ellos. Notamos una tendencia generalizada en los lingüistas a incluir en su resumen demasiada información del apartado de *Introducción*. Creemos que esto es debido a su falta de formación en medicina, la cual les lleva a seleccionar definiciones, datos históricos, trabajos anteriores al del artículo que se quiere resumir, etc. En cambio, los médicos no incluyen este tipo de informaciones debido a que éstas se sobreentienden al poseer una formación específica en este ámbito. Otro dato que nos llama la atención es que los médicos seleccionan información importante de cada uno de los apartados, y no así los lingüistas, que en algunos casos

descuidan información de alguno de ellos. Observamos además que los tres médicos suelen seleccionar información numérica, sobre todo en el apartado de *Pacientes y métodos* y *Resultados*, mientras que los lingüistas tienden a seleccionar información no numérica, con más explicaciones y menos cifras. Finalmente, vemos que los lingüistas seleccionan bastantes contenidos del apartado de *Conclusión*, mientras que los médicos ofrecen normalmente una conclusión mucho más breve.

Todas estas observaciones nos hacen ver que la formación concreta del colectivo que vaya a resumir un texto influye a la hora de hacerlo. Es decir, que aunque podamos creer de antemano que la formación de un lingüista pueda ser suficiente a la hora de seleccionar la información importante (ya que se supone que éste posee capacidades para la correcta visualización de la estructura del discurso, para encontrar conectores discursivos que guíen dicha estructura, para entender y asimilar la estructura textual e, incluso, sintáctica) finalmente esto no es así. Puede ser que los lingüistas seleccionen información relevante, pero ésta no es la necesaria para este tipo concreto de resumen, teniendo en cuenta el receptor. Son los resúmenes de los tres médicos los que se asemejan entre ellos y se acercan más al del autor, considerado éste como el resumen adecuado en este ámbito. Como conclusión del análisis, llegamos a la constatación de que para la realización de esta tarea se necesita ser un profesional en el ámbito al cual pertenecen los textos que se desee resumir. Los médicos, mediante su conocimiento y experiencia, son los más indicados para llevar a cabo un resumen correcto sobre un texto de medicina. De todas maneras debemos tener en cuenta que habrá ocasiones en las que el resumen del autor pueda no ser el más indicado, ya que hay autores que no tienen suficiente capacidad para condensar las ideas más relevantes de sus propios artículos, y por tanto, el resumen podrá ser de una calidad inferior a la esperada. Esta situación no es la habitual, pero debemos tenerla en cuenta para futuros experimentos.

4. CONCLUSIONES

El estudio realizado pretendía demostrar que los resúmenes que redactan los autores de artículos médicos publicados en revistas de investigación son adecuados para este tipo de textos. El experimento indica que otros tres profesionales del ámbito de la

medicina seleccionan informaciones muy similares a ellos a la hora de resumir. No así el otro colectivo del experimento, los lingüistas.

Este trabajo se ha llevado a cabo con la intención de contrastar en el futuro los resúmenes de los médicos con los *output* de nuestro sistema de resumen automático de artículos médicos en español. Creemos que este trabajo muestra la viabilidad de nuestro planteamiento, aunque somos conscientes de que para constatarlo definitivamente debemos ampliar el estudio, lo cual nos deja trabajo futuro por delante.

5. AGRADECIMIENTOS

Este estudio no podría haberse llevado a cabo sin la estrecha colaboración de los seis participantes del experimento. A ellos queremos agradecer su constancia y esfuerzo. Se trata de tres lingüistas del Institut Universitari de Lingüística Aplicada de la Universitat Pompeu Fabra de Barcelona (Carme Bach, Anna Joan y Rogelio Nazar) y de tres médicos de A Coruña: Eduardo Barge (Hospital Juan Canalejo), Iria Glez (Hospital Teresa Herrera) y Daniel López (Hospital Juan Canalejo).

6. BIBLIOGRAFÍA

- Bhatia, V. 1993. *Analyzing genre: Language Use in Professional Settings*. Londres: Longman.
- Cabré, M. T. 2002. "Análisis textual y terminología, factores de activación de la competencia cognitiva en la traducción". En Alcina & Gamero (2002): *La traducción científico-técnica y la terminología en la sociedad de información*. Castellón: Univ. Jaume I. 87-105.
- da Cunha, I. y L. Wanner. 2005. "Towards the Automatic Summarization of Medical Articles in Spanish: Integration of textual, lexical, discursive and syntactic criteria". En actas del Workshop "Crossing Barriers in Text Summarization Research". RANLP-2005 (*Recent Advances in Natural Language Processing*). Borovets (Bulgaria).
- Llopis, J. 1996. *La estadística: una orquesta hecha instrumento*. Barcelona: Ariel Ciencia.

- Mani, I. y M. Maybury. 1999. *Advances in Automatic Text Summarization*. The MIT Press.
- Mani, I. 2001. *Automatic Summarization*. Amsterdam: John Benjamins.
- Teufel, S. y M. Monees. 2002. "Summarizing Scientific Articles: Experiments with Relevance and Rhetorical Status". *Computational Linguistics* 28. 409-445.
- van Dijk, T. A. 1989. *La ciencia del texto*. Barcelona: Paidós.