

Desarrollo de la capacidad de resumen y procesamiento de los resúmenes de textos escolares

Development of the capacity of summarization and scholars' texts summaries processing

Unai Atutxa

Euskal Herriko Unibertsitatea (UPV/EHU) Ixa Taldea
atutxaunai@gmail.com

Resumen: El resumen tiene como base la comprensión, de ahí la importancia que tiene en la educación. El alumno o alumna en la escuela tiene que ser capaz de entender lo que lee, oye o ve; para después expresar lo entendido a su modo resumiendo y jerarquizando el contenido. Nuestro objetivo en la tesis es la de estudiar la capacidad de resumir de los alumnos y alumnas e incidir en ella. Para conseguirlo emplearemos técnicas de Procesamiento del Lenguaje Natural desde el punto de vista textual y discursivo. El acercamiento teórico es la *Rhetorical Structure Theory* (RST), válido también para generar y evaluar resúmenes automáticamente. Las etapas en las que investigaremos la capacidad de resumir de los alumnos y alumnas serán desde la Educación Primaria hasta la universidad, de este modo, queremos identificar las necesidades que tienen en etapas diferentes cuando tratan de elaborar un resumen; y así poder ofrecer recursos didácticos y una metodología o recursos para responder a dichas necesidades.

Palabras clave: Resumen automático, evaluación, educación, análisis discursivo

Abstract: Summarization is closely related to comprehension, this makes summarization very important in education. Students have to be capable of understanding what they are reading, seeing or listening; to understand, link and underline the most important ideas into words. Our main aim in this PhD is to analyse the summarizing skills of the student and provide them resources to develop these skills. To do so, we are going to use Natural Language Processing Techniques and Rhetorical Structure Theory (RST), which is useful to create and evaluate summaries automatically. The target of this study goes from Primary Education to university, in this way, we want to identify the needs that they have on these stages when they summarize a text; and in this way, we will have the opportunity to offer educational resources and a sound methodology to answer those needs we detect.

Keywords: automatic summarization, evaluation, education, discourse analysis

1 Justificación de la investigación propuesta

En el entorno educativo, la tarea de resumir textos es de gran importancia ya que refleja la capacidad de síntesis y de comprensión de la información de quien realiza el resumen. Hoy en día, el alumno o la alumna puede tener entre manos una gran cantidad de información, sin embargo, eso no garantiza que entienda y que interiorice toda esa información. Antes de nada, el alumno o la alumna tiene que ser capaz de entender lo que lee, oye o lo que ve; y después,

mediante sus palabras (o imágenes) deberá dar a conocer lo más importante de todo lo que ha entendido. Resumir un texto es un proceso en el que se han de extraer las ideas más relevantes, mantener la coherencia entre dichas ideas en un espacio más reducido al del texto original. Sin embargo, la información que los alumnos toman por relevante no siempre concuerda con la información más importante del texto.

Nuestra intención en la tesis es estudiar la competencia de resumir y proponer métodos para su evaluación. Para ello, tiene especial importancia trabajar a nivel textual y del discurso; ya que interviniendo a nivel

discursivo seremos capaces de entender cómo entienden y jerarquizan la información los alumnos y alumnas.

Suele ser habitual no trabajar el resumen en la escuela por falta de tiempo para su corrección o por falta de herramientas que ayuden en tareas de evaluación. Si nos detenemos a observar el material escolar, por ejemplo, en nuestro entorno, en el País Vasco, nos damos cuenta que en euskera, el resumen de abstracción se trabaja poco en general. En consecuencia, competencias tan fundamentales como los relacionados con la gramática, la progresión temática, la conexión, la cohesión y la coherencia no se adquieren adecuadamente.

En el Decreto educativo 236/2015 de la Comunidad Autónoma del País Vasco, al desglosar las competencias transversales y disciplinares, podemos darnos cuenta que el resumen debe tener un papel fundamental a la hora de desarrollar tanto las competencias transversales como las disciplinares.

Para poder avanzar en el desarrollo de la competencia del resumen en la escuela, consideramos que es necesario desarrollar herramientas que usen tecnologías del procesamiento del lenguaje natural (PLN). Pensamos que si se desarrolla un resumidor automático para el euskera (usando técnicas superficiales) y la evaluación automática de los resúmenes, nos brindará un contexto adecuado para poder estudiar las necesidades que tienen profesorado como alumnado (de primaria, secundaria y universidad). Para ello, en esta tesis utilizaremos técnicas del Procesamiento del Lenguaje y partiremos desde la Teoría de la Estructura Retórica o *Rhetorical Structure Theory* (RST) (Mann y Thompson, 1987). La RST es una teoría que nos permite describir la coherencia entre fragmentos textuales combinando la idea de nuclearidad, o importancia de un fragmento del discurso, con la identificación de las relaciones retóricas que unen los fragmentos del texto.

Con esta teoría se han descrito las relaciones retóricas de diversas lenguas, entre ellas el euskera (Iruskieta et al., 2013). Además, se ha empleado en diferentes tareas en el Procesamiento del Lenguaje (Taboada y Mann, 2006) y también en la generación de analizadores que describen automáticamente la estructura discursiva; el analizador CODRA (Joty et al., 2015) es ejemplo de ello.

2 Origen y trabajo relacionado

Algunos investigadores han analizado la aportación que le supone la estructura discursiva al resumen automático, por ejemplo, Wang et al. (2015) y Cohan y Goharian (2017) utilizaron la estructura discursiva para resumir textos. Por otra parte, Bosma (2008), Chencheng (2010), y Cardoso et al. (2015, 2011) trabajaron el resumen basándose en RST. Además, es reseñable que también se hayan empleado para el resumen de textos técnicas de las redes neuronales (Sarda y Kulkarni 2015).

En palabras de Molina (2013), la generación automática de resúmenes es una tarea desafiante. Al clasificar tipos de resúmenes, diferencia dos tipos de resúmenes según la metodología empleada: la extracción y la abstracción. El resumen por extracción ha sido el más desarrollado dada la dificultad de resumir por abstracción, ya que a la tarea de tener que identificar las ideas principales se le suma la de hacerlo mediante frases escritas de diferente manera. Según Molina, la comprensión de frases establece un puente entre ambas familias. Presenta una investigación sobre la comprensión de frases y propone un modelo lineal que es aplicable al resumen.

Además, apreciamos que hay diversos trabajos en el ámbito del resumen automático. Vivaldi et al. (2010) presentan un nuevo algoritmo para el resumen automático de textos especializados que combinan recursos terminológicos y semánticos: un extractor de términos y una ontología. El extractor de términos proporciona la lista de los términos que están presentes en el texto junto con su término correspondiente. La ontología se utiliza para calcular la similitud semántica entre los términos encontrados en el cuerpo del texto y los que se encuentran en el título del documento. Las frases con la puntuación más alta se eligen para formar parte del resumen final.

A su vez, Barberi et al. (2015) describen la demostración de una serie de aplicaciones de resumen automático y extracción de informaciones integradas en una tableta. En este trabajo presentan funcionalidades para resumir las noticias publicadas en la Web, extraer información sobre eventos y resumir textos que desee el usuario tanto en español como en inglés.

Pero, no solo se trata de la generación de resúmenes automáticos, también es importante

su evaluación. Zipitria (2008) analiza la toma de decisiones en la evaluación de resúmenes en un marco computacional. Llevó a cabo un estudio empírico para analizar las decisiones que subyacen en el comportamiento humano a la hora de clasificar resúmenes.

En cuanto a la evaluación de los resúmenes, otro trabajo de Zipitria et al. (2013), aborda el tema de la obtención automática de puntuaciones de discurso general a partir de medidas de discurso de superficie para el euskera.

Por otro lado, Saggion et al (2010) estudian la correlación de las clasificaciones de los sistemas de resumen de texto utilizando métodos de evaluación con y sin modelos humanos.

Como hemos dicho previamente, nos basaremos en la estructura discursiva para poder llevar a cabo esta tesis. Para ello, nuestra base será RST. Además de utilizarlo para segmentar los textos y determinar la estructura jerárquica, será uno de los ejes que emplearemos para evaluar los resúmenes.

3 Descripción de la investigación

Nuestro objetivo es desarrollar la competencia de resumir de los alumnos, para poder incidir en ella, y a la vez, proponer recursos didácticos y metodología basados en el Procesamiento del Lenguaje; y cuya finalidad sea posibilitar el desarrollo de dicha competencia en las diferentes etapas del proceso de aprendizaje. Es por ello que en esta tesis intentaremos responder a las necesidades teóricas y metodológicas que pueda haber en las diferentes etapas. Para ello, hemos establecido unos objetivos más concretos, debemos tener en cuenta que dichos objetivos no son secuenciales y se alimentarán unos a otros:

3.1 Proponer una metodología y recursos didácticos para el resumen

La metodología deberá de tener como eje una perspectiva transdisciplinar o global y, a la vez, deberá permitir un aprendizaje cooperativo.

Trabajaremos sobre todo con textos en euskera, y una vez hayamos desarrollado la metodología, investigaremos la capacidad de los alumnos en otras lenguas y si los resultados son comparables.

A la hora de proponer recursos didácticos (unidades didácticas o actividades), será de suma importancia poder ofrecer al alumno o

alumna textos o actividades que estén un poco más allá de su nivel actual de competencia lingüística, tal y como propone Krashen (1985) en su teoría de la adquisición del segundo idioma.

3.2 Construir un corpus adecuado

La metodología que vamos a proponer estará basada en datos, y se utilizarán técnicas de PLN; por lo tanto, tendremos que elaborar un corpus adecuado para ello. Utilizaremos textos que usan los alumnos y las alumnas en la escuela. A partir de esos textos, reuniremos modelos elaborados por personas expertas y resúmenes generados por los alumnos. Es de mencionar que a día de hoy tenemos un acuerdo con la Confederación de escuelas Euskal Herriko Ikastolen Elkarte (<https://www.ikastola.eus/>) . Gracias a ello podemos trabajar con textos que se emplean en las escuelas y obtener resúmenes reales en contexto escolar elaborados a partir de esos textos.

I) Textos que se deberán resumir: como nuestro objetivo es recolectar resúmenes elaborados por los alumnos y las alumnas, utilizaremos textos que se empleen en el proceso de aprendizaje. Utilizaremos textos de Educación Primaria, Secundaria y de la universidad. Trabajaremos con distintos géneros de textos, ya que cualquier texto producido de un estudiante y su evaluación están condicionados por el género textual. II) Gold Standard: una vez tomada la decisión de cuáles serán los textos a resumir, crearemos su Gold Standard. Serán resúmenes realizados por expertos: de extracción y abstracción. Los Gold Standard tendrán dos funciones principales. Por una parte los utilizaremos para compararlos con los resúmenes de los alumnos; por otra parte, los emplearemos para que la herramienta tenga con que comparar los resúmenes realizados por los alumnos y así poder ofrecer una generación y evaluación automática.

II) Gold Standard: una vez tomada la decisión de cuáles serán los textos a resumir, crearemos su Gold Standard. Serán resúmenes realizados por expertos: de extracción y abstracción. Los Gold Standard tendrán dos funciones principales. Por una parte los utilizaremos para compararlos con los resúmenes de los alumnos; por otra parte, los emplearemos para que la herramienta tenga con que comparar los resúmenes realizados por los

alumnos y así poder ofrecer una generación y evaluación automática.

III) Resúmenes de los alumnos y alumnas: para finalizar, reuniremos los resúmenes de los alumnos y alumnas. Hasta la fecha hemos obtenido 1121 resúmenes (685 de alumnos universitarios y 436 de alumnos de primaria) de los cuales 1067 se reunieron digitalmente mediante COMPRESS-EUS (Atutxa et al., 2017).

3.3 Crear criterios de evaluación para evaluar los resúmenes

Para evaluar los resúmenes propondremos criterios de evaluación. Dichos criterios, por una parte, serán significativos para la evaluación que llevará a cabo el profesor o profesora; y por otra parte, serán diseñados para poder aplicar con ellos técnicas de Procesamiento del Lenguaje. Utilizaremos técnicas superficiales desde las bolsas de palabras hasta las discursivas.

3.4 La evaluación de los alumnos y alumnas

I) Evaluación automática: A parte de las técnicas estándar ROUGE y BLEU que se emplean para la evaluación de los resúmenes, propondremos y utilizaremos otras técnicas basadas en el discurso para evaluar el corpus recogido. En este momento evaluamos los resúmenes mediante ROUGE y BLEU y nuestro siguiente paso será evaluar la fiabilidad de dichas evaluaciones y examinar qué podemos implementar para mejorarlas.

II) Evaluación humana: evaluaremos los resúmenes de los alumnos y alumnas con los criterios de evaluación establecidos previamente.

III) Comparación entre la evaluación automática y la humana: evaluaremos la fiabilidad de la evaluación automática y nos dará pie a saber en qué aspectos debemos alimentarla. La herramienta COMPRESS-EUS estará amoldada a los criterios de evaluación que hayamos establecido previamente. COMPRESS-EUS está disponible para quien quiera utilizarlo en la siguiente página web: <http://ixa2.si.ehu.es/compress-eus/>.

3.5 Diseñar el feedback para los alumnos, alumnas, profesores y profesoras

Después de haber evaluado los resúmenes con distintas técnicas, nuestro objetivo será proporcionar un *feedback* automático del resumen realizado por el alumnado; y a la vez, proporcionarle al profesorado un *feedback* añadido para que tengan la opción de interpretar lo mejor posible el resumen realizado por el alumnado, de forma individual o colectiva.

3.6 Desarrollar una herramienta para reunir resúmenes

Crear una interfaz que nos dé la opción de reunir resúmenes (extracción y abstracción) elaborados por los alumnos y alumnas. Para eso elaboramos y mejoramos la herramienta COMPRESS-EUS. Esta herramienta, nos facilitará evaluar los resúmenes teniendo en cuenta su estructura discursiva, dándonos la mayor información posible de manera automática. Además, nos da la opción de saber la opinión del usuario.

3.7 Difusión del trabajo realizado

Para finalizar, dejaremos los resultados más significativos en manos de la comunidad educativa y documentamos todo el trabajo realizado de manera sistemática, para que sea beneficioso para otros investigadores o usuarios de ámbito internacional.

4 Cuestiones específicas de investigación

Las principales cuestiones de investigación a las que pretendemos responder con esta tesis son las siguientes:

- Nuestra investigación tendrá como base el nivel discursivo. ¿Qué otros aspectos o niveles del lenguaje se deben armonizar junto al discurso para conseguir un resumen o una evaluación automática que sea lo más completa posible? ¿Cómo se puede llevar a cabo esa armonización de niveles diferentes del lenguaje?
- ¿Cómo podemos comparar dos abstracciones descritas mediante RST? ¿Qué técnicas o métodos se pueden emplear para poder equiparar el

contenido de una posposición, que después nos permita comparar textos descritos mediante RST?

- La evaluación en el resumen de texto puede ser extrínseca o intrínseca. En una evaluación extrínseca, los resúmenes se evalúan en el contexto de una tarea específica que debe realizar un ser humano o una máquina. En una evaluación intrínseca, los resúmenes se evalúan en referencia a algún modelo ideal, en nuestro caso el Gold Standard. ¿Qué información obtendremos de cada una de las evaluaciones? ¿Es posible que cada tipo de evaluación ayude a obtener cierta información y que eso nos posibilite ofrecer un *feedback* más completo?
- El resumen requiere de comprensión y jerarquización, y factores como la utilización de imágenes afectan de manera directa a los dos componentes. ¿Qué factores debemos tener en cuenta a la hora de potenciar la comprensión y la jerarquización? ¿Cómo debemos adaptar los factores a cada etapa de aprendizaje? ¿Qué efecto causan en cada una de ellas?
- A la hora de proporcionarles el feedback a los alumnos, debemos ofrecer actividades o unidades didácticas que estén un poco más allá del nivel actual de su competencia lingüística. ¿Cómo podemos saber el nivel actual de cada alumno utilizando técnicas de PLN? ¿Cómo podremos ofrecer a cada alumno o alumna automáticamente recursos didácticos que estén un poco más allá del nivel actual de su competencia lingüística?
- ¿Qué influencia puede llegar a tener un generador de resúmenes automáticos en un aula? ¿Cómo le puede beneficiar al alumno o al profesor la obtención automática de un resumen? ¿Qué características tendrá que tener un generador de resúmenes automático, para que contribuya lo máximo posible al desarrollo de la capacidad de resumir?

Bibliografía

- Atutxa, U; Iruskieta, Mikel; Ansa, O; Molina, A. 2017. COMPRESS-EUS: I(ra)kasleen laburpenak lortzeko tresna. EUDIA: Euskararen bariazioa eta bariazioaren irakaskuntza-III. 87-98.
- Barbieri, F., Ronzano, F., & Saggion, H. (2015). Summarization and Information Extraction in your Tablet. *Procesamiento del Lenguaje Natural*, (55)
- Bosma, Wauter E. (2008). Discourse oriented summarization. Enschede: University of Twente Thesis Typ
- Cardoso, P. C., Jorge, M. L. D. R. C., & Pardo, T. A. S. (2015). Exploring the Rhetorical Structure Theory for multi-document summarization. In Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural, XXXI. Sociedad Española para el Procesamiento del Lenguaje Natural-SEPLN.
- Cardoso, P. C., Maziero, E. G., Jorge, M. L., Seno, E. M., Di Felippo, A., Rino, L. H., ... & Pardo, T. A. (2011, October). CSTnews-a discourse-annotated corpus for single and multi-document summarization of news texts in Brazilian Portuguese. In Proceedings of the 3rd RST Brazilian Meeting (pp. 88-105).
- Chengcheng, Li. (2010). Automatic text summarization based on Rhetorical Structure Theory. Proceedings of International Conference on Computer Application and System Modeling (ICCSM 2010). (pp. V13-595-598). Taiyuan, China
- Cohan, A., & Goharian, N. (2017). Scientific article summarization using citation-context and article's discourse structure. arXiv preprint arXiv:1704.06619.
- Iruskieta, M., Aranzabe, M. J., de Ilarraza, A., Gonzalez, I., Lersundi, M., & de Lacalle, O. L. (2013). The RST Basque TreeBank: an online search interface to check rhetorical relations. In 4th workshop RST and discourse studies (pp. 40-49).
- Mann, W. C. Thompson, S. A. 1987. Rhetorical structure theory: A theory of text

- organization (pp. 87-190). University of Southern California, Information Sciences Institute.
- Molina, A. (2013). Compresión automática de frases: un estudio hacia la generación de resúmenes en español. *Inteligencia Artificial*, 16(51), 41-62.
- Saggion, H., Torres-Moreno, J. M., Cunha, I. D., & SanJuan, E. (2010, August). Multilingual summarization evaluation without human models. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters* (pp. 1059-1067). Association for Computational Linguistics.
- Sarda, A., & Kulkarni, A. (2015). Text summarization using neural network and rhetorical structure theory. *International Journal of Advanced Research in Computer and Communication Engineering*, 4(6), 49-52.
- Shafiq Joty, Giuseppe Carenini, and Raymond Ng. 2015. CODRA: A Novel Discriminative Framework for Rhetorical Analysis. *Computational Linguistics*, Volume 41:3, MIT Press.
- Wang, X., Yoshida, Y., Hirao, T., Sudoh, K., & Nagata, M. (2015). Summarization based on task oriented discourse parsing. *IEEE Transactions on Audio, Speech, and Language Processing*, 23(8), 1358-1367.
- Zipitria, I. Arruarte, A. Elorriaga, J. 2013. Discourse measures for Basque summary grading. *Interactive Learning Environments*, 21(6), 528-547.
- Zipitria, I., Larrañaga, P., Armañanzas, R., Arruarte, A., & Elorriaga, J. A. (2008). What is behind a summary-evaluation decision?. *Behavior Research Methods*, 40(2), 597-612