Procesamiento del Lenguaje Natural, Revista nº 61, septiembre de 2018, pp. 101-108 recibido 01-03-2018 revisado 30-04-2018 aceptado 03-05-2018

Legibilidad del texto, métricas de complejidad y la importancia de las palabras

Text readability, complexity metrics and the importance of words

Rocío López-Anguita, Arturo Montejo-Ráez
Fernando J. Martínez-Santiago, **Manuel Carlos Díaz-Galiano**
Centro de Estudios Avanzados en TIC
Universidad de Jaén
{rlanguit, amontejo, dofer, mcdiaz}@ujaen.es

Resumen: El presente trabajo expone un estudio sobre la determinación de la edad recomendada de lectura sobre un conjunto de textos infantiles. Se ha evaluado el mismo con 12 medidas de complejidad propuestas por distintos autores. Usando estas medidas como características, hemos modelado los textos y aplicado una validación cruzada con varios clasificadores automáticos. Los resultados se han comparado con otras formas de representación de los textos, como vectores de palabras y vectores TF.IDF. Nuestras conclusiones indican que el rasgo más determinante para la determinación de la edad de lectura recomendada no radica tanto en factores como la complejidad sintáctica o léxica, sino en el uso de determinado vocabulario.
Palabras clave: Legibilidad, complejidad textual, modelado del lenguaje

Abstract: This article describes our study on the identification of the recommended age for readers in texts written for children. They have been evaluated over 12 complexity metrics proposed by different authors. By using these metrics as features, we have trained several automatic classifiers and cross-validated their performances to detect recommended reader level. The results have been compared with the classification performance obtained from other document models, like word embeddings and TF.IDF vectors. Our conclusiones are that the most relevant facet to identify the recommended reader age is not on lexical or syntactical complexities, but strongly related with the vocabulary involved.
Keywords: Readability, text complexity, language modelling

1 Introducción

Conocer cómo de adecuado es un texto para un persona es algo que no está resuelto. Mucho se ha investigado en la legibilidad de textos según el lector. En educación primaria podemos encontrar desde la primera fórmula de Spache (1953) hasta la visión más holística de Larson y Marsh (2014). En el caso de trastornos del lenguaje y la adecuación de los textos a personas con dificultades cognitivas también hay un extenso trabajo realizado, como veremos.

Determinar la legibilidad de un texto no es una tarea sencilla, pues cada lector presenta destrezas diferentes (Cain, Oakhill, y Bryant, 2004) o, incluso, limitaciones como la dislexia (Rello et al., 2013) o el autismo. Entendemos por legibilidad la facilidad de lectura de comprensión de un texto. Si no tenemos en consideración aspectos de forma, color, maquetado y tipografía (Ripoll, 2015), la legibilidad se determina por los rasgos lingüísticos, que suelen agruparse en aquellos relacionadas con la gramática (o lo que es lo mismo, la sintaxis) y aquellos relacionados con el léxico (es decir, el vocabulario) (Alliende González, 1994). Son múltiples los autores que han establecido métricas para la legibilidad en ambos ámbitos. En concreto, este trabajo visita 12 de las métricas más utilizadas para la legibilidad léxica y sintáctica y estudia su idoneidad como características para la determinación de la edad recomendada de lectura en textos de educación primaria.

El uso de algoritmos de aprendizaje automático para determinar la edad recomendada para un texto a partir de estas medidas de complejidad se ha comparado median-

ISSN 1135-5948. DOI 10.26342/2018-61-11

te validación cruzada con otras dos formas de modelar el lenguaje: mediante vectores de palabras (Mikolov et al., 2013) y mediante el modelo de espacio vectorial clásico (Salton, Wong, y Yang, 1975). Los resultados muestran que el último modelo, con una selección de características adecuada y determinados algoritmos de aprendizaje, resulta muy apropiado para esta tarea, destacando de esta forma la relevancia del vocabulario en la comprensión lectora. Los datos utilizados para la experimentación han sido extraídos de la web y consisten en diversas lecturas recomendadas para distintos niveles en Educación Primaria.

El artículo está organizado como sigue: en la Sección 2 se introducen las medidas de complejidad utilizadas para el modelado de los textos; la Sección 3 describe el conjunto de textos que hemos utilizado como base para los experimentos, que se detallan en la Sección 4. En la Sección 5 analizamos los resultados obtenidos y en la Sección 6 se cierra el trabajo con las conclusiones del mismo.

2 Medidas de Complejidad

En este apartado, vamos hacer un recorrido por las diferentes métricas de complejidad que se han propuesto por diversos autores. Si bien algunas de estas medidas proporcionan directamente la edad recomendada, como la medida de García López (2001), basada, a su vez, en la de Flesch (1948), otras ofrecen índices de más difícil interpretación, como la complejidad léxica de Anula (2008), el índice de complejidad de oraciones o la profundidad del árbol de dependencia de (Saggion et al., 2015), entre otras.

Complejidad del léxico. Esta medida de complejidad fue propuesta por Anula (2008) para medir la complejidad léxica de un texto, determinada por la frecuencia de uso y la densidad léxica. Se considera que a mayor densidad léxica (mayor número de palabras diferentes por textos) mayor dificultad para la comprensión.

Legibilidad de Spaulding. La legibilidad de Spaulding, comúnmente conocido como el índice SSR, fue propuesta por Spaulding (1956). Se centra en medir el vocabulario y la estructura de oraciones para predecir la dificultad relativa de legibilidad de un texto.

Complejidad de oraciones. El índice de complejidad de oraciones fue propuesto por Anula (2008). Mide el número de palabras por oración, obteniendo así el índice de longitud oracional, y el número de frases complejas que hay por oración, a partir de un índice de frases complejas.

Índice de legibilidad automatizado. Senter y Smith (1967) nos proponen uno de los índice más utilizados debido a su facilidad de cálculo. Mide la dificultad de un texto a partir del número medio de caracteres (letras y números) por palabra y del número medio de palabras por oración.

Altura del árbol de dependencia. Esta medida fue propuesta por Saggion et al. (2015). Es una métrica muy útil para capturar la complejidad sintáctica: las oraciones largas pueden ser sintácticamente complejas o contener una gran cantidad de modificadores (adjetivos, adverbios o frases adverbiales). Estos últimos no aumentan la complejidad sintáctica y no dan lugar a árboles muy profundos, mientras que los primeros tienen una fuerte tendencia a producir árboles profundos.

Marcas de Puntuación. Esta medida fue también propuesta por Saggion et al. (2015). En la complejidad de un texto, el número promedio de signos de puntuación se utiliza como uno de los indicadores de complejidad del mismo.

Lecturabilidad de Fernández-Huerta. Blanco Pérez y Gutiérrez Couto (2002) y Ramírez-Puerta et al. (2013) nos proponen esta medida de complejidad como una adaptación al español de la prueba de legibilidad de Flesch (Flesch (1948)). Parte de que en español las palabras en promedio tienen más sílabas y las oraciones también son más largas. Mide el promedio de sílabas por palabra y el promedio de palabras por oración que hay en el texto.

Legibilidad de Flesch-Szigrist (IFSZ). Los trabajos De Granada Barrio-Cantalejo et al. (2008) y Ramírez-Puerta et al. (2013) nos proponen el índice de legibilidad de Flesch-Szigristzt como una modificación de la fórmula de Flesch (Flesch, 1948) adaptada al

castellano. El índice de legibilidad IFSZ es considerado de referencia para la lengua española. Mide el número de sílabas por palabra y el número de palabras por oración que hay en el texto.

Comprensibilidad de Gutiérrez. Fue creada para el castellano (Rodríguez, 1980) y consiste en una fórmula matemática, generada por métodos de regresión múltiple, que incluye ciertas características lingüísticas del material cuya dificultad se pretende evaluar. Se centra en medir el promedio de letras por palabra y el promedio de palabras por oración.

Legibilidad μ. La Legibilidad μ propone una fórmula para calcular la facilidad lectora de un texto. Proporciona un índice comprendido entre 0 y 100 y fue desarrollada por Muñoz (2006). Esta medida se centra en medir el número de palabras, la media del número de letras por palabra y su varianza.

Edad mínima de comprensibilidad. En el trabajo de García López (2001) podemos encontrar otra fórmula para medir la edad necesaria para entender un texto. Es, de nuevo, una adaptación al castellano de la fórmula original de Flesch (Flesch (1948)) para el inglés. Mide el promedio de sílabas por palabra y el promedio de palabras por oración para obtener la edad mínima necesaria para entender un texto.

SOL. Contreras et al. (1999) nos propone la métrica SOL como una adaptación al español de la fórmula SMOG propuesta por Mc Laughlin (1969). Mide la legibilidad de un texto mediante el nivel de grado, que viene a ser el número de años de escolaridad necesarios para entender el texto.

3 Descripción del corpus

Nuestro objetivo es evaluar la idoneidad de un sistemas de clasificación automático que utilice estas métricas para determinar la edad recomendada de lectura para un texto.

El corpus que utilizamos en este trabajo está compuesto por 300 textos de lecturas en español, orientados a alumnos de Educación Primaria. Dichos textos están clasificados por el curso a los que van dirigidos. En consecuencia, tenemos 6 grupos (1º, 2º , 3º , 4º , 5º y 6º de primaria) y cada grupo consta de 50 textos.

Los textos han sido obtenidos, principalmente, de un trabajo realizado por un grupo de profesores de distintos centros de Sevilla, coordinados por María José Moya Bellido y Antonio Ruiz y Martín (Inspectores de Educación del Servicio de Sevilla), en el curso escolar 2011/2012[1].

Además hemos completado nuestro corpus con otras lecturas y con los primeros capítulos de algunos cuentos, obtenidos de distintos sitios web, donde encontramos recursos educativos accesibles y gratuitos, como por ejemplo "Orientación Andújar"[2] o "Rincón de lecturas"[3].

En la Tabla 1, podemos observar las características más relevantes del corpus compilado.

4 Experimentos

Como se ha indicado anteriormente, hemos evaluado la aportación de estas métricas de complejidad en una tarea de determinación de la edad de lectura recomendada mediante aprendizaje automático. Además, hemos estudiado la dependencia de las distintas métricas con el nivel de Educación Primaria mediante un análisis de los valores Chi-cuadrado que obtenemos al estudiar los pares de distribuciones <*valores de la característica, nivel*>

Para evaluar los sistemas hemos configurado distintos conjuntos de datos con diferente granularidad en los niveles de Educación Primaria considerados, tal y como pasamos a detallar a continuación.

4.1 Conjuntos de datos

Para llevar a cabo los experimentos hemos definido los dos siguientes conjuntos de datos:

- `123456`: A cada curso de primaria le asignamos un nivel (de 1º a 6º de educación primaria).
- `110022`: Tomamos dos grupos, donde los cursos 1º y 2º pertenecen al nivel 1, 5º y 6º pertenecen al nivel 2 y los cursos 3º y 4º no son considerados.

La razón de esta configuración de las clases (es decir, los niveles), es para propiciar

[1] `http://sosprofes.es`

[2] `https://www.orientacionandujar.es/`

[3] `http://rincondelecturas.com/`

Curso	1º	2º	3º	4º	5º	6º	Total
Nº Textos	50	50	50	50	50	50	300
Tamaño Vocabulario: palabras distintas en el texto	1966	2648	2799	3759	4979	5333	21484
Longitud Media en palabras	150.54	204.6	222.22	370.26	489.16	514.26	1978.04
Longitud Mín en palabras	36	41	31	61	38	73	280
Longitud Máx en palabras	454	692	1508	1640	1797	1572	7663
Media de palabras raras	49.58	64.88	69.82	114.98	147.3	161.04	607.6
Nº palabras por oración	12.68	16.43	16.14	22.52	20.27	22.05	110.09
Nº sílabas por palabra	1.88	1.98	1.95	1.96	1.98	1.995	11.75
Media de oraciones complejas	9.20	12.34	12.65	19.11	17.14	17.52	87.96
Nº total palabras	7533	10202	11081	18478	24429	27038	98761

Tabla 1: Caracterización del Corpus

mayor separabilidad y dilucidar de forma clara qué características son las más significativas cuando de determinar el nivel de lectura se trata. De esta forma, la segunda configuración debería ser más separable a nivel de clases.

4.2 Evaluación mediante aprendizaje supervisado

La hipótesis es que, dado que las métricas enumeradas anteriormente capturan distintos aspectos de la complejidad del texto, deberían ser válidas como características en un modelo que represente cada documento en este proceso de clasificación del nivel. Para ello vamos a aplicar validación cruzada sobre los distintos conjuntos (`123456` y `110022`) y algoritmos (LinearSVC, Multilayer Perceptron, Random Forest y Naive Bayes). Adicionalmente, vamos a comparar estos resultados con los obtenidos con otros modelos del texto pero no asociados a la complejidad, como los vectores de palabras (*word embeddings* mediante Word2Vec) o el modelo de espacio vectorial clásico con TF.IDF.

Analizaremos, de esta forma, la pertinencia o no de estas medidas de complejidad de manera indirecta a partir de la evaluación de los clasificadores entrenados con ellas. Obtenemos, por validación cruzada y macropromediado, las medidas de exactitud (*accuracy*), F1, cobertura (*recall*) y precisión (*precision*) de cada algoritmo para cada combinación de clases.

Una vez obtenidos dichos valores para cada clasificador, realizamos una prueba de Chi-cuadrado a las muestras y volvemos aplicar la validación cruzada para ver cuáles son las medidas de complejidad más relevantes.

Por último, también probamos con las representaciones de Word2Vec y TF.IDF.

Para todo esto hemos usado la herramienta *Freeling* (Padró y Stanilovsky, 2012) para procesar los textos y las bibliotecas de Python *SciKit-Learn* (Pedregosa et al., 2011) para aprendizaje automático y *Gensim* (Rehurek y Sojka, 2011) para trabajar con vectores de palabras.

4.3 Experimentos

Hemos representado los textos según cuatro modelos diferentes:

1. Los textos son representados por las medidas de complejidad indicadas anteriormente, sobre las que hemos aplicado una normalización L2 de los valores.
2. Word2Vec siguiendo el método de aglomeración propuesto por Montejo-Ráez y Díaz-Galiano (2016).
3. TF.IDF.
4. Medidas de Complejidad + Word2Vec + TF.IDF.

5 Análisis de resultados

Como era de esperar, independientemente de los algoritmos y las representaciones utilizadas para los textos, obtenemos el peor resultado cuando consideramos el conjunto de datos `123456`. Esto se debe a que no hay tanta diferencia entre ellos, puesto que son cursos consecutivos y van aumentando su dificultad gradualmente. El mejor resultado lo obtenemos con el conjunto 110022, puesto que la diferencia entre los cursos considerados de primer ciclo de primaria y tercer ciclo de primaria es más significativa. En cualquier caso,

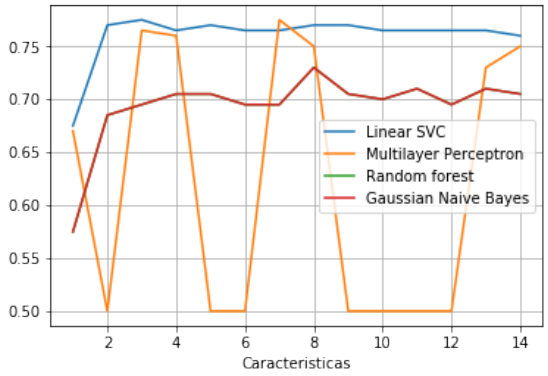

Figura 1: Resultados filtrando por Chi-cuadrado las medidas de complejidad

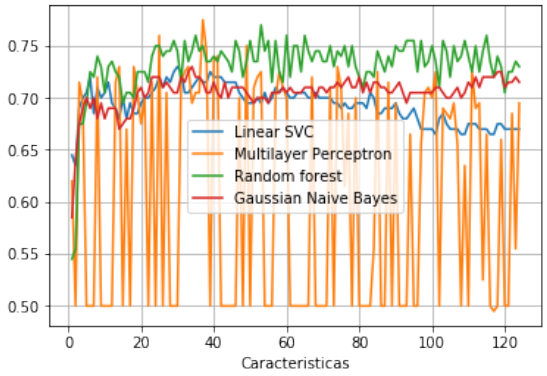

Figura 2: Resultados filtrando por Chi-cuadrado en word2Vec

es importante remarcar que para el objeto de nuestra investigación nos interesa sobre todo analizar las diferencias entre modelos de representación de los textos bajo las mismas condiciones de dificultad de clasificación (clases) y algoritmos usados.

5.1 Resultados con medidas de complejidad

En la Tabla 2 presentamos la estadística del aprendizaje automático supervisado con las medidas de complejidad para las distintas configuraciones de clases.

En la Tabla 2, podemos observar que el máximo se alcanza con un *accuracy* de 0.76 con Multilayer Perceptron. En cualquier caso, los resultados arrojados por los distintos clasificadores son comparables. Las diferencias entre ellos pueden responder más a cuestiones de ajuste de los parámetros del clasificador que a su conveniencia o no en esta tarea. El uso de distintos algoritmos no es sino para comprobar si nuestra hipótesis se mantiene independientemente del tipo de algoritmos de clasificación utilizado.

Cuando aplicamos la Chi-cuadrado y después la validación cruzada obtenemos la Figura 1. Podemos observar que el máximo se alcanza con una *accuracy* de 0.775 con tres características con el clasificador Linear SVC.

5.2 Resultados con Word2Vec

Los resultados los podemos observar en la Tabla 3 para las dos tareas de clasificación y los distintos algoritmos utilizados.

En la Tabla 3, podemos observar que el máximo se alcanza con una *accuracy* de 0.73 con el clasificador Random Forest. Cuando le aplicamos la Chi-cuadrado antes de la validación cruzada para seleccionar características obtenemos la Figura 2. Siendo el máximo 0.775 y se alcanza con 37 características en el clasificador Multilayer Perceptron.

Como es de esperar en el uso de una red neuronal como es el perceptrón multicapa, cuando el número de características es grande el clasificador puede requerir de una mayor cantidad de ejemplos para llegar a estimar un buen modelo. Esta sería la razón para su buen comportamiento al filtrar características por Chi-cuadrado (a 37) y al usar medidas de complejidad (12) frente a vectores de palabras o vectores TF.IDF que cuentan con centenares de características.

5.3 Resultados con TF.IDF

Podemos observar los resultados obtenidos en la Tabla 4, de nuevo, para las dos tareas de clasificación y los distintos algoritmos utilizados.

En la tabla 4, podemos observar que el máximo se alcanza con una *accuracy* de 0.76 con el clasificador SVC Linear. Cuando le aplicamos la Chi-cuadrado al conjunto 110022 con los mismos clasificadores sobre la representación de TF.IDF, obtenemos la Figura 3. Siendo el máximo 0.92 y se alcanza con 199 características en el clasificador Gaussian Naive Bayes.

5.4 Resultados con medidas de complejidad + Word2Vec + TF.IDF

Los resultados basados en las medidas de complejidad junto a Word2Vec y TF.IDF se presentan en la Tabla 5.

clases	algoritmo	accuracy	F1	precision	recall
123456	Linear SVC	0.26	0.19	0.07	0.18
123456	Multilayer Perceptron	0.27	0.23	0.07	0.20
123456	Random Forest	0.29	0.29	0.34	0.33
123456	Gaussian Naive Bayes	0.26	0.22	0.22	0.25
110022	Linear SVC	**0.76**	0.74	0.59	0.64
110022	Multilayer Perceptron	**0.76**	**0.75**	0.25	0.5
110022	Random Forest	0.72	0.69	0.79	0.74
110022	Gaussian Naive Bayes	0.7	0.68	0.76	0.72

Tabla 2: Resultados con Medidas de Complejidad

clases	algoritmo	accuracy	F1	precision	recall
123456	Linear SVC	0.30	0.28	0.32	0.30
123456	Multilayer Perceptron	0.17	0.05	0.03	0.17
123456	Random Forest	0.26	0.25	0.32	0.26
123456	Gaussian Naive Bayes	0.30	0.27	0.30	0.30
110022	Linear SVC	0.72	**0.70**	0.74	0.72
110022	Multilayer Perceptron	0.50	0.33	0.25	0.50
110022	Random Forest	**0.73**	0.69	0.76	0.77
110022	Gaussian Naive Bayes	0.71	0.69	0.74	0.71

Tabla 3: Resultados con Word2Vec

clases	algoritmo	accuracy	F1	precision	recall
123456	SVC Linear	0.27	0.25	0.27	0.27
123456	Multilayer Perceptron	0.18	0.08	0.08	0.18
123456	Random Forest	0.27	0.31	0.34	0.31
123456	Gaussian Naive Bayes	0.19	0.17	0.21	0.19
110022	SVC Linear	**0.76**	**0.74**	0.79	0.76
110022	Multilayer Perceptron	0.50	0.33	0.25	0.50
110022	Random Forest	0.71	**0.74**	0.81	0.74
110022	Gaussian Naive Bayes	0.61	0.57	0.64	0.61

Tabla 4: Resultados con TF.IDF

clases	algoritmo	accuracy	F1	precision	recall
123456	Linear SVC	0.31	0.27	0.29	0.31
123456	Multilayer Perceptron	0.21	0.14	0.12	0.21
123456	Random Forest	0.28	0.28	0.32	0.28
123456	Gaussian Naive Bayes	0.16	0.12	0.15	0.16
110022	Linear SVC	**0.77**	**0.74**	0.78	0.77
110022	Multilayer Perceptron	0.68	0.64	0.76	0.68
110022	Random Forest	0.75	0.73	0.80	0.75
110022	Gaussian Naive Bayes	0.55	0.44	0.56	0.55

Tabla 5: Resultados con Medidas de Complejidad + Word2Vec + TF.IDF

En la Tabla 5, podemos observar que el máximo se alcanza con una *accuracy* de 0.77 con el clasificador SVC Linear. Cuando le aplicamos la Chi-cuadrado al conjunto 110022 con los mismos clasificadores sobre las medidas de complejidad y las representaciones de Word2Vec y de TF.IDF, obtenemos la siguiente gráfica que se muestra en la Figura 4. Siendo el máximo 0.90 y se alcanza con 209 características en el clasificador Multilayer Perceptron.

En resumen, obtenemos resultados muy

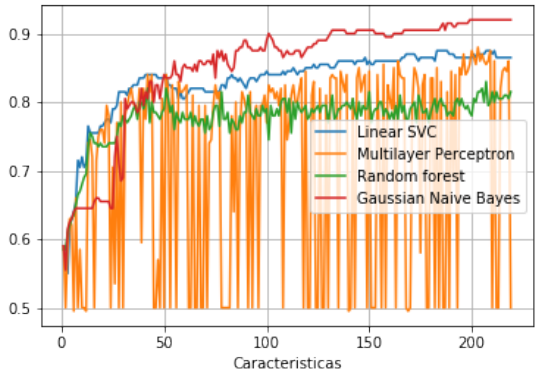

Figura 3: Resultados filtrando por Chi-cuadrado en TF.IDF

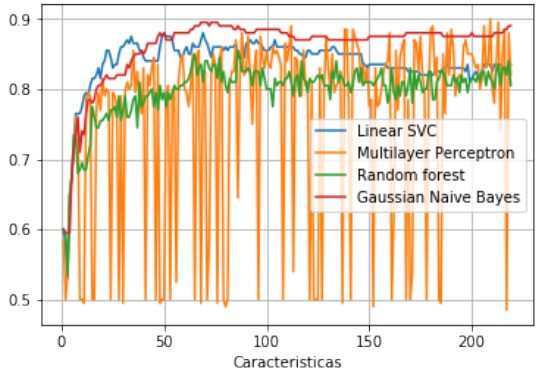

Figura 4: Resultados filtrando por Chi-cuadrado las medidas de complejidad + Word2Vec + TF.IDF

parecidos para las medidas de complejidad y Word2vec. Sin embargo, se mejoran los resultados si consideramos la unión de las medidas de complejidad con Word2Vec y TF.IDF. En cualquier caso, el mejor resultado es el obtenido TF.IDF con filtrado de términos mediante Chi-cuadrado. Los clasificadores Gaussian Naive Bayes y Multilayer Perceptron son sensibles al número de características, y una reducción de las mismas llegar a arrojar resultados destacados.

6 Conclusiones

Hemos visto cómo los indicadores clásicos de complejidad textual sobre el español pueden ayudarnos a resolver una tarea de clasificación de textos en distintos niveles o cursos del ciclo formativo de primaria. En cualquier caso, los modelos de representación basados en contenido resultan mejores para esta tarea, si bien las medidas de complejidad pueden complementarles. Destacan los resultados utilizando un modelo clásico de representación de los textos como TF.IDF, frente a representaciones muy utilizadas en la actualidad como los vectores de palabras. Parece no haber duda de la relevancia del vocabulario a la hora de determinar qué es más recomendable que lean.

Nuestro estudio, de una forma empírica, corrobora el de Stahl (2003), al reflejar la fuerte influencia de la riqueza de vocabulario del lector en la comprensión lectora, más allá de los símbolos y la gramática. No obstante, consideramos que las medidas de complejidad pueden ser una forma conveniente para modelar el lenguaje natural en determinadas aplicaciones, como la detección de autoría, la selección de textos para personas con dificultades asociadas a trastornos del lenguaje (autismo, parálisis cerebral...), o la detección temprana de deterioros cognitivos, como el Alzheimer. Otra línea futura de trabajo es el analizar no sólo el lenguaje formal, sino también el informal en entornos como los medios sociales en Internet.

Agradecimientos

Este trabajo ha sido parcialmente financiado por el Gobierno de España a través del proyecto REDES (TIN2015-65136-C2-1-R).

Bibliografía

Alliende González, F. 1994. La legibilidad de los textos. *Santiago de Chile: Andrés Bello*, 24.

Anula, A. 2008. Lecturas adaptadas a la enseñanza del español como l2: variables lingüísticas para la determinación del nivel de legibilidad. *La evaluación en el aprendizaje y la enseñanza del español como LE L*, 2:162–170.

Blanco Pérez, A. y U. Gutiérrez Couto. 2002. Legibilidad de las páginas web sobre salud dirigidas a pacientes y lectores de la población general. *Revista española de salud pública*, 76(4):321–331.

Cain, K., J. Oakhill, y P. Bryant. 2004. Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of educational psychology*, 96(1):31.

Contreras, A., R. Garcia-Alonso, M. Echenique, y F. Daye-Contreras. 1999. The sol

formulas for converting smog readability scores between health education materials written in spanish, english, and french. *Journal of health communication*, 4(1):21–29.

De Granada Barrio-Cantalejo, D. S., P. Simón-Lorda, M. Melguizo, I. Escalona, M. Marijuán, P. Hernándo, y others. 2008. Validación de la escala inflesz para evaluar la legibilidad de los textos dirigidos a pacientes.

Flesch, R. 1948. A new readability yardstick. *Journal of applied psychology*, 32(3):221.

García López, J. 2001. Legibilidad de los folletos informativos. *Pharmaceutical Care España*, 3(1):49–56.

Larson, J. y J. Marsh. 2014. *Making literacy real: Theories and practices for learning and teaching*. Sage.

Mc Laughlin, G. H. 1969. Smog grading-a new readability formula. *Journal of reading*, 12(8):639–646.

Mikolov, T., I. Sutskever, K. Chen, G. S. Corrado, y J. Dean. 2013. Distributed representations of words and phrases and their compositionality. En *Advances in neural information processing systems*, páginas 3111–3119.

Montejo-Ráez, A. y M. C. Díaz-Galiano. 2016. Participación de sinai en tass 2016. En *TASS@ SEPLN*, páginas 41–45.

Muñoz, M. 2006. Legibilidad y variabilidad de los textos. *Boletín de Investigación Educacional, Pontificia Universidad Católica de Chile, 21*, 2:13–26.

Padró, L. y E. Stanilovsky. 2012. Freeling 3.0: Towards wider multilinguality. En *LREC2012*.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, y others. 2011. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830.

Ramírez-Puerta, M., R. Fernández-Fernández, J. Frías-Pareja, M. Yuste-Ossorio, S. Narbona-Galdó, y L. Peñas-Maldonado. 2013. Análisis de legibilidad de consentimientos informados en cuidados intensivos. *Medicina Intensiva*, 37(8):503–509.

Rehurek, R. y P. Sojka. 2011. Gensim–python framework for vector space modelling. *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic*, 3(2).

Rello, L., R. Baeza-Yates, S. Bott, y H. Saggion. 2013. Simplify or help?: text simplification strategies for people with dyslexia. En *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility*, página 15. ACM.

Ripoll, J. C. 2015. Font legibility in first year primary students/legibilidad de distintos tipos de letra en alumnos de primero de primaria. *Infancia y Aprendizaje*, 38(3):600–616.

Rodríguez, T. 1980. Determinación de la comprensibilidad de materiales de lectura por medio de variables lingüísticas. *Lectura y vida*, 1(1):29–32.

Saggion, H., S. Štajner, S. Bott, S. Mille, L. Rello, y B. Drndarevic. 2015. Making it simplext: Implementation and evaluation of a text simplification system for spanish. *ACM Transactions on Accessible Computing (TACCESS)*, 6(4):14.

Salton, G., A. Wong, y C.-S. Yang. 1975. A vector space model for automatic indexing. *Communications of the ACM*, 18(11):613–620.

Senter, R. y E. A. Smith. 1967. Automated readability index. Informe técnico, CINCINNATI UNIV OH.

Spache, G. 1953. A new readability formula for primary-grade reading materials. *The Elementary School Journal*, 53(7):410–413.

Spaulding, S. 1956. A spanish readability formula. *The Modern Language Journal*, 40(8):433–441.

Stahl, S. A. 2003. Vocabulary and readability: How knowing word meanings affects comprehension. *Topics in Language Disorders*, 23(3):241–247.