

Máster Universitario en Inteligencia
Artificial Avanzada: Fundamentos,
métodos y aplicaciones. UNED.
Curso 2018-2019

Pedagogía poética en sistemas recomendadores afectivos para entornos de e-learning

Trabajo de Fin de Máster

Alberto Barbado González

Directores:

Ángeles Manjarrés Riesco

Víctor Fresno Fernández

Índice

| | |
|---|------------|
| 1. Introducción | 12 |
| 1.1. Preámbulo..... | 12 |
| 1.2. Motivación | 13 |
| Pedagogía poética | 16 |
| Poesía y psicología..... | 18 |
| 1.3. Objetivos de investigación | 19 |
| 1.4. Metodología | 20 |
| 1.4.1. Experimentos realizados | 20 |
| 1.4.2. Metodología CRIPS-DM | 23 |
| 1.4.3. Otras metodologías: TORMES | 24 |
| 1.5. Estructura del documento..... | 25 |
| 2. Descripción del problema..... | 28 |
| 2.1. Estado del Arte | 28 |
| 2.1.1. Sistemas Recomendadores Sensibles al Contexto Afectivo | 28 |
| 2.1.2. Ontologías para entornos de E-Learning..... | 38 |
| 2.1.3. Tratamiento automático del contenido afectivo..... | 42 |
| 2.1.4. Etiquetado Automático de Repositorios..... | 54 |
| 2.1.5. Word Embeddings..... | 56 |
| 2.1.6. Composición de textos | 92 |
| 2.4. Planteamiento de la solución..... | 97 |
| 2.4.1 Estudio tecnológico preliminar | 97 |
| 3. Descripción de la experimentación..... | 111 |
| 3.1. Detalle de los objetivos | 111 |
| 3.2. Detalle de los experimentos | 114 |
| POSTDATA..... | 114 |
| AULA | 117 |
| 3.3. Detalle de la metodología aplicada: CRISP-DM | 119 |
| 4. Descripción de la aplicación..... | 121 |
| 4.1. Introducción | 121 |
| 4.2. Descripción del código..... | 121 |
| Procesado de sonetos | 122 |
| Aplicación para AULA | 124 |
| Aplicación para POSTDATA | 127 |
| Excel para anotaciones de POSTDATA | 127 |
| 4.3. Diagramas de procesos..... | 128 |
| 5. Planificación | 133 |

| | |
|--|------------|
| 6. Resultados..... | 135 |
| 6.1. Estudio del corpus de sonetos | 135 |
| 6.2 Experimentos con POSTDATA | 140 |
| Análisis exploratorio de datos sobre las anotaciones | 141 |
| Experimento 1 | 153 |
| Experimento 3 | 156 |
| Experimento 2 y 4..... | 159 |
| 6.3. Resultados en AULA | 167 |
| 7. Conclusiones y Líneas Futuras | 179 |
| 7.1. Conclusiones | 179 |
| 7.2. Líneas futuras | 181 |
| 7.2.1. Minería de textos y NLP en poesía | 181 |
| 7.2.2. RecSys afectivos | 183 |
| 7.2.3. Trusted AI | 185 |
| 8. Referencias..... | 188 |
| 9. Anexos | 194 |
| 9.1. Campos de las ontologías..... | 194 |
| 9.2. API endpoints | 197 |
| 9.3. JSON con la información para cada soneto | 198 |
| 9.4. JSON con la información de uso de AULA | 199 |
| 9.5. Resultados de Aula 2019..... | 200 |
| Datos de uso de los usuarios de la aplicación | 200 |
| Número de veces que se han elegido las distintas palabras | 201 |
| 9.6. Diccionario de siglas | 202 |
| 9.7. Rasgos afectivos utilizados | 203 |
| 9.8. Resultados completos del Experimento 2 y 4 | 204 |
| Artículo: DISCO PAL: Diachronic Spanish Sonnet Corpus with Psychological and Affective Labels | 206 |

Índice de Ilustraciones

| | |
|---|----|
| Ilustración 1 Fases de CRISP-DM..... | 23 |
| Ilustración 2 Clasificación de los RecSys según distintas categorías, en las que se engloba la clasificación según sus componentes y se listan los parámetros diferenciadores (Ricci, Rokach, & Shapira, 2011) | 29 |
| Ilustración 3 Ejemplo de matriz de usuarios e ítems (destinos de viaje) donde las recomendaciones se harán en función de elementos contextuales, como es el caso del "tipo de compañero de viaje". (Aggarwal, 2016) | 33 |
| Ilustración 4 Esquema de las distintas técnicas posibles de modelado contextual | 34 |
| Ilustración 5 Esquema de inferencia de estados de un latent variable context model..... | 35 |
| Ilustración 6 Arquitectura de un RecSys sensible al contexto afectivo para el ámbito de e-learning. En verde, los complementos propuestos a (Nurjanah, 2016). | 37 |
| Ilustración 7 Esquema de relación supervisada | 43 |
| Ilustración 8 Esquema de relación no supervisada. | 44 |
| Ilustración 9 Ejemplo de construcción de un vector numérico usando BoW. Con esta técnica se representa un texto con un vector de unos y ceros, técnica conocida como One-Hot Encoding. | 56 |
| Ilustración 10 Ejemplo de construcción de BoW que tiene en cuenta la frecuencia de las palabras | 57 |
| Ilustración 11 Ejemplo de representación con vectores en 3 dimensiones. | 58 |
| Ilustración 12 Ejemplo de obtención de probabilidades usando una ventana de tamaño dos | 59 |
| Ilustración 13 Ejemplo de red neuronal para el modelo Skip-Gram..... | 61 |
| Ilustración 14 Paso de matriz de pesos a tabla lookup..... | 62 |
| Ilustración 15 Ejemplo sencillo de vector de dimensión 5 y su correspondiente table lookup..... | 62 |
| Ilustración 16 Visualización de embeddings con t-SNE..... | 63 |
| Ilustración 17 Arquitectura alternativa para dos vectores de entrada | 64 |
| Ilustración 18 Comparativa de CBOW vs Skip-Gram; uno predice el contexto alrededor de una focus word y otro predice la focus word de un determinado contexto. | 65 |
| Ilustración 19 Así, con CBOW se parte de un contexto X y con él se intenta predecir una palabra concreta Y utilizando una red neuronal | 65 |
| Ilustración 20 Arquitectura de CBOW..... | 66 |
| Ilustración 21 Matriz de coocurrencias para distintas palabras junto con los ratios de co-ocurrencias entre dos casos. | 68 |
| Ilustración 22 Representación de la función de peso. | 69 |
| Ilustración 23 Arquitectura de ELMo | 72 |
| Ilustración 24 Representación matricial de los caracteres usando 4 dimensiones para cada uno | 73 |
| Ilustración 25 Matriz kernel o de filtrado que se utilizará en este ejemplo | 73 |
| Ilustración 26 A medida que se recorre la matriz de entrada con un kernel concreto se van extrayendo valores en función del producto de la matriz de filtrado por esa subsección indicada | 74 |
| Ilustración 27 Se aplica un pooling sobre el vector extraído (en este caso max pooling)..... | 74 |
| Ilustración 28 Aplicación iterativa de otro filtro sobre los caracteres. | 75 |
| Ilustración 29 Vector final resultante tras aplicar varias matrices kernel y varias iteraciones | 75 |
| Ilustración 30 Representación vectorial final para la palabra | 76 |
| Ilustración 31 Bias para ajustar el vector resultante..... | 76 |
| Ilustración 32 Estructura interna de una LSTM..... | 77 |
| Ilustración 33 Conexión entre varias LSTM..... | 78 |
| Ilustración 34 RNN para modelado de lenguaje. | 78 |
| Ilustración 35 Las RNN en la etapa forward intenta predecir la palabra siguiente dada una secuencia previa de palabras mientras que la backward intenta predecir la palabra previa a una secuencia dada | 79 |
| Ilustración 36 Ejemplo para el caso forward. | 79 |
| Ilustración 37 Expansión de los componentes internos de la RNN | 80 |

| | |
|--|-----|
| Ilustración 38 Obtención de embeddings y suma ponderada de los mismos. Se suman para ello los vectores de forward y backwards combinados de cada capa junto con la entrada del embedding resultante del CharCNN..... | 80 |
| Ilustración 39 Comparativa de las soluciones de BERT base frente a otras técnicas del estado del arte para distintas tareas de NLP..... | 82 |
| Ilustración 40 Ejemplo de enmascaramiento del input en BERT. | 83 |
| Ilustración 41 Transformer en BERT aplicado al análisis de una frase..... | 83 |
| Ilustración 42 Existen dos posibles modelos, de tamaño sensiblemente distinto. | 84 |
| Ilustración 43 Además de optimizar la predicción de las palabras enmascaradas BERT intenta optimizar la predicción de la frase siguiente a un texto determinado. | 84 |
| Ilustración 44 Aplicación de un Transformer para una tarea de traducción automática de secuencias de textos..... | 85 |
| Ilustración 45 En la versión base de BERT se usan 12 encoders en serie con 768 parámetros de entrada/salida en cada uno. En la versión large son 24 encoders con vectores de entrada/salida de 1024 componentes. | 86 |
| Ilustración 46 Distintas posibilidades de combinación de los embeddings de los encoders para la representación final junto con la efectividad de estas. | 87 |
| Ilustración 47 Estructura interna de un Encoder..... | 88 |
| Ilustración 48 Ejemplo de cómo con la capa Self-Attention se encuentra que el pronombre 'it' hace referencia en este caso a la palabra 'animal'..... | 89 |
| Ilustración 49 Estructura de metadatos de BERT..... | 89 |
| Ilustración 50 Esquema de BERT..... | 90 |
| Ilustración 51 Comparativas de métricas entre bert large y bert base en los distintos datasets de SQuAD..... | 91 |
| Ilustración 52 Distintas técnicas de agregación para la definición de "aeropuerto" en inglés. “a place where planes take off and land” (Amigó Cabrera & Fresno Fernández, 2018)..... | 93 |
| Ilustración 53 Para el ejemplo anterior se muestra como para el caso de LR-DEP se va descendiendo por el gráfico de dependencia agregando las palabras con respecto al vector resultante de las agregaciones previas..... | 94 |
| Ilustración 54 Estructura sintáctica de la frase de ejemplo obtenida con el parser de Stanford. | 95 |
| Ilustración 55 Comparativas de las distintas funciones de composición para las distintas técnicas de agregación donde se ven los resultados de cada una de ellas (Amigó Cabrera & Fresno Fernández, 2018)..... | 96 |
| Ilustración 56 Mejores resultados para el caso de icm 1.2 sin stopwords (Amigó Cabrera & Fresno Fernández, 2018)..... | 96 |
| Ilustración 57 Ejemplo de definición del XML-TEI con su ampliación con RDF y su correspondiente visualización en modo RDF (Ruiz Fabo et al., 2017)..... | 99 |
| Ilustración 58 Ejemplo de metadatos para el XML de un soneto de DISCO (Ruiz Fabo et al., 2017)..... | 100 |
| Ilustración 59 Ejemplo de algunos atributos anotados siguiendo la especificación de TEI sobre el XML de un soneto..... | 101 |
| Ilustración 60 LOM extendido para incluir información sobre accesibilidad y función cognitiva (Bustos et al., 2016). | 103 |
| Ilustración 61 Modelo de usuario de MIRLO (Raffenne, 2013)..... | 106 |
| Ilustración 62 Modelo de Actividad (Raffenne, 2013)..... | 108 |
| Ilustración 63 Modelo de Competencias (Raffenne, 2013)..... | 109 |
| Ilustración 64 Esquema de alto nivel de la API desarrollada para la experiencia de AULA. Recibe las consultas del usuario y que devuelva el soneto más similar semánticamente..... | 114 |
| Ilustración 65 La primera pantalla consiste en una presentación del sistema, una descripción de en qué va a consistir la interacción con él, y un breve declaración de responsabilidad sobre los datos que se van a guardar de la interacción del usuario..... | 124 |

| | |
|---|-----|
| <i>Ilustración 66 Se muestran dos columnas con categorías de nivel de excitación/valencia para que el usuario eligiese una de cada columna</i> | 125 |
| <i>Ilustración 67 Después se presenta el IR. El usuario introduce una palabra de consulta y el sistema muestra dos sonetos recuperados usando la función coseno como métrica y F joint y F sum como funciones de composición. El orden en que se muestran es aleatorio para no introducir sesgos. El usuario elegiría uno de los dos sonetos y pasaría a la pantalla final.</i> | 125 |
| <i>Ilustración 68 Haciendo scroll sobre el soneto se pueden ver además los datos del título, autor y siglo, tal y como aparecen en los XML TEI de DISCO o de forma análoga en el json con los sonetos adicionales del siglo XX</i> | 126 |
| <i>Ilustración 69 Finalmente el usuario puntúa su experiencia con el sistema</i> | 126 |
| <i>Ilustración 70 Esquema de la lógica para el procesado de los textos.</i> | 129 |
| <i>Ilustración 71 Flujograma del proceso esquematizado en la Ilustración 68</i> | 130 |
| <i>Ilustración 72 Aplicación de escritorio utilizada para las pruebas de AULA</i> | 130 |
| <i>Ilustración 73 Aplicación de escritorio para las pruebas con POSTDATA.</i> | 131 |
| <i>Ilustración 74 Diseño de la API que de entrada tiene un endpoint para que el usuario envíe un mensaje y reciba el soneto más similar semánticamente usando BERT y F Joint RL SEQ con métrica de comparación de coseno</i> | 131 |
| <i>Ilustración 75 Flujograma genérico para cualquiera de las aplicaciones desarrolladas</i> | 132 |
| <i>Ilustración 76 Excel que usaron los anotadores de POSTDATA para llevar a cabo las anotaciones para cada soneto de los rasgos de los que se dispone de información a nivel de palabras.</i> | 132 |
| <i>Ilustración 77 Diagrama de Gantt para el TFM</i> | 134 |
| <i>Ilustración 78 Box-plot de los distintos rasgos</i> | 139 |
| <i>Ilustración 79 A la derecha se muestra el top de palabras sin lematizar. A la izquierda el de las lematizadas</i> | 140 |
| <i>Ilustración 80 Histograma para las anotaciones de la variable arousal</i> | 149 |
| <i>Ilustración 81 Histograma para las anotaciones de la variable valence</i> | 149 |
| <i>Ilustración 82 Histograma para las anotaciones de la variable Ansiedad</i> | 150 |
| <i>Ilustración 83 Diferencias de anotaciones por soneto para la variable arousal y los anotadores 1 y 2</i> | 150 |
| <i>Ilustración 84 Diferencias de anotaciones por soneto para la variable arousal y los anotadores 1 y 3</i> | 150 |
| <i>Ilustración 85 Diferencias de anotaciones por soneto para la variable arousal y los anotadores 2 y 3</i> | 151 |
| <i>Ilustración 86 Diferencias de anotaciones por soneto para la variable valence y los anotadores 1 y 2</i> | 151 |
| <i>Ilustración 87 Diferencias de anotaciones por soneto para la variable valence y los anotadores 1 y 2</i> | 151 |
| <i>Ilustración 88 Diferencias de anotaciones por soneto para la variable valence y los anotadores 1 y 2</i> | 152 |
| <i>Ilustración 89 Ejemplo de datos recogidos por usuario</i> | 169 |
| <i>Ilustración 90 Selección según la función de composición</i> | 170 |
| <i>Ilustración 91 Histograma con las selecciones de nivel de excitación; se puede observar cómo los estados intermedios son los más habituales</i> | 170 |
| <i>Ilustración 92 Se puede apreciar en las selecciones de valencia cómo los estados claramente más habituales son los de alegría y entusiasmo, habiendo datos muy minoritarios en comparación para los otros casos</i> | 171 |
| <i>Ilustración 93 Para la puntuación se aprecia en el histograma cómo la mayoría de los usuarios han puntuado a su experiencia con el uso de la aplicación con un 5 o un 4, siendo así la experiencia de uso extremadamente positiva</i> | 171 |

| | |
|--|-----|
| Ilustración 94 Ejemplo de los datos introducidos por cada usuario junto con el id del soneto seleccionado correspondiente. Los datos completos aparecen como archivos adjuntos y además hay una referencia a ellos en el Anexo 7.5 | 172 |
| Ilustración 95 Correlación entre los valores declarados, la puntuación emitida y las variables afectivas (I) | 173 |
| Ilustración 96 Correlación entre los valores declarados, la puntuación emitida y las variables afectivas (II) | 174 |
| Ilustración 97 Correlación entre los valores declarados, la puntuación emitida y las variables afectivas (III)..... | 175 |
| Ilustración 98 Correlación entre las variables seleccionadas por los usuarios..... | 177 |
| Ilustración 99 Histogramas con las puntuaciones según las funciones de composición..... | 178 |
| Ilustración 100 JSON ejemplo de respuesta del RecSys desarrollado | 198 |
| Ilustración 101 JSON con los campos del TEI XML de los sonetos de DISCO incluyendo la extensión para los campos afectivos | 199 |

Índice de Tablas

| | |
|--|-----|
| Tabla 1 Matriz de concurrencia | 67 |
| Tabla 2 Campos principales para el modelo de contenidos | 104 |
| Tabla 3 Campos principales para el modelo de usuario..... | 107 |
| Tabla 4 Solo se rechaza la hipótesis 6 si el MAP medio para todos los términos es menor que 0.25 y de manera individual dicho MAP no es consistentemente menor que 0.25..... | 118 |
| Tabla 5 Palabras presentes del corpus en los distintos conjuntos de datos..... | 136 |
| Tabla 6 Rasgos psicológicos | 141 |
| Tabla 7 Rasgos afectivos | 141 |
| Tabla 8 Valores de k-alpha para todos los anotadores por cada variable | 142 |
| Tabla 9 Comparativa entre los anotadores 1 y 2 | 143 |
| Tabla 10 Comparativa entre los anotadores 1 y 3 | 143 |
| Tabla 11 Comparativa entre los anotadores 2 y 3 | 144 |
| Tabla 12 One-way ANOVA para cada variable por par de anotadores | 146 |
| Tabla 13 Métricas para comparar las anotaciones de variables binarias entre los distintos anotadores | 147 |
| Tabla 14 Comparativa del soneto con los valores medianos frente al anotador 1 | 147 |
| Tabla 15 Comparativa del soneto con los valores medianos frente al anotador 2 | 148 |
| Tabla 16 Comparativa del soneto con los valores medianos frente al anotador 3 | 148 |
| Tabla 17 Correlación de los valores de media y desviación típica de “anger”, “sadness”, “disgust”, “happiness” y “fear” (eje Y) inferidas del texto usando (Ferré, Guasch, Martínez-García, Fraga, & Hinojosa, 2017) con los anotados por POSTDATA. | 154 |
| Tabla 18 Correlación de algunos rasgos usados en (Ullrich et al., 2016) (eje X) inferida del texto usando (Stadthagen-Gonzalez, Imbault, Pérez Sánchez, & Brysbaert, 2017) con los anotados por POSTDATA (eje Y) | 154 |
| Tabla 19 Correlaciones de los rasgos afectivos anotados (eje Y) con los rasgos afectivos inferidos (eje X) para “concreteness”, “context availability”, “imageability”, “valence y “arousal” con el conjunto de datos (Guasch, Ferré, & Fraga, 2016) | 155 |
| Tabla 20 Correlaciones de los términos psicológicos con los rasgos afectivos anotados por POSTDATA..... | 158 |
| Tabla 21 Correlaciones de los términos psicológicos con los rasgos afectivos inferidos desde las palabras individuales..... | 158 |
| Tabla 22 Resultados de los Experimentos 2 y 4: métricas MAP para los distintos escenarios..... | 162 |
| Tabla 23 Comparativa de escenarios para la métrica coseno..... | 163 |
| Tabla 24 Comparativa de escenarios para la métrica icm..... | 163 |
| Tabla 25 Contrastes de hipótesis comparando las métricas de recuperación por anotador frente a las obtenidas con los valores medianos de los 3..... | 164 |
| Tabla 26 Top 10 de las mejores métricas de recuperación usando los datos de los anotadores individuales | 165 |
| Tabla 27 Métricas de los distintos modelos de ML supervisado por categoría psicológica | 166 |
| Tabla 28 Comparativas entre los modelos de ML y los de IR..... | 166 |
| Tabla 29 Comparativas con un escenario ficticio de sonetos balanceados por etiquetas | 167 |
| Tabla 30 Lista de campos en FOAF Core..... | 194 |
| Tabla 31 Metadatos de autores en DISCO..... | 195 |
| Tabla 32 Campos en xAPI | 196 |
| Tabla 33 Campos para el método GET de la API desarrollada | 197 |
| Tabla 34 Resultados de los usuarios de AULA | 200 |
| Tabla 35 Frecuencia de las palabras usadas por los usuarios de AULA..... | 201 |
| Tabla 36 Tabla completa de resultados para los Experimentos 2 y 4 | 205 |

Agradecimientos

La presente memoria de Trabajo de Fin de Máster del Máster Universitario de Inteligencia Artificial Avanzada: Fundamentos, Métodos y Aplicaciones de la Universidad Nacional de Educación a Distancia se ha realizado bajo la supervisión de la Dra. Ángeles Manjarrés Riesco (Dto. de Inteligencia Artificial, ETSI Informática. UNED) y del Dr. Víctor Fresno Fernández (Dto. de Lenguajes y Sistemas Informáticos, ETSI Informática. UNED).

A su vez, en el desarrollo de esta memoria se ha colaborado con el grupo de investigación POSTDATA (*Poetry Standardization and Linked Open Data*, (POSTDATA, 2015) de la UNED a través del Dr. Salvador Ros Muñoz (ETSI Informática. UNED). “POSTDATA es una apuesta totalmente interdisciplinar en el ámbito de humanidades digitales que tiene como finalidad el estudio y análisis digital de la poesía. El proyecto propone un doble sistema de estandarización en el ámbito filológico y tecnológico cuyo objetivo es convertir el estudio de la poesía en las diferentes lenguas en un sistema accesible por ordenador, que permita establecer vías de comunicación entre las diferentes tradiciones literarias y culturales; facilitar y abrir el acceso a los usuarios de diferentes perfiles (investigadores, estudiantes, aficionados); y transformar el análisis poético en un lenguaje de datos enlazados, multilingüe, inteligible para el ordenador y enlazable con la información de la nube de datos de la web semántica”¹. POSTDATA ha proporcionado los distintos conjuntos de datos utilizados en esta memoria, en particular un conjunto de sonetos a través de la publicación de DISCO (Ruiz Fabo et al., 2017) y las anotaciones que sobre ellos han llevado a cabo Laura Alises, Marie Olivier y Aroa Rabdán.

Finalmente, en esta memoria se ha colaborado con IFEMA² (*Institución Ferial de Madrid*). “IFEMA es una entidad que organiza ferias, salones y congresos en sus instalaciones de Madrid, España. En sus eventos se dan cita empresas españolas y extranjeras para generar relaciones comerciales, multiplicar sus contactos y presentar todas las novedades”³. La colaboración con IFEMA se ha hecho a con la feria de AULA a través de su directora, María Dolores González Barbado (Directora Grupo Ferial 3. IFEMA). “AULA es un salón que ayuda al estudiante a decidir eficazmente su futuro formativo y profesional. En Aula, se muestra una significativa oferta de centros educativos y formativos, entidades y empresas de servicios que dan respuesta a las necesidades de formación de estudiantes (de 4º de E.S.O en adelante) que acuden a elegir el centro en el que formarse para acceder al mercado laboral. En 2019 se atendieron más de 4.000 personas de manera individual. Este espacio tendrá zonas especializadas en función de los destinatarios del servicio: estudiantes, centros, profesionales, familias, posgraduados y formación continua”⁴.

¹ POSTDATA. Web. 29/07/2019.

http://portal.uned.es/portal/page?_pageid=93,53372334&_dad=portal&_schema=PORTAL

² IFEMA. Web. 29/07/2019. <http://www.ifema.es>

³ Wikipedia. “Ifema”. Web. 29/07/2019. <https://es.wikipedia.org/wiki/Ifema>

⁴ IFEMA. “¿Qué es Aula?”. Web. 29/07/2019. <https://www.ifema.es/aula/que-es>

Resumen

Los sistemas de recomendación (RecSys) para aplicaciones de e-learning han supuesto un gran desarrollo en la educación personalizada, ya que pueden adaptar y ajustar los contenidos y el ritmo pedagógico a las necesidades de cada estudiante. Esto puede mejorarse aún más cuando estos sistemas consideran no solo el nivel de conocimiento de los estudiantes, sino también el contexto afectivo en un momento dado. A pesar de esto, el enfoque tradicional que incluye el contexto afectivo dentro de un RecSys normalmente reduce su aplicación a detectar un conjunto de estados del usuario y ajustar el mismo contenido con esa información. Otras soluciones clásicas incluyen el uso de información afectiva para enviar algunos mensajes predefinidos para, por ejemplo, reforzar la motivación. Sin embargo, como se ha demostrado en estudios filosóficos, psicológicos y pedagógicos, las dimensiones humanas están todas entrelazadas, lo que significa que el proceso de aprendizaje no sólo debe reducirse a las dimensiones racional e intelectual, sino que también debe abordarse directamente desde las afectivas. Esto es algo que sostiene la corriente de la pedagogía poética, defendiendo que la poesía puede usarse como un método complementario a los procesos de aprendizaje tradicionales, ya que puede contribuir al aprendizaje de conceptos, valores... y la forma de lograrlo es a través de la conexión afectiva establecida entre el alumno y el poema.

La presente memoria muestra una investigación y propuesta de cómo incluir la poesía pedagógica dentro de un RecSys. Inicialmente se revisa la estructura genérica de un RecSys afectivo. Con ello se indican posibles extensiones de algunas de las ontologías bien establecidas en el campo de los RecSys para incorporar la poesía pedagógica dentro de ellos con el objetivo de facilitar la interoperabilidad. Después de eso, la memoria analiza uno de los componentes principales que deben incluirse, el análisis automático de poesía con técnicas de procesamiento del lenguaje natural (NLP) para recomendar poesía personalizada a los estudiantes. Esto se hace considerando conjuntamente dos enfoques para generar las recomendaciones de cada poema, la información semántica y la afectiva.

El primer enfoque estudia diferentes técnicas de representación semántica del texto que abarcan desde los enfoques tradicionales, basados en semántica distribucional, hasta las soluciones contextuales más recientes, como BERT. La investigación en esta área también incluye la exploración de técnicas de composición semántica de textos, abordando aproximaciones de combinación de representaciones vectoriales de términos aislados (*word embeddings*) para encontrar representaciones semánticas de fragmentos de texto más largos. El segundo enfoque incluye, a partir de un corpus público en el que se asignan diferentes valores afectivos a palabras individuales, investigar la posibilidad de combinar esa información a nivel de palabra dentro de rasgos genéricos a nivel de texto, para usarlos después y poder seleccionar poemas según el estado afectivo del estudiante en un momento dado. Toda la investigación se realiza para el caso de uso específico de sonetos en castellano de los últimos cinco siglos.

La valoración de propuestas se analiza gracias a un equipo de expertos en el dominio poético de la UNED (POSTDATA). Ellos proporcionarán anotaciones afectivas a nivel de texto completo para los sonetos. Esto servirá para ver la relación de las anotaciones supervisadas globales respecto a las resultantes de agregar el valor afectivo de las palabras individuales. Junto con ello se incluyen anotaciones para estudiar la posible correlación entre la información afectiva de los sonetos, determinados trastornos psicológicos y si coinciden con la recuperación semántica. También proporcionarán anotaciones para estudiar las distintas alternativas de funciones de composición semántica. Las propuestas también se analizan con usuarios reales durante la feria AULA en IFEMA (Madrid) en un taller llamado "Te receto un soneto" mediante el uso de un asistente virtual que actúa como un sistema de recuperación de información que recomienda poesía personalizada dado una consulta de texto de entrada.

Abstract

Recommender systems (RecSys) for e-learning applications lead to a huge development in personalized education because they can tailor and adjust both contents and pedagogical pace to the needs of every student. This could be further enhanced when these systems consider not only the level of knowledge of the students but also the affective context in a particular moment. Despite this, the traditional approach that includes the affective context within a RecSys normally reduces its application to detect a set of user states and adjust the same content with that information. Other classical solutions include the usage of affective information to send some predefined messages to, for example, reinforce motivation. Nevertheless, as it has been shown from philosophical, psychological and pedagogical studies, human dimensions are all intertwined, which means that the learning process should not only be reduced to the rational and intellectual dimensions, but it should also be approached directly from the affective ones. This is something that the pedagogical poetry current claims, defending that poetry can be used as a complementary method to traditional learning processes since it can contribute to the learning of concept, values... but the way to achieve it is through the affective connection established between the student and the poem.

The present work shows a research and proposal of how to include the pedagogical poetry within a RecSys. This is initially done by reviewing the generic structure of an affective RecSys and indicating possible ways to extend some well-established ontologies used by these systems in order to allow interoperability. After that, the work analyses one of the main components that should be included, the automatic analysis of poetry with Natural Language Processing (NLP) techniques in order to recommend personalized poetry to students. This is done by considering jointly two approaches to generate those recommendations, the semantic information within each poem and the affective one.

The first approach is studied analysing different word embedding techniques that span from traditional approaches based on distributional semantics to the most recent contextual solutions such as BERT. Research in this area also includes the exploration of semantic text composition techniques, addressing approaches to combine vector representations of isolated terms (word embeddings) to find semantic representations of longer text fragments. The second approach includes the usage of public corpora to assign different affective values to individual words and investigate the possibility of combining that word-level information within generic text-level features. This is done in order to be able to select poems according to the student affective state in a given moment. This research is done for the specified use case of Spanish sonnets from the last five centuries.

The proposals are checked against a team of poetry domain experts of the UNED (POSTDATA). They will provide text-level annotations for the affective features using a group of those sonnets. The purpose of this is to analyse the relation of sonnet-level annotations with the result of the composition of the individual words using those public corpora. Together with that, they will include annotations to study the potential correlation between affective information and psychological disorders, and to see if the recovery of poems based on semantics match those terms. They will also provide annotations to analyse the different alternatives of semantic composition.

The proposals are also checked against a group of real users during the AULA fair at IFEMA (Madrid) in a workshop called "*Te receto un soneto*" ("*I prescribe you a sonnet*") by using a virtual assistant that acts as an information retrieval system recommending personal poetry given an input text query.

1. Introducción

En este primer capítulo se presenta la motivación del trabajo, así como las hipótesis y objetivos de investigación. Se presentará, a continuación, la metodología que se llevará a cabo, describiendo en detalle el planteamiento experimental; por último, se presentará la estructura de la presente memoria.

1.1. Preámbulo

Decir que la Inteligencia Artificial (IA) es una realidad presente ya en el día a día de la mayoría de las personas no es decir nada nuevo. Efectivamente esta tecnología aparece tras gran parte de las tareas que se realizan de manera cotidiana sin que muchas veces los seres humanos sean incluso conscientes de ello (edición de fotos, etiquetado automático de imágenes, autocompletado de textos...).

Algunos ejemplos de aplicaciones de este tipo, presentes de manera transversal en varios sectores, son los sistemas que personalizan su contenido y las acciones que llevan a cabo en función de la información que posean del usuario con el que se están relacionando. Así, se pueden personalizar la experiencia de compra online, la recomendación de contenidos audiovisuales adecuados a los gustos particulares de cada usuario, o incluso las sugerencias de patrones de hábito saludable en función de la información recogida por distintos dispositivos electrónicos.

Estas aplicaciones de IA que recomiendan un contenido o acción concreta para un determinado usuario se conocen con el nombre de Sistemas de Recomendación (RecSys), y han determinado una de las líneas de trabajo e investigación más prolíficas de las últimas décadas.

Los ejemplos citados son algunos de los múltiples escenarios en los que se puede utilizar un RecSys, pero no son los únicos. Otra de las grandes aplicaciones de estos es al aprendizaje de los estudiantes en entornos de e-learning, de manera que el contenido e itinerario formativo esté totalmente personalizado y adecuado a sus necesidades y ritmo formativo, tratando de maximizar siempre la experiencia de aprendizaje al reconocer la unicidad de cada persona.

Existen tantas propuestas de diseñar estos RecSys para entornos de e-learning como propuestas pedagógicas, de manera que se han desarrollado sistemas de este tipo para formaciones en diversas clases de disciplinas y áreas del conocimiento, al tiempo que se han diseñado teniendo en cuenta tanto la realidad de la *teoría de las inteligencias múltiples* del ser humano (y personalizando una formación que no siempre ha de estar ligada al ámbito lógico-matemático) como la realidad de que el ser humano es una unidad de razón, voluntad y afectos, de manera que una formación íntegra deberá estar enfocada a interactuar con todas sus dimensiones.

Es por este último motivo que una prolífica línea de investigación en el ámbito de los RecSys en general, y en los aplicados a e-learning en particular, es el de los RecSys sensibles al contexto afectivo de la persona. En el caso de sistemas de aprendizaje no sólo se considera el progreso académico para personalizar los contenidos, sino que también se enfocan y adaptan estos teniendo en cuenta el amplio espectro de afectos que pueden estar presentes en el

estudiante en un momento dado. Así, por ejemplo, la recomendación para el aprendizaje será diversa según se detecte que el estudiante está alegre, aburrido o cansado.

Esta consideración de la dimensión afectiva dentro del ámbito de los RecSys para entornos de e-learning puede ser llevada a cabo de distintas maneras. Una propuesta interesante y que ofrece una atractiva línea de investigación es la aplicación de la teoría educativa de la pedagogía poética. Según esta teoría la comprensión del estudiante se puede reforzar, especialmente para determinados contenidos, cuando la vía de diálogo se establece mediante textos poéticos específicos. Esto es así debido a que los poemas propician la conexión con la dimensión afectiva de la persona, y esto permite complementar y reforzar el aprendizaje más clásico, focalizado en el trabajo intelectual y puramente racional. De esta manera, un RecSys que tuviese en cuenta el contexto afectivo en un entorno de e-learning recomendaría un poema en particular para distintas casuísticas que apareciesen durante el aprendizaje (necesidad de refuerzo motivacional, enseñanza de ciertos conceptos...).

Además de contribuir al aprendizaje, un RecSys de estas características podría también recomendar poesía de una manera más abierta mediante un diálogo directo con el usuario. Es decir, podría recibir una determinada consulta, interés, frase de entrada... y responder mediante poesía al mensaje (del estudiante o incluso del profesor), de modo que la respuesta también tuviese en cuenta la dimensión afectiva de la persona, muchas veces detectable en sus propias palabras. La idea de desarrollar RecSys de contenidos literarios que no se centren en recomendar un contenido concreto simplemente mediante una coincidencia con unas etiquetas determinadas sino que sea por un diálogo desde el punto de vista semántico con el propio texto que se recomienda es algo que por ejemplo ha realizado Google con su herramienta *Talk to Books*⁵. Esta herramienta permite escribir consultas libremente en texto plano de forma que el sistema recomienda obras literarias en las que determinados pasajes tienen una relación semántica con la consulta introducida. De este modo se conocen las distintas obras literarias no sólo por resúmenes o reseñas sino por lo que la propia obra dice de sí misma ante una consulta específica. Explorar e investigar entorno a esta idea llevado hacia la poesía y considerando la dimensión afectiva del usuario es una opción muy interesante para incluir también en el campo de e-learning. El ciclo de aprendizaje se enriquece aún más; no sólo se recomiendan determinados poemas para refuerzo motivacional o para la comprensión de determinados conceptos, sino que se abre la posibilidad de que alumnos y profesores dialoguen directamente con todo el corpus y que sea la propia poesía la que hable por sí misma.

Todas estas ideas plantean retos claves para las aplicaciones de la IA en la educación que pueden redundar en un mayor grado de personalización de la oferta educativa que al mismo tiempo considere aún más a la persona en su conjunto y en toda la riqueza de sus dimensiones. Este preámbulo y contextualización se concreta en el capítulo siguiente en tareas específicas que se van a describir en la presente memoria.

1.2. Motivación

Tradicionalmente, los Sistemas Adaptativos en la Educación (SAE) se han centrado casi exclusivamente en el aprendizaje de los alumnos entendiendo este desde una perspectiva reducida que sólo implica una formación a nivel cognitivo que no tiene en cuenta más dimensiones de la persona, como se advierte en (Manjarrés-Riesco, Santos, Boticario, &

⁵ Google. "Talk to Books". Web. 09/06/2019. <https://books.google.com/talktobooks/>

Saneiro, 2013), y más recientemente en (Santos, 2016). Esta perspectiva es muy simplista y se requiere ampliar la visión sobre la realidad humana para entender el proceso pedagógico como algo íntegro y unitario respecto a todas las dimensiones del ser humano.

Ciertamente este no es un problema nuevo, sino que ya ha sido estudiado ampliamente desde el punto de vista de la antropología filosófica. Según qué noción filosófica subyazca al investigador de los SAE, ésta condicionará el desarrollo y propuestas que haga de los mismos. ¿Qué nociones podrían subyacer entonces? Clásicamente se ha visto al ser humano desde la visión monista, que lo considera como un mero ente material o un ente espiritual, hasta la visión dualista defendida por numerosos autores (Descartes, 2013) en donde se puede ver al ser humano como un ser fracturado en una *res cogitans* y una *res extensa* (que rara vez interactúan entre ellas) de manera que en última instancia se pueda ponderar o dar más importancia a una de ellas que a la otra. Es así que, durante mucho tiempo, se han minusvalorado el cuerpo y su afectividad frente al intelecto, entendido en sentido reducido como pura capacidad de comprensión y abstracción a nivel de las ciencias empíricas newtonianas.

Esta interpretación de la realidad humana se heredó en todo el idealismo posterior y fue recogida también en la reflexión de la razón pura de Kant (Kant, 1998) donde la visión fracturada del hombre evoluciona hasta incluso el desprecio explícito de los propios afectos y sentimientos por pervertir estos la dinámica de la acción humana (Kant, 1999) entendiendo así que el cuerpo sería, efectivamente, un sepulcro del alma (Platón, 1997).

Con esta visión no sería importante que un SAE no tuviese en cuenta los afectos, pero esta visión está errada de inicio. Ya clásicamente muchos autores, como Aristóteles (Nussbaum & Rorty, 1992) han defendido un hilemorfismo en el que se ve al hombre como una unidad cuerpo-alma (afectos- intelecto), de forma que en todas sus acciones (incluido el aprendizaje) estarían presente estas dimensiones. Esta visión clásica ha sido renovada y defendida también por muchos autores modernos (Eberl, 2004).

Es por esto que autores como (Boticario, Jesus et al., Jul, 2017) consideran muy importante y urgente fomentar la investigación en computación afectiva en el campo de la educación apoyada en técnicas de IA de forma que las interacciones educativas sean verdaderamente personalizadas, entiendan la casuística real del alumno y por ello incorporen, en la medida de lo posible, la visión del ser humano desde el principio de la razón vital donde todo él está siendo tenido en cuenta.

Junto con intentar dar respuesta a esta necesidad educativa desde una visión antropológica completa, aparece también otra cuestión en relación a la educación personalizada, y es el poder tratar a cada persona como ser irrepetible que es (González Fernández, 2016), en su categoría de único, dando respuesta a sus necesidades educativas, sean las que sean, pudiendo, efectivamente, democratizar la educación dando acceso a ella a potencialmente cualquier persona, tenga las circunstancias vitales que tenga.

Así pues, la investigación antropológico-filosófica aporta una sólida justificación sobre lo crucial que es incorporar dicha dimensión afectiva dentro de la formación pedagógica en general, y por tanto también particularmente para los SAE. Pero la manera de incorporar los afectos dentro del itinerario formativo no es una tarea evidente, como también se menciona en (Manjarrés-Riesco et al., 2013). En dicho artículo se señalan muchos de los retos que se tienen que afrontar al desarrollar un sistema de IA para la educación que incluya la componente afectiva, condensándose en cinco áreas principales:

1. La falta de casos de uso y de información al respecto de esta disciplina.
2. La dificultad de obtener la información afectiva y de interpretarla correctamente para poder inferir estados desde ella.
3. Lo complejo que es poderse comunicar con el propio alumno de forma que se aborden estas cuestiones afectivas.
4. En relación con el punto anterior, que se pueda dar un soporte personalizado y afectivo que vaya más allá del simple refuerzo motivacional (que es lo único en lo que clásicamente se han centrado muchos de estos sistemas).
5. La falta de formación del profesorado en estas nuevas tecnologías, de forma que puedan usarlas eficientemente y estar desde el principio involucrados en su desarrollo, pudiendo contribuir muy positivamente a su evolución, como también se detalla en (Santos, Saneiro, Boticario, & Rodriguez-Sanchez, 2016).

Aunque el artículo referenciado data del 2013, el panorama descrito no ha cambiado sensiblemente desde entonces, como se pone de evidencia en la revisión (Santos, 2016). De esta forma, aparece como punto fundamental encontrar casos de uso que incluyan la dimensión afectiva para algo más que dar refuerzos motivacionales a los alumnos.

A pesar de que la motivación principal fue inicialmente enriquecer los RecSys educativos afectivos, también aparece como motivación dar soporte a la implantación de la pedagogía poética en las aulas. Para este propósito es muy interesante disponer de herramientas de autor que permitan a los docentes de cualquier nivel educativo recopilar los poemas idóneos para el desarrollo de unas competencias concretas, para su uso en contextos no solo virtuales sino también presenciales.

Finalmente, también se menciona como motivación el hecho de que hay muy pocas investigaciones en el campo de NLP enfocado a poesía, en particular para el castellano. El campo de la poesía plantea muy interesantes desafíos, dada la idiosincrasia del lenguaje poético. Para otros idiomas se mencionan como relevantes las recientes investigaciones de (Ullrich, Aryani, Kraxenberger, Jacobs, & Conrad, 2016; Walshe, 2017; Wassiliwizky, Koelsch, Wagner, Jacobsen, & Menninghaus, 2017).

Respecto al castellano, uno de los grupos de investigación más prolíficos es POSTDATA. Según el propio proyecto, éste “concibe desde un punto de vista interdisciplinario y está estructurado en tres dimensiones diferentes para reflejar la necesidad de construir puentes entre las disciplinas, los usuarios y las metodologías: académicas y filológicas, tecnológicas, sociales y orientadas al usuario”.

Con él se busca la creación de una plataforma digital para la edición de poesía dirigida a diferentes tipos de usuarios: académicos que desean trabajar en ediciones digitales críticas, usuarios no experimentados que desean leer, compartir y aprender más sobre tradiciones poéticas, y también empresas que van a utilizar este recurso para diferentes aplicaciones en campos como la educación, la psicología, el turismo o con fines culturales. La accesibilidad se considera uno de los requisitos clave para garantizar el éxito de la plataforma, por lo que todas las tecnologías utilizadas son transparentes para la mayoría de los usuarios. De esta manera,

ofrecerá un entorno fácil de usar para las personas que no poseen ninguna capacidad técnica de computación.

Desde el punto de vista filológico, el objetivo principal de esta propuesta es desarrollar una estandarización conceptual para construir un modelo abstracto para la representación de la poesía basado en conceptos filológicos existentes tomados de proyectos, manuales y corpus de las diferentes tradiciones. Además, este se divide en tres objetivos secundarios:

- Llevar a cabo un análisis comparativo de proyectos digitales y repertorios de diferentes tradiciones poéticas y extraer los principales elementos conceptuales y metadatos compartidos por la mayoría de ellos.
- Analizar la evolución de los manuales y estudios académicos sobre poesía para comprender la evolución de las teorías y elegir los conceptos más adecuados para su representación digital.
- Crear un modelo abstracto de representación con los elementos comunes esenciales extraídos en los dos primeros objetivos.”⁶

Pedagogía poética

Continuando con la reflexión filosófica, (Lapiedra Gutiérrez, 1997) recuerda el concepto clásico de Razón Poética de la filósofa malagueña María Zambrano. Dicho concepto es usado para designar cómo razón y poesía se complementan y cómo gracias a esta última es posible conocer la intimidad de las cosas, sirviendo así como vía epistemológica para descubrir una dimensión profunda del ser que de otra manera sería imposible de alcanzar. La poesía es conocida por su interlocución y vinculación con la dimensión afectiva del ser humano, y es por este motivo que en pedagogía se han desarrollado corrientes que la incluyen explícitamente dentro del proceso de aprendizaje, dando lugar a conceptos como la denominada *pedagogía poética*.

El concepto de *pedagogía poética* aparece claramente detallado en (Andújar, 2016) donde el autor define y muestra ejemplos para entender y ver cómo aplicarla dentro de un itinerario formativo. Dicha pedagogía poética busca fomentar y contribuir a un aprendizaje integral del estudiante al complementar otros tipos de formaciones más intelectuales con una formación dirigida a la comprensión de determinados conceptos mediante la dimensión afectiva del ser humano, gracias al uso de la poesía como elemento conductor del diálogo con dicha dimensión.

A continuación se muestran algunos ejemplos de explicaciones poéticas de conceptos que se muestra en (Andújar, 2016) frente a la definición de esos mismos conceptos en la RAE⁷.

⁶ POSTDATA. “Project objectives”. Web. 29/07/2019. <http://postdata.linhd.uned.es/project/>

⁷ Real Academia de la lengua Española. “verdad”. Web. 01/08/2019. <https://dle.rae.es/srv/fetch?id=bbdGpd4>

Verdad:

- Poesía:
*La verdad es lo que es,
y sigue siendo verdad aunque se piense al revés*
-- Anónimo
- RAE [1]: Conformidad de las cosas con el concepto que de ellas forma la mente
- RAE [4]: Juicio o proposición que no se puede negar racionalmente

Infancia:

- Poesía:
*Mi infancia son los recuerdos de un patio de Sevilla
y un huerto claro donde madura el limonero;
mi juventud, 20 años en tierra de Castilla;
mi historia, algunos casos que recordar no quiero.*
-- Antonio Machado
- RAE [1]: Período de la vida humana desde el nacimiento hasta la pubertad.

Como se puede apreciar con estos ejemplos, gracias a la pedagogía poética se consiguen dos efectos positivos. Por un lado, como ilustra el primer ejemplo, si bien el poema da una definición no tan rigurosa o exhaustiva del concepto en cuestión, esta definición lo aproxima de una manera simple y que deja una impronta en la memoria más difícil de borrar, por lo que es más sencillo recordarla y volver a ella, e incluso dialogar con ella y completarla más adelante gracias a definiciones no poéticas de más peso intelectual.

En el segundo ejemplo se ilustra el segundo efecto: la poesía sirve de complemento a la definición no poética de corte intelectual. Si bien se puede entender sin dificultad la definición de la RAE de la palabra *infancia*, el breve poema de Machado complementa dicha definición con una riqueza estética que estimula la imaginación y trae la dimensión afectiva al proceso cognoscitivo sobre dicho concepto.

Se reivindica así con esta pedagogía que la poesía es fuente de conocimiento, tanto para la comprensión de conceptos específicos como de determinadas ideas, por su gran capacidad de diálogo con la dimensión afectiva del ser humano, capaz de abrirse, utilizando la terminología tomista, sobre los trascendentales del ser, desde la contemplación del *pulchrum* a la comprensión del *verum* de la realidad.

Esta comprensión y educación alcanzada por la poesía no es solamente para una comprensión intelectual, como se ha mencionado haciendo referencia al *verum*, sino que también sirve para fomentar una educación moral, de manera que gracias a ella el alumno aprenda, utilizando la terminología de fenomenólogos como Scheler, la importancia de determinados valores para él mismo y para la sociedad o el bien común. Así, dicha contemplación del *pulchrum* también lleva a descubrir la existencia del *bonum*.

Volviendo a tomar un ejemplo de (Andújar, 2016):

Aprendizaje:

- Poesía:
*Aquí y allá aprendemos cautamente
en el río, en la noche, en la fatiga.
A vivir glorias o a morir de pena
en el rumbo mejor o a la deriva.*
-- Mario Benedetti
- RAE [1]: Adquirir el conocimiento de algo por medio del estudio o de la experiencia

En este ejemplo se puede ver cómo la aproximación poética sirve no sólo para comprender mejor el concepto, sino también para integrar dicho conocimiento teórico en la propia vida, de manera que el conocimiento contribuya al desarrollo personal.

Además de lo mencionado previamente, en la literatura aparecen numerosas referencias sobre el desarrollo de competencias en distintos ámbitos gracias al uso de la poesía. Un ejemplo es el de (Shapiro & Rucker, 2003) donde se ha definido y validado un programa de humanidades y arte para el ámbito médico. Se ha podido observar cómo la contribución humanista de la poesía ha ayudado a estimular en los médicos la capacidad de observación y de análisis del lenguaje de los pacientes y de su comportamiento, ha servido para aumentar la imaginación y curiosidad sobre las experiencias que han vivido los pacientes, ha aumentado su empatía y reforzado el vínculo emocional con los pacientes, y en definitiva ha servido para contribuir a que los médicos vean a los pacientes como un todo, como una persona única. La incorporación de asignaturas y formación humanista dentro del currículum médico ha servido para educar de forma más amplia y transversal a los facultativos.

Otro ejemplo más reciente es el de (Walshe, 2017) donde se estudia, en particular, la contribución de la poesía para reforzar la formación en asignaturas de geografía (concretamente asignaturas sobre desarrollo sostenible), consiguiendo que los estudiantes adquieran una mayor vinculación afectiva con la materia, sean más críticos y aumente su interés en ella. En la referenciada memoria se presentan poemas que sirven como vía alternativa para entender determinados conceptos de la asignatura al tiempo que suscitan el debate entre alumnos.

De esta manera aparece como línea de trabajo prometedora explorar la aplicación y utilidad de la pedagogía poética para los SAE como caso de uso adicional en que se tiene en cuenta la dimensión afectiva dentro de los itinerarios formativos y, para el caso concreto de los RecSys aplicados a e-learning, dentro de las propias recomendaciones proporcionadas en base al perfil del estudiante.

Poesía y psicología

Diversas publicaciones científicas hablan de la potencial influencia de la poesía en el tratamiento de algunos trastornos psicológicos (Czernianin, 2016; Parastoo, Amenehsadat, & Shahla, 2016; Wassiliwizky et al., 2017). Otras publicaciones relevantes mencionan el potencial de considerar la dimensión afectiva dentro de un SAE para la inclusión de personas con determinados trastornos psicológicos, como el autismo (Santos, 2016). Por este motivo es

interesante que un sistema de recuperación de información (IR) para poesía personalizada pudiese vincular los objetos de aprendizaje con determinadas categorías psicológicas. Esto podrá servir para un doble propósito: la comprensión de los distintos trastornos (propósito educativo) y su tratamiento (propósito terapéutico).

También es de interés mencionar que muchos poetas afamados han padecido trastornos mentales, lo que sugiere utilizar su obra para explorar el universo conceptual de las personas que padecen estos trastornos mediante minería de textos.

1.3. Objetivos de investigación

Como se mencionó en el apartado anterior, la motivación principal de esta investigación es dar soporte a la pedagogía poética en los RecSys educativos afectivos. Para que esto sea posible es necesaria la selección de poemas en base a características afectivas y semánticas. Esto ha hecho necesario que la investigación se centrara también en los campos de la minería de textos y NLP, que adquirieron una gran importancia y se constituyeron finalmente en la parte central del trabajo.

De esta manera, el punto central de la memoria será, tras analizar el Estado del Arte (*State of the Art*, SOTA) de los RecSys afectivos, del uso de poesía para computación afectiva, del tratamiento de poesía mediante técnicas de NLP y del modelado semántico mediante *word* y *text embeddings*, abordar las siguientes hipótesis de investigación:

- **Hipótesis 1:** Los rasgos afectivos que tienen las palabras que componen un poema, y cuyo valor se toma de estudios sobre poesía en otros idiomas, como (Ullrich et al., 2016), son igual de relevantes para el castellano y servirán para predecir el gusto (*liking*) estético del poema.
- **Hipótesis 2:** El valor de los rasgos afectivos de los poemas en castellano se infiere del valor afectivo individual derivado de las palabras que lo componen, de modo similar a como se hace en (Ullrich et al., 2016) para poesía en alemán.
- **Hipótesis 3:** La composición semántica de un poema a partir de los *word embeddings* individuales de las palabras que contiene servirá al proceso de selección y recuperación de poemas dentro de un proceso de aplicación de la pedagogía poética. Esto es coherente con los resultados presentados (Amigó Cabrera & Fresno Fernández, 2018).
 - **Hipótesis 3.1:** *Embeddings* contextuales como BERT (*Bidirectional Encoder Representations from Transformers*) aportarán más que *word2vec* dentro de un sistema de recuperación de información (*Information Recovery*, IR) para poesía.
- **Hipótesis 4:** El contenido de los sonetos está identificado según esté o no relacionado con un término psicológico, concretamente un tipo de desorden psicológico. Por este motivo, como la literatura (Parastoo et al., 2016; Wassiliwizky et al., 2017) menciona que hay una correlación entre trastornos psicológicos y el estado afectivo, esa correlación debe aparecer también reflejada en los sonetos.
- **Hipótesis 5:** Es posible etiquetar poemas utilizando metadatos coherentes con las ontologías de uso común en los RecSys educativos afectivos, de modo que sean

recuperables como objetos de aprendizaje personalizados, haciendo uso de técnicas de NLP.

La especificación de dichas hipótesis de investigación lleva a definir los siguientes objetivos como marco para su evaluación. Estos objetivos se expresan en este apartado de forma sucinta, remitiéndose al Capítulo 3 (Descripción de la Experimentación) para más detalle.

Objetivo primario: Realizar una primera aproximación al soporte de la pedagogía poética en los SAE y, más concretamente, en los RecSys sensibles al contexto afectivo aplicados a entornos de e-learning.

- **Objetivo 1:** Delimitación de los retos tecnológicos planteados en el objetivo primario.
- **Objetivo 2:** Aproximación a la cuestión principal: la utilización de poemas como objetos de aprendizaje personalizados, atendiendo a vincularlos con el contexto afectivo, y a capturar su significado semántico gracias a la aplicación NLP, en concreto aplicando aproximaciones basadas en semántica distribucional.

1.4. Metodología

En este apartado sobre la metodología se van a cubrir 3 aspectos en concreto: las etapas y especificaciones de los experimentos que se van a llevar a cabo, la metodología usada como referencia para el proyecto y otras metodologías que se deberían considerar en el futuro para el desarrollo de un SAE al completo que integrase el enfoque de la pedagogía poética.

1.4.1. Experimentos realizados

POSTDATA

Con la colaboración de POSTDATA se abordan 3 experimentos para analizar tres de las hipótesis planteadas en el apartado anterior.

- Experimento 1:
 - Sirve para comprobar la hipótesis 2.
 - La hipótesis 2, en la que se recogía que se puede componer el valor afectivo de palabras individuales para obtener rasgos afectivos a nivel de soneto
 - El experimento consiste en comparar las anotaciones afectivas del subconjunto de sonetos etiquetados con las derivadas de agregar la información individual.
 - La hipótesis se aceptará si hay correlación entre los rasgos inferidos y los correspondientes rasgos anotados.
- Experimento 2:
 - Sirve para comprobar la hipótesis 3.1, en la que se va a estudiar si BERT como *word embedding* contextual da mejores resultados que word2vec

- Se dan dos alternativas experimentales. La opción 1 es desarrollar una aplicación prototipo que muestre ante una consulta de entrada dos poemas recuperados usando una misma función de composición y una misma métrica pero distinto *word embedding*. La opción 2 es comparar los resultados de efectividad en la recuperación de sonetos usando como referencia los rasgos psicológicos anotados por POSTDATA.
 - Aun planteándose estas dos alternativas y desarrollándose la aplicación para abordar la Opción 1, en esta memoria se incluirá el análisis para esta Opción 2, remitiéndose el uso de la herramienta mencionada en la Opción 1 para líneas futuras.
 - Con la Opción 2, el experimento evalúa si hay diferencias significativas en la recuperación de poemas según el *embedding* utilizado.
- Experimento 3:
 - Para comprobar la hipótesis 4 se analiza la correlación de los rasgos psicológicos (0 o 1 dependiendo de si el soneto los representa o no) procedentes de (García Franco & Manjarrés Riesco, 2016) con los rasgos afectivos.
 - La hipótesis se aceptará si hay correlación entre el vector de rasgos afectivos y el de psicológicos.
- Experimento 4:
 - Sirve para comprobar la hipótesis 5, que recogía que es posible etiquetar poemas para que sean recuperables como LO.
 - En determinadas ocasiones las etiquetas de los LO serán manuales, pero en muchas otras se querrán recuperar LO de temáticas específicas de las que no se dispone de etiquetas previas.
 - Para que la hipótesis 5 se acepte se deben poder recuperar sonetos de distintas temáticas de forma no supervisada.
 - El experimento que se lleva a cabo consiste en utilizar técnicas de NLP y de IR y desarrollar un sistema que devuelva sonetos recuperados ante consultas de entrada.
 - Como caso de uso se analizan las salidas de dicho sistema ante consultas de términos psicológicos simulando un escenario en el que se recomendasen sonetos como LO de un curso de psicología a sus estudiantes.
 - Para el experimento se comparan los sonetos recuperados con consultas con respecto a las anotaciones de los expertos de POSTDATA ante esos mismos términos de consulta (que indicaban cuando el contenido del soneto tenía que ver con dicho concepto o no).

AULA

Con la colaboración de AULA se abordan 2 experimentos para analizar dos de las hipótesis planteadas en el apartado anterior.

- Experimento 5:

- Para comprobar la hipótesis 1, que menciona que los rasgos afectivos sirven para predecir lo que a un usuario va a gustar un soneto se van a utilizar los rasgos afectivos derivados de las palabras individuales en caso de que la hipótesis 2 sea cierta.
 - Se desarrolla una aplicación para AULA mediante la cual los usuarios escriben una consulta y el sistema les recomiende sonetos semánticamente similares.
 - Para el experimento se analiza la correlación entre el valor afectivo de los sonetos seleccionados y la puntuación emitida.
- Experimento 6:
 - Para comprobar la hipótesis 3 se usa la misma aplicación indicada en el apartado anterior.
 - Se muestran 2 sonetos recuperados según la proximidad semántica ante una consulta del usuario, acotada a una única palabra.
 - La función de composición usada para uno de los sonetos es RL_SEQ y para otro F sum. La métrica para ambos casos será el coseno. Los *word embedding* usados serán BERT.
 - La idea es comparar el vector de elecciones de los sonetos de un tipo frente a los del otro, así como las puntuaciones que se les ha dado, para ver si hay diferencias significativas.

1.4.2. Metodología CRISP-DM

Aunque la metodología CRISP-DM (*Cross-Industry Process for Data Mining*) descrita en (Shearer, 2000) se usa generalmente para el desarrollo de proyectos de minería de datos donde se utilizan modelos de aprendizaje automático (generalmente de aprendizaje supervisado), se puede utilizar como referencia para la especificación y guía de las distintas etapas abordadas en esta memoria.

De manera general, las fases de esta metodología se recogen en la siguiente ilustración:

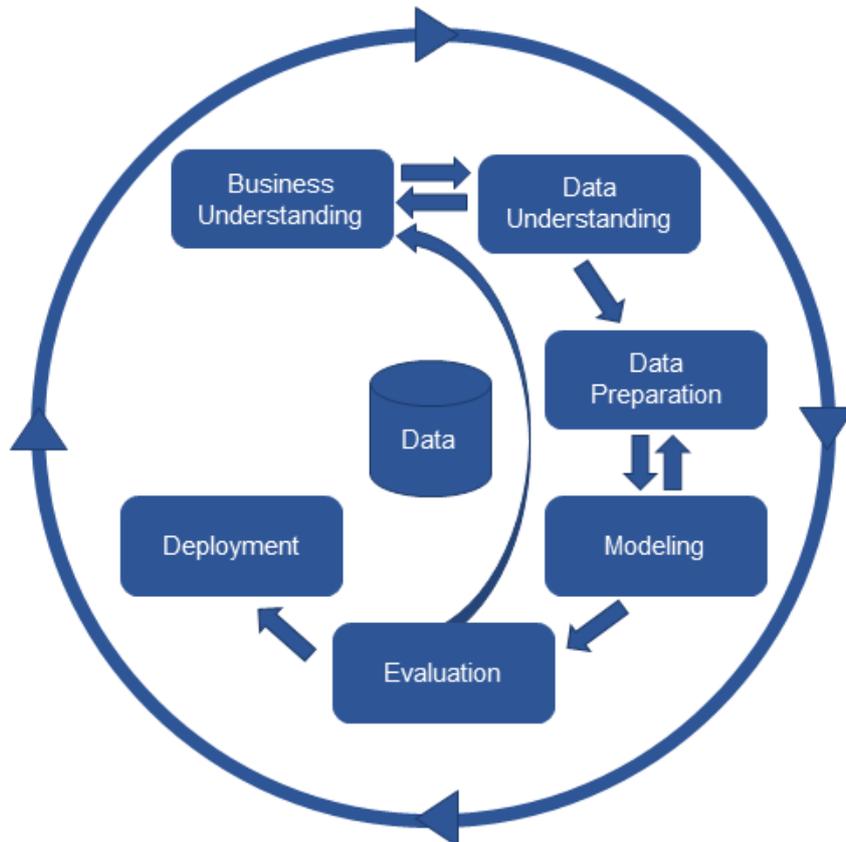


Ilustración 1 Fases de CRISP-DM⁸

⁸ Taylor, James. "Four Problems in Using CRISP-DM and How To Fix Them". Web. 07/06/2019. <https://www.kdnuggets.com/2017/01/four-problems-crisp-dm-fix.html>

A continuación se describen brevemente las distintas etapas, dejándose su detalle y aplicación a este TFM para el Capítulo 3.

1. **Comprensión de Negocio:** La primera fase apunta a comprender el contexto de negocio, formular las preguntas que se quiere resolver y especificar los entregables del proyecto. Se determinan así objetivos y reglas de negocio que indiquen cuando se han cumplido dichos objetivos.
2. **Comprensión de Datos:** Analizar el conjunto de datos disponible y conocer los campos con los que se puede trabajar y cómo se relacionan entre sí. Por ejemplo, llevar a cabo un Análisis Exploratorio de Datos para comprender el conjunto (o conjuntos) de datos del que se dispone.
3. **Preparación de Datos:** Preprocesar adecuadamente los datos para poder aplicar sobre ellos modelos (p.e. de ML) que infieran matemáticamente relaciones dentro del conjunto de datos. Entre las tareas de preprocesado aparece la transformación de campos de texto a valores numéricos, la limpieza de datos (p.e. eliminación de datos nulos), el balanceo de datos (p.e. para tener cantidades de datos similares para todas las combinaciones categóricas) ...
4. **Modelado:** Elegir el modelo a aplicar sobre los datos (p.e. un árbol de decisión mediante el algoritmo C5.0) para dar respuesta a los objetivos definidos en el punto 1.
5. **Evaluación:** Evaluar los resultados obtenidos en el punto 4 mediante el uso de métricas, analizando si se han alcanzado los objetivos del punto 1.
6. **Iteraciones:** En caso de que no se hayan alcanzado aún los objetivos, iterar y volver al punto 1. Es posible que en este punto haya aumentado el conocimiento de negocio y la comprensión de los datos, de manera que siguiendo un esquema de metodología ágil, se reevalúen los objetivos y, si es necesario, se redefinan.
7. **Despliegue:** Una vez satisfechos los objetivos propuestos, se pasa a producción la propuesta.

1.4.3. Otras metodologías: TORMES

Otra metodología que mencionar es TORMES (*Tutor-Oriented Recommendations Modelling for Educational Systems*), que si bien no ha sido aplicada explícitamente en esta investigación, ha servido de inspiración para alguno de los puntos. Además, es relevante referenciarla ya que sería una metodología muy indicada para aplicar en caso de que el presente trabajo se integrase dentro del desarrollo de un SAE completo. Dicho SAE integraría el enfoque de la pedagogía poética, lo que significaría que la poesía podría ser objeto de recomendación tanto con propósitos de aprendizaje como de soporte afectivo.

TORMES es una metodología especialmente útil en proyectos de analítica del aprendizaje relacionados con la educación. Dicha metodología aparece publicada en el artículo (Santos & Boticario, 2011) y se basa en la ISO 9241-210, buscando fomentar el desarrollo de los SAE centrándose en el usuario. Se pretende que en el desarrollo tengan un papel clave la analítica y

minería de datos, y que haya una estrecha colaboración entre todos los implicados en el SAE, incluyendo así desde el principio del desarrollo a los propios educadores.

La metodología TORMES define cuatro actividades principales para estos desarrollos iterativos centrados en el usuario:

1. Entender y especificar el contexto de uso
2. Especificar los requerimientos del usuario
3. Producir el diseño de soluciones de las soluciones para satisfacer los requerimientos
4. Evaluar el diseño frente a los requerimientos

Además de los puntos mencionados se propone la incorporación de diversos análisis y minería de datos sobre los distintos datos de los que se disponga (por ejemplo, interacciones de usuarios). Con ello se busca identificar situaciones potencialmente interesantes para ir incorporando más adelante, o detectar problemáticas que hayan ido surgiendo en las interacciones con el sistema, y con todo ello afinar los procesos del SAE y su toma de decisiones. Esto sirve, por ejemplo, para perfilar mejor al usuario de modo que si el SAE fuese un RecSys para e-learning se ajustasen más las recomendaciones y personalizaciones del sistema. De esta manera se ofrece una solución evolutiva que se va ajustando a las necesidades detectadas a lo largo del curso.

El primer punto de la metodología es definir el problema que se va a abordar. Tras ello, se define un escenario donde se deben satisfacer unas metas, y después se reescribe dicho escenario para convertirse en un escenario con soluciones integradas a los desafíos a través de recomendaciones interactivas, ya sea dando una respuesta directa o explícitamente ignorando algunas casuísticas. Esos escenarios pueden ser múltiples y tener distintas soluciones posibles, que pueden ser complementarias.

Una vez que se detectan los requerimientos del usuario, se procede al diseño de la solución, en el que se comienza viendo qué acción es deseable que el usuario emprenda y debe por tanto recomendarse, qué datos son relevantes para identificar este escenario, cuál es la justificación para esta acción, cómo comunicar la recomendación al alumno sin interrumpir su proceso cognitivo y cuáles son los atributos recomendados.

TORMES incluye dentro de su proceso iterativo la verificación constante del desarrollo gracias a expertos en el dominio, donde se suele incluir al personal docente. Para la presente memoria esto ha servido ya de inspiración para consultar a profesores experimentados, concretamente a los tutores de este TFM, de cara a definir los metadatos relevantes para la recuperación de poemas.

1.5. Estructura del documento

Tras presentar la introducción, motivación, objetivos y metodología empleada para el proyecto, el documento que sigue se estructura de la siguiente manera.

Capítulo 2: Descripción del problema.

El Capítulo 2 estudia y presenta el Estado del Arte (*State of the Art*, SOTA) de las áreas científicas relacionadas con la presente memoria, comenzando por el ámbito de los RecSys sensibles al contexto afectivo, especialmente para las aplicaciones de e-learning, así como de las posibles ontologías para su representación. Además, se aborda el estudio de la literatura sobre tratamiento automático del contenido afectivo en textos poéticos, así como de los distintos *corpora* disponibles para la construcción de rasgos afectivos en castellano. Finalmente se analiza el SOTA de las representaciones vectoriales del valor semántico de las palabras gracias al uso de *word embeddings*, mostrando un breve recorrido histórico por la evolución de dichas técnicas hasta llegar a las soluciones más novedosas indicando sus potenciales ventajas. Junto con el análisis de las representaciones vectoriales de palabras individuales se detallan algunas técnicas modernas de representación de textos al completo mediante el uso de técnicas de composición.

Este estudio sobre el SOTA servirá para definir el contexto donde se enmarca el presente TFM y servirá de marco y justificación para las propuestas de investigación ya introducidas en el capítulo previo. Con ello se describe a continuación tanto un estudio tecnológico preliminar más general como una propuesta de desarrollo de un sistema prototipo que sirva para analizar las hipótesis de investigación.

Capítulo 3: Descripción de la experimentación.

El Capítulo 3 retoma los objetivos, experimentos y metodología introducidos de forma sucinta en el Capítulo 1 para detallar de qué se componen. Los objetivos generales se desglosan en subobjetivos concretos a abordar. Los experimentos se definen en sus distintas fases y las tareas que se realizan en cada una de ellas. La metodología se concreta a las tareas llevadas a cabo en este TFM.

Capítulo 4: Descripción de la aplicación.

El Capítulo 4 muestra una descripción sobre las distintas aplicaciones de software prototipo que desarrolladas durante la realización del TFM tanto a nivel esquemático y usando flujogramas como a nivel más específicamente técnico, indicando las librerías usadas y funciones principales y haciendo referencia al repositorio de código público que alberga los desarrollos.

Capítulo 5: Planificación.

El Capítulo 5 menciona el cronograma y planificación temporal llevada a cabo en el proyecto.

Capítulo 6: Resultados.

El Capítulo 6 detalla los experimentos realizados y los resultados obtenidos de los mismos para el caso del experimento en AULA y para el caso de POSTDATA. Este capítulo incluye a su vez una referencia a futuras líneas de investigación que están planteadas para seguir abordando tras este proyecto.

Capítulo 7: Conclusiones.

El Capítulo 7 sintetiza las conclusiones que se han alcanzado.

Capítulo 8: Referencias.

El Capítulo 8 incluye las referencias de la literatura consultadas para esta memoria.

Capítulo 9: Anexos.

Por último, tras el Capítulo de referencias, en el Capítulo 9 se incluye un anexo con información detallada sobre algunos de los puntos anteriores.

2. Descripción del problema

2.1. Estado del Arte

En este apartado se introduce el SOTA de los ámbitos de investigación principales de la presente memoria. Se comienza en el punto 2.1.1 por el ámbito de los RecSys sensibles al contexto afectivo, especialmente para las aplicaciones de e-learning, y en el 2.1.2. se revisan las ontologías relevantes en este dominio. Tras ello, en el punto 2.1.3. se referencia la literatura sobre tratamiento automático del contenido afectivo en textos poéticos, así como los distintos *corpora* disponibles para la construcción de rasgos afectivos en castellano. En el siguiente punto, el 2.1.4, se analiza el SOTA de las representaciones vectoriales del valor semántico de las palabras gracias al uso de *word embeddings*, mostrando un breve recorrido histórico por la evolución de dichas técnicas hasta llegar a las soluciones más novedosas indicando sus potenciales ventajas. Por último, el punto 2.1.5. introduce desde la literatura algunas técnicas modernas de representación de textos al completo mediante el uso de técnicas de composición.

2.1.1. Sistemas Recomendadores Sensibles al Contexto Afectivo

El primer punto por resolver es investigar el SOTA referente a los propios sistemas de recomendación (RecSys) en entornos de e-learning especialmente cuando tienen en cuenta el contexto y éste incluye la dimensión afectiva de los usuarios. Este tipo de RecSys se enmarca en el ámbito de los Sistemas Adaptativos para Educación (SAE), siendo un caso particular de estos.

Existen distintas formas de acercarse y comprender la realidad de un SAE, como se detalla en (Ricci, Rokach, & Shapira, 2015), pudiendo ser las clasificaciones según:

- Las tareas que soporten:
 - Aquí aparecen aspectos como sugerir recursos novedosos, sugerir actividades, sugerir compañeros (*peers*), predecir el aprendizaje o el rendimiento de un alumno, proponer secuencias de contenidos o caminos a seguir para llegar a metas concretas...
- A nivel operacional:
 - Según su **arquitectura**:
 - Centralizada
 - Distribuida
 - Según su **localización**: *¿dónde está y se ejecuta el algoritmo?*
 - En el lado del usuario (cuando el sistema se almacena localmente, como ocurre con un *spam mail filtering* donde las recomendaciones se producen localmente)

- En la fuente de información (el proveedor o la fuente de información da las recomendaciones a los usuarios. El perfil de usuario se almacena en el lado del proveedor de información)
 - Recomendaciones en el servidor (están en un tercero)
 - Otros
- Según su **modo**: Hace referencia a quien inicia el proceso de la recomendación
- *Push* (activo, se hace un *push* al usuario aunque este no busque la recomendación; p.e. mandando un mail)
 - *Pull* (activo, se producen recomendaciones pero solo se presentan al usuario cuando este lo permite)
 - Modo pasivo (las recomendaciones se producen como parte de las operaciones regulares del sistema; por ejemplo, recomendaciones de productos cuando se navega por una web)

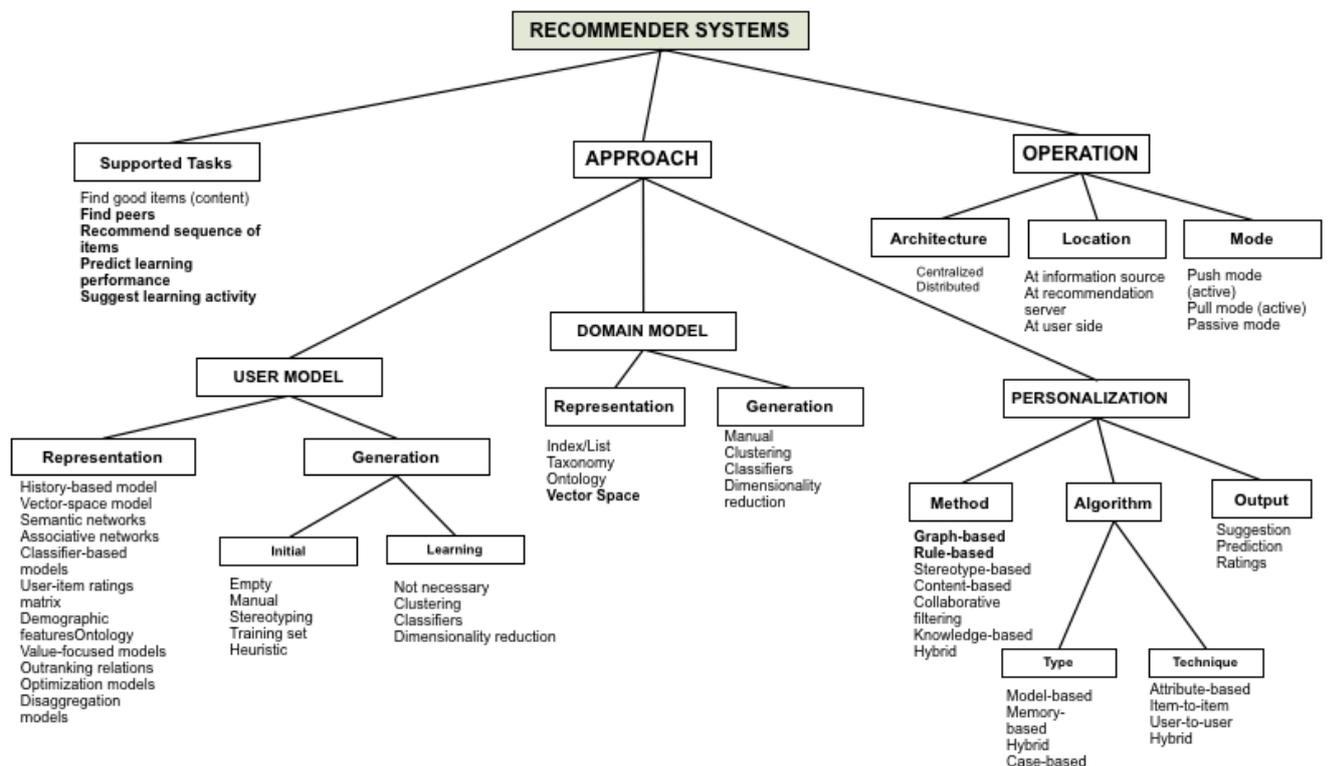


Ilustración 2 Clasificación de los RecSys según distintas categorías, en las que se engloba la clasificación según sus componentes y se listan los parámetros diferenciadores (Ricci, Rokach, & Shapira, 2011)⁹

Ahora bien, existe una clasificación más a nivel de **aproximación** en la que se identifica los SAE según las especificidades de los elementos que lo componen, que aparecen ampliamente descritos en la literatura clásica y en la reciente (Beck, Stern, & Haugsjaa, 2016). Estos sistema se componen de un **modelo de alumno**, que contiene información que hace referencia al propio estudiante (información que puede ser implícita e inferida de, por ejemplo sus comportamientos, o explícita como puede ser una preferencia que éste haya manifestado

⁹ (Ricci et al., 2011). "Recsystem". Web. 10/04/2018. <https://sites.google.com/site/recsystem/>

abiertamente), un **modelo de dominio** que contiene información sobre los propios contenidos del sistema educativo con los que se relacionarán los usuarios del sistema y con el que se representa el conocimiento de una o unas determinadas áreas, un **modelo pedagógico** con el que se lleva a cabo el diseño instruccional marcando las pautas de interacción del alumno con el sistema en aras a poderle guiar de forma eficiente por el mismo, y una **interfaz** que sirve de intermediaria entre el usuario y los contenidos y que podrá ser desarrollada mediante un compendio de distintas tecnologías.

A continuación se procede a describir de forma general el SOTA de los RecSys afectivos, y se indica cómo se tendrían que relacionar entre ellos y cuáles son algunas de las investigaciones que se están llevando a cabo. Por este motivo, en el siguiente apartado se analizarán algunas investigaciones recientes entorno a la computación afectiva en entornos de e-learning y a continuación que caracteriza de forma general a un RecSys contextual. Después de ello se estudiará en base a las propuestas del punto anterior cómo se podría integrar la pedagogía poética dentro de su arquitectura.

Computación afectiva en e-learning

Existen en la literatura varias aplicaciones de computación afectiva a los SAE en general, y a los RecSys para e-learning en particular, que tratan de enriquecer y aumentar el nivel de personalización de la recomendación de los objetos de aprendizaje (*Learning Objects*, LO) utilizando como información contextual adicional el estado afectivo de los estudiantes.

A modo de ilustración, a continuación se describe brevemente el proyecto BIG-AFF (Boticario, Jesus et al., Jul, 2017). Este parte de la importante premisa de que es necesario incluir en los SAE el estado afectivo de los estudiantes de cara a ofrecer recomendaciones pedagógicas que tengan en cuenta esta dimensión. De esta forma los sistemas no se limitan a tener sólo, por ejemplo, un modelo *overlay* donde todo el proceso de aprendizaje se regula en relación al nivel de conocimientos que un estudiante tenga en función de un patrón de referencia teóricamente óptimo.

BIG-AFF parte así de que es posible enriquecer la experiencia formativa de los alumnos aumentando el nivel de personalización, usando dispositivos de bajo coste (sensores) y no intrusivos que sean capaces de capturar datos afectivos que incluyen información cognitiva, sobre el comportamiento y fisiológica, procesando dichos datos mediante etapas de detección de afectos, modelado y *feedback*.

Para BIG-AFF se ha reutilizado un sistema ITS encargado de enseñar aritmética (Arevalillo-Herraez, Arnau, & Marco-Gimenez, 2013) que ya permitía recoger información afectiva a través de la información que reportasen autónomamente los usuarios. Este sistema se ha adaptado para incluir un nivel de estado afectivo más rico, deducido de las interacciones pasadas del usuario con el sistema y de los datos recogidos por el cuestionario SAM (*Self Assessment Manikin*) (Bradley & Lang, 1994). Estos datos sobre el usuario (que no provienen de SAM) se recogen tanto de una plataforma Arduino encargada de monitorizar rasgos fisiológicos, basada en (Santos et al., 2016), como de forma no intrusiva desde las interacciones del usuario con el sistema (Salmeron-Majadas, Santos, & Boticario, 2014).

Con toda la información se crea el modelo de usuario, de tal manera que la personalización del sistema podrá tener en cuenta esos aspectos subjetivos (afectividad) junto con los objetivos (rendimiento académico).

Continuando con el SOTA de los RecSys que tienen en cuenta la dimensión afectiva en los entornos de e-learning cabe mencionar el artículo de (Santos, 2016), que lleva a cabo una revisión bastante detallada de las últimas publicaciones científicas dentro de este ámbito.

El artículo comienza por recordar la importancia de la computación afectiva para enriquecer el perfilado de estudiantes, considerando como parámetros las emociones y el tipo de personalidad. La personalidad es algo más o menos estable, de manera que se suele detectar con *tests* psicológicos (por ejemplo, el modelo Five Factor). Puede influir en los afectos y en el rendimiento académico, por lo que es útil tenerla en cuenta. Ahora bien, las emociones cambian mucho y por eso se deben medir constantemente.

Como emociones se entienden las respuestas rápidas ante un evento percibido (interna o externamente) que van acompañadas de cambios mentales, cambios en el comportamiento o cambios psicológicos.

El artículo menciona también cómo trabajar con afectos en e-learning es aún más difícil, ya que los afectos no cambian tanto durante el aprendizaje, en comparación con otros ámbitos donde sí hay más variación. Además, no hay una única teoría sobre cómo los afectos influyen en el aprendizaje. Aún sin esta teoría cerrada, existen diferentes modelos para e-learning en los que se relacionan los afectos y emociones con factores como el rendimiento académico.

Por lo tanto, aun siendo una etapa temprana para la definición de la teoría sobre qué influencia tienen los afectos en el aprendizaje, es bueno ir proponiendo soluciones que los integren, incluso por contribuir a la propia investigación científica entorno a ello.

En cualquier caso, los pasos que propone como pasos generales para cualquier entorno que quiera incorporar la dimensión afectiva como parámetro adicional son:

- 1) Definir el fenómeno (si se quiere mirar un afecto, emoción, sentimiento...)
- 2) Elegir métodos de estudio, conjuntos de datos, interpretación de datos...
- 3) Preparar e integrar datos
- 4) Derivar un significado de los datos (y, por ejemplo, llevar a cabo algo con la información obtenida)

Tras ello, el artículo detalla 26 trabajos de sistemas de e-learning que tienen en cuenta la dimensión afectiva. En ellos se puede ver que la mayoría de las fuentes consideradas para extraer la información afectiva suelen ser sensores, cámaras... o en algunos casos determinados cuestionarios. Además de ello, las métricas más habituales suelen considerar confusión, aburrimiento, curiosidad, enfado, alegría, valencia, satisfacción o nivel de excitación. Aparecen pocos casos de sistemas que consideren la información léxica y semántica, pero aparece citado un caso (Mavrikis et al., 2010) en el que se usan modelos supervisados para clasificar el estado afectivo de un usuario dentro de determinadas categorías (aburrimiento, confusión, *flow*, frustración y neutralidad) usando como rasgos de entrada determinadas características lingüísticas de los mensajes de respuesta escritos por los alumnos.

Finalmente se mencionan varias cuestiones abiertas como:

- Comprobar si existe vinculación entre el **estilo de aprendizaje** (medido por ejemplo con el índice de *Learning Styles*, ILS) y el **estado afectivo** en un momento dado (a diferencia de la relación con el tipo de personalidad, que aparece mucho más probada).
- Investigar sobre la importancia y contribución de las **emociones en el aprendizaje colaborativo**, de manera que, por ejemplo, se creen grupos de trabajo en base a las emociones que se detecta en los estudiantes.
- **Obtener los afectos de manera no intrusiva**. Se menciona cómo es importante para la automatización y facilidad de detección de afectos que se implementen técnicas que no sean intrusivas para el estudiante.
- Aplicar procesamiento de Big Data para flujos de afectos multimodales. Debido a la potencial gran cantidad de información que se procesa en la computación afectiva (especialmente para los casos de uso de *Internet of Things*, IoT), se propone trabajar con infraestructuras Big Data principalmente hospedadas en *cloud* (o en local si se trabaja en escenarios de *edge computing*).
- Identificar el **refuerzo o soporte afectivo adecuado** para el estudiante teniendo en cuenta la información contextual, y proporcionar así lo que se conoce como feedback afectivo contextualizado.
- Definir de entornos con computación afectiva que respondan a un **aprendizaje ubicuo** y donde además se usen estándares que favorezcan la **interoperabilidad** entre dichos sistemas y otros.
- Vinculación de la información afectiva no sólo con el aprendizaje cognitivo o afectivo sino también con el **aprendizaje motor**, fomentándose así un desarrollo de sistemas de aprendizaje holísticos en los que se tengan en cuenta todas esas dimensiones.
- Profundizar en la importancia de las emociones y la personalidad en el **aprendizaje inclusivo**. Se cita por ejemplo la importancia de detectar los estados afectivos de personas con autismo para regular la relación con ellos. Además de esto, se menciona otro caso de uso para personas con trastorno bipolar, a las cuales puede servir mucho un sistema que detecte cómo se encuentran y les ayude a conocer por sí mismos su estado afectivo.

Siguiendo con la línea del aprendizaje inclusivo, cabe destacar que la contribución positiva de los afectos a la relación con usuarios que poseen algún tipo de diversidad funcional es algo ampliamente estudiado en la literatura, como por ejemplo en (Czernianin, 2016; Parastoo et al., 2016) donde se menciona precisamente la contribución positiva que genera la poesía en pacientes con diversos tipos de discapacidades o trastornos psíquicos.

De esta manera, se detecta una gran oportunidad de investigación en el uso de distintas fuentes de datos que permitan identificar el valor de variables afectivas, para con ello fomentar la construcción de RecSys afectivos que tengan en cuenta el contexto, en línea con la literatura científica. Al margen de su potencial uso, el simple hecho de crear estas herramientas podrá contribuir a la investigación científica sobre las aplicaciones de contextos afectivos en los entornos de e-learning.

Sistemas de recomendación contextuales

Hay que mencionar que los RecSys sensibles al contexto afectivo son en definitiva sistemas que entre todos los elementos contextuales a considerar para adecuar sus criterios de recomendación de contenidos consideran (además de la propia información sobre contenidos y usuarios) el peso afectivo de los contenidos y el estado afectivo de los usuarios en un momento dado. En dichos RecSys (Aggarwal, 2016) se tienen en cuenta dimensiones adicionales con lo que se pasan del esquema clásico de los RecSys de una matriz R de dimensión $m \times n$, donde aparecen los distintos usuarios (m) y los ratings (n) que dan a cada uno de los ítems del sistema, a un esquema de dimensionalidad mayor al generar una matriz F de dimensión $m \times n \times p$ donde se tienen en cuenta dimensiones adicionales como pueden ser espaciales, temporales... De igual forma existen los sistemas que incluyen el *social media tagging*, de forma que construyen una *folksonomy* con la que aparece un campo de tags que definen los usuarios sobre los propios ítems, y que permitirá definir recomendaciones, búsquedas e indexaciones en base a dichos tags.

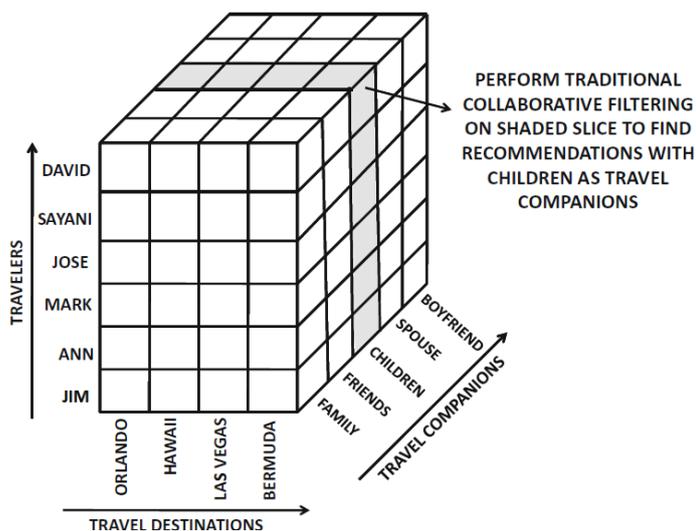


Ilustración 3 Ejemplo de matriz de usuarios e ítems (destinos de viaje) donde las recomendaciones se harán en función de elementos contextuales, como es el caso del "tipo de compañero de viaje". (Aggarwal, 2016)

Aplicado a un RecSys en entornos de e-learning, se podría por ejemplo pensar en un sistema en el que la tercera dimensión representase las posibles ubicaciones de un estudiante en un escenario multidispositivo (en casa, en la escuela, con móvil, pc...). Las otras dos dimensiones serían los alumnos (usuarios) y los ejercicios (ítems) de un mismo curso. Los valores de la matriz serían las calificaciones que han obtenido en los ejercicios que han hecho. Un filtrado colaborativo clásico inferiría qué calificaciones van a obtener los alumnos en los ejercicios que aún no han realizado en función de la similitud en las notas que han sacado respecto al resto de alumnos. Pues bien, con la dimensión adicional de la ubicación se estimaría la nota en función de donde se hagan los ejercicios, algo que puede tener sentido ya que ciertos ejercicios podrían ser más eficiente para realizarse en un entorno más relajado (por ejemplo, en metro de camino a la escuela y con el móvil) mientras que otros requerirían de mayor concentración (por ejemplo, hacerlos en casa con el pc). Así, recomendaría el ejercicio más adecuado (en un modelo pedagógico sencillo podría ser en el que se vaya a sacar más nota) en función de cómo

son las calificaciones de alumnos de un nivel similar y en función de dónde esté realizando el ejercicio.

Esta tercera dimensión del contexto se puede utilizar de formas distintas siguiendo las aproximaciones que se describen a continuación:

- **Métodos de pre-filtrado:** En este caso primero se seleccionaría un contexto determinado y luego se harían las recomendaciones pertinentes siguiendo las líneas habituales mencionadas antes. El único hándicap es que quizá se pueda acabar teniendo muy pocos datos para la dimensión/condición seleccionada.
- **Métodos de post-filtrado:** En este caso se ignoraría inicialmente el contexto y se generarían todas las recomendaciones que posteriormente se filtrarán según la relevancia que tengan en el contexto. Hay varias formas de ver esa relevancia: por ejemplo, se puede calcular un set de recomendaciones para un usuario, pero luego se haría un re-ranking de ellas según la relevancia a su perfil contextual.
- **Modelado de contexto:** En este caso se utiliza directamente el contexto como dimensionalidad extra mediante el uso de factorizaciones de tensores.

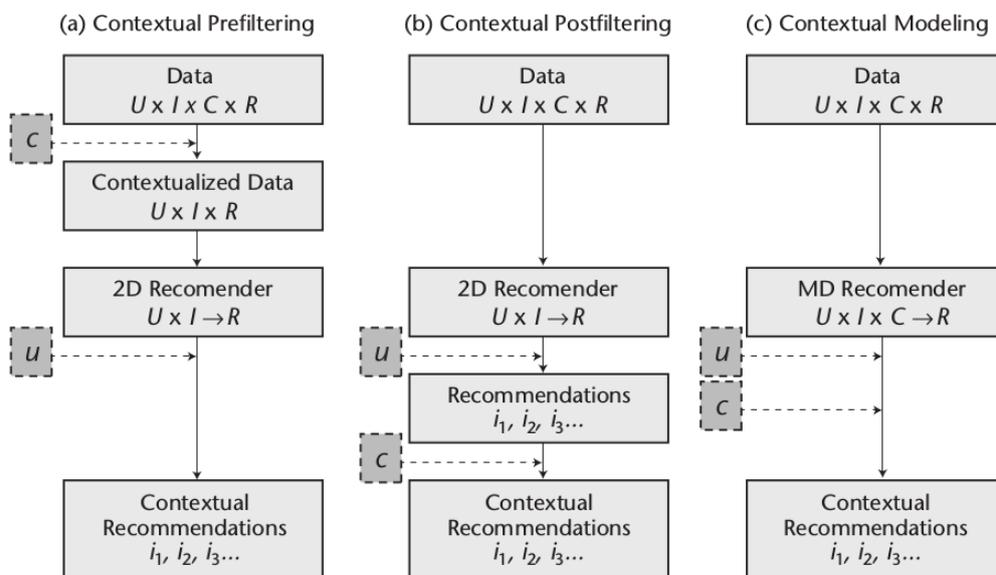


Ilustración 4 Esquema de las distintas técnicas posibles de modelado contextual¹⁰

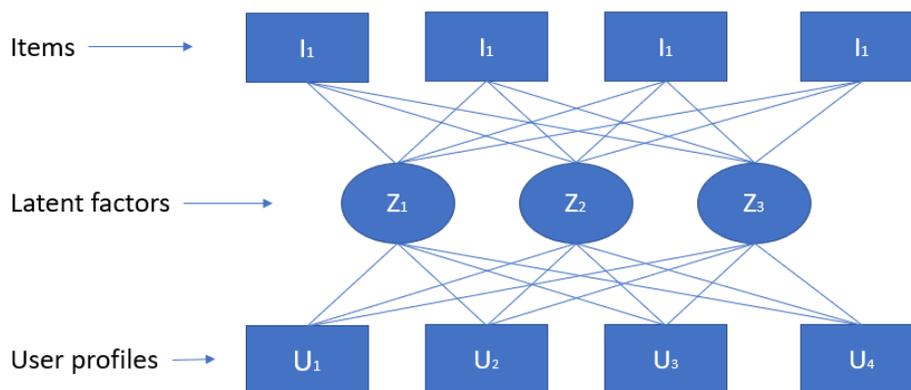
Un problema evidente de estos sistemas es que, al incluir la dimensión del contexto, se aumenta la dimensionalidad del sistema, resultando un sistema donde hay más dispersión de datos, con el detrimento que esto supone a la hora de realizar los cálculos de las recomendaciones, debido a que la mayoría de los posibles valores estarán sin rellenar. Esto, que ya de por sí es un problema en los RecSys, es especialmente problemático en los RecSys para entornos de aprendizaje (*technology enhanced learning, TEL*) debido a que la cantidad de datos de la que se dispone no es demasiado numerosa (los usuarios, los estudiantes, no suelen ser demasiados

¹⁰ Web. 04/04/2018. <https://www.coursera.org/specializations/recommender-systems>

y por ello el conjunto de datos será más reducido). El aumento de dimensionalidad con la inclusión del contexto se puede paliar con el uso de algoritmos de minería de datos para la reducción de dimensionalidad como el *análisis principal de componentes* (PCA), los *autoencoders* (AE), o las *redes neuronales convolucionales* (CNN) (Zhang, X., Zhao, & LeCun, 2015). Estos algoritmos se encargarán bien de hacer una selección de componentes (*feature selection*), como también se hace en las técnicas de *differential context modelling* donde se identifica un subconjunto de las variables contextuales más importantes y se asocia a ellas un peso como parte del proceso de modelado, o bien de generar un menor número de dimensiones de mayor nivel que agrupen a las originales (*feature extraction*).

Un contexto se caracteriza mediante rasgos observables y rasgos no directamente observables, inferidos de ciertos comportamientos. Cuando no se puede observar directamente el contexto existen distintas maneras de inferirlo. Por ejemplo, deduciendo el contexto físico de un usuario de una aplicación educativa en función del uso que haga de la misma.

La ventaja de la inferencia del contexto es que en muchas ocasiones no será necesario ni siquiera predefinir dichas variables pues se podrán modelar de forma automática desde las variables observables. Esto ocurre con los *latent variable context models* donde, con una aproximación generativa, se asume que las interacciones de usuarios implican un pequeño conjunto de estados contextuales latentes que pueden explicar su comportamiento en diferentes puntos a lo largo de sus interacciones. La ventaja de estas propuestas es que la información de las variables contextuales se recoge en unas variables latentes generadas de manera automática.



$$\Pr(U_i, I_j) = \sum_k \Pr(U_i) \times \Pr(Z_k|U_i) \times P_r(I_j|Z_k)$$

Ilustración 5 Esquema de inferencia de estados de un latent variable context model.¹¹

Investigación desarrollada en el ámbito descrito

En este apartado se particulariza la investigación del SOTA a la arquitectura de los RecSys sensibles al contexto afectivo que integran el enfoque de la pedagogía poética. Este

¹¹ Web. 04/04/2018. <https://www.coursera.org/specializations/recommender-systems>

constituye el marco de la propuesta expuesta en la presente memoria, tras detectar una ausencia de propuestas de uso de la pedagogía poética dentro de los RecSys, aun cuando ésta constituye un potente paradigma educativo, de enfoque integral y experiencial.

Efectivamente, se ha comprobado en revisiones recientes de la literatura del campo (Santos, 2016) las oportunidades de la de investigación en torno a los RecSys afectivos para, entre otras cosas, modelar afectivamente al estudiante de forma no intrusiva o contextualizar el feedback y refuerzo motivacional que se le proporciona. Esto, que es útil para cualquier estudiante, se ha probado especialmente importante para el aprendizaje inclusivo. Resulta igualmente de enorme interés la consideración de la dimensión afectiva del aprendizaje bajo el enfoque de la pedagogía poética, de manera que el modelado afectivo sea tanto de alumnos como de contenidos, de cara a personalizar más la relación entre ambos. Para integrar este paradigma los poemas se tratarán como objetos de aprendizaje anotados con metadatos (tono emocional, valencia, familiaridad, competencias conceptuales...).

Este es el motivo por el que en este apartado se propone una primera aproximación para incluir la pedagogía poética dentro de la arquitectura de un RecSys afectivo. De esta forma, con esta propuesta se aborda el objetivo 1.1 de la memoria, indicado en el capítulo anterior.

En la Ilustración 6 se muestra la arquitectura de un RecSys en educación basado en la propuesta de (Nurjanah, 2016) e incorporando el contexto afectivo. Una posibilidad para incluir el valor afectivo de los LO (por ejemplo, poemas) es que el sistema RecSys tenga integrada una herramienta de autoría de cursos con utilidades para la caracterización afectiva de los LO por parte de los docentes. Esta posibilidad de etiquetado afectivo de manera supervisada por los autores es interesante. Sin embargo, en apartados posteriores se exploran propuestas de etiquetado automático del contenido afectivo de los poemas mediante el uso de *corpus* públicos que sirve para modelar el contenido afectivo de un texto en base a las palabras que lo componen. Además, las anotaciones afectivas realizadas por expertos del proyecto de POSTDATA (POSTDATA, 2015) abren distintas vías de investigación de soluciones de modelado automático en base al contenido de los textos (por ejemplo su contenido semántico) en función de su relación con las etiquetas dadas por los expertos. Una forma de abordar esto podría ser mediante técnicas de aprendizaje automático (*ML*, *machine learning*).

Los poemas podrán servir, en línea con el paradigma de la pedagogía poética, como LO en sí o como refuerzo motivacional para otro tipo de cursos (en la Ilustración 7 se asume que los únicos LO son poemas, pero se podría generalizar incluyendo otro tipo de LO). Las ontologías sirven para definir las estructuras tanto de los LO como del modelo de usuario.

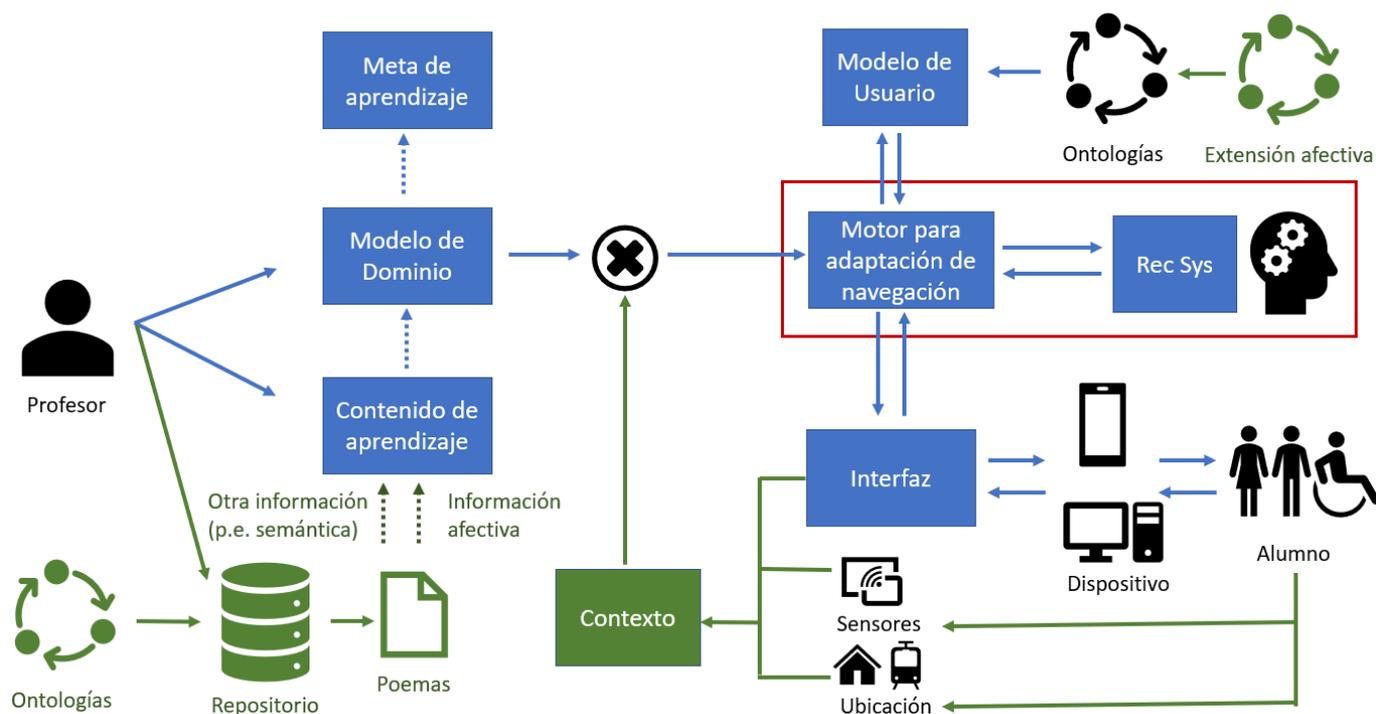


Ilustración 6 Arquitectura de un RecSys sensible al contexto afectivo para el ámbito de e-learning. En verde, los complementos propuestos a (Nurjanah, 2016).

Efectivamente, en la arquitectura que mostrada en la Ilustración 6 aparecen los clásicos modelos de usuario/estudiante y de dominio, pero con dos consideraciones adicionales. Por un lado se contempla la posibilidad de incluir el contexto como dimensión adicional, ya sea adaptando el contenido a la ubicación donde se realiza el curso, al dispositivo, a si el alumno tiene o no diversidad funcional, y a su estado afectivo, el cual se puede monitorizar, por ejemplo, gracias al uso de sensores o incluso gracias a los textos que este escriba. Más aún, la propia interacción del usuario con el sistema sirve para el modelado (Salmeron-Majadas et al., 2014). Toda la información de contexto debe tenerse en cuenta al definir la ontología. Así, como en la presente memoria se contempla la extensión contextual con la dimensión afectiva, en la Ilustración 6 se representa una extensión a las ontologías convencionales que incluye los rasgos afectivos.

Además, en el marco de la pedagogía poética se atribuye una dimensionalidad afectiva a los poemas que permite usarlos como refuerzo motivacional además de cómo LO.

Por lo tanto, hasta donde se ha podido comprobar por medio de los artículos revisados, la literatura científica más reciente sobre RecSys afectivos en entornos de e-learning hace referencia al uso del contexto afectivo para el refuerzo motivacional (Santos et al., 2016) pero no al uso de poesía. Además, no se refieren sistemas que incluyan el enfoque de la pedagogía poética, recomendando poesía como LO personalizado. Así, en el contexto de investigación descrito, el objetivo primario de investigación es extender la arquitectura de un RecSys afectivo sensible al contexto con el fin de contemplar a los poemas como LO y como recursos para el soporte afectivo. Por ello, una de las aportaciones del presente trabajo es desarrollar un prototipo que tenga identificados los contenidos poéticos a nivel semántico y afectivo para que puedan ser recomendados de manera personalizada por un RecSys afectivo.

2.1.2. Ontologías para entornos de E-Learning

En este apartado se muestra la investigación relacionada con el SOTA de ontologías para RecSys en el ámbito de e-learning, en particular cuando éstos tienen en cuenta el contexto afectivo. Este análisis se particularizará para el caso de la presente investigación en la que se consideran los poemas como LO de un RecSys.

La revisión de ontologías de uso común es imprescindible por dos razones: para conocer los datos que habitualmente se registran en los RecSys en entornos de e-learning que tienen en cuenta el contexto afectivo, y para favorecer la integración de la propuesta de esta memoria con dichas ontologías. La integración con las ontologías existentes es fundamental para que los componentes del RecSys propuesto sean interoperables y sus datos estén estandarizados.

En los siguientes subapartados se introduce una reflexión general sobre las ontologías en los entornos de e-learning. En esta memoria se presenta (ver Capítulo 4) una primera propuesta para incluir el contenido afectivo tanto en las ontologías de usuario como en las de LO.

Descripción de ontologías en e-learning

Las ontologías en e-learning se diferencian según la tarea a la que sirvan. En (Al-Yahya, George, & Alfaires, 2015) se refieren las siguientes tareas:

1) Gestión y modelado curricular (*curriculum modeling and management*):

El modelado curricular facilita el acceso y recuperación de información relacionada con los currícula. Por currículum se entiende el camino que sigue un estudiante para alcanzar una determinada meta educativa.

Esto implica la especificación de lo siguiente:

- Contenidos: Teorías, conceptos, habilidades...
- Pedagogías: Métodos de enseñanza, métodos de entrega de información...
- Estructuras: Cómo se organiza el contenido y cómo se secuencia

De esta manera, se ofrece un marco en el que vincular las unidades de aprendizaje (*learning units*, LU) con los resultados obtenidos tras el proceso de aprendizaje (*outcomes*) y con los distintos LO.

Un currículum contempla LO o tareas, y esto es lo que debe reflejar en la ontología. Las tareas incluyen un diseño de programa, una revisión y una evaluación.

Ejemplo de esto es la ontología CURONTO, que permite la especificación y gestión de todo el currículum. Esta ontología se usa para vincular las salidas del proceso de aprendizaje con los LO y con los objetivos generales del programa de estudios, en línea con lo que se espera de una ontología curricular.

La gestión curricular puede involucrar también a información de usuario (esto se aborda en el punto 3 de manera más detallada) que ha de vincularse a la información con la información sobre las competencias adquiridas (en función del currículum abordado). Esto sirve para ir comunicando todo ello al estudiante de manera temprana para guiar su aprendizaje (Raffenne, 2013).

2) Descripción de dominios de aprendizaje:

Se definen a su vez ontologías del dominio de aprendizaje en cuestión desde distintas perspectivas, para enriquecer su descripción y facilitar la recuperación de información. Se distingue entre ontología del dominio (*object-domain ontologies*; historia, geografía, programación...) y ontología de las tareas de aprendizaje (*task-ontologies*; lecciones, actividades, objetos de evaluación, ejercicios, LO, *feedback*)

Las **ontologías de dominio** modelan una materia en particular en función de determinados objetos de conocimiento. Así, cuando se despliegan estas ontologías, se anotan los LO con los conceptos relacionados, según al dominio al que pertenecen.

En la literatura se muestran varios ejemplos de ontologías de dominio. Así, en (Ming-Che, Ding, & Tzone, 2005) se reporta una ontología para el aprendizaje del lenguaje de programación Java. En ella aparecen los conceptos y características del lenguaje, así como la relación entre ellos.

Las **ontologías de tareas de aprendizaje** definen los elementos estructurales de las tareas (*tasks*) de *e-learning*, como actividades o ejercicios. Se centran en modelar sus componentes principales y la relación entre ellos, además de cómo se pueden combinar esos componentes para generar componentes más grandes. En un sistema de *e-learning* pueden encontrarse uno de estos tipos de ontologías, o los dos.

En las ontologías de tareas de aprendizaje se modela el proceso de evaluación, el *feedback*, el diseño pedagógico, la búsqueda y recuperación de información:

- **Evaluación:** Sirve para tener una visión objetiva del progreso del alumno. En (Boticario, Jesús G. & Gaudioso Vázquez, 2003) se indica cómo se puede tener un mapa de conceptos donde se vea el progreso de los estudiantes y el conocimiento que han ido adquiriendo al ver que temáticas han ido cubriendo. Se tiene un mapa conceptual del estudiante que se va creando a medida que este avanza y se convierte en una ontología; este mapa conceptual del estudiante es referenciado frente a uno de referencia que crea el profesor para definir su progreso y evaluarle
Otro ejemplo es la ontología de *e-learning* *OeLE* (Fernández-Breis, Cañavate-Cañavate, Castellanos-Nieves, Valencia-García, & Vivancos-Vicente, 2005) que es una *ontology-based assesment system* que de forma automática marca las respuestas de texto libre de los estudiantes a preguntas de naturaleza conceptual. Las respuestas de los estudiantes se mapean en forma de mapa conceptual usando esa ontología.
- **Feedback:** Es importante dar *feedback* a los estudiantes durante su estudio mediante el uso de recomendaciones. Se tiene un modelo del conocimiento aprendido (dominio) y de cómo se debería aprender el conocimiento (tarea) y con ello se genera el *feedback*. El *feedback*

puede ser tanto durante el aprendizaje como durante la generación del curso con la herramienta de autor.

Existen ejemplos de generación *feedback* a preguntas abiertas en *tests* usando ontologías y anotaciones semánticas. Este *feedback* se genera calculando la similitud semántica entre las anotaciones de la respuesta a un ítem y las anotaciones de la respuesta del estudiante (por ejemplo, a partir de los vectores de palabras de ambas respuestas).

En (Manjarrés Riesco, Santos, & Boticario, 2013) se define una categorización de recomendaciones con dimensión afectiva para incluir *feedback* dentro de un SAE. En dicho artículo, tras un proceso de elicitación con varios expertos en educación, se definen varios escenarios afectivos (por ejemplo, que la fecha de examen de un alumno esté próxima) y vinculados a ellos aparecen unas dimensiones afectivas (emociones como depresión o ansiedad). A su vez, se plantean las recomendaciones pertinentes en cada caso (como recordar al alumno los logros que ha alcanzado para reforzarle afectivamente) y el tono a aplicar (empático, convincente...). Esta propuesta es resultado del proyecto MAMIPEC (Enfoques Multimodales para el Modelado de Aspectos Emocionales en Escenarios de Educación Personalizados e Inclusivos en Contextos Inteligentes) del grupo de investigación aDeNu de la UNED (Santos et al., 2012).

- **Diseño pedagógico:** Un ejemplo es RSPP (*Recommendation System of Pedagogical Patterns*) descrito en (Cobos et al., 2013); este sistema usa ontologías para representar los patrones pedagógicos. Como patrón pedagógico se entiende una descripción detallada de una situación educacional que alcance un fin específico y sea repetible dadas circunstancias similares.

Existen casos de patrones pedagógicos que buscan ayudar a los profesores con el aprendizaje colaborativo (CL).

- **Búsqueda y recuperación:** Para que un usuario pueda navegar y buscar un contenido en concreto es necesario disponer de una ontología que defina los metadatos y semántica de los recursos de aprendizaje. Eso se puede combinar con el perfil del estudiante (*learner profile*) para encontrar la mejor combinación.

Como ejemplo de ello aparece el proyecto CUBER (*Curriculum Builder in the Federated Virtual University of the Europe of Regions*) descrito en (Bernd J. Krämer, 2001). En él se define una ontología con metadatos y elementos semánticos relacionados con distintos cursos. Los estudiantes puedan buscar distintos tipos de cursos de educación superior introduciendo consultas. CUBER basa su propuesta de ontología en el estándar de e-learning LOM.

3) Descripción de los datos del estudiante:

Es necesario tener información sobre el estudiante para personalizar los contenidos según su perfil, así como monitorizarle para poder actualizar esta información. Ejemplos de esta información son los datos de rendimiento académico y sus datos históricos (unidades completadas, tiempo en la aplicación...).

Los datos del perfil del usuario abarcan datos personales, progreso realizado y datos sobre el rendimiento académico. En general, para su conceptualización se suelen usar estándares como

IMS LIP¹² (*Learner Information Package*), y en ocasiones se usan metadatos de varios estándares. IMS LIP es un estándar muy utilizado que define una ontología para la especificación de datos del usuario muy en línea con lo que sería la especificación de un CV clásico. Otro ejemplo de estándar es IEEE PAPI (*Public And Private Information*), ontología que incluye el progreso y rendimiento del estudiante.

Las ontologías de estudiante se han destinado a distintas aplicaciones y casos de uso. Ejemplos de ello son desde la construcción de redes bayesianas con ontologías hasta modelos más estándar que incluyen características básicas del estudiante como información demográfica, conocimiento del dominio, intereses, estilo de aprendizaje y preferencias de interacción.

Se pueden plantear también ontologías en dos partes: una de información general del estudiante y otra de su comportamiento en lo referente al dominio de aprendizaje.

La información sobre el estudiante es un elemento clave de cara a personalizar la interacción con el sistema. La personalización consiste en adecuar los contenidos de estudio en base a las necesidades y preferencias (estilo de aprendizaje, progreso en el aprendizaje, trasfondo individual, requerimientos...).

Un tipo de personalización es la basada en contenidos; se especifican contenidos según la elección previa del estudiante o según su perfil concreto. También existe la personalización de navegación, enlaces, presentación...

4) Descripción de modelos de e-learning:

Se busca también facilitar la interoperabilidad entre servicios de e-learning de forma que puedan comunicarse entre sí, utilizar los LO o los datos unos de otros...

Para que se dé esto se suele recomendar el uso de estándares educativos, que facilitan la accesibilidad, reusabilidad e interoperabilidad entre sistemas de aprendizaje heterogéneos (p.e. IMS consortium¹³). Cabe mencionar también los conversores entre ontologías (*ontology mapping*) o el uso de una ontología común (*common ontology*, CO) que incorpora todos los metadatos conocidos de los estándares de E-Learning.

Ontologías para poesía

Dentro del ámbito de las ontologías específicas para dominios, se ve la necesidad de analizar las propuestas existentes para el dominio poético.

El propio grupo de investigación POSTDATA ha publicado una ontología para el dominio poético (Díez Platas & Bermúdez Sabel,) que es la utilizada para el proyecto DISCO (Ruiz Fabo et al., 2017), que dispone de un repositorio de sonetos en castellano del siglo XV al XIX. Esta ontología contiene clases y propiedades referentes a información relacionada con obras poéticas, sus manifestaciones y otras características. También representa información útil de

¹² *Learner Information Package (LIP)*. Web. 14/09/2019.

https://www.msglobal.org/accessibility/acclipv1p0/imsacclip_infov1p0.html

¹³ *IMS (Instructional Management Systems) Global Consortium*. Web. 14/09/2019. <https://www.msglobal.org/>

dominios relacionados relevantes para las obras poéticas, a través de clases como Persona, Organización, Evento, Lugar y Rol.

La presente investigación toma como ontología de referencia para los LO esta ontología de POSTDATA, desde la especificación propuesta en DISCO.

Ontologías en computación afectiva

Aunque no se va a profundizar en la presente memoria sobre ontologías para la computación afectiva, dado que la investigación se enmarca en los RecSys sensibles al contexto afectivo se ve necesario indicar sucintamente alguna referencia bibliográfica en este ámbito. Así se remite a la reciente revisión bibliográfica de (Abaalkhail, Guthier, Alharthi, & El Saddik, 2018) para profundizar en una propuesta ontológica completa.

Investigación desarrollada en el ámbito descrito

La propuesta de este TFM hace referencia al análisis de ontologías y estándares de representación para etiquetar y representar un corpus de poesía, tras procesarlo con distintas técnicas de NLP para su utilización en un RecSys afectivo. Por este motivo, en lugar de analizar exhaustivamente el SOTA para las ontologías de todos los componentes de un sistema de e-learning, se ha proporcionado en este caso una reflexión genérica más sucinta de un sistema.

Para el soporte de tareas de búsqueda y recuperación de poemas es necesario una propuesta ontológica para definir estos LO incluyendo la información afectiva. Esta ontología se enmarca en las ontologías de dominio. Para ello, como ya se ha comentado, se usa como base la ontología para poesía de POSDATA, y dicha ontología se amplía con atributos que representan los rasgos afectivos. Sería también necesario incluir la información semántica, pero este punto no se aborda en esta memoria.

De igual manera es necesario una ontología para el modelo de dominio pedagógico para poder personalizar la propuesta de LO a cada estudiante. La propuesta ontológica para ello se define en la presente memoria de manera genérica, quedando para futuros trabajos la implementación de ella.

El detalle de las ontologías que se proponen se proporciona en el apartado 3 de este Capítulo (punto 2.3).

2.1.3. Tratamiento automático del contenido afectivo

En el tratamiento automático del contenido afectivo se engloban de forma genérica las distintas técnicas encargadas de inferir estados afectivos o anímicos usando textos como punto de partida.

Una de las tareas clásicas en este ámbito es el análisis de sentimiento. El análisis de sentimiento, también llamado minería de opinión, es el campo de estudio que analiza las opiniones,

sentimientos, evaluaciones, valoraciones, actitudes de las personas, y emociones hacia entidades como productos, servicios, organizaciones, individuos, problemas, eventos, temas y sus atributos (Liu, 2012).

Dicho tratamiento automático del sentimiento se puede llevar a cabo en base a documentos de texto. Así, esto es aplicable al ámbito de la minería de textos donde dichos sentimientos se pueden extraer, por ejemplo, mediante el uso de algoritmos de aprendizaje supervisado en los que las etiquetas sean las distintas categorías de sentimientos (por ejemplo, *alegría*) y los rasgos de entrada sea la información semántica de un texto en concreto para el que se ha anotado dicho sentimiento.

Ahora bien, esta no es la única manera de enfocar el tratamiento automático del contenido afectivo en textos. De manera general, lo que se busca es relacionar una información afectiva global del texto (GAM, *general affective meaning*) con otra información del texto, que puede ser información semántica, información afectiva de las palabras individuales que lo componen, tipo de texto y sus características sintácticas... De igual manera, la relación entre la información del texto y el GAM puede establecerse mediante el uso de distintas funciones, desde funciones de composición sencillas (como agregar directamente la información individual de las palabras del texto) a soluciones más sofisticadas (como el uso de algoritmos de ML supervisado para el caso de que se cuente con etiquetas que reflejen el GAM).

En las Ilustraciones 7 y 8 se muestran dos maneras posibles de extraer etiquetas para el GAM de un texto. De manera general se definen una serie de rasgos que representen al texto en particular y esos rasgos sirven de entrada para una función que dé como salida la/s etiqueta/s GAM correspondientes. En el proceso se puede contar con etiquetas GAM manuales (relación supervisada) o no (relación no supervisada). Si se cuenta con etiquetas de partida se puede usar dicha información para verificar el uso de determinados rasgos para obtener el GAM de manera general.

Relación supervisada

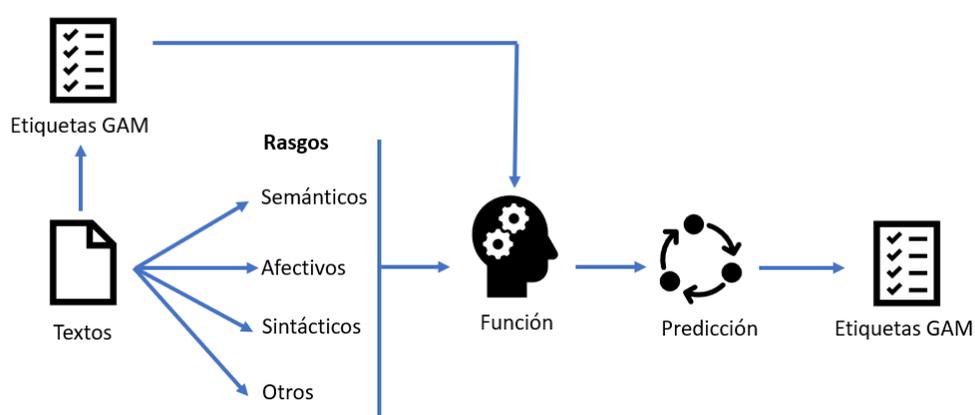


Ilustración 7 Esquema de relación supervisada

Relación no supervisada

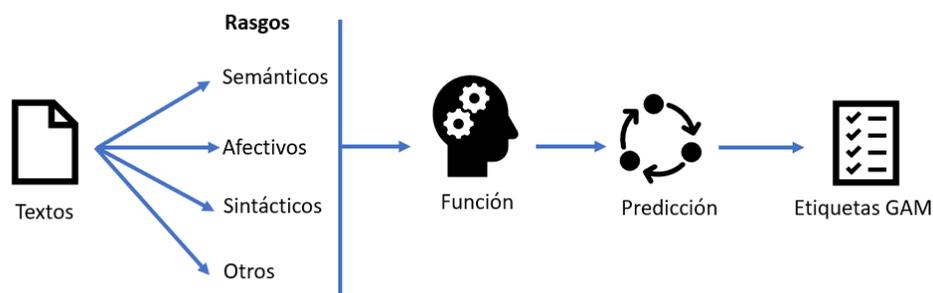


Ilustración 8 Esquema de relación no supervisada.

A. Tratamiento automático del contenido afectivo en textos poéticos

En este apartado se particulariza la reflexión introducida previamente para textos en general, para el caso concreto de poemas. En primer lugar se menciona literatura reciente sobre extracción de rasgos de distinto tipo para obtener las etiquetas GAM generales, destacando aspectos que se deben tener en cuenta para ello. En un segundo apartado se introducen estudios psicológicos que prueban la influencia de la poesía sobre el estado afectivo de las personas, justificando que se puedan representar los poemas con un GAM general en base a la información contenida en ellos.

Rasgos para obtener el GAM en textos poéticos

Como se ha indicado previamente, los textos en general, y en particular los textos poéticos, contienen una información afectiva que es posible extraer utilizando distintas técnicas (a nivel de palabras individuales, análisis de sentimiento...). Es importante cuantificar este valor afectivo de los textos poéticos de cara a operativizar la pedagogía poética. Así, la idea es detectar qué elementos poéticos son especialmente relevantes para calcular a través de ellos el valor afectivo de los poemas. En los artículos mostrados a continuación se analizan distintas maneras de extraer y cuantificar aspectos afectivos de textos poéticos.

Un ejemplo reciente de tratamiento del contenido afectivo en textos poéticos se describe en el artículo (Ullrich et al., 2016) para el alemán. En este artículo se explora cómo influyen los rasgos de un texto poético (a nivel subléxico, léxico e inter-léxico) en el GAM que se percibe. Así, este artículo muestra qué rasgos afectivos son relevantes para un texto en función de cómo de relacionados estén con el GAM, y cómo calcularlos. Para calcular los rasgos de un texto

poético se usa la base de datos BAWL (*Berlin Affective Word List*), descrita en (Võ et al., 2009) en su revisión, que proporciona los valores para las distintas palabras en alemán.

Para anotar el GAM en los sonetos, el artículo realiza un estudio sobre un grupo de lectores que puntúan un subconjunto de 57 poemas del autor alemán H.M. Enzenberg dentro de una escala afectiva que incluye dar una puntuación a las siguientes variables:

1. Puntuación en una escala de 7 de la valencia o nivel de afecto positivo o negativo del texto, donde -3 sería muy negativo, 0 neutral y 3 muy positivo (*valence*).
2. Puntuación en una escala de 5 del nivel de excitación del texto del poema, que pasa de textos que inspiran tranquilidad o calma a otros que buscan motivar o que son más excitantes, donde 1 es muy tranquilo y 5 muy excitante (*arousal*).
3. Puntuación en una escala de 1 a 5 del nivel de amigabilidad, donde 1 indica que el texto no es nada amigable y 5 que lo es mucho (*friendliness*).
4. Puntuación en una escala de 1 a 5 del nivel de tristeza, donde 1 sería que el texto no es nada triste y 5 que lo es mucho (*sadness*).
5. Puntuación de 1 a 5 del nivel de malevolencia, con 5 siendo mucho el nivel (*spitfulness*).
6. Puntuación de 1 a 5 indicando si ha gustado mucho o poco el poema, siendo 5 mucho (*liking*).
7. Puntuación de 1 a 5 indicando el nivel de poeticidad, donde 5 indicaría que el poema es muy poético y 1 que lo es poco (*poeticity*).
8. Puntuación de 1 a 5 indicando el nivel de onomatopeya (nivel que cuantifica el uso de este recurso literario). 5 indicaría mucha presencia de onomatopeyas (*onomatopeias*).

Esas anotaciones realizadas por los usuarios para expresar el GAM sirven como variables dependientes desde las que se analizan la correlación con las variables extraídas de los textos que cuantifican distintos parámetros afectivos. De este modo se identifican las variables de los textos que influyen más y están más relacionadas con las variables dependientes que registran la percepción de los usuarios tanto del texto como de cómo les ha parecido su lectura (punto 6).

Los rasgos extraídos de los textos se definen a nivel subléxico, léxico e inter-léxico. El nivel léxico captura la media de valores de *valencia* y nivel de excitación de las palabras presentes en el texto, el nivel inter-léxico cuantifica picos, rangos y cambios dentro del contenido afectivo léxico, y el nivel subléxico mide el tono afectivo base de acuerdo con la correspondencia a ese nivel subléxico. El análisis arroja resultados en los distintos niveles.

En el nivel **léxico** se pueden capturar métricas relacionadas con la *valencia* y el *nivel de excitación* de las palabras individuales del texto usando la información de BAWL. Gracias al *nivel de excitación* y *valencia* se pueden definir variables o predictores como la *media de nivel de excitación* y *valencia* en los textos. También se define un factor denominado *factor sigma*

dividiendo la media del nivel de excitación de las palabras del poema entre el nivel de excitación del corpus general (1 - referencia de neutralidad) para ver cómo se desvía. Esa media de 1 se multiplica por la raíz cuadrada del número de palabras del poema que están en el corpus.

Aparece también en el GAM la influencia de los efectos **inter-léxicos**, aunque no es trivial de cuantificar. Una misma frase puede estar, por ejemplo, dominada por una palabra fuertemente afectiva que anule las demás. También puede ocurrir que una composición concreta de dos afectos consecutivos lleve a un afecto distinto (por ejemplo, ¡*Alégrate, morirás!*). Así, aquí se analizan el nivel de excitación y valencia de las palabras que están en el texto. Se toman como predictores la desviación de esas métricas para cada poema y el *intervalo* (*span*). El *intervalo* es la diferencia entre los valores máximos y mínimos de esas métricas dentro del poema, algo muy relevante en determinadas obras, como las de suspense que, precisamente, juegan con esos saltos afectivos. Se toman además como rasgos esos máximos y mínimos, y se ve también como rasgo el valor de la correlación entre el vector de posiciones (de principio a final del poema) y el vector de las correspondientes métricas.

En última instancia se define la influencia **sub-léxica** en el GAM. Existen estudios sobre cómo se relaciona el nivel fonético de las palabras con el significado afectivo de un poema (por ejemplo, las palabras *oclusivas* suelen aparecer más en poemas alegres mientras que las *nasales* aparecen más en los tristes). Además, ciertos clústeres fonológicos aparecen más habitualmente con determinados significados afectivos (por ejemplo, suele ser habitual encontrar alto nivel de excitación junto con valencias negativas).

El estudio presentado en el artículo omite, por tanto, toda influencia afectiva super-léxica por su naturaleza altamente subjetiva, como la influencia en el estado afectivo derivada del conocimiento del lector sobre el poema o autor, su capacidad de comprensión, y otras variables de personalidad... Tampoco se tienen en cuenta características sintácticas o estructurales de los poemas. Solo se consideran los fonemas individuales, las palabras y relaciones básicas entre las mismas.

Todas estas especificaciones dan lugar a la definición de 55 rasgos afectivos (entre los 3 niveles antes descritos) de los que los estudios muestran que aproximadamente el 50% de la varianza explicada para las variables de salida de las puntuaciones afectivas se puede alcanzar con los rasgos del nivel léxico, de manera que estas variables son especialmente interesantes, seguidas de las del nivel inter-léxico; junto con los otros dos niveles se alcanza un 75% de varianza explicada aproximadamente. En conclusión, los mejores predictores serían la *media de la valencia* y la *media del nivel de excitación* de entre las distintas palabras.

Eso sí, considerar sólo estos dos rasgos supondría que es irrelevante el orden de las palabras en el texto para el impacto afectivo cuando realmente no es así; el orden importa, y que se experimenten *crescendos* o *decrecendos* afectivos es algo fundamental, de manera que el intervalo (*span*) del nivel de excitación es otro aspecto clave a considerar.

Junto con lo anterior aparece la consideración de cómo evolucionan el *nivel de excitación* y la *valencia* a medida que se da una sucesión de palabras, ya que generalmente se perciben los poemas como más **tristes** cuando la valencia de las palabras va siendo cada vez menor (más negativa) y cuando el nivel de excitación al final es menor, y se perciben los poemas como más

amigables cuando la valencia de las últimas palabras del texto es más positiva. De esta manera, es importante considerar el coeficiente de correlación entre el vector de afectividad (nivel de excitación /valencia) de las palabras individuales con el vector de posiciones de éstas en el texto.

Un artículo que también arroja luz sobre el ámbito poético y que puede ayudar a plantear una solución de cara a la computación afectiva es (Gloria Schultz Eastman, 2015). En este artículo se abordan dos cuestiones relevantes al respecto.

Por un lado, se recuerda cómo la poesía usa ampliamente las metáforas y el lenguaje figurativo (palabras abiertas a muchos significados e interpretaciones). Esto hace que la extracción de información afectiva no sea siempre tan evidente como simplemente asignar a cada palabra un valor contenido en un repositorio y luego componer todos los valores individuales. Las metáforas se interpretan en gran parte también desde la subjetividad del lector, su experiencia personal... por lo que no es trivial e inmediato incorporar toda la información posible. Esto es también aplicable al ámbito semántico (es decir, a representar vectorialmente una palabra individual en base a una información fija de un repositorio).

Por otro lado se menciona también que la comprensión de las palabras de un texto poético no se debe hacer sólo en base al propio texto sino que un poema de un autor determinado se puede comprender en mayor profundidad si se compara con otros poemas de ese autor o con poemas de otros autores. Se mencionan así dos tipos de transferencias de aprendizaje poético:

- *Near transfer*: aprender a comprender un tipo de poema de un autor para así comprender otros textos de ese autor.
- *Far transfer*: aprender a comprender poemas de un autor y que ese aprendizaje sirva para comprender cualquier otro tipo de documento.

Como lo que se buscaría es un aprendizaje que sea *Far transfer*, y que por tanto no se caiga en una comprensión sólo de un nicho muy específico, es necesario un aprendizaje en varios contextos diferentes, y con textos diferentes, de manera que se obtenga un conocimiento mucho más sintético y generalizable. Por este motivo los rasgos relevantes cuya excitación se pueda automatizar deben abarcar cuantos más textos mejor, y no hacerse sobre subconjuntos demasiado particulares.

Como último comentario se menciona el trabajo de (Obermeier et al., 2013), donde se plantea un estudio de la influencia de la poesía en los afectos gracias a determinados elementos estéticos y emocionales como son la métrica del poema y su rima.

Así, la hipótesis de partida es que la métrica y la rima tienen impacto en la percepción estética, en la involucración emocional y en la valencia. Para comprobar esto se analiza la influencia de la métrica y la rima en la percepción estética y emocional de la poesía, así como su interacción con el léxico, usando como referencias para el estudio las estrofas de los poemas.

Aunque podrían también influir la forma léxica y los rasgos estructurales (por ejemplo, forma poética) en la respuesta emocional provocada. Mencionan que existe mucha literatura sobre

cómo influye la valencia de las palabras en la percepción del texto, pero hay menos información sobre cómo influyen otros rasgos formales.

Aristóteles ya habla de cómo la *dulzura* de la métrica y la armonía del sonido en una tragedia hacen que, incluso teniendo en cuenta el drama de ésta, experimentemos placer (*hedoné*) al contemplarla gracias a que existe una belleza en la armonía estructural de la obra. La métrica y la rima son distintos rasgos pero ambas sirven para estructurar la poesía: representan patrones de recurrencia o similitud (por ejemplo, ambas sirven para estructurar la poesía en el tiempo).

- Rima:
 - Representa palabras fonológicamente idénticas desde la última vocal acentuada hasta el final de la palabra.
 - Influye en la memoria y comprensión de las palabras. Contribuye también a la organización y estructuración mental léxico-semántica.
 - Estructura un poema a nivel de verso marcando el final de versos individuales o la cesura dentro de un mismo verso separándolo en hemistiquios.
 - Las rimas finales de la poesía métrica estructuran temporalmente tanto los versos individuales como su configuración dentro del conjunto de versos que forman una unidad como, por ejemplo, una estrofa.

- Acento (métrica):
 - Hace referencia a la percepción de alternancia de sílabas acentuadas (fuertes) o no acentuadas (débiles).
 - El acento puede atraer la atención de un lector ante un estímulo concreto.
 - La rima estructura el poema de forma simétrica a gran escala; en cambio el acento da un marcado asimétrico a pequeña escala.
 - Las estructuras métricas regulares son más fáciles de recordar que las irregulares, y por ello ayudan al proceso cognitivo y colaboran para la comprensión semántica y sintáctica.
 - Diversos estudios afirman que requiere menos esfuerzo mental la lectura de poemas con una estructura métrica definida que los que carecen de ella.

Con todo ello, la métrica y la rima contribuyen a la facilidad de comprensión y retención, requieren menor esfuerzo intelectual, causan un mayor impacto estético, y una mayor capacidad para envolver emocionalmente. Esto es lo que el estudio busca probar empíricamente.

En la investigación se analiza la posible correlación entre las propiedades estructurales de un poema (rima y acento) y las respuestas estéticas y emocionales que suscitan. Para ello se plantea una experiencia con 60 adultos y poemas en alemán (100 poemas del s. XIX y s. XX), con estrofas de 4 versos en las que hay conjuntos con diferencias léxicas (palabras reales vs pseudopalabras, manteniendo las vocales pero cambiando alguna consonante en esas pseudopalabras garantizando que sean pronunciables), rima vs no rima, acento vs no acento.

Los sujetos escuchaban audios de las poesías y puntuaban los poemas en 4 escalas:

- *Liking*: apreciación estética
- *Intensity*: potencia de la respuesta emocional

- *Perceived emotion*: emoción que expresa la estrofa
- *Felt emotion*: emoción experimentada por los sujetos (puede ser distinta de la anterior)

Según la hipótesis la métrica y la rima deberían influir en el rating estético y emocional (todos los ratings deberían ser mayores para poemas con rima y para poemas con acento/métrica). La rima es mucho más marcada al final de los poemas, y por ello cabe esperar que sus resultados sean más robustos que los del acento/métrica.

Eso sí, se sabe que el léxico influye en la respuesta (sobre todo en la valencia), por lo que éste afectará a los resultados. Afectará principalmente a la respuesta emocional, más que al gusto estético que se sabe que es más independiente del contenido semántico en sí.

Los resultados obtenidos confirman las hipótesis de partida:

- *Liking*: Mejores ratings estéticos para poesía con métrica y para estrofas con rima.
- *Intensity rating*: Tanto para poesía con o sin métrica, con o sin rima, se trata de ver cómo el léxico influyen en esta puntuación. Mayor respuesta con las estrofas de palabras reales.
- *Perceived emotion*: Influencia de léxico, métrica y rima (sobre todo de las dos últimas). Mejor puntuación para estrofas con pseudopalabras si no tienen métrica que cuando sí la tienen. Esta última diferencia no aparece en los casos de palabras reales. Para la emoción percibida afectan sobre todo los casos de pseudo palabras y, curiosamente, afectan positivamente en los casos de pseudopalabras cuando no hay métrica (de cara a detectar la emoción que quiere trasladar el texto).
- *Felt emotion*: principalmente afecta la rima, y hay una interacción triple entre léxico-métrica-rima. Cuando hay rima la emoción percibida es más fuerte, y esto es aún más intenso en los casos de pseudopalabras.

Estudios psicológicos sobre la influencia de la poesía en el estado afectivo

Cabe mencionar también los estudios que analizan y sirven de garantía para detectar y justificar la relación entre poesía y afectos mencionada previamente. En (Wassiliwizky et al., 2017) se plantea una serie de estudios psico-fisiológicos sobre cómo la poesía (recitada) puede actuar como un poderoso estímulo que arranque respuestas emocionales potentes (escalofríos, alegría...). Esto tiene similitudes con cómo se producen las respuestas con la música, pero también hay diferencias significativas. En el caso de la poesía puede ocurrir que un pico de placer estético ocurra junto con marcadores fisiológicos de afectos negativos (placer estético en un poema, por ejemplo, de dolor o tristeza).

La distribución de cómo se experimentan los escalofríos a lo largo de un poema ayuda a comprender aspectos sobre los principios de composición poéticos.

El artículo muestra a la poesía como elemento presente desde el principio de la escritura literaria, mostrando la fuerte conexión entre cognición y emoción. Dicho artículo recalca que existe aún mucho desconocimiento sobre cómo funcionan los mecanismos neuronales respecto a la poesía (a diferencia de otras composiciones, como la música) y que no se han investigado

aún el impacto emocional del lenguaje poético y el placer estético asociado (aspectos fundamentales para enganchar con la poesía). Por este motivo investiga entonces:

- Si la poesía puede causar respuestas emocionales fuertes
- Qué estructuras del cerebro son las responsables de estas respuestas y si son similares a las de la música
- Qué rasgos en concreto de la poesía son las responsables de estas respuestas

Para analizarlo se usa un grupo de control donde los participantes escuchan audios de poesía y deben elegir poemas que les causen sentimientos de escalofríos o similares.

Durante este proceso se plantea una serie de mediciones como: cambios en la piel, con una cámara... Los participantes daban a un botón con su mano dominante cuando sentían escalofríos. Con ello se busca analizar las variaciones fisiológicas, comparar las respuestas de escalofríos medidas frente a las que reportan los usuarios, y ver la relación entre las variables medidas y las respuestas. Los resultados obtenidos confirman que la poesía puede desencadenar respuestas intensas afectivas a nivel fisiológico. Esto ocurre no sólo con las obras que son conocidas por una persona sino también ocurre con las obras que no les son familiares.

Con los datos recogidos se analiza la correlación de los escalofríos (indicados por los momentos en los que el usuario presiona el botón) con el efecto de 'piel de gallina' medido por los sensores (*goosecam*). Entre lo que se detecta se ve cómo la tristeza es atributo que produce respuestas más fuertes, y cómo se detecta una *pre-relajación* (*prechill*) de forma general unos segundos antes de dar al botón. Algunas investigaciones dicen que las experiencias más intensas suelen ser las vinculadas a la anticipación, más que a la propia degustación posterior. Sin embargo, en la música es al revés: la mayor intensidad es la *relajación* (*chill*) durante la experiencia, no antes. Interesa ver en cuál de las dos situaciones se sitúa la poesía.

En poesía, esa *pre-relajación* aparece como anticipación de lo que va a ocurrir y depende tanto del contenido semántico como del fonológico y de aspectos estructurales como rimas y métrica; estas estructuras ayudan a ver cuándo acabará una frase, una parte del poema... o incluso, para composiciones muy estructurales como sonetos y haikus, cuando acabará el poema. Esta información la descubren incluso los no conocedores de poesía, de manera que en ellos también se mide esa *pre-relajación* (algo lógico ya que la *relajación* se suele experimentar de forma concentrada al final de las unidades textuales, de manera que el cerebro es consciente de ello y genera por tanto dicho afecto). Los experimentos muestran cómo los lectores recuerdan bien los pasajes que han producido el efecto de *relajación*. Aunque experimentan más *relajación* los lectores habituales de poesía, también lo experimenta un gran número (77%) de los lectores no conocedores de poesía. En cuanto a los *escalofríos*, *estos* son un elemento externo, visible (pelos erizados...) a diferencia de la *relajación* que es un elemento íntimo y subjetivo, difícil de comprobar desde fuera. Esto posiblemente guarda relación con el hecho de que dichas experiencias responden de manera opuesta ante situaciones de repetición de exposición frente al foco afectivo.

Los resultados obtenidos son coherentes con la actividad del músculo corrugado en momentos de *relajación*, lo que los vincula con afectos negativos (tristeza, disgusto, miedo o ira), mucho más consistentes que los del músculo cigomático (que se vincula a afectos positivos, como la

alegría). A su vez se detecta y confirma cómo la *relajación* que se experimenta con la poesía está vinculada al circuito neuronal responsable del *sistema de recompensa (reward system)* (parte de los elementos del circuito responsable aparecen en la *pre-relajación* y parte en la *relajación*), de manera que se puede hablar, como hacen ciertos autores, de *búsqueda de la experiencia estética*. Hay así una mezcla entre el placer estético y los afectos negativos, en línea con la experiencia de la propia catarsis griega.

Se da también una búsqueda de placer estético gracias a afectos negativos, ya que estos producen emociones fuertes y que perduran en el tiempo (eso sí, aparecen en casos como estos, de poesía, donde se enmarcan en una situación de fondo de seguridad del lector. El lector es siempre consciente de que es una ficción y de que puede retirarse de ella en cualquier momento). Se comprueba cómo la *relajación* aparece sobre todo al final de los poemas (que son el culmen del proceso afectivo por el que dirige el poema) o en el cierre de alguno de los diferentes niveles del poema (de una línea, de una estrofa...).

Hay un aumento de la actividad del núcleo *accumbens* (NAcc) y por ello de la excitación psicológica cuando se acerca el final de una línea o de un poema, pudiéndose interpretar como la expectación ante el cumplimiento o no cumplimiento de las predicciones del lector. La expectación y la intensidad afectiva es mayor antes del culmen del poema que antes del culmen de cualquier otra etapa, remarcando cómo es más impactante todo el conjunto de la obra que cualquiera de sus partes por separado, viéndose así cómo es fundamental tratar a los poemas siempre como una unidad global textual y no de forma fragmentada.

Conclusión

A modo de conclusión, los artículos mostrados justifican, gracias a mediciones a nivel fisiológico, que existe una relación entre el contenido afectivo de un poema y su respuesta afectiva en el ser humano. A su vez, en esta revisión de artículos se ha observado cómo existen diversas maneras de extraer rasgos relevantes de un texto poético para definir el GAM esperado. Dichos rasgos pueden ser afectivos, pero también a nivel de estructura del soneto, sintácticos o semánticos.

Los rasgos que se han analizado más en detalle son los relacionados directamente con el plano afectivo. Estos rasgos pueden ser desde cuantificaciones a nivel léxico de la media de valencia/nivel de excitación de un determinado texto hasta el análisis a nivel inter-léxico, que considera además la posición de las palabras en el texto o la diferencia máxima entre nivel de excitación/valencia que se da dentro de él.

Estos rasgos poéticos condicionarían el sentimiento que se genera en el lector según qué valores tuviesen de valencia o nivel de excitación y se podrían completar con la importancia que se ha visto que tienen otros elementos poéticos como la rima o la métrica en la percepción estética que los lectores tienen de la obra poética. La recomendación de un poema debería pues tener en cuenta qué impacto afectivo se quiere conseguir para elegir el poema más adecuado tanto a nivel del valor de los rasgos inferidos como a nivel de la métrica o rima que este tenga.

La importancia del análisis de la literatura planteado se debe a lo siguiente: partiendo de la base de que los poemas tienen un fuerte contenido afectivo y que precisamente por eso podrán contribuir a un diálogo afectivo con el estudiante, ¿cuál sería el mejor poema que se debería presentar a un alumno de entre varios posibles? Por ejemplo, si se quiere educar al estudiante en el valor de la “*Justicia*” (usando el esquema de valores de Scheler), ¿qué poema se recomendaría de entre varios poemas que traten esta temática?.

La respuesta a ello no es inmediata ni fija, sino que es una respuesta viva y dinámica que dependerá del estado afectivo del estudiante en cada momento, y del criterio pedagógico establecido. Por este motivo, y para plantear una solución agnóstica aplicable a cualquier escenario posible, será necesario conocer ciertos parámetros afectivos tanto del estudiante como del poema para adaptar la propuesta en cada momento y contexto.

Las relaciones que existen en el plano afectivo entre una persona y un poema han de cuantificarse matemáticamente de cara a operativizar un sistema que tuviese en cuenta dicha información afectiva para que de manera automática recomendar a un usuario el mejor y más personalizado poema para la casuística en cuestión que se quisiese tratar. Este es el motivo por el que, dentro del ámbito de este TFM, se han investigado distintas soluciones y técnicas para cuantificar numéricamente distintos rasgos afectivos a partir de una composición de texto en castellano.

Hay una gran carencia en la literatura en cuanto al tratamiento automático de poesía en castellano, así que una propuesta de investigación es analizar si los rasgos afectivos destacados como relevantes en textos poéticos de otros idiomas (Ullrich et al., 2016), y que se han construido gracias a repositorios con los valores individuales a nivel de las distintas palabras, son igualmente útiles para poesía en castellano. Por este motivo se calcularán para un corpus poético acotado y se plantea validar tanto con usuarios finales como con expertos de dominio si dicha información es verdaderamente útil. Además se comprueba si la valoración de los rasgos de dichos para un soneto a partir de los valores individuales de las distintas palabras que lo componen se corresponde realmente con la valoración según el criterio de un grupo de expertos.

B. Corpora para la construcción de rasgos afectivos en castellano

Como se ha mencionado en la conclusión del punto anterior, de entre todas las aproximaciones para modelar el GAM de un texto se plantea comprobar la validez de los rasgos propuestos en (Ullrich et al., 2016) para el caso del castellano. A modo de recordatorio, este artículo parte de valores afectivos individuales de palabras en alemán, obtenidas de BAWL y con ello compone unos rasgos generales para el texto. BAWL es una base de datos donde las palabras de un idioma (alemán en ese caso) están expresadas según su valor afectivo a través de valores de valencia y nivel de excitación inferidos de experimentos con varios grupos de control.

Ahora bien, para poder llevar el estudio a la poesía en castellano será necesario utilizar un equivalente a BAWL para este idioma. A continuación se indican distintos *corpora* que se han utilizado en la presente experimentación. En ellos se tiene un diccionario con distintas palabras

en castellano y asociados a ellas aparecen distintos parámetros relacionados con la dimensión afectiva y que servirán para componer los rasgos mencionados en (Ullrich et al., 2016).

En (Ferré, Guasch, Martínez-García, Fraga, & Hinojosa, 2017) se anotan 2267 palabras en castellano (junto con su traducción al inglés) con los siguientes campos¹⁴:

- *Spanish_Word*: Palabra en castellano.
- *English_Translation*: Traducción de dicha palabra al inglés.
- *Hap_Mean*: Valor medio asociado al sentimiento de *felicidad* gracias al conjunto de usuarios.
- *Hap_SD*: Desviación típica asociada al sentimiento *felicidad* gracias al conjunto de usuarios.
- *Ang_Mean*: Ídem para el sentimiento de *enfado*
- *Ang_SD*: Ídem para el sentimiento de *enfado*.
- *Sad_Mean*: Ídem para el sentimiento de *tristeza*.
- *Sad_SD*: Ídem para el sentimiento de *tristeza*.
- *Fear_Mean*: Ídem para el sentimiento de *miedo*.
- *Fear_SD*: Ídem para el sentimiento de *miedo*.
- *Disg_Mean*: Ídem para el sentimiento de *asco*.
- *Disg_SD*: Ídem para el sentimiento de *asco*.
- *N*: Número de sujetos usados en la muestra

En (Guasch, Ferré, & Fraga, 2016) se anotan 1400 palabras en castellano con los siguientes campos:

- *ID*: Mero campo auto incremental.
- *Word*: Palabra en castellano.
- *English Trans.*: Traducción de las palabras al inglés.
- *POS*: Etiquetado PoS (*Part of Speech*) para esa palabra.
- *VAL_M*: Valor medio de la valencia para el conjunto de sujetos.
- *VAL_SD*: Desviación típica de la valencia para el conjunto de sujetos.
- *VAL_N*: Número de sujetos usados para obtener los valores de valencia.
- *ARO_M*: Ídem para nivel de excitación.
- *ARO_SD*: Ídem para nivel de excitación.
- *ARO_N*: Ídem para nivel de excitación.
- *CON_N*: Ídem para *concreteness*¹⁵.
- *CON_SD*: Ídem para *concreteness*.
- *CON_N*: Ídem para *concreteness*.
- *IMA_M*: Ídem para *imageability*¹⁶.
- *IMA_SD*: Ídem para *imageability*.
- *IMA_N*: Ídem para *imageability*.
- *AVA_M*: Ídem para *context availability*¹⁷.

¹⁴ Todos los campos mencionados se miden con una escala de 1 a 5.

¹⁵ *Concreteness* se define como el grado de especificidad de la palabra, valiendo 1 cuando la palabra es muy abstracta y 7 cuando es muy concreta. Palabras como 'objeto' son más abstractas que otras como 'mesa'

¹⁶ *Imageability* se define como la facilidad o dificultad de construir una imagen mental asociada a dicha palabra, valiendo 1 cuando la palabra es muy difícil de imaginar y 7 cuando es muy fácil. Es más fácil imaginarse algo con palabras como 'bandera' que con otras como 'caridad'.

¹⁷ *Context availability* se define como la facilidad o dificultad en asociar dicha palabra con un contexto en el que pudiese aparecer, valiendo 1 cuando la palabra es muy difícil de asociar a un contexto y 7 cuando es muy

- *AVA_SD*: Ídem para *context availability*.
- *AVA_N*: Ídem para *context availability*.
- *FAM_M*: Ídem para *familiarity*¹⁸.
- *FAM_SD*: Ídem para *familiarity*.
- *FAM_N*: Ídem para *familiarity*.

En (Stadthagen-Gonzalez, Imbault, Pérez Sánchez, & Brysbaert, 2017) se recogen los siguientes campos para 14031 palabras¹⁹:

- *Word*: Palabra del diccionario.
- *ValenceMean*: Valor medio de la valencia inferido de los distintos sujetos.
- *ArousalMean*: Valor medio del nivel de excitación inferido de los distintos sujetos.
- *ValenceSD*: Desviación típica de los valores de valencia dados por los distintos sujetos.
- *ArousalSD*: Desviación típica de los valores de nivel de excitación dados por los distintos sujetos.
- *%ValenceRaters*: Porcentaje del total de sujetos que ha dado un valor a la valencia.
- *%ArousalRaters*: Porcentaje del total de sujetos que ha dado un valor al nivel de excitación.

Por último (Alonso, Fernandez, & Díez, 2015) describe para 7040 palabras otras características como la media de edad a la que se suele aprender esa palabra (*averageAoA*), la edad mínima a la que se aprende (*Min*) o máxima (*Max*) y la desviación en esos datos de edades (*SD*), así como la frecuencia literaria con la que se suele encontrar.

2.1.4. Etiquetado Automático de Repositorios

Un ámbito adicional donde ubicar la presente investigación es el de las anotaciones automáticas de los LO. Así, efectivamente, en esta memoria se propone una forma de anotar un tipo particular de LO (sonetos en castellano) de manera automática en base a su contenido (texto). Con ello se está llevando a cabo una anotación automática para textos poéticos en castellano.

Revisión de la literatura

En la literatura aparecen varias referencias a tareas de anotación automática de LO. Un caso es (Cechinel, da Silva Camargo, Sicilia, & Sánchez-Alonso, 2016) donde se menciona la importancia de automatizar la anotación del nivel de calidad de los LO que utiliza un sistema de IR. Efectivamente, los objetos que devuelve un sistema de IR pueden tener su ranking ajustado (si se devuelve una lista de LO candidatos) en función de criterios como la calidad de dichos LO. El problema es que los repositorios que contienen estos LO (*LOR*, *Learning Objects Repository*) suelen ir incorporando continuamente LO y no es factible revisarlos

fácil. Es más fácil construir frases o buscar ejemplos de uso para palabras como 'mesa' que para otras como 'ciudadanía'

¹⁸ *Familiarity se define como el grado de familiaridad, valiendo 1 cuando la palabra es muy poco familiar y 7 cuando es mucho. Una palabra como 'pescado' es más familiar que otra como 'quark'*

¹⁹ En este caso, las escalas para las dos métricas estaban entre 1 y 9

manualmente para asociarles unas etiquetas de calidad u otras. Por este motivo el artículo describe cómo aplicar técnicas de ML supervisado (concretamente *Redes Neuronales Artificiales*, ANN) para etiquetar los LO del LOR en base a una serie de rasgos y a las etiquetas de algunos LO que se han anotado manualmente en el pasado. Estos rasgos es importante que se puedan extraer de manera automática para que los nuevos LO puedan ser etiquetados con la categoría de calidad correspondiente de manera automática. La propuesta se ha estudiado sobre los conjuntos de datos de MERLOT y de Connexions.

Otro ejemplo de propuesta es (Sabitha, Mehrotra, & Bansal, 2015) en la que se combina la información estática asociada a un LO con unas anotaciones dinámicas denominadas KO (knowledge object). En esos KO aparecen metadatos (palabras clave, una introducción, un resumen...), información sobre la meta educativa o información sobre el contenido. Así, la necesidad es similar a la del caso anterior; se puede disponer de una serie de etiquetas en un LO pero cuando es necesario enriquecer dicha información con otras anotaciones es preciso recurrir a técnicas que implementen esto de manera automática. Las anotaciones automáticas pueden hacerse gracias a la información del propio LO, como en el artículo anterior para las etiquetas de calidad, o como en este ejemplo si se quieren anotar resúmenes o palabras clave en base al contenido de los LO. También pueden hacerse en base a información externa, como ocurre en este ejemplo, si lo que se quiere anotar es información sobre la meta educativa.

Más específicamente en el ámbito de la minería de textos destaca el artículo (Saquicela et al., 2018). En este artículo se presenta una propuesta en la de generación de anotaciones dinámicas que contienen información sobre el programa académico de diversas asignaturas en base al contenido del LO de las mismas. Estas anotaciones se generan tras aplicar distintas técnicas de minería de textos (eliminación de *stopwords*, lematización, reducción de dimensionalidad, *clustering* para encontrar LO similares semánticamente...).

Investigación desarrollada en el ámbito descrito

En la presente investigación se plantea la necesidad de anotar automáticamente el corpus de sonetos con rasgos adicionales de cara a incrementar la interoperabilidad de un RecSys sensible al contexto afectivo que incorpore el enfoque de la pedagogía poética. Para este RecSys los sonetos son los LO y que contienen las etiquetas incluidas por DISCO sobre la ontología de POSTDATA (Díez Platas & Bermúdez Sabel). Surge pues la necesidad de anotar sobre dichas etiquetas una serie de rasgos afectivos y semánticos, de manera automática, y utilizando la propia información que contiene el LO (el poema). Las anotaciones afectivas se incluirán dentro de la extensión de la ontología, como se detalla en el Capítulo 3. Las anotaciones semánticas se persisten en un fichero aparte sin incluirse dentro de campos extendidos sobre la ontología de POSTDATA.

Así, en los capítulos sucesivos se detalla cómo llevar a cabo una anotación automática de rasgos semánticos y afectivos sobre unos LO que son sonetos en castellano.

2.1.5. Word Embeddings

A. Introducción

La representación de un texto en formato numérico es necesaria para que éste pueda ser procesado por un computador. Dentro del ámbito del NLP se han propuesto a lo largo de los años distintas alternativas para ello. Las propuestas más básicas han considerado las palabras de manera aislada, como por ejemplo los *vector space model*, como la *Bag of Words* (BoW). Posteriormente se han planteado representaciones que en lugar de considerar las palabras de manera aislada han intentado capturar dentro de la representación numérica la dependencia existente entre términos. Este es el caso de la *semántica distribucional*, que se basa en la *hipótesis distribucional: elementos lingüísticos con distribuciones similares tienen significados similares*²⁰. En ella se apoyan las representaciones de *word embeddings*, muy utilizadas en la actualidad.

El SOTA que se muestra a continuación se centra especialmente en este último tipo de representaciones, atendiendo a la evolución histórica de dichas técnicas especificando qué se ha mejorado en las sucesivas propuestas respecto de técnicas previas y por qué era necesaria dicha evolución. Se acabará este SOTA indicando las técnicas más recientes, remarcando las ventajas que estas aportan.

B. Representaciones no densas

El modelo clásico original para abordar esta tarea es el denominado *Bag of Words* (BoW), modelo con el que, como recogen artículos que lo detallan como (Zhang, Y., Jin, & Zhou, 2010), consiste en construir una matriz con todas las palabras existentes en los textos y cada vector asociado a uno de ellos tendría un 1 o un 0 dependiendo que palabras del conjunto total contiene. El texto de entrada se reproduciría de la misma manera con la salvedad de que para las potenciales palabras que apareciesen que no estuviesen contenidas en el texto original se utilizaría un valor *dummy* (generalmente una etiqueta UNK) para representarlas por defecto.

the dog is on the table

| | | | | | | | |
|-----|-----|-----|----|-----|----|-------|-----|
| 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 |
| are | cat | dog | is | now | on | table | the |

Ilustración 9 Ejemplo de construcción de un vector numérico usando BoW²¹. Con esta técnica se representa un texto con un vector de unos y ceros, técnica conocida como One-Hot Encoding.

Uno de los problemas que aparecen con las BoW simples es que no se ponderan más las palabras que aparecen más frecuentemente, y por eso se plantean también soluciones que en

²⁰ Extraído de https://es.wikipedia.org/wiki/Sem%C3%A1ntica_distribucional Última visita: 31/07/2019

²¹ “Text classification and prediction using the Bag Of Words approach”. Web. 25/04/2019.

<https://medium.freecodecamp.org/text-classification-and-prediction-using-bag-of-words-8aeb1396cded>

lugar de simplemente asignar 1 o 0 según si aparece o no la palabra se asigne un número con la frecuencia de aparición.

| | Document 1 | Document 2 | Document 3 | Document 4 | Document 5 | Document 6 | Document 7 | Document 8 |
|-----------|------------|------------|------------|------------|------------|------------|------------|------------|
| Term(s) 1 | 10 | 0 | 1 | 0 | 0 | 0 | 0 | 2 |
| Term(s) 2 | 0 | 2 | 0 | 0 | 0 | 18 | 0 | 2 |
| Term(s) 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Term(s) 4 | 6 | 0 | 0 | 4 | 6 | 0 | 0 | 0 |
| Term(s) 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Term(s) 6 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| Term(s) 7 | 0 | 1 | 8 | 0 | 0 | 0 | 0 | 0 |
| Term(s) 8 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 |

Document Vector

Word Vector
(Passage Vector)

Ilustración 10 Ejemplo de construcción de BoW que tiene en cuenta la frecuencia de las palabras²²

Pero en este caso surge otra dificultad; se ponderan más las palabras muy habituales frente a algunas menos habituales pero que pueden ser muy relevantes en el texto, cuando podría ocurrir que una palabra poco frecuente fuese muy importante dentro del texto. Una solución, por supuesto, es eliminar las denominadas *stopwords* y construir una BoW sin ellas (sin determinantes, conjunciones, preposiciones...) pero esto solo sirve para paliar el problema ya que se sigue sin ponderar lo suficiente a las palabras poco frecuentes pero que son fundamentales.

Para solventar esta dificultad apareció el método denominado TF-IDF (*term frequency inverse document frequency*) expresado en la siguiente ecuación:

$$w_{i,j} = tf_{i,j} \times \log \left(\frac{N}{df_i} \right)$$

Donde $w_{i,j}$ es el valor numérico asociado a una determinada palabra i (su peso) en un documento j en concreto, N es el número total de documentos, df_i es el número de documentos que contienen la palabra i y $tf_{i,j}$ es la frecuencia de la palabra i en el documento j .

Con ello se penalizan las palabras frecuentes que aparecen en muchos documentos distintos, de manera que se conseguiría poner en más valor a las palabras que aparecen en pocos documentos, remarcando su valor especial por su originalidad de cara a señalar las singularidades de dicha obra.

Uno de los problemas que aun aparecen con el método de TF-IDF es la gran *sparsity* de los vectores de los documentos por tener en su mayoría valor 0 en sus componentes. Esto es así porque aun filtrando *stopwords*, cuando se tenga un gran número de documentos y el vocabulario generado en base a ellos sea muy extenso, la mayoría de los documentos tendrán asociada una gran cantidad de valores nulos en sus vectores correspondientes a todas esas

²² Godbout, Camron. "Creating a Search Engine". Web. 25/04/2019. <https://medium.com/@camrongodbout/creating-a-search-engine-f2f429cab33c>

palabras que no han aparecido en ellos; esta *sparsity* reduce significativamente el rendimiento de los sistemas de IR y es necesario buscar soluciones que generen vectores sin tantos nulos.

C. Word2Vec

Para solventar este problema de la *sparsity* (representaciones no densas) se desarrolló una técnica conocida como word2vec (Mikolov, Chen, Corrado, & Dean, 2013) que consiste en asignar un valor numérico a las palabras en base a la semántica de las mismas. Es decir, cada palabra tendrá asociado un vector numérico representado en un espacio de determinadas dimensiones de manera que dicho vector recoja el valor semántico de cada una de las palabras en función de la relación contextual que dicha palabra tenga con el resto de las palabras del corpus.

En la Ilustración 11 se muestra el ejemplo clásico de representación con vectores en 3 dimensiones. Gracias a que los valores numéricos se infieren por el contexto se pueden hacer operaciones del siguiente tipo: King – Man = Queen – Woman. Es decir, que los vectores representan el valor semántico de forma numérica hasta el punto de que si a “Rey” se le resta la contribución de “Hombre” y a “Reina” de “Mujer” se obtendría un mismo valor que representaría el valor semántico genérico de “Monarca”

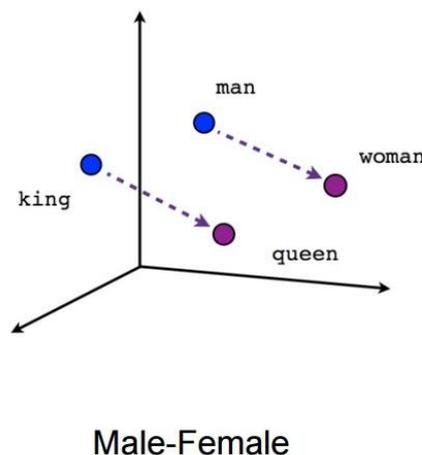


Ilustración 11 Ejemplo de representación con vectores en 3 dimensiones²³.

Una forma de conseguir estos *embeddings* es mediante una técnica denominada *Skip-Gram* (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013). De forma resumida y en su versión más sencilla, esta técnica consiste en entrenar una red neuronal en la que el parámetro de entrada de esta será una palabra en cuestión y su salida será las probabilidades asociadas al resto de palabras de aparecer juntamente con dicha palabra de entrada. Una vez entrenada la red neuronal se usarán los pesos de la capa intermedia como valores del vector de *embedding* para dicha palabra de entrada.

²³ Critchlow, Will. “A Beginner’s Guide to word2vec AKA What’s the Opposite of Canada?”. Web. 26/04/2019. <https://www.distilled.net/word2vec-examples/>

Para ello, en primer lugar se parte del conjunto de textos disponibles y se define, de partida, el número de palabras vecinas que se van a considerar. Con ello se van recorriendo todas las palabras de los textos, se va viendo qué palabras aparecen juntamente con respecto a cada una de las palabras en particular y con ello se obtiene un vector de probabilidades para cada palabra que contenga la probabilidad de que cada palabra del vocabulario aparezca en su entorno vecino.

En la Ilustración 12 aparece un ejemplo de obtención de probabilidades usando una ventana de tamaño 2 para los vecinos (2 palabras previas y 2 palabras posteriores a la palabra elegida, *focus word*, en cada momento). Por ejemplo, para la palabra 'fox' un primer conjunto de vecinos es 'quick', 'brown', 'jumps', 'over'. Se puede ver también como la palabra 'the' (omitiendo que sea mayúscula) tiene de vecinos 'quick' y 'brown', pero también 'jumps', 'over', 'lazy' y 'dog'.

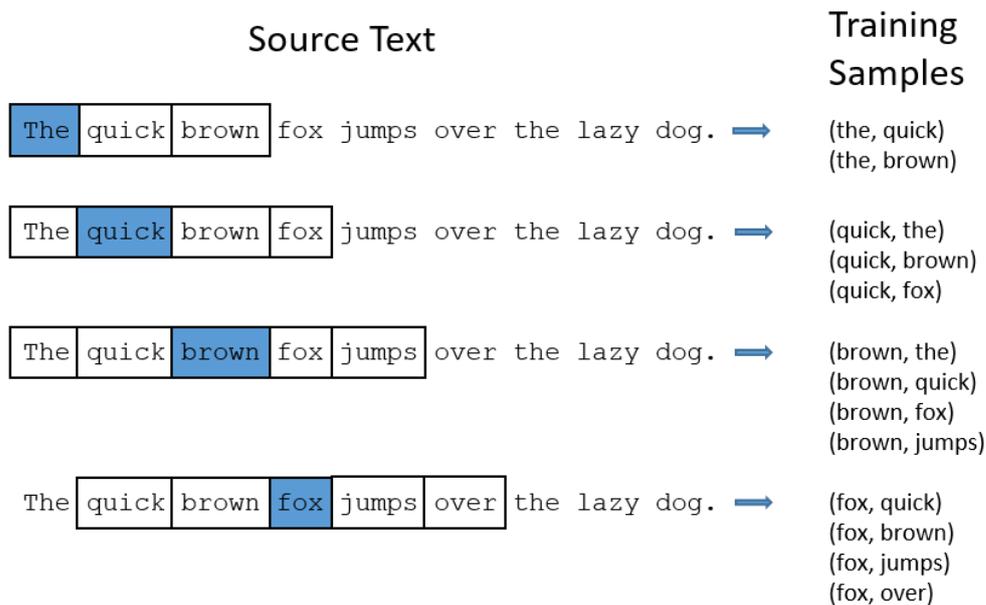


Ilustración 12 Ejemplo de obtención de probabilidades usando una ventana de tamaño dos ²⁴

Para determinar el vector de salida se calcula la probabilidad conjunta de que dada la palabra de entrada se tenga una palabra de salida en concreto.

$$P_{ij} = P(j|i) = \frac{X_{ij}}{\sum_k X_{ik}}$$

Siendo X_{ij} el número de veces que la palabra j aparece en el contexto de i y con el denominador representando el total de pares de palabras entre la palabra X_i y el resto de palabras presentes en su contexto.

Si el Skip-Gram se hiciese sobre un corpus que fuese solo la frase de la ilustración anterior entonces se tendrían los siguientes vectores de probabilidad:

- Para el caso de la palabra 'fox' aparecen 4 pares con las palabras de su contexto. Así, por ejemplo:

²⁴ McCormick, Chris. "Word2Vec Tutorial - The Skip-Gram Model". Web. 27/04/2019. <http://mccormickml.com/2016/04/19/word2vec-tutorial-the-skip-gram-model/>

$$P(\text{quick}|\text{fox}) = \frac{X_{\text{quick.fox}}}{\sum_k X_{k,\text{fox}}} = \frac{1}{4}$$

Y con ello:

$$\text{fox} = \begin{bmatrix} \text{the} = 0 \\ \text{quick} = \frac{1}{4} \\ \text{brown} = \frac{1}{4} \\ \text{jumps} = \frac{1}{4} \\ \text{over} = \frac{1}{4} \\ \text{lazy} = 0 \\ \text{dog} = 0 \end{bmatrix}$$

- Otro ejemplo. Para la palabra 'the' en la que aparecen 2+4 pares de bigramas:

$$\text{the} = \begin{bmatrix} \text{fox} = 0 \\ \text{quick} = 1/6 \\ \text{brown} = 1/6 \\ \text{jumps} = 1/6 \\ \text{over} = 1/6 \\ \text{lazy} = 1/6 \\ \text{dog} = 1/6 \end{bmatrix}$$

Con dichos vectores representando la probabilidad de que cada palabra del corpus aparezca entre los vecinos de dicha palabra de entrada.

Estos vectores de probabilidad se usarán como vectores de salida para entrenar una red neuronal que tenga como entrada una palabra en concreto expresada mediante su *one-hot encoding*.

En la Ilustración 13 se muestra un ejemplo de red neuronal para el modelo *Skip-Gram*. Ante una palabra de entrada expresada mediante su *one-hot encoding* se entrenan 2 capas, una capa oculta de un tamaño predefinido concreto y una capa *softmax* que de las probabilidades de que cada una de las palabras del corpus pueda aparecer como vecino de la palabra de entrada. La red neuronal se entrenará de manera que el vector de salida real será el de las probabilidades inferidas desde las coocurrencias de las palabras en el corpus mediante el método explicado previamente.

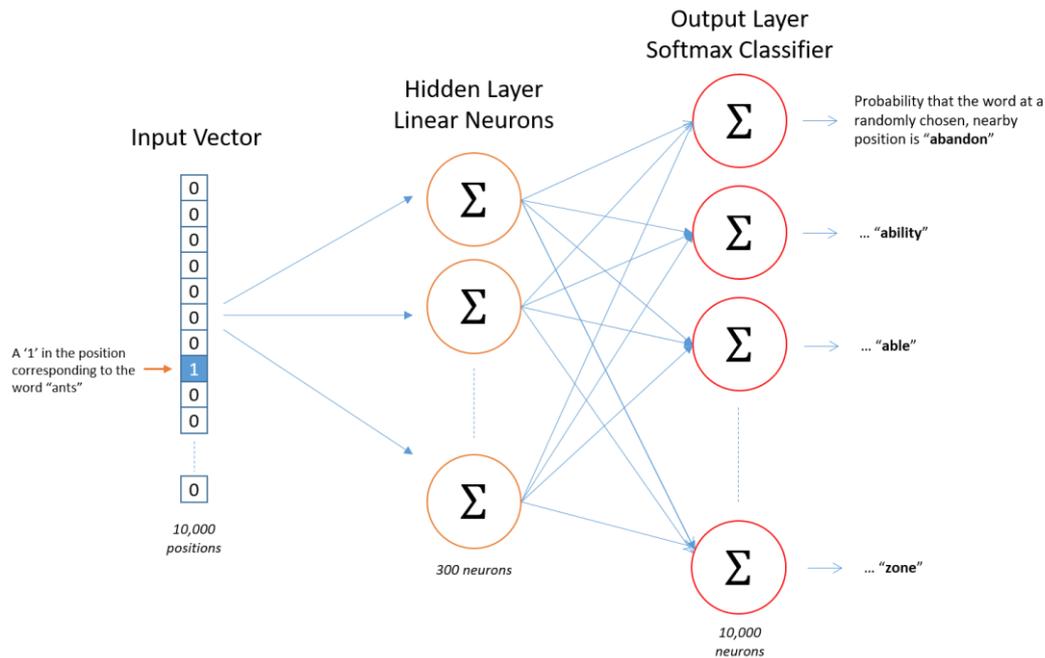


Ilustración 13 Ejemplo de red neuronal para el modelo Skip-Gram.²⁵

Tras entrenar esta red neuronal se extrae la *capa oculta* mencionada previamente de manera que los vectores de *embeddings* para cada palabra corresponderán a los pesos correspondientes asociados a esa capa oculta, por lo que la dimensionalidad de estos dependerá del número de neuronas utilizada en la misma.

Esta capa oculta daría lugar a lo que se conoce como *look-up table* ya que daría lugar a una matriz en la que se tendrían tantos pesos como palabras x número de neuronas de manera que cada palabra tendría asociada una fila en dicha matriz, siendo precisamente esa fila el *word embedding* para esa palabra.

En la Ilustración 14 muestra como para el ejemplo de la imagen anterior en el que se tenía un corpus de 10000 palabras la matriz para la capa oculta tendría esa dimensionalidad donde cada fila representaría el *embedding* de una palabra en concreto y cada columna el valor de esa componente del vector dentro de un espacio dimensional de dimensión coincidente con el número de neuronas utilizadas

²⁵ McCormick, Chris. "Word2Vec Tutorial - The Skip-Gram Model". Web. 27/04/2019. <http://mccormickml.com/2016/04/19/word2vec-tutorial-the-skip-gram-model/>

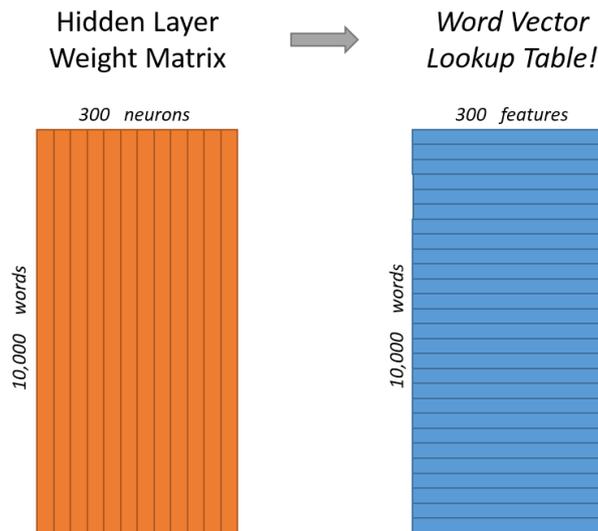


Ilustración 14 Paso de matriz de pesos a tabla lookup.²⁶

En la Ilustración 15 se muestra un sencillo ejemplo en el que se tuviese un corpus de 5 palabras y se usasen 3 neuronas en la capa de *embedding*, tras entrenar el sistema se tendría una look-up table de tamaño 5x3 con la que, ante un vector de entrada asociado a una palabra y expresado mediante su *one-hot encoding* se obtiene el vector de *embedding* que a diferencia del de entrada es un vector denso, no *sparse*, y que por supuesto tiene incluida la información de frecuencia de palabras, además de representar adecuadamente las palabras menos comunes gracias a que al margen de su frecuencia se tiene siempre información de su contexto.

$$[0 \ 0 \ 0 \ 1 \ 0] \times \begin{bmatrix} 17 & 24 & 1 \\ 23 & 5 & 7 \\ 4 & 6 & 13 \\ 10 & 12 & 19 \\ 11 & 18 & 25 \end{bmatrix} = [10 \ 12 \ 19]$$

Ilustración 15 Ejemplo sencillo de vector de dimensión 5 y su correspondiente table lookup.

Así, palabras de significado similar, como el ejemplo de antes con *King* y *Queen*, tendrán palabras vecinas similares y por lo tanto los vectores asociados a las mismas tendrán unas componentes de valor similar, ya que los pesos asociados a ellos en la red neuronal para obtener las probabilidades de coocurrencias de palabras serán prácticamente los mismos.

En la Ilustración 16 se muestra un ejemplo de visualización. Utilizando una técnica de representación denominada t-SNE se pueden visualizar los embeddings de un espacio multidimensional en 2D. Se puede observar cómo palabras de contextos muy similares, como 'scripture' y su plural 'scriptures' aparecen muy cercanas en el espacio vectorial al tener contextos similares

²⁶ McCormick, Chris. "Word2Vec Tutorial - The Skip-Gram Model". Web. 27/04/2019. <http://mccormickml.com/2016/04/19/word2vec-tutorial-the-skip-gram-model/>

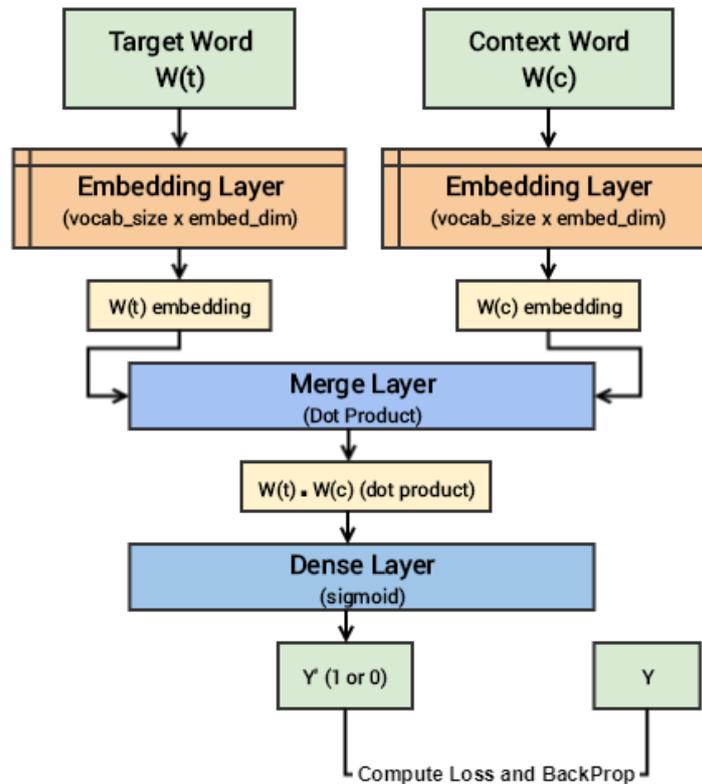


Ilustración 17 Arquitectura alternativa para dos vectores de entrada²⁸

Otra propuesta alternativa a Skip-Gram es el método denominado CBOW (*Continuous Bag of Words*) (Mikolov et al., 2013), que funciona de manera inversa al método anterior; en lugar de intentar predecir el contexto de una palabra partiendo de la palabra en cuestión se parte de un contexto para intentar predecir la palabra central del mismo.

²⁸ Fuente: <https://www.kdnuggets.com/2018/04/implementing-deep-learning-methods-feature-engineering-text-data-skip-gram.html> Última visita: 27/04/2019

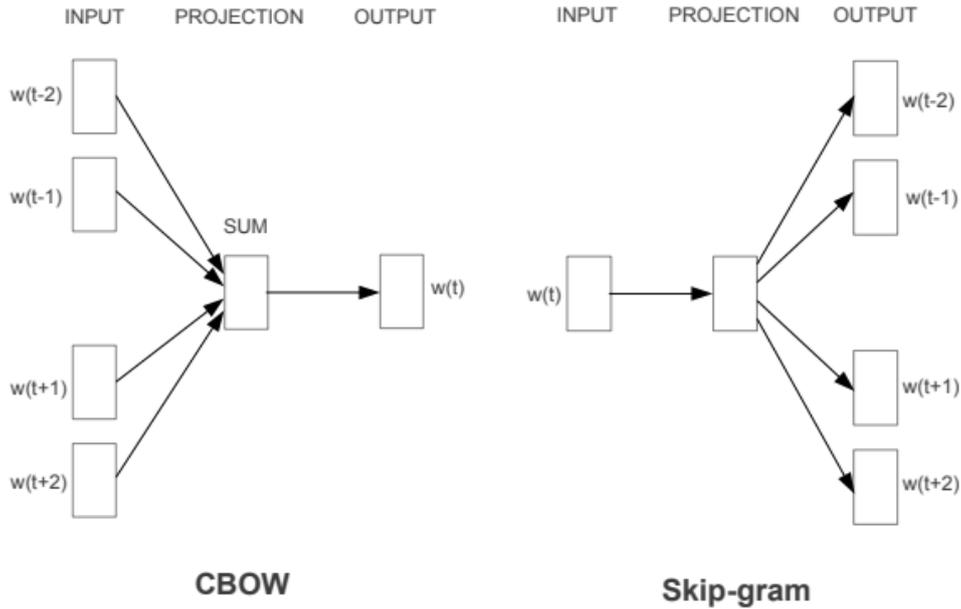


Ilustración 18 Comparativa de CBOW vs Skip-Gram; uno predice el contexto alrededor de una focus word y otro predice la focus word de un determinado contexto.²⁹

Context (X): ['old', 'testament', 'james', 'bible'] -> Target (Y): king
 Context (X): ['first', 'book', 'called', 'genesis'] -> Target(Y): moses
 Context(X): ['beginning', 'god', 'heaven', 'earth'] -> Target(Y): created
 Context (X): ['earth', 'without', 'void', 'darkness'] -> Target(Y): form
 Context (X): ['without', 'form', 'darkness', 'upon'] -> Target(Y): void
 Context (X): ['form', 'void', 'upon', 'face'] -> Target(Y): darkness
 Context (X): ['void', 'darkness', 'face', 'deep'] -> Target(Y): upon
 Context (X): ['spirit', 'god', 'upon', 'face'] -> Target (Y): moved
 Context (X): ['god', 'moved', 'face', 'waters'] -> Target (Y): upon
 Context (X): ['god', 'said', 'light', 'light'] -> Target (Y): let
 Context (X): ['god', 'saw', 'good', 'god'] -> Target (Y): light

Ilustración 19 Así, con CBOW se parte de un contexto X y con él se intenta predecir una palabra concreta Y utilizando una red neuronal³⁰

La arquitectura detallada de este modelo se muestra a continuación.

²⁹ Fuente: <https://www.kdnuggets.com/2018/04/implementing-deep-learning-methods-feature-engineering-text-data-skip-gram.html> Última visita: 27/04/2019

³⁰ Fuente: <https://www.kdnuggets.com/2018/04/implementing-deep-learning-methods-feature-engineering-text-data-cbow.html> Última visita: 27/04/2019

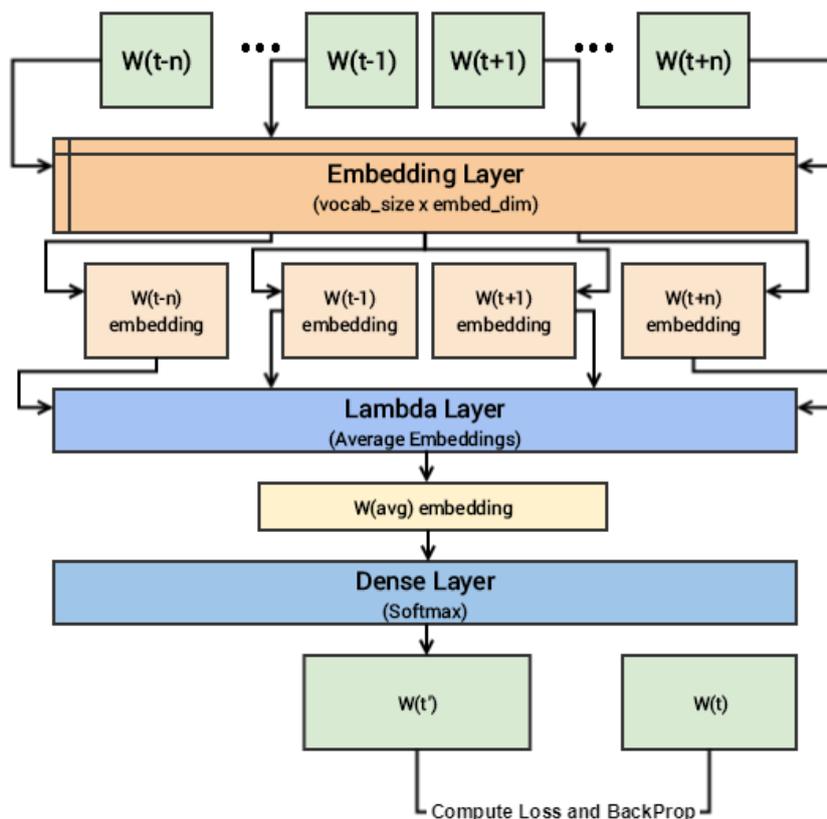


Ilustración 20 Arquitectura de CBOW³¹

Para el caso de CBOW el vector de entrada sería un vector similar al del *one-hot encoding* de *Skip-Gram* con la salvedad que ahora contendrá un uno las palabras que aparezcan como contexto para una determinada palabra de salida. Dicho vector se pasa a una capa oculta que representa de igual forma el *embedding layer* pero que ahora dará como salida no un único vector sino un vector de *embedding* por cada palabra del vocabulario. Dichos vectores obtenidos se agregarán en el *Lambda Layer* generalmente haciendo la media de ellos obteniendo así un vector de *embedding* para el contexto de la palabra foco (*focus word*). Así, como ocurría con *Skip-Gram*, el orden de las palabras en el contexto para la construcción de estos *embeddings* no se tiene en cuenta: al agregar las palabras del contexto se pierde la información sobre la posición relativa de las mismas respecto a la palabra foco.

Dicho vector se pasará a la capa softmax para tener como salida la probabilidad de que dicho contexto se asocie a cada una de las palabras posibles del vocabulario, y con ello se irá entrenando el sistema.

De esta manera se tiene una matriz de *embeddings* que funcionaría de la misma manera que en el caso de *Skip-Gram*; una vez entrenado el sistema se consultará una palabra en particular y se verá su *embedding* asociado.

Los autores de estos métodos mencionan como diferencia entre ellos que *Skip-Gram* suele funcionar mejor cuando el conjunto de datos de entrenamiento para obtener las probabilidades

³¹ Sarkar, Dipanjan. "Implementing Deep Learning Methods and Feature Engineering for Text Data: The Continuous Bag of Words (CBOW)". Web. 27/04/2019. <https://www.kdnuggets.com/2018/04/implementing-deep-learning-methods-feature-engineering-text-data-cbow.html>

de las palabras de los contextos no es muy amplio y cuando se quiere dar importancia a las palabras que ocurren raras veces.

En cambio CBOW es un método más rápido para entrenar (ya que no hay que construir los vectores de probabilidades para la salida analizando todo el corpus, como ocurría en *Skip-Gram*; simplemente hay que ir pasando los contextos de cada palabra foco según se vayan encontrando al recorrer las palabras) y que suele dar unos resultados algo mejores para representar las palabras que ocurren frecuentemente.

Un claro hándicap que tienen los modelos de word2vec descritos previamente es que no tienen en cuenta de forma global lo que ocurre en el corpus sino que se limitan a analizar lo que ocurre en las ventanas de vecinos escogidas para cada palabra, pudiéndose perder información o patrones relevantes de los textos. Esto es relativamente importante ya que como word2vec analiza contextos acotados, para ejemplos como “el” y “gato” que pueden ser palabras que tengan un alto nivel de coocurrencia, word2vec no es capaz de distinguir si esto es porque “el” es una palabra muy común (y que por tanto dará altos niveles de coocurrencia) o si es porque las palabras “el” y “gato” tienen algún tipo de relación entre sí.

D. GLoVe

GLoVe (Pennington, Socher, & Manning, 2014) intenta solventar esos problemas mencionados en word2vec mediante una aproximación global a los textos disponibles y no restringida exclusivamente al número de vecinos predefinido aplicado sobre cada palabra. Para ello lo primero que hace es construir una matriz de coocurrencia de palabras de la siguiente manera.

Para una frase como “*the cat sat on the mat*” la matriz de coocurrencia si se considerase una ventana de vecinos de tamaño 2 (2 palabras previas y posteriores a cada palabra) sería la siguiente:

| | the | cat | sat | on | mat |
|------------|------------|------------|------------|-----------|------------|
| the | 2 | 1 | 2 | 1 | 1 |
| cat | 1 | 1 | 1 | 1 | 0 |
| sat | 2 | 1 | 1 | 1 | 0 |
| on | 1 | 1 | 1 | 1 | 1 |
| mat | 1 | 0 | 0 | 1 | 1 |

Tabla 1 Matriz de concurrencia

Siendo por supuesto una matriz simétrica (ya que las coocurrencias da igual desde que palabra se mire ya que no se está teniendo en cuenta el orden de aparición, de manera que la coocurrencia de (cat, sat) es la misma que la de (sat, mat)) y en la que la diagonal principal tiene el número de veces que aparece como tal el término de manera aislada (es decir, el unigrama para cada palabra).

A partir de esto, GLoVe plantea la siguiente hipótesis: *los cocientes o ratios entre dos palabras en un contexto están fuertemente conectadas con su significado.*

Esto quiere decir, siguiendo el ejemplo que cita el propio artículo, que palabras como “ice” (hielo) y “steam” (vapor) tienen significados diferentes pero tienen un elemento de su

significado común y es que ambas son posibles estados para el *agua*. Por este motivo las palabras relacionadas con el agua, como “*water*” (agua) deberían aparecer en ambos contextos, mientras que otras específicas de cada uno de los estados, como “*cold*” (frío) y “*solid*” (sólido) sólo aparecerían en el caso del estado *hielo* pero no en el de *vapor*. De igual manera, otras como “*fahion*” (moda) no aparecerían en ninguno de ellos por no tener relación con ellos a nivel particular ni con ellas a nivel común.

| Probability and Ratio | $k = solid$ | $k = gas$ | $k = water$ | $k = fashion$ |
|-----------------------|----------------------|----------------------|----------------------|----------------------|
| $P(k ice)$ | 1.9×10^{-4} | 6.6×10^{-5} | 3.0×10^{-3} | 1.7×10^{-5} |
| $P(k steam)$ | 2.2×10^{-5} | 7.8×10^{-4} | 2.2×10^{-3} | 1.8×10^{-5} |
| $P(k ice)/P(k steam)$ | 8.9 | 8.5×10^{-2} | 1.36 | 0.96 |

Ilustración 21 Matriz de coocurrencias para distintas palabras junto con los ratios de co-ocurrencias entre dos casos.³²

Como se puede ver en el ejemplo anterior, al construir los ratios entre los dos pares de co-ocurrencias los valores de los mismos serán muy altos o muy cercanos a 0 cuando se haga referencia las partes no en común de los significados de las palabras mientras que serán cercanas a 1 cuando se haga referencia los aspectos comunes (en el ejemplo anterior se ve que tanto *water* como *fashion* son elementos comunes a ambas, por aparecer o por no tener nada que ver por igual, y por este motivo los cocientes son cercanos a 1 para esos casos).

A partir de este punto el artículo propone una ecuación aritmética (a diferencia de los modelos previos de redes neuronales) que sea capaz de llegar a relacionar el ratio de coocurrencia entre 2 pares de palabras que tengan una de sus palabras en común con los vectores de palabras correspondientes (sus embeddings), de manera que:

$$F(w_i, w_j, \overline{w_k}) \sim \frac{P_{i,j}}{P_{jk}}$$

Siendo las probabilidades P las inferidas en base al número de ocurrencias de una determinada palabra respecto a una *focus word* dividido entre el número total de coocurrencias posibles para dicha *focus word*, como ya se contó en puntos anteriores.

Finalmente, el artículo concluye que la función F se puede aproximar de la siguiente forma:

$$\text{dot}(w_i, \overline{w_k}) + b_i + \overline{b_k} = \log(X_{ik})$$

siendo $\text{dot}(w_i, \overline{w_k})$ el producto escalar de esos dos vectores

y con X_{ik} el número de co – ocurrencias para esas dos palabras

La ecuación anterior se pondera con una función de peso que los autores estiman empíricamente con la que pretenden penalizar las palabras que tienen pocas coocurrencias, tratándolas como un elemento de ruido a mitigar, al tiempo que no se ponderen de más las ocurrencias extremadamente habituales y que quitarían valor a las demás (como con palabras como “él es”). Definen así la siguiente función de peso:

³² Fuente: el propio artículo

$$\text{weight}(x) = \min\left(1, \left(\frac{x}{x_{\max}}\right)^{\frac{3}{4}}\right)$$

Gráficamente esta función de peso quedaría:

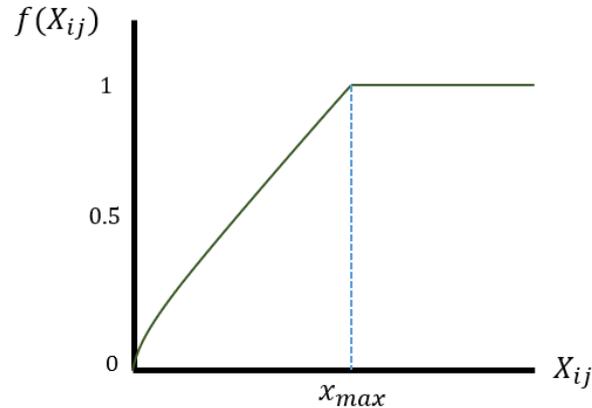


Ilustración 22 Representación de la función de peso.³³

Y con esto se define una función de error en base al error cuadrático medio que intente minimizar el ratio predicho mediante la ecuación en relación al ratio real entre bigramas. Esta ecuación de error tendría el siguiente aspecto considerando ya en ella el factor de peso:

$$\sum_{ij} \text{weight}(X_{ij})(\text{dot}(w_i, \tilde{w}_j) + b_i + \tilde{b}_k - \log(X_{ij}))^2$$

Los autores en el artículo concluyen como GLoVe es muy similar a word2vec en cuanto a la función de error que se quiere minimizar (que toma como base las coocurrencias de palabras de forma análoga) pero que la ventaja de GLoVe es que el tiempo de computación que necesita para generar los *embeddings* es menor que el de word2vec, especialmente cuando hay muchas palabras que ocurren frecuentemente.

En cuanto a la eficiencia de sus *embeddings*, aunque el artículo señala que son mejores que los de word2vec, la práctica ha demostrado que para la mayoría de las tareas los resultados son extremadamente similares.

Es más, tanto para este proyecto como para la mayoría de los proyectos en general, se va a partir de un conjunto de *embeddings* pre-entrenados, con lo que el tiempo de generación de estos acaba siendo un elemento casi irrelevante.

³³ “Paper Dissected: “Glove: Global Vectors for Word Representation” Explained”. Web. 27/04/2019. <http://mlexplained.com/2018/04/29/paper-dissected-glove-global-vectors-for-word-representation-explained/>

E. Embeddings con n-gramas - FastText

Los métodos anteriores de obtención de *embeddings* tienen un importante hándicap en común, y es cómo tratar las palabras que no estuviesen presentes en el vocabulario original para tareas como clasificación supervisada, IR... Por ejemplo, ¿qué hacer cuando se recibe un texto de consulta como “*quiero información sobre tiburones*” y en nuestro corpus original sólo se tienen referencias al *tiburón* en singular? En la mayoría de los casos cuando las palabras nuevas que se encuentran los modelos no aparecen en el corpus original o bien se descartan o bien se identifican con etiquetas genéricas del tipo UNK, de manera que todas esas palabras caerían en un mismo cajón de sastre y no estarían aportando información al sistema. Es cierto que para evitar estos problemas muchas veces se opta por lematizar/truncar las palabras de manera que los verbos se expresen en infinitivo, los sustantivos o adjetivos estén en singular... pero esta solución, que ha sido la aproximación que se ha utilizado de manera más habitual durante mucho tiempo, no está exenta de algunos problemas de cierta importancia.

Al representarse todas las palabras del corpus original con sus lemas, por un lado, se pierde mucha información relevante sobre ellas que hace que se pierda a su vez información del texto original. Así, por ejemplo, y más en obras de gran riqueza lingüística como la poesía, expresar todos los verbos con su infinitivo en aras a simplemente poder maximizar el reconocimiento de palabras nuevas que lleguen al sistema hace que se tenga una interpretación de las obras muy reduccionista.

Más grave aún que todo eso es la inclusión de sesgos en todos los sistemas de NLP, especialmente en lo que respecta a sesgos de género. Al lematizarse las palabras, en particular para el castellano, todos los sustantivos se expresarán en su forma singular que en muchas ocasiones es en masculino. Esto hace que toda la información particular que pudiese hacer referencia a la mujer desaparezca tras la lematización. Por ejemplo, en un sistema de IR que lematizase su corpus de textos y también los mensajes de consulta ocurriría lo siguiente.

- Ante el mensaje de consulta: “*quiero información sobre las ingenieras más importantes del S. XX*” se construiría el siguiente mensaje de consulta tras lematizar (y eliminando *stopwords*): “querer información sobre **ingeniero** importante S.XX”³⁴. De esta manera, no se podría consultar información específica para un determinado género si se lematizasen los corpora y los mensajes de entrada.

Estas casuísticas son las que se han intentado solventar con soluciones como FastText (Joulin et al., 2016). FastText es un método con el que se busca incluir información a nivel de caracteres en los *embeddings* al incorporar en estos no sólo las palabras de los textos originales sino también los distintos n-gramas de una misma palabra, usando de fondo alguno de los métodos de word2vec descritos previamente.

Así, tomando el ejemplo anterior de la palabra *ingenieras* y con el uso de n-gramas de hasta $n=3$, FastText obtendría los *embeddings* para $\langle in, ing, nge, eni, nie, ier, era, ras, as \rangle$ donde los símbolos \langle y \rangle se utilizan para designar el principio y fin de la palabra. Estos n-gramas se

³⁴ Caso real replicable con la versión 2.1.0 de spaCy y el modelo para castellano es_core_news_sm 2.1.0 Última fecha: 29/04/2019

utilizarían, junto con la palabra original, como palabras del corpus y tendrían así sus propios *embeddings*.

Es decir, la frase “*Las ingenieras más relevantes en España son...*” obtendría *embeddings* para *ingenieras* y para $\langle in, ing, nge, eni, nie, ier, era, ras, as \rangle$ en base a esa frase, y si además se tuviese una frase como “*Estudiar ingeniería implica...*” se representaría la palabra *ingeniería* como ella misma y como sus n-gramas $\langle in, ing, nge, gen, eni, nie, ier, erí, ría, ía \rangle$, de manera que el *embedding* de algunos n-gramas se obtendría con la información de ambas frases (por ejemplo el de *ing*) consiguiéndose así además de solventar los problemas antes citados (sobre los sesgos y la pérdida de información) enriquecer aún más el sistema con la información de afijos en general.

Así pues, volviendo al sistema de IR antes citado, todos los textos del corpus se representarían con las palabras originales y sus n-gramas, y lo mismo se haría para el mensaje de consulta, el cual tendría las palabras originales y sus n-gramas, de manera que el texto que se escogería como respuesta para esa consulta sería el que mayor similitud tuviese entre dicho mensaje de consulta y dichos textos originales, todos ellos ampliados con los *strings* adicionales de los n-gramas.

Así pues, el *embedding* final de una palabra será la suma de los *embeddings* de todos los n-gramas que la componen tras analizar todo el corpora. Para las nuevas palabras que llegasen al sistema (a través de los mensajes de consulta) se utilizaría directamente el valor del word *embedding* resultante de la suma de los n-gramas con el corpus de entrenamiento. En caso de no existir dicha palabra en el corpus se podría descomponer dicha palabra en sus n-gramas, consultar el valor del *embedding* de los n-gramas en los resultados de los *embeddings* del corpus de entrenamiento con FastText, y sumar los *embeddings* de los n-gramas que hayan sido encontrados, ignorando los n-gramas que no tuviesen representación. De esta manera se consigue aproximar una representación para virtualmente todo tipo de palabras, a pesar de que no apareciesen en el corpus original.

Otra ventaja adicional de FastText con la representación de *embeddings* por la suma de los *embeddings* de los n-gramas es que se obtienen buenas representaciones incluso para palabras que aparezcan poco en el texto (o incluso hapaxes). Así pues, aunque haya una única aparición de la palabra *ingenieros* y muchas otras de la palabra *ingeniero*, en vez de tratarlas como dos palabras independientes y construir sus *embeddings* en base a los contextos en los que aparecen (que puede que sean bastante diferentes, más aún si una de ellas aparece poco) gracias a FastText se relacionarán ambos contextos al tener gran parte de sus n-gramas en común.

Se consigue además con relación a esto otra mitigación de sesgo por género. En los corpora de algunos idiomas puede ocurrir, por ejemplo, que haya muchas más referencias a *enfermeras* que a *enfermeros*, de manera que se tenga un sesgo en la construcción del *embedding* de la profesión por el género utilizado. Gracias a que los *embeddings* se construyen con los n-gramas la información de ambas palabras se relacionará ya que casi todos los n-gramas que tienen son n-gramas comunes.

Ahora bien, todos estos métodos de representación de *embeddings* tienen una carencia común y es que las representaciones de los *embeddings* son representaciones únicas para cada palabra donde no se hace ningún tipo de desambiguación semántica. Por lo tanto, por ejemplo, la

representación para *banco* será un único *embedding* que no distingue entre si es un *banco* como entidad financiera, un *banco* para sentarse o un *banco* de peces.

F. Embeddings contextualizados - ELMo

Para tratar de resolver el problema de la desambiguación anteriormente descrito, se han planteado soluciones dentro de lo que se conoce como ***embeddings contextualizados*** que precisamente buscan analizar los textos, ver los distintos usos de una misma palabra y representar cada uso distinto de la misma con *embedding* diferente como hace ELMo, *embeddings from language model design*, (Peters et al., 2018).

Para hacer esto, en vez de tener una tabla de consulta de palabras para ver su *embedding* como ocurría en el caso de las técnicas anteriores, ELMo define una red neuronal con la que obtendrá las representaciones de cualquier vector de *embedding* al vuelo a medida que se reciban textos nuevos.

Para ello se construye una arquitectura de *deep learning* (DL) con dos capas en la que cada una contiene dos niveles de subcapas con múltiples redes neuronales recurrentes (RNN) con *long-term short memory* (LSTM) para construir los *embeddings* leyendo el texto tanto de derecha a izquierda como de izquierda a derecha, utilizando lo que se denomina modelo bidireccional de lenguaje (biLM).

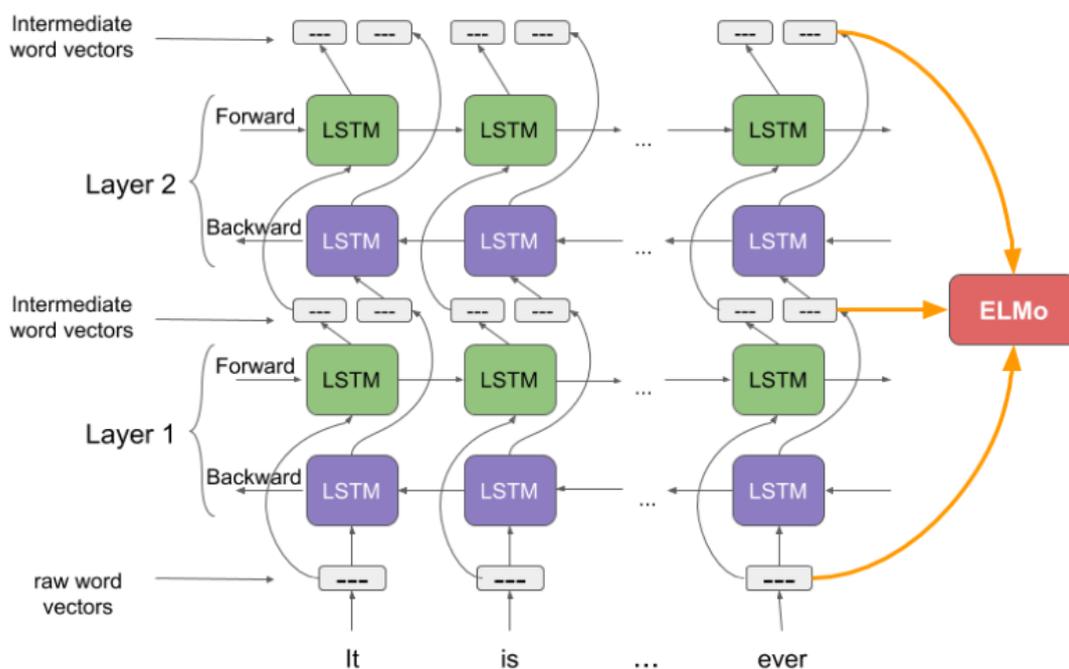


Ilustración 23 Arquitectura de ELMo³⁵

³⁵ Joshi, Prateek. "A Step-by-Step NLP Guide to Learn ELMo for Extracting Features from Text". Web. 29/04/2019. <https://www.analyticsvidhya.com/blog/2019/03/learn-to-use-elmo-to-extract-features-from-text/>

En el primer nivel de ELMo se lleva a cabo un *encoding* a nivel de caracteres con una técnica denominada Char CNN utilizando redes neuronales convolucionales, CNN, como describe el artículo (Zhang, X., Zhao, & Lecun, 2015). La idea consiste en partir de un vector que identifique a cada uno de los caracteres presentes en una palabra (por ejemplo, un *one-hot encoding* para cada caracter) y que así cada palabra se represente con una matriz con los vectores de cada caracter.

Sobre dicha matriz se aplicaría una CNN de la siguiente manera³⁶.

- En primer lugar se parte de la matriz de *encodings* por carácter (como la de la imagen siguiente para la palabra *absurdity*).

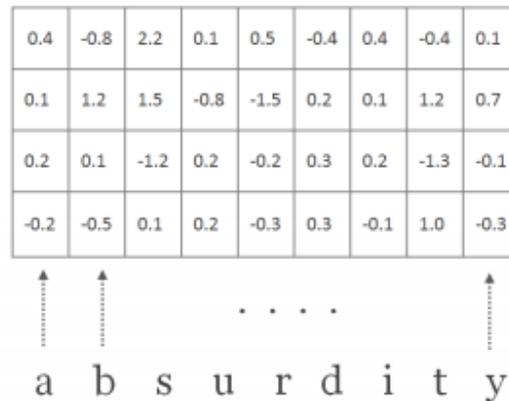


Ilustración 24 Representación matricial de los caracteres usando 4 dimensiones para cada uno

- Posteriormente se aplica una matriz de filtrado sobre la matriz de representación de caracteres. Esta matriz de filtrado tiene tantas columnas como caracteres consecutivos a considerar.

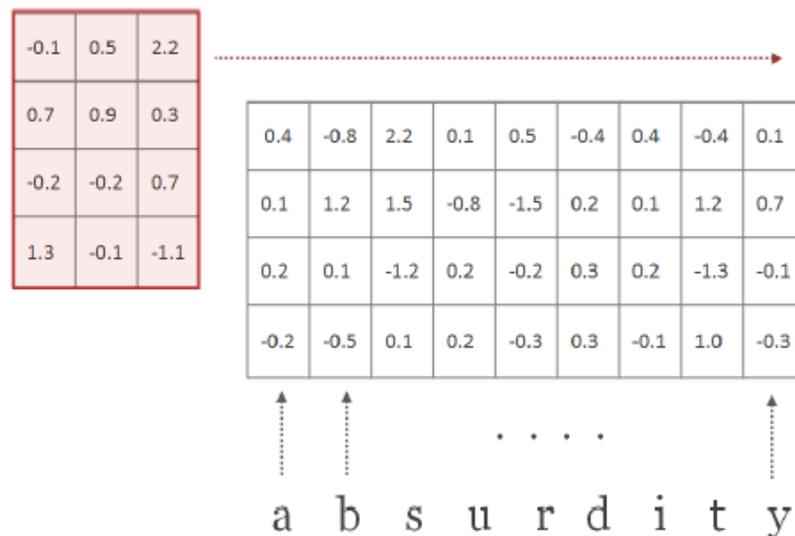


Ilustración 25 Matriz kernel o de filtrado que se utilizará en este ejemplo

³⁶Kim, Yoon et al. "Character-Aware Neural Language Models". Web. 29/04/2019. <https://nlp.seas.harvard.edu/slides/aaai16.pdf>

- A continuación se aplica el filtrado de las CNN de manera que se va recorriendo la matriz de entrada en función del *padding* y utilizando la matriz de filtrado mostrada previamente.³⁷



Ilustración 26 A medida que se recorre la matriz de entrada con un kernel concreto se van extrayendo valores en función del producto de la matriz de filtrado por esa subsección indicada

- A su vez, los vectores extraídos se pueden reducir aún más en dimensionalidad aplicando sobre ellos un *pooling*

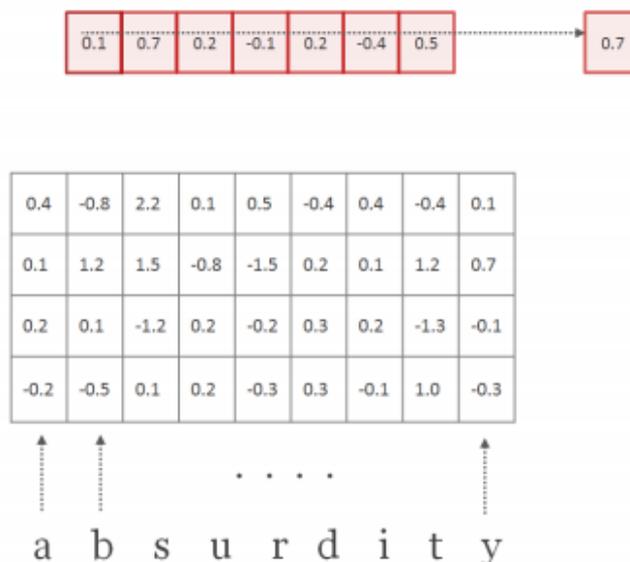


Ilustración 27 Se aplica un pooling sobre el vector extraído (en este caso max pooling)

³⁷ Nota: El valor resultante de sumar la matriz obtenida de multiplicar la matriz de caracteres y el kernel (multiplicación a nivel elementos) no se corresponde con lo que aparece en las imágenes. Por ejemplo, en vez de 0.7 como aparece para ese kernel debería ser 2.83, y en vez de 0.1 ser 4.78. Así pues, esta imagen es meramente ilustrativa de cómo funciona el proceso.

- Se extraen varias matrices de filtrado usando distintos kernels

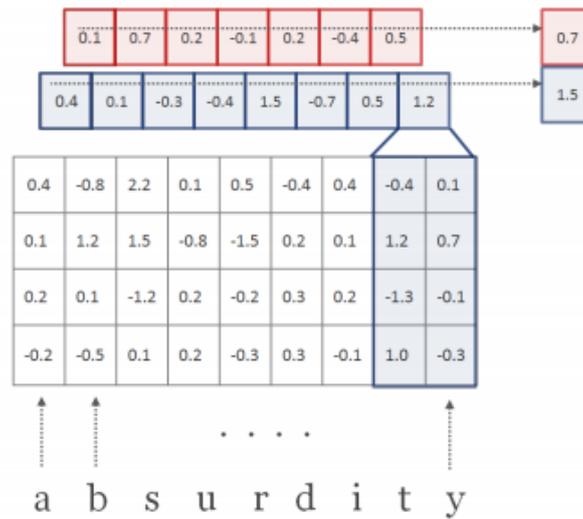


Ilustración 28 Aplicación iterativa de otro filtro sobre los caracteres.

- Para tener con ello un vector final al hacer el *pooling* sobre los resultados de todas ellas

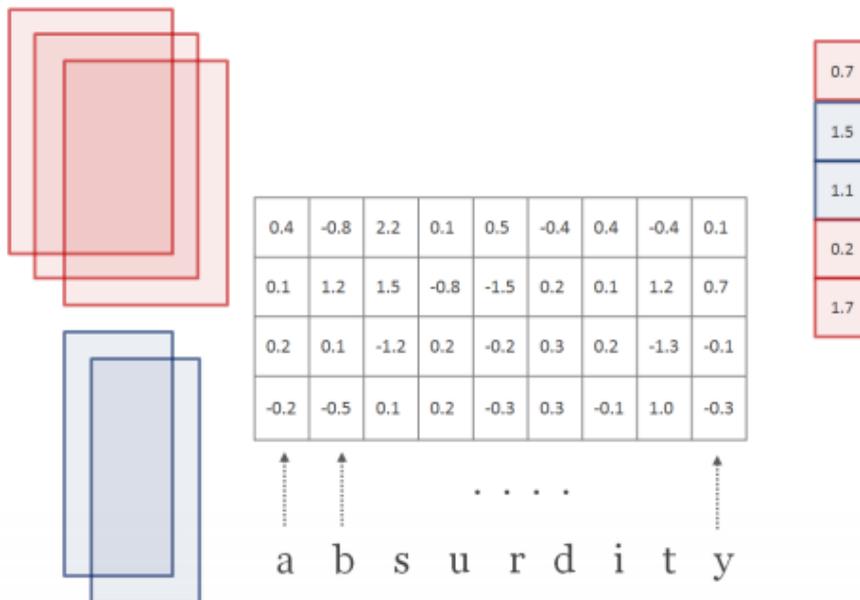


Ilustración 29 Vector final resultante tras aplicar varias matrices kernel y varias iteraciones

- Y así, se tiene finalmente una representación de la palabra tras haber partido de las representaciones de sus caracteres

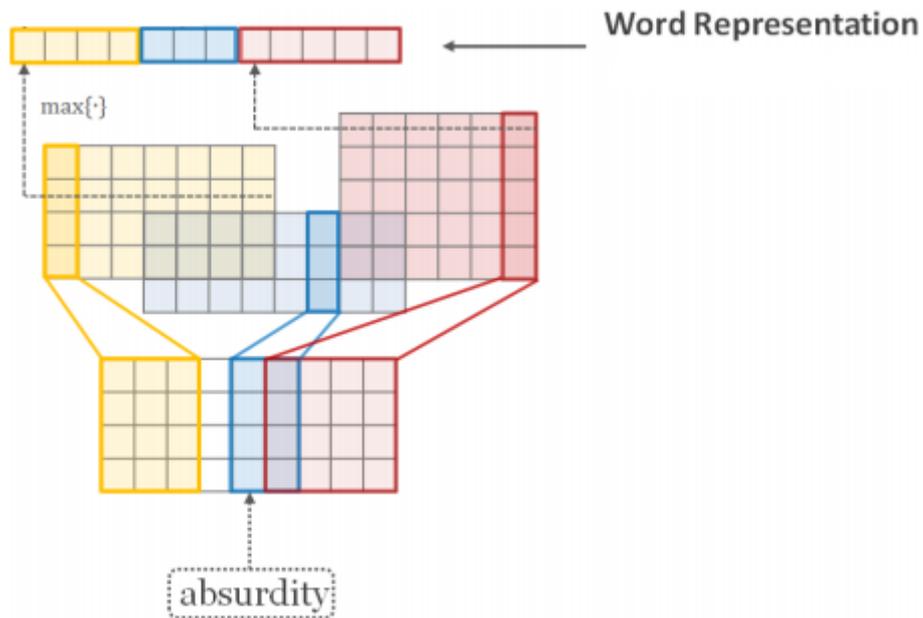


Ilustración 30 Representación vectorial final para la palabra

- Por último, sobre la capa final que tiene la representación de palabras se puede aplicar un término de *bias* y una función de transformación no lineal para ganar generalidad

$$\tanh \left(\begin{array}{c} 0.7 \\ 1.5 \\ 1.1 \\ 0.2 \\ 1.7 \end{array} + b \right) = \begin{array}{c} 0.8 \\ 1.0 \\ 0.9 \\ 0.5 \\ 1.1 \end{array}$$

Ilustración 31 Bias para ajustar el vector resultante

La ventaja de construir *embeddings* en base a los caracteres de las palabras individuales (y no partiendo directamente de las palabras) es que se reduce el problema ya mencionado de trabajar con palabras que no aparezcan en el vocabulario inicial (OOV, *out of vocabulary*), ya que al construir los *embeddings* partiendo de la agregación de los *embeddings* de los caracteres es más fácil llegar a tener algún tipo de representación siempre (de nuevo, si todos los textos hablasen sólo de *ingenieros*, sin ninguna mención a *ingenieras*, que la representación inicial de *ingenieras* fuese la que aportan los caracteres *ingeniers*, más el de *a*, no está demasiado mal ya que será muy próxima a la de *ingenieros*).

La obtención de estos *embeddings* iniciales para cada palabra se hace en base a los *embeddings* de cada carácter que la compone de manera aislada. Es decir, se parte de una matriz con los *embeddings* por carácter y se reduce la dimensionalidad de esta obteniéndose un único vector gracias al uso de CNN, de manera análoga al funcionamiento de estas para la extracción de las *features* más relevantes en una imagen. Esto tiene sus ventajas: los vectores de partida para

palabras similares, como *bello* y *belleza* serán muy similares básicamente por el gran número de caracteres que tienen en común (y la posición de estos, en este caso, la raíz), relacionándose así significados antes ya de enriquecer los *embeddings* con un análisis por el contexto.

Estos vectores obtenidos tras la Char CNN se utilizan como entrada para la biLM tanto en su parte *backward* como en su parte *forward*.

La idea es sencilla; por ejemplo, para el caso *forward*, el *embedding* de la primera palabra se utiliza para obtener la entrada de la LSTM. Dicho vector de entrada (x_t) se multiplicará por sus correspondientes pesos de la matriz de pesos para obtener la entrada de la LSTM (i_t), que se combinará con la salida generada por la capa anterior (h_{t-1}) para generar tanto el vector de salida (*embedding*) de dicha palabra como la entrada h_t que se usará para obtener el *embedding* de la siguiente.

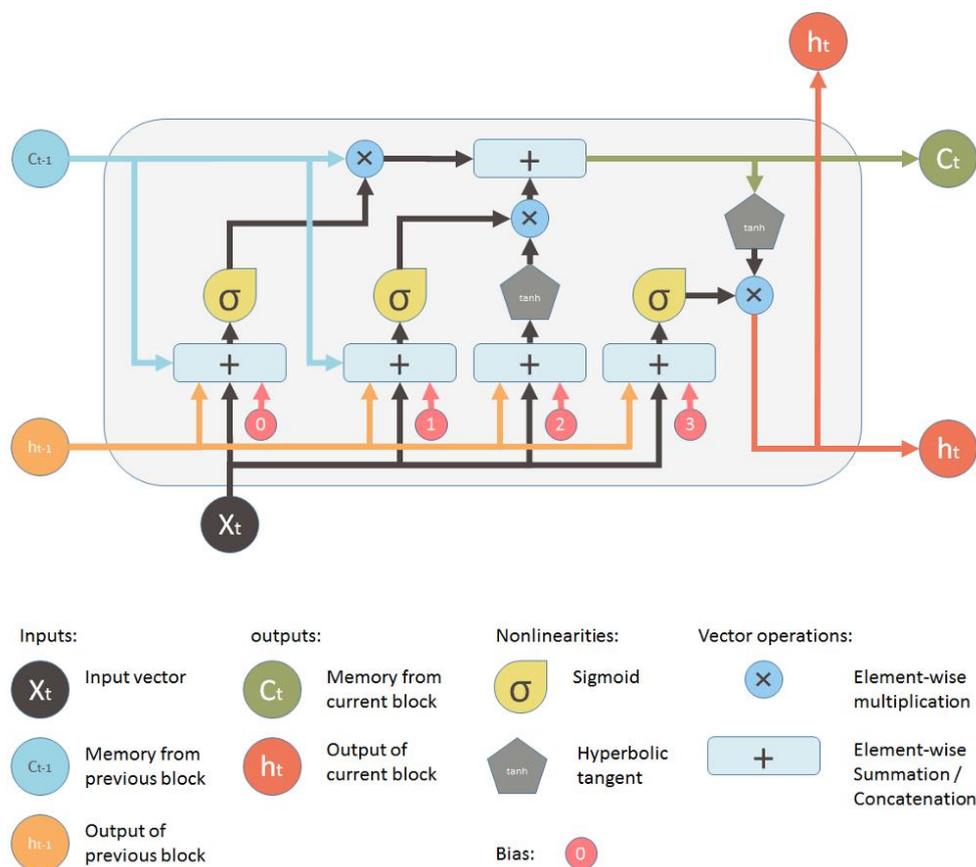


Ilustración 32 Estructura interna de una LSTM³⁸

³⁸ Yan, Shi. "Understanding LSTM and its diagrams". Web. 02/05/2019. <https://medium.com/mlreview/understanding-lstm-and-its-diagrams-37e2f46f1714>

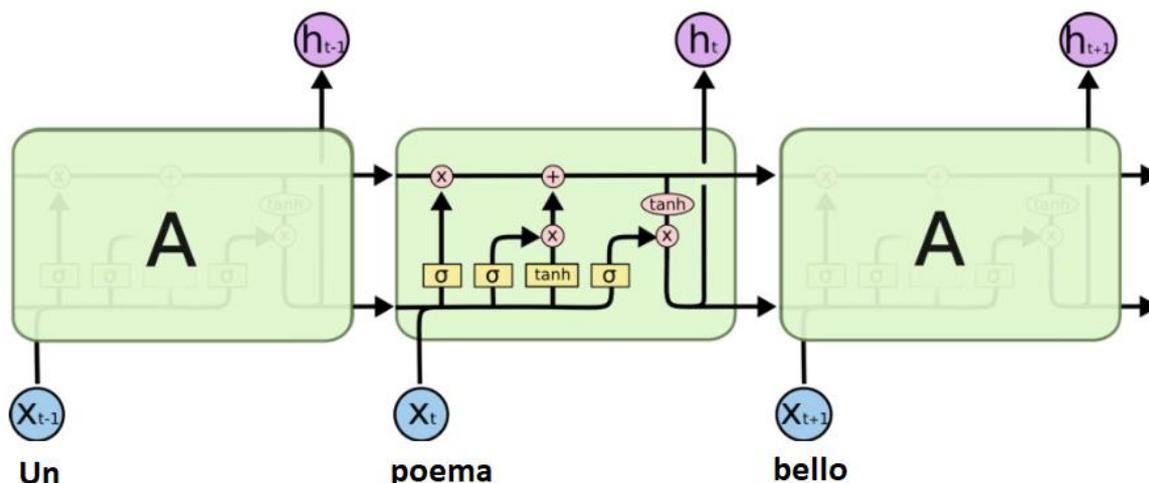


Ilustración 33 Conexión entre varias LSTM³⁹

En la Ilustración 34 se muestra como usando las distintas etapas de conexión de las RNN, para el modelado de lenguaje se intenta predecir la siguiente palabra usando una capa *softmax* sobre el vector resultante de la capa oculta. El sistema intenta optimizar los pesos de toda la red neuronal para predecir lo mejor posible esa siguiente palabra dada toda la información inicial.

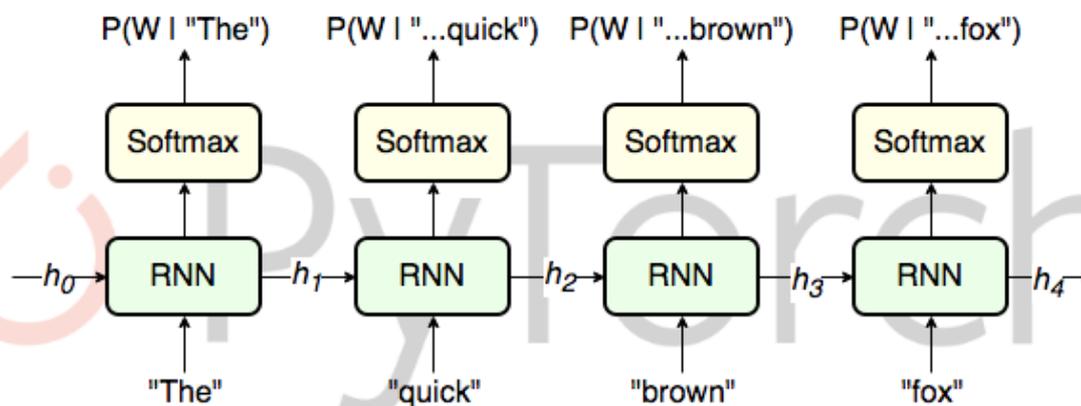


Ilustración 34 RNN para modelado de lenguaje.⁴⁰

Así, el problema que define ELMo es un problema de *modelado de lenguaje* en el que la red neuronal del sistema es entrenada en un corpus enorme de textos que usará para intentar predecir las palabras siguientes o previas (dependiendo de si es la capa *forward* o *backwards*) dadas el resto de las palabras.

³⁹ “Understanding LSTM Networks”. Web.02/05/2019. <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

⁴⁰ Stamenković, Florijan. “RNN Language Modelling with PyTorch—Packed Batching and Tied Weights”. Web. 02/05/2019. https://medium.com/@florijan.stamenkovic_99541/rnn-language-modelling-with-pytorch-packed-batching-and-tied-weights-9d8952db35a9

$$\text{Forward: } p(t_1, t_2, \dots, t_N) = \prod_{k=1}^N p(t_k | t_1, t_2, \dots, t_{k-1})$$

$$\text{Backward: } p(t_1, t_2, \dots, t_N) = \prod_{k=1}^N p(t_k | t_{k+1}, t_{k+2}, \dots, t_N)$$

Ilustración 35 Las RNN en la etapa forward intenta predecir la palabra siguiente dada una secuencia previa de palabras mientras que la backward intenta predecir la palabra previa a una secuencia dada

Así, por ejemplo para el caso *forward*, la LSTM partirá del resultado de la *Char CNN* para la primera palabra e intentará predecir el resultado del *embedding* del *Char CNN* de la segunda palabra. Los pesos de las LSTM se ajustarán en aras de optimizar la predicción de la secuencia de palabras.

En la Ilustración 36 se ve cómo en el caso forward se intenta predecir el vector de entrada de la siguiente palabra de la secuencia dada la información de las palabras anteriores. La entrada para la primera palabra de la secuencia suele ser el identificador de un token que especifique que comienza ahí el texto y que la siguiente palabra es la primera (p.e. CLS)

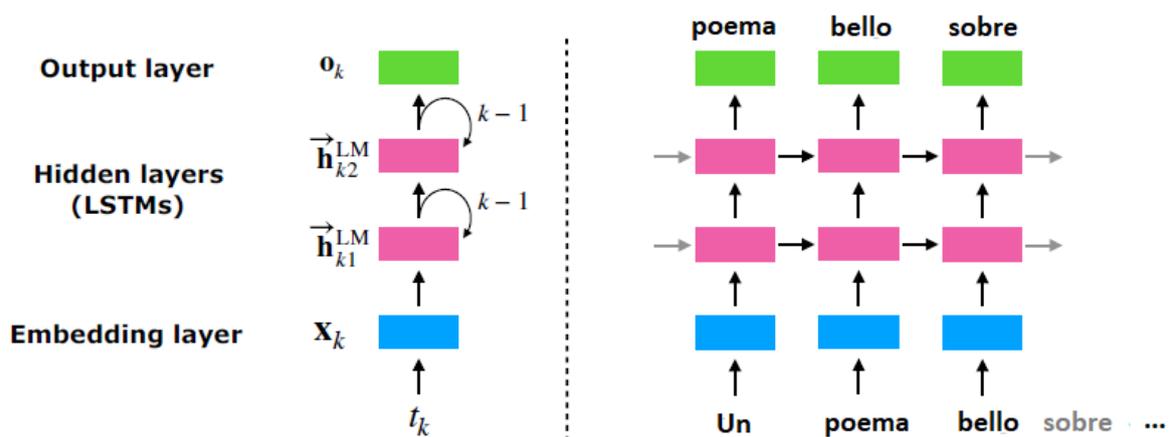


Ilustración 36 Ejemplo para el caso forward⁴¹.

En la ilustración 37 aparece cómo la RNN tiene para cada etapa una LSTM pero que comparte los pesos para cada etapa/palabra; esto es así para que el número de parámetros no se dispare con el número de etapas (que es lo que ocurriría si cada etapa tuviese sus propios pesos) y para que se pueda utilizar el sistema para cualquier tamaño de la secuencia de entrada (en donde simplemente se especificaría con un token especial cuando empiezan y cuando acaban las secuencias, pero como los pesos son compartidos por cada etapa podrían tener cualquier tamaño).

⁴¹ Fuente: Propia

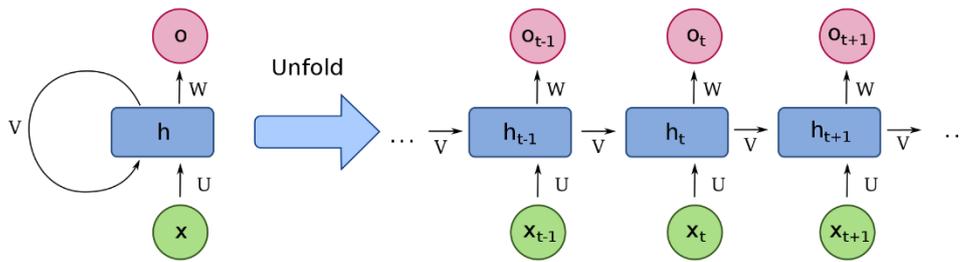


Ilustración 37 Expansión de los componentes internos de la RNN⁴²

ELMo finalmente obtendrá el vector de embedding para cada palabra concatenando los vectores resultantes en la *capa oculta* para *forward* y *backward* y haciendo una suma ponderada de los vectores compuestos para cada nivel de la arquitectura, como se mostró en la Ilustración 23 y como se detalla en la siguiente imagen.

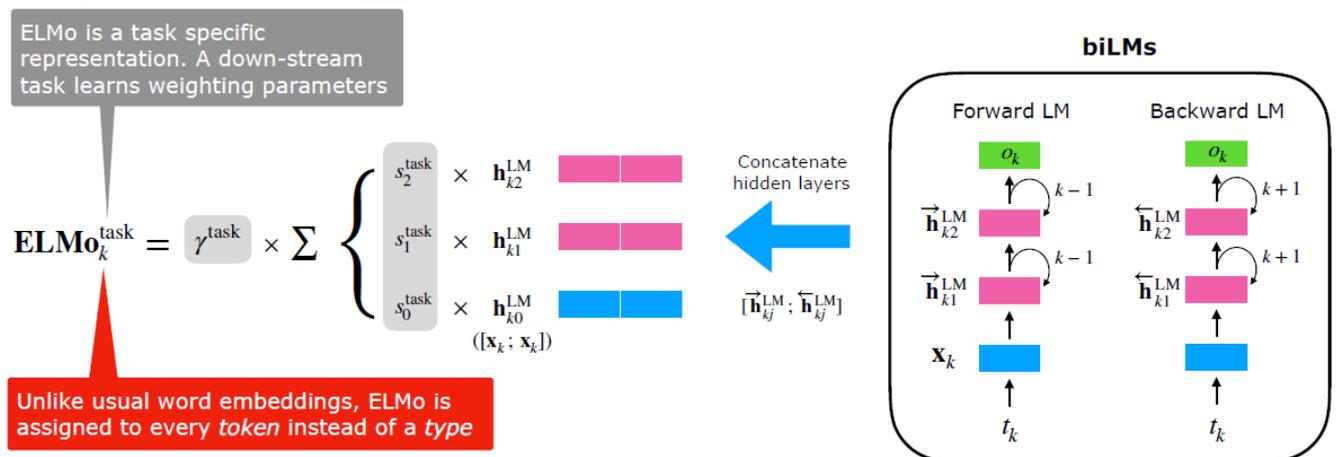


Ilustración 38 Obtención de embeddings y suma ponderada de los mismos.⁴³ Se suman para ello los vectores de forward y backwards combinados de cada capa junto con la entrada del embedding resultante del CharCNN

Como se puede ver, este sistema no define una tabla de consulta (*lookup table*) sino que define una arquitectura de redes neuronales que irá recibiendo frases de entrada y obteniendo, de forma contextualizada, el *embedding* para cada palabra que la compone.

La parte positiva de estos sistemas es que, además de ofrecer un *embedding* contextualizado, a diferencia de word2vec no se está usando un contexto local para obtener el *embedding*, sino que se utiliza el texto en su totalidad. Esto tiene varias ventajas, entre las que destaca el poder incorporar información de dependencias más lejanas entre palabras.

⁴² “Recurrent neural network”. Web. 02/05/2019. https://en.wikipedia.org/wiki/Recurrent_neural_network

⁴³ Peters, Matthew E. et al. “Deep contextualized word representations”. Web. 02/05/2019. https://www.slideshare.net/shuntaroy/a-review-of-deep-contextualized-word-representations-peters-2018?from_action=save

El esquema de ELMo ha sido uno de los primeros que ha planteado un nuevo tipo de soluciones para la operativización de la comprensión del lenguaje natural. Así, en vez de obtenerse unas tablas de consulta para *embeddings* fijas sobre la que extraer un vector de representación numérico, se construyen modelos de aprendizaje profundo que son ajustados a un conjunto de palabras y que incluso luego permiten un ajuste fino (*fine tuning*) para cada caso de uso en concreto. De hecho, se considera que esto ha supuesto un hito fundamental en el desarrollo de soluciones para NLP comparable a lo que fue la creación de ImageNet para el avance de la visión artificial, ya que se ofrecen todos estos modelos pre-entrenados, llevándose con ello a cabo un efectivo *transfer learning*, que permite trabajar con ellos directamente o adecuarlos a los escenarios particulares para cada caso, pero con todo ese conocimiento de partida ya embebido en ellos.

Esto ha dado lugar a otras soluciones, además de ELMo, como son Transfer, la solución de OpenAI, ULMFiT o, como se comentará a continuación más en detalle por ser la última técnica propuesta y que intenta recoger todas las aportaciones anteriores, BERT.

G. Embeddings contextualizados - BERT

BERT (*Bidirectional Encoder Representations from Transformers*), recogido en el artículo de (Devlin, Chang, Lee, & Toutanova, 2018), plantea otro esquema de modelado de lenguaje para distintas tareas de NLP (entre las que aparecería la generación de *embeddings*) con una idea que continúa algo que ya se planteó en ELMo, la bidireccionalidad en el análisis de las frases para generar los *embeddings* contextualizados. La idea de ELMo era usar una capa *forward* y otra *backwards* para así contextualizar mejor las palabras al no caer en el esquema habitual de modelado de lenguaje en donde el análisis siempre se centra en estudiar la secuencia de palabras en una única dirección (de izquierda a derecha generalmente). Esto es relativamente importante, ya que en muchas ocasiones lo que ayuda a contextualizar y desambiguar una palabra no es toda la secuencia de palabras previa sino muchas veces la o las palabras que vienen a continuación.

La idea aparece de nuevo en BERT pero con una aproximación distinta: en vez de analizar por separado la secuencia de izquierda a derecha y viceversa para luego combinar los resultados, BERT plantea un análisis *no secuencial* en el que se recoge la información global de todas las palabras de la secuencia al mismo tiempo, sin tener que analizarlas una a una, de manera que se solventa de forma implícita la cuestión de la bidireccionalidad.

Esta idea para capturar la información y dependencias de las palabras en los textos se lleva a cabo con una estructura denominada *Transformer*, de la que se menciona que es más eficiente en capturar dicha información que las estructuras secuenciales utilizadas en soluciones, como ELMo, que se basan en ese uso de RNNs. Para ello han comprobado los resultados obtenidos frente a otras soluciones como ELMo u otras posteriores sobre distintos corpus y para distintas tareas de NLP, obteniéndose resultados sorprendentemente buenos.

| Tasks | Dev Set | | | | |
|----------------------|-----------------|---------------|---------------|----------------|---------------|
| | MNLI-m (Acc) | QNLI (Acc) | MRPC (Acc) | SST-2 (Acc) | SQuAD (F1) |
| BERT _{BASE} | 84.4 | 88.4 | 86.7 | 92.7 | 88.5 |
| No NSP | 83.9 | 84.9 | 86.5 | 92.6 | 87.9 |
| LTR & No NSP | 82.1 | 84.3 | 77.5 | 92.1 | 77.8 |
| + BiLSTM | 82.1 | 84.1 | 75.7 | 91.6 | 84.9 |

Ilustración 39 Comparativa de las soluciones de BERT base frente a otras técnicas del estado del arte para distintas tareas de NLP.⁴⁴

El modelado de lenguaje que se lleva a cabo, por lo tanto, no es el de predecir la siguiente palabra dadas las palabras previas o posteriores, sino que la solución que se implementa en BERT es la de enmascarar una serie de palabras (un 15% de las palabras del texto original) e intentar predecir dichas palabras enmascaradas usando conjuntamente toda la información previa y posterior a las mismas (es decir, todo el contexto de la frase, a diferencia de los contextos limitados usados en word2vec).⁴⁵

En la Ilustración 40 aparece un ejemplo donde se ve como se enmascara una palabra del input y BERT obtiene vectores para todas las palabras de manera que se intente optimizar que la predicción de la palabra enmascarada sea la palabra correcta. En la Ilustración 41 se muestra un ejemplo de análisis de una frase con el *Transformer* de BERT. En el esquema del *Transformer* con palabras enmascaradas de BERT se analiza de forma conjunta toda la frase, a diferencia de ELMo donde se analizan por separado con las LSTM secuenciales las secuencias previas o posteriores a la palabra a predecir.

⁴⁴ “Paper Dissected: “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding” Explained”. Web. 03/05/2019. <https://mlexplained.com/2019/01/07/paper-dissected-bert-pre-training-of-deep-bidirectional-transformers-for-language-understanding-explained/>

⁴⁵ Los autores del artículo, en realidad, hablan de que ese 15% de palabras enmascaradas se descompone en un 80% de él con palabras enmascaradas (p.e. “El poema es <MASK>”), un 10% de ellas con un token aleatorio (p.e. “El poema es *salta*”), y un 10% de ellas con el token real (“El poema es *bello*”); de esta manera afirman que se obliga a la estructura a que aprenda el contexto de *todas* las palabras que aparecen en los textos, y no sólo de las que están enmascaradas. Ciertamente se corre el peligro de introducir ruido al incorporar los tokens aleatorios, pero debido a que el porcentaje de ellos es muy pequeño (un 1.5%) el ruido no es casi significativo y según los autores no afecta al rendimiento del modelo.

Use the output of the masked word's position to predict the masked word

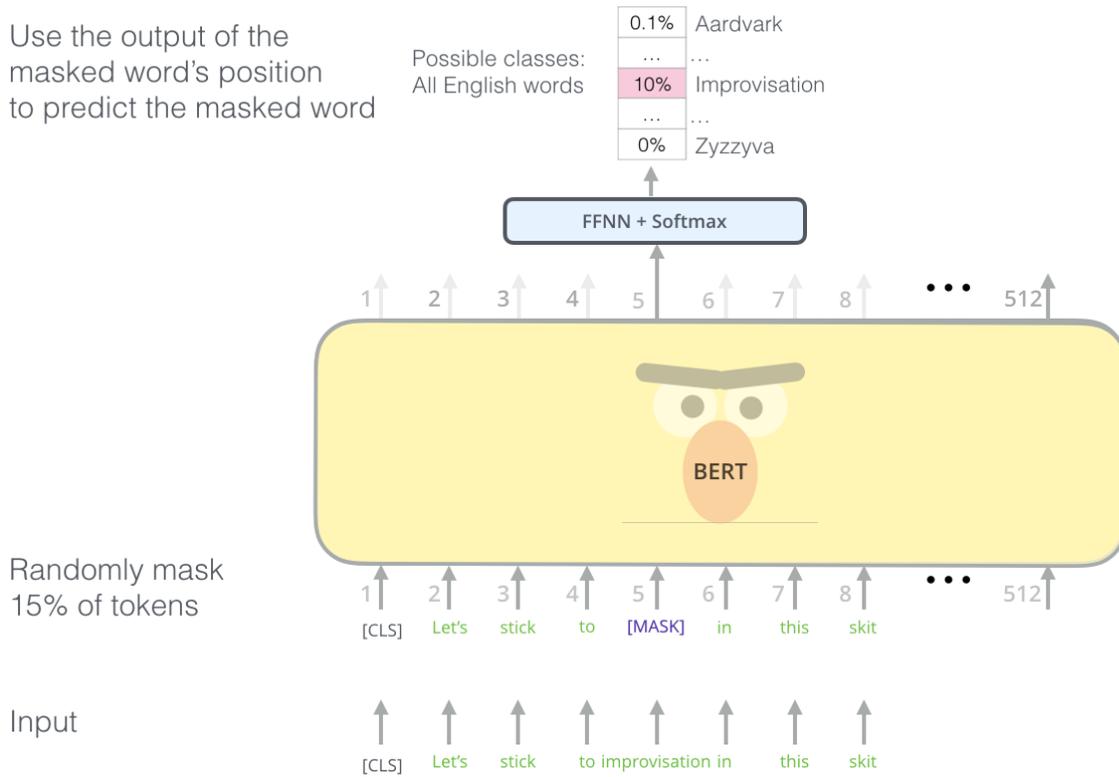


Ilustración 40 Ejemplo de enmascaramiento del input en BERT.⁴⁶



Ilustración 41 Transformer en BERT aplicado al análisis de una frase.⁴⁷

En la Ilustración 42 se indican los dos posibles modelos existentes. Estos son de tamaño sensiblemente distinto. En el esquema *large BERT* contiene 24 bloques con Transformers, 1024 capas ocultas, lo que da lugar a 340 millones de parámetros. El modelo es entrenado usando Wikipedia (sólo en su versión inglesa son 2.5 billones de palabras) y *BookCorpus* (800 millones de palabras). El esquema base tiene vectores de 768 parámetros, 12 capas y 12 *heads* para la capa *MultiHead Attention* (frente a las 16 *heads* en *large*).

⁴⁶ Almmar, Jay. "The Illustrated BERT, ELMo, and co. (How NLP Cracked Transfer Learning)". Web. 03/05/2019. <http://jalammar.github.io/illustrated-bert/>

⁴⁷ "Paper Dissected: "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding" Explained". Web. 03/05/2019. <https://mlexplained.com/2019/01/07/paper-dissected-bert-pre-training-of-deep-bidirectional-transformers-for-language-understanding-explained/>

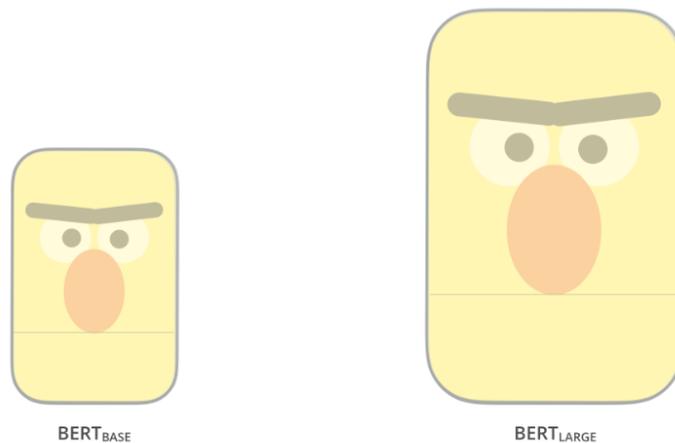


Ilustración 42 Existen dos posibles modelos, de tamaño sensiblemente distinto.⁴⁸

Además de todo esto se plantea en el sistema una tarea de *predicción de frases siguientes* para capturar, junto con todo lo anterior, las dependencias que existen entre distintas frases.

En la Ilustración 43 se muestra cómo además de optimizar la predicción de las palabras enmascaradas BERT intenta optimizar la predicción de la frase siguiente a un texto determinado; ante una secuencia del texto se usa un clasificador binario para detectar si esa frase es la siguiente a otra frase dada o no lo es.

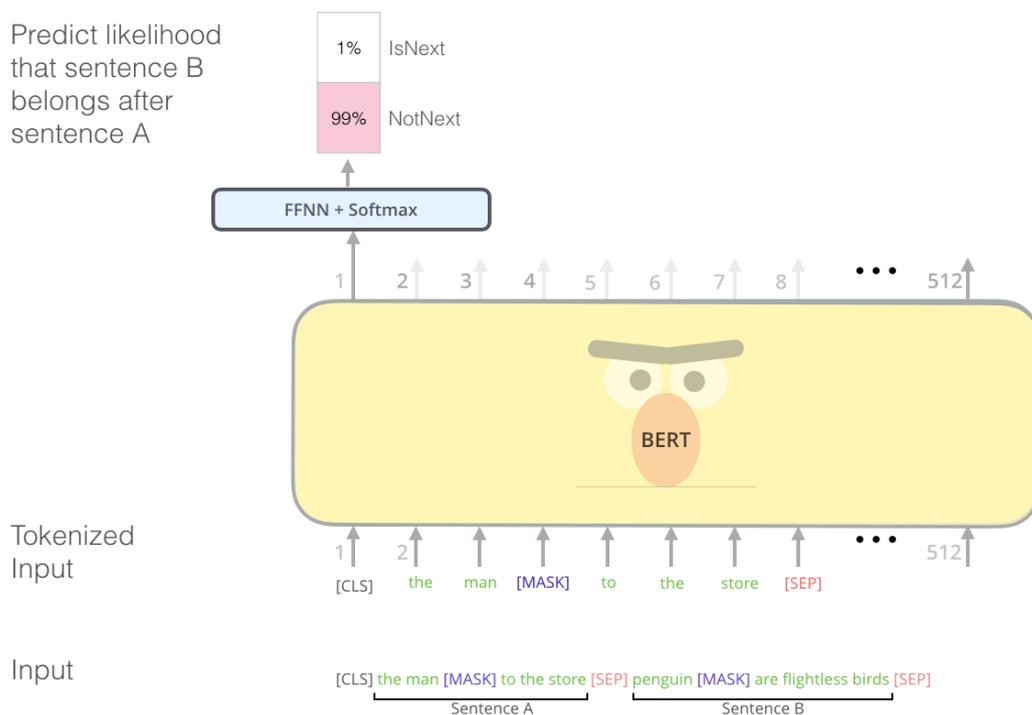


Ilustración 43 Además de optimizar la predicción de las palabras enmascaradas BERT intenta optimizar la predicción de la frase siguiente a un texto determinado.⁴⁹

⁴⁸ Almmar, Jay. "The Illustrated BERT, ELMo, and co. (How NLP Cracked Transfer Learning)". Web. 03/05/2019. <http://jalammar.github.io/illustrated-bert/>

⁴⁹ Almmar, Jay. "The Illustrated BERT, ELMo, and co. (How NLP Cracked Transfer Learning)". Web. 03/05/2019. <http://jalammar.github.io/illustrated-bert/>

Bajando a más detalle, la estructura que aparece en BERT para el encoding de palabras es el uso de distintos *encoders* en serie, cada uno de los cuales daría un vector numérico resultante para el vector numérico de entrada a él. Estos encoders son, de hecho, una de las estructuras internas que aparecen en el *Transformer* y que crean el *embedding* necesario en base a la secuencia de entrada para obtener la secuencia de entrada necesaria en otro de sus componentes, el *decoder*, de manera que se puedan realizar tareas de transformación de una secuencia a otra, algo que sirve para aplicaciones de NLP como la *traducción automática*.

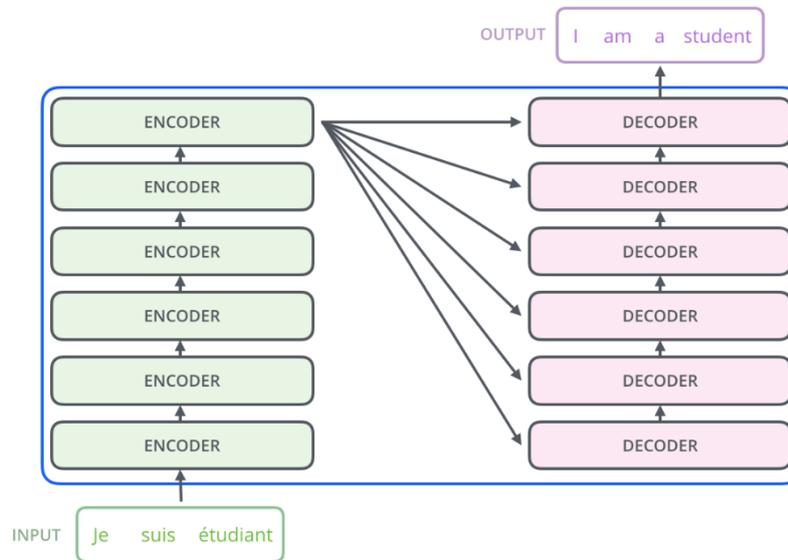


Ilustración 44 Aplicación de un Transformer para una tarea de traducción automática de secuencias de textos.⁵⁰

Para el caso de BERT se aprovecha esta idea pero para llevar a cabo el encoding de las frases iniciales a una representación numérica eficiente y contextualizada de las palabras que lo componen se usa exclusivamente la parte del encoder, utilizándose para los *embeddings* de las palabras las distintas salidas que aparecen tras cada encoder.

⁵⁰ Fuente: Última visita: 03/05/2019 Almmar, Jay. “The Illustrated Transformer”. Web. 03/05/2019. <https://jalammr.github.io/illustrated-transformer/>

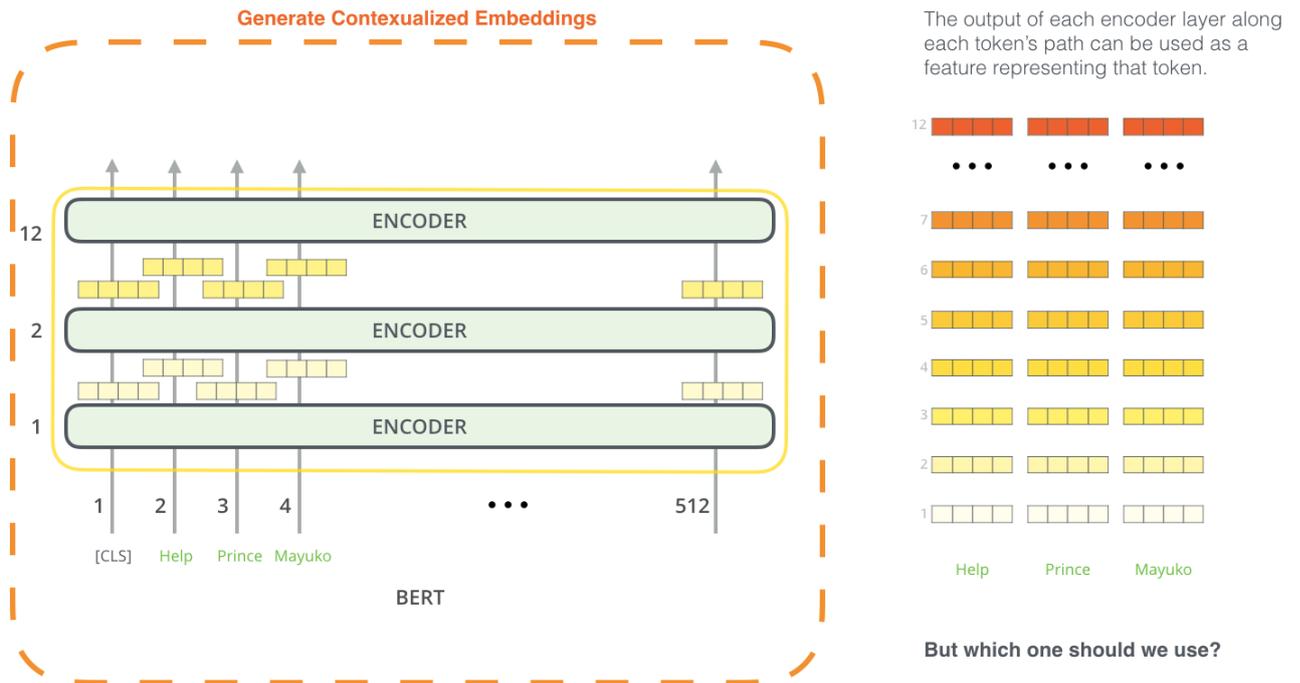


Ilustración 45 En la versión base de BERT se usan 12 encoders en serie con 768 parámetros de entrada/salida en cada uno. En la versión large son 24 encoders con vectores de entrada/salida de 1024 componentes.⁵¹

Ahora bien, la cuestión es qué salida de los *encoders* usar para la representación como *embedding* de cada palabra dentro de una secuencia. Los autores del artículo plantean distintas soluciones, como muestra la Ilustración 46.

⁵¹ Almmar, Jay. “The Illustrated BERT, ELMo, and co. (How NLP Cracked Transfer Learning)”. Web. 03/05/2019. <http://jalmmar.github.io/illustrated-bert/>

What is the best contextualized embedding for “Help” in that context?
 For named-entity recognition task CoNLL-2003 NER

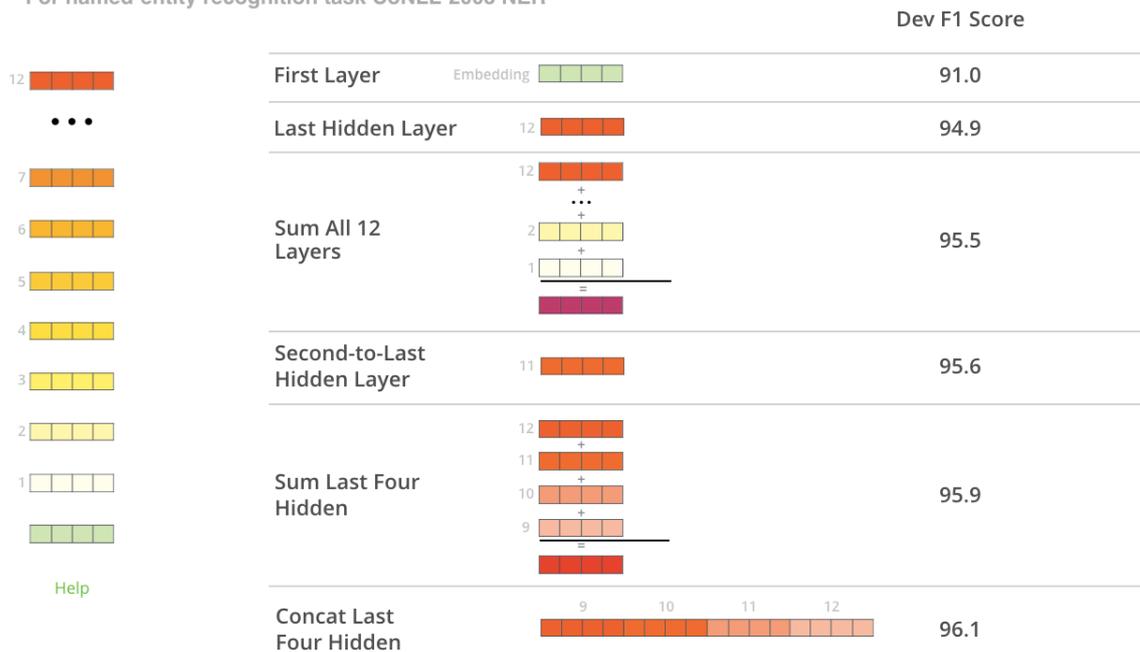


Ilustración 46 Distintas posibilidades de combinación de los embeddings de los encoders para la representación final junto con la efectividad de estas.⁵²

Como se puede ver en la Ilustración 46, los mejores resultados para el caso de BERT base (12 encoders) se consiguen concatenando los *embeddings* de los 4 últimos *encoders*. Los resultados son muy parecidos, eso sí, al de hacer la suma de las cuatro últimas capas. En general, la mayoría de las librerías que incorporan este modelo, como Gluon NLP⁵³, utilizan por simplicidad esta última aproximación por tenerse con ella una dimensionalidad de los vectores de *embeddings* mucho más reducida que además coincide con las dimensiones de entrada del modelo (768 para el caso Bert base, 1024 para el BERT large).

Cabe mencionar que la estructura interna de cada uno de esos encoders se compone de unas ANN junto con una capa *Self-Attention*, como muestra la Ilustración 47.

⁵² Almmar, Jay. “The Illustrated BERT, ELMo, and co. (How NLP Cracked Transfer Learning)”. Web. 03/05/2019. <http://jalanmar.github.io/illustrated-bert/>

⁵³ Reference: Devlin, Jacob, et al. “Bert: Pre-training of deep bidirectional transformers for language understanding.” arXiv preprint arXiv:1810.04805 (2018).

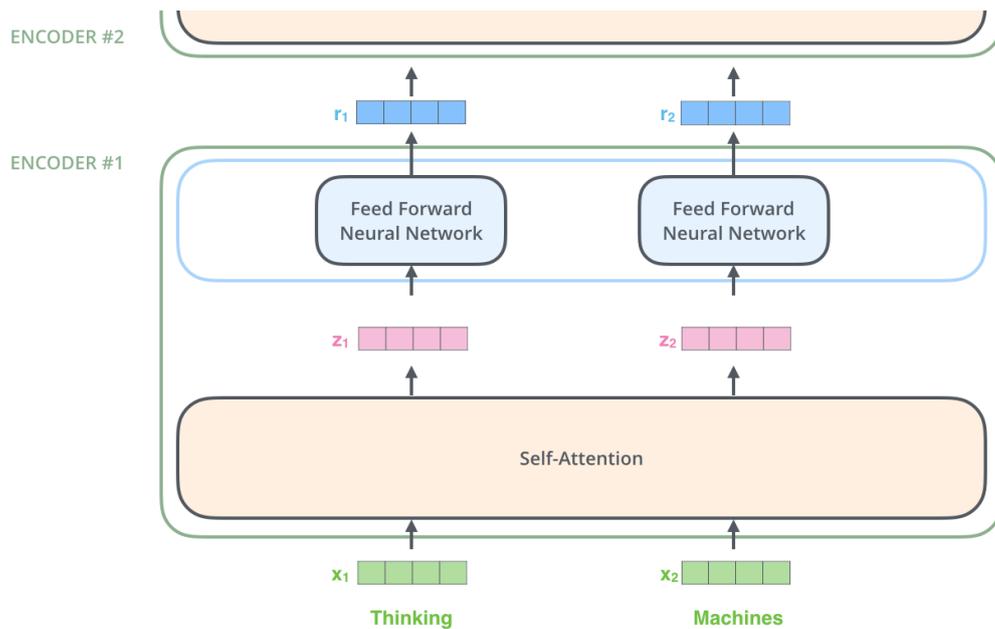


Ilustración 47 Estructura interna de un Encoder.⁵⁴

Dicha capa *Self-Attention* es una de las mayores aportaciones del *Transformer* y gracias a ella se consigue la comprensión de cada palabra de la secuencia con relación a todas las demás (anteriores y posteriores). Esto sirve para, por ejemplo, encontrar las dependencias entre determinadas palabras, como pronombres, con los sustantivos a los que hacen referencia, de manera que se mejore la comprensión del contexto (ya que el *embedding* asociado a dichos pronombres coincidirá con el del sustantivo correspondiente y dicha información se usará para la obtención de los *embeddings* del resto de palabras de la secuencia).

⁵⁴ Fuente: Última visita: 03/05/2019 Almmar, Jay. "The Illustrated Transformer". Web. 03/05/2019. <https://jalammar.github.io/illustrated-transformer/>

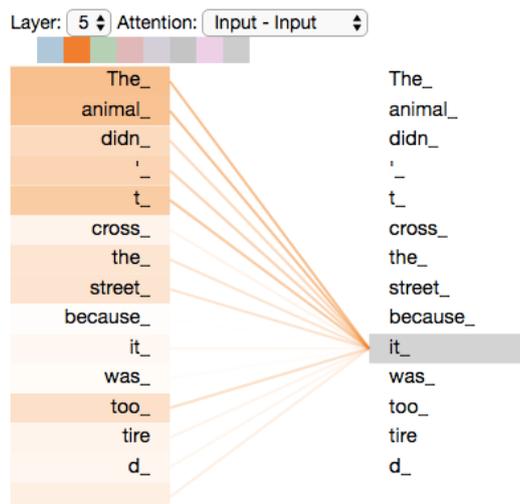


Ilustración 48 Ejemplo de cómo con la capa Self-Attention se encuentra que el pronombre 'it' hace referencia en este caso a la palabra 'animal'⁵⁵

El problema que aparece con el uso de los *encoders* es que, aunque estos analizan la frase en su conjunto, si una misma palabra aparece en distintas posiciones dentro de la misma la analizará como un mismo token, sin distinguir sobre la posición de ésta. Por este motivo BERT incorpora algunos metadatos para indicar cuándo comienza o acaba una frase, así como *embeddings* posicionales que especifiquen la ubicación concreta de cada palabra de la secuencia. Toda esta información es la que define el vector de entrada al sistema para cada palabra.

En la Ilustración 49 se muestra la estructura de metadatos de BERT donde se añade la información sobre la posición de cada palabra de la secuencia. Este es el esquema de entrada a toda la estructura de BERT. El token *embedding* corresponde a algún *embedding* inicial como por ejemplo word2vec.

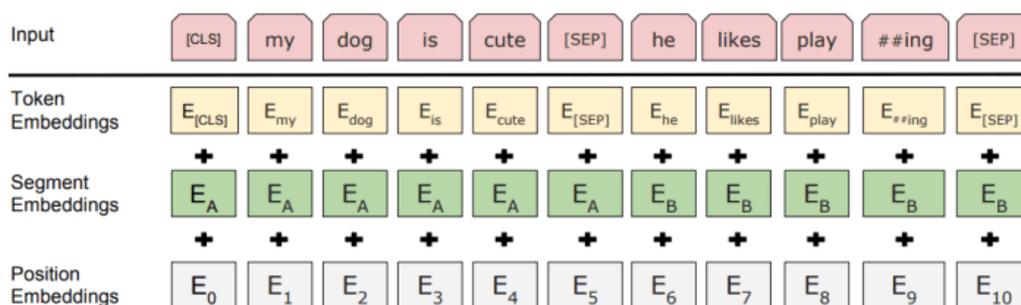


Ilustración 49 Estructura de metadatos de BERT

⁵⁵ Fuente: Última visita: 03/05/2019 Almmar, Jay. “The Illustrated Transformer”. Web. 03/05/2019. <https://jalammar.github.io/illustrated-transformer/>

Por último, hay que mencionar que la arquitectura de BERT ofrece muchas soluciones para distintas tareas de NLP; se puede usar como entrada directamente para una tarea de aprendizaje supervisado (clasificación), para analizar la similitud entre dos textos... o como preentrenamiento (*pre-training*) sirviendo para simplemente extraer los *embeddings* de las palabras de una secuencia y utilizarlos para otras tareas posteriores, como es el caso de lo que interesa en esta memoria. BERT permite incluso hacer un ajuste fino (*fine tuning*) con el que se parta del modelo preentrenado y se usen unas ANN para ajustar los *embeddings* de partida (que son fruto de los datos de Wikipedia principalmente) al contexto con el que se esté trabajando (esto es especialmente útil cuando los textos del contexto con los que se trabaje difieran mucho en expresión de lenguaje respecto a los textos de Wikipedia).

En la Ilustración 50 se muestra el esquema de BERT. Por un lado aparece el modelo preentrenado con el corpus de Wikipedia, de manera que dicho modelo se puede usar directamente para la resolución de muchos problemas que involucren NLP, como por ejemplo obtener *embeddings* para palabras concretas. Además del modelo preentrenado aparece la posibilidad de usar BERT directamente como clasificador para problemas de aprendizaje supervisado, como por ejemplo la detección de spam en correos en base al contenido de estos, como aparece en el ejemplo de la Ilustración 50.

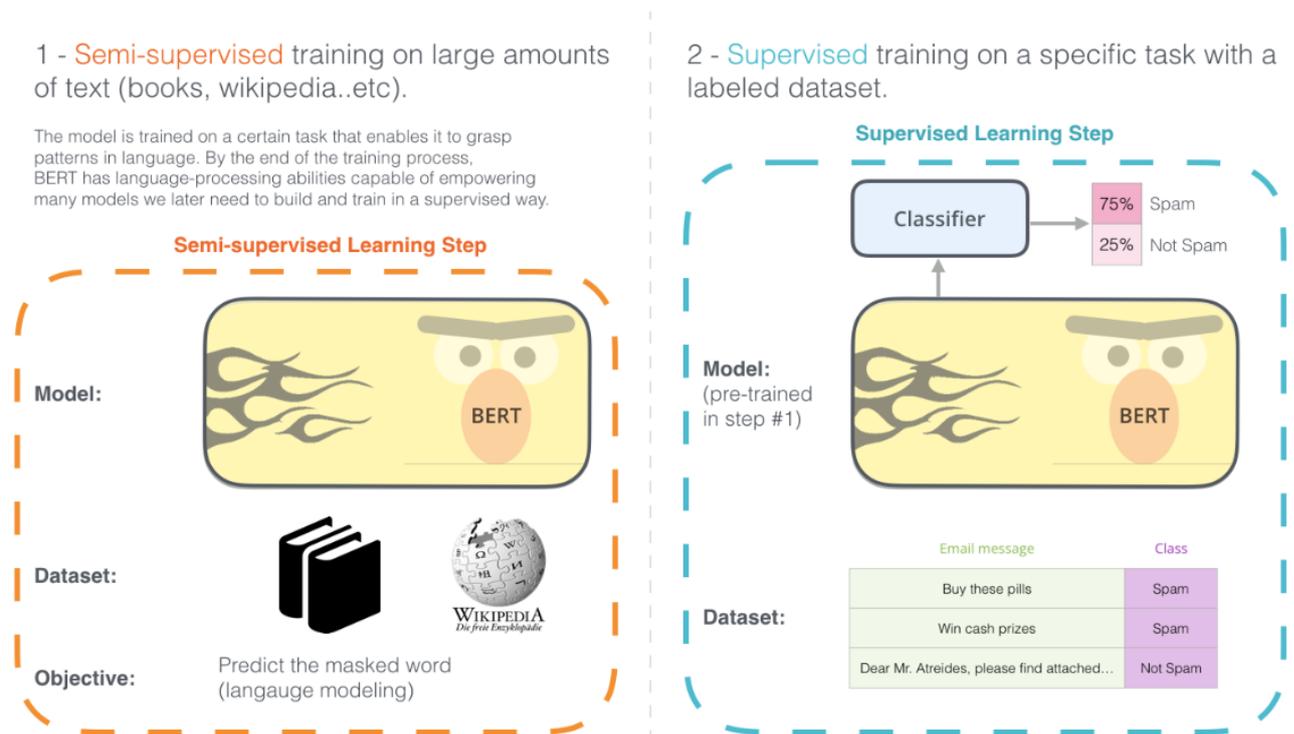


Ilustración 50 Esquema de BERT.⁵⁶

En la Ilustración 51 se ven unas Comparativas de métricas entre *bert large* y *bert base* en los distintos *datasets* de SQuAD (pares de preguntas-respuestas). EM es la métrica de *exact match*

⁵⁶ Almmar, Jay. "The Illustrated BERT, ELMo, and co. (How NLP Cracked Transfer Learning)". Web. 03/05/2019. <http://jalmmar.github.io/illustrated-bert/>

que mide el porcentaje de predicciones que encajan con alguna de las respuestas que se tienen en los datos de test (*ground truth*) de manera exacta.

| Dataset | SQuAD 1.1 | SQuAD 1.1 | SQuAD 2.0 |
|---------|----------------|-----------------|-----------------|
| Model | bert_12_768_12 | bert_24_1024_16 | bert_24_1024_16 |
| F1/EM | 88.53/80.98 | 90.97/84.05 | 77.96/81.02 |

Ilustración 51 Comparativas de métricas entre bert large y bert base en los distintos datasets de SQuAD.⁵⁷

De cara a reducir el vocabulario utilizado, BERT descompone las palabras generalmente en sus lemas y sus morfemas, de manera que en vez de tener todas las combinaciones de las palabras y, por ejemplo, sus plurales, se tiene en el vocabulario el *encoding* de la palabra base y del valor genérico del plural, de manera que el vector final de una palabra es la combinación de todos esos factores (p.e. *dragón + ##es*).

H. Conclusión

Junto con el aspecto computacional afectivo, la solución debe permitir buscar poemas específicos mediante la realización de consultas en texto plano, al estilo habitual de los sistemas de recuperación de información (IR). Esto es así ya que en muchas ocasiones los profesores querrán consultar qué poemas existen en el repositorio mediante un buscador, o incluso los alumnos podrían querer consultar poemas en base a sus propias inquietudes.

Para que esto sea posible se deben representar numéricamente las distintas palabras tanto del repositorio de poesía como del texto de consulta, para encontrar la poesía que más se adecue a cada texto introducido. Una vez representadas las palabras con un valor numérico, se obtiene un vector para la consulta de entrada y otro para cada poesía del corpus y usando métricas de comparación (similitud coseno, distancia euclídea...) se selecciona la obra poética cuyo vector representante estuviese más cerca del (fuese más similar al) vector de entrada.

Es por este motivo que es importante conocer el SOTA de las técnicas de representación numérica de palabras. BERT aparece como la solución más novedosa y que aparentemente mejora las técnicas previas de *embeddings* al ser más precisa y general, tanto desde el punto de vista teórico como por los resultados que se han obtenido en distintas pruebas con el inglés como lengua de base, pero no se ha estudiado si realmente su aportación es mucho más significativa para el caso de la poesía y en particular para el caso del castellano. No hay hasta la fecha, hasta donde se ha podido comprobar, estudios que trabajen con textos poéticos en castellano usando BERT como *embeddings* y comparando los resultados con otros *embeddings* como word2vec, de manera que este TFM contribuirá a la investigación en este punto.

⁵⁷ “GluonNLP”. Web. 03/05/2019. <https://www.ctolib.com/article/releases/82604>

2.1.6. Composición de textos

Además de la cuestión ya tratada del *embedding* de las palabras de un texto y las distintas formas de abordar esta tarea, especialmente en el ámbito poético, aparece otra cuestión a plantear, el *embedding* de secuencias de textos enteras.

En un sistema de IR se considera, como ya se ha mencionado, la similitud entre un texto de entrada (que puede ser desde una palabra hasta un texto extenso) y los distintos documentos disponibles, para encontrar el documento que más se aproxima a dicha entrada. Ahora bien, esto plantea dos cuestiones: ¿Qué información usar de la entrada y de los documentos para llevar a cabo la comparación? y ¿qué métrica utilizar para determinar que dos textos son similares?.

Estas cuestiones sobre la información a considerar de los textos y las métricas a utilizar aparecen recogidas en (Amigó Cabrera & Fresno Fernández, 2018), donde además se plantean y comparan distintas soluciones para ambas, utilizando para ello las entradas del diccionario on-line de Cambridge, evaluando la similitud entre la palabra de entrada y la definición de la misma. Para ello usan como *embedding* el modelo word2vec entrenado en el corpus de Google News (300 dimensiones y 3 billones de palabras).

En lo que respecta a funciones de composición, en dicho artículo se define un esquema general que sirve para instanciarse en distintos tipos de funciones de composición. A grandes rasgos, el artículo habla de cómo partiendo de una función genérica como:

$$f(v1, v2) = \hat{f}(v1, v2) \|f(v1, v2)\|$$

con:

$$\hat{f}(v1, v2) = \frac{\gamma v1 + v2}{\|\gamma v1 + v2\|}$$

$$\|f(v1, v2)\| \approx \sqrt{(\alpha - \beta)(\|v1\|^2 + \|v2\|^2) + \beta \langle v1, v2 \rangle}$$

Se pueden especificar los parámetros α, β, γ de manera que se obtengan con ello distintas funciones de composición entre los dos vectores $v1$ y $v2$ (que representarían el embedding de 2 palabras distintas).

Por ejemplo, si $(\alpha, \beta, \gamma) = (3, 2, 1)$ se obtendría la función básica, que consistiría en sumar dichos vectores ya que:

$$\begin{aligned} F_{Sum} &= \frac{(v1 + v2) \left(\sqrt{(3 - 2)(\|v1\|^2 + \|v2\|^2) + 2 \langle v1, v2 \rangle} \right)}{\|v1 + v2\|} \\ &= \frac{(v1 + v2) \left(\sqrt{(3 - 2)(\|v1\|^2 + \|v2\|^2) + 2 \|v1\| \|v2\| \cos \theta} \right)}{\|v1 + v2\|} \\ &= \frac{(v1 + v2) \|v1 + v2\|}{\|v1 + v2\|} = v1 + v2 \end{aligned}$$

Otro ejemplo de función de agregación, que según el artículo da muy buenos resultados, es la denominada *joint*, la cual se expresa de la siguiente forma:

$$F_{\text{Joint}}(v_1, v_2) = \frac{(v_1 + v_2)}{\|v_1 + v_2\|} \cdot \sqrt{\|v_1\|^2 + \|v_2\|^2 - \langle v_1 \cdot v_2 \rangle}$$

Estas funciones de agregación dan como resultado un vector de la misma dimensionalidad que los dos vectores de entrada, pasándose así de dos vectores a uno. Ahora bien, este caso está aplicado a únicamente 2 vectores (es decir, 2 palabras). ¿Qué hacer para agregar la información de toda una secuencia de textos? Ante esto se plantean varias alternativas en el artículo, que pasan desde ir agregando las palabras secuencialmente nada más aparecen en el texto hasta aplicar funciones de composición más complejas que tienen en cuenta las relaciones sintácticas entre las palabras. Estas posibles agregaciones se muestran en la Ilustración 52:

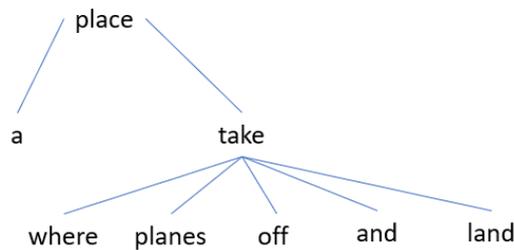
| | |
|--------|---|
| LR-SEQ | [a[place[where[planes[take[off[and land]]]]]]]]] |
| RL-SEQ | [[[[[[[[a]place]where]planes]take]off]and]land]]]]]]] |
| LR-SYN | [a[place[where[planes[take[off [and land]]]]]]]]] |
| RL-SYN | [[a place][where[planes[[take off][and land]]]]]]] |
| LR-DEP | [place[a[take[where[planes[off[and land]]]]]]]]]]] |
| RL-DEP | [place[[a take][[[[where planes]off]and]land]]]]]]] |

Ilustración 52 Distintas técnicas de agregación para la definición de "aeropuerto" en inglés. "a place where planes take off and land" (Amigó Cabrera & Fresno Fernández, 2018)

La primera estrategia que se menciona, **LR-SEQ**, simplemente busca agregar las palabras de izquierda a derecha según van apareciendo en el texto. Análogamente, la técnica **RL-SEQ** agrega las palabras pero en orden inverso al anterior, de derecha a izquierda.

Estas dos técnicas agregan las palabras sin tener en cuenta ninguna información adicional a nivel sintáctico o semántico. Distinta es la técnica **LR-DEP**, donde la agregación de palabras se hace desde el nodo superior a medida que se va descendiendo por el gráfico de dependencia, y agregando en cada nivel los resultados previos con las palabras de ese nivel de izquierda a derecha (hasta completar el nivel y seguir descendiendo).

LR-DEP



place
place a
place a take
place a take where
place a take where planes
place a take where planes off
place a take where planes off and
place a take where planes off and land

V1: place
V2: a
V3: (V1, V2)
V4: (V3, take)
V5: (V4, where)
V6: (V5, planes)
V7: (V6, off)
V8: (V7, and)
Vfinal: (V8, land)

Ilustración 53 Para el ejemplo anterior se muestra como para el caso de LR-DEP se va descendiendo por el gráfico de dependencia agregando las palabras con respecto al vector resultante de las agregaciones previas⁵⁸

De forma genérica, estas agregaciones a nivel de DEP (RL-DEP, LR-DEP) se hacen recorriendo la estructura del gráfico de dependencia y agregando las palabras de cada nodo con sus nodos hijos en orden secuencial.

Otro tipo de agregaciones son las tipo SYN; el proceso de agregación es similar al anterior solo que en este caso las agregaciones se hacen según el análisis sintáctico del texto. En caso de que un elemento tenga varias palabras se agregarán de derecha a izquierda o viceversa (según si es **RL-SYN** o **LR-SYN**).

En la Ilustración 54 se muestra la Estructura sintáctica de la frase de ejemplo obtenida con el *parser* de Stanford. Se va recorriendo el árbol sintáctico agregando las palabras en cada nivel y a medida que se desciende se agregan los resultados del nivel anterior con las palabras de ese nivel.

⁵⁸ Fuente: Propia

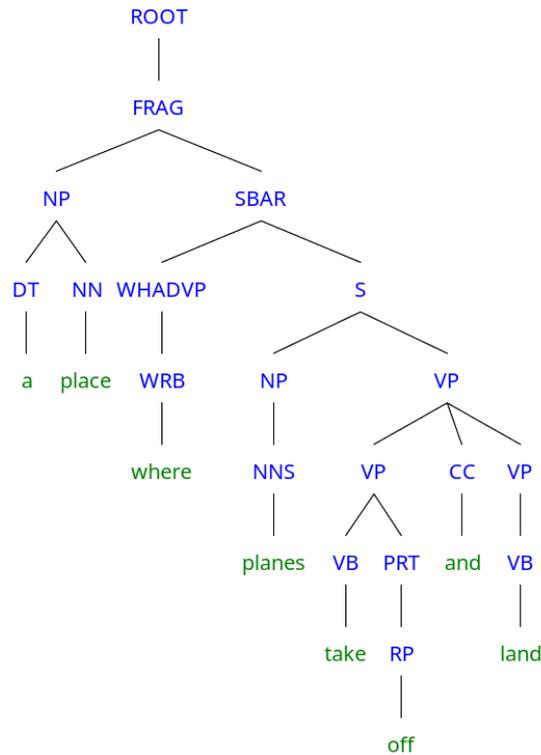


Ilustración 54 Estructura sintáctica de la frase de ejemplo obtenida con el parser de Stanford.⁵⁹

La evaluación de resultados se usa comparando el *embedding* de las distintas palabras que aparecen en el diccionario de sinónimos de WordSim-353⁶⁰ con el resultado de la composición de *embeddings* de las definiciones del diccionario online de Cambridge correspondiente a dichas palabras, usando como *word embedding* de base para todos los casos word2vec. Si la función de composición de cada una de esas definiciones diese buenos resultados, el resultado semántico del vector resultante debería ser similar al de la palabra a la que hacen referencia (a la entrada del diccionario correspondiente), y dicha similitud entre los dos vectores se puede medir con alguna de las funciones de comparación que se mencionan, como el coseno o icm. El caso ideal es que el resultado de cualquiera de esas métricas sea bajo (esto indica que los vectores son muy similares), de manera que se obtiene para todas las palabras y se ordenan después de mayor a menor resultado (de peores representaciones a mejores), aplicando además la función logaritmo como función de penalización sobre la lista de resultados obtenidos (para así penalizar más si los primeros resultados son muy malos) y se calcula la media de dichos valores para usar el resultado de ella como valor de métrica comparativa entre los distintos experimentos.

De entre todas estas funciones de composición el artículo indica que los mejores resultados se obtienen con Fjoint para SYN_LR.

⁵⁹ “Stanford Parser”. Web. Stanford Parser. 07/05/2019. <http://nlp.stanford.edu:8080/parser/index.jsp>:

⁶⁰ “WordSim-353”. Web. 12/06/2019. <https://www.wikidata.org/wiki/Q31845205>

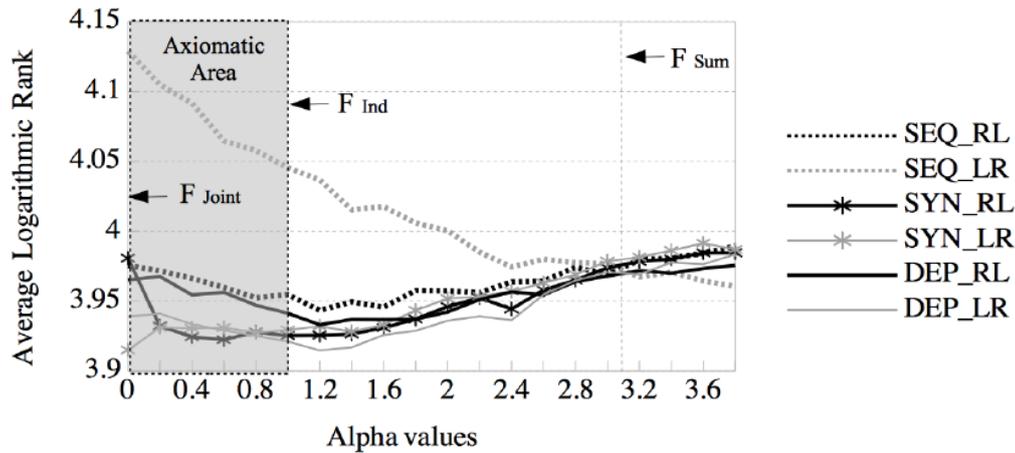


Ilustración 55 Comparativas de las distintas funciones de composición para las distintas técnicas de agregación donde se ven los resultados de cada una de ellas (Amigó Cabrera & Fresno Fernández, 2018)

Además de las funciones de agregación se definen y evalúan distintas métricas de comparación para analizar el vector final resultante de una determinada composición de un texto con el de otro para ver su similitud.

Entre las métricas que se mencionan, además de algunas clásicas como la distancia euclídea o el coseno de los vectores, se define una métrica genérica que recoja la similitud de los vectores no sólo con respecto a su ángulo (coseno) sino también con respecto a su magnitud (como haría la distancia euclídea), denominada ICM (*Information Contrast Model*), y que en su expresión general se define como:

$$ICM \approx (c_1 - c) \|v_1\|^2 + (c_2 - c) \|v_2\|^2 + c \langle v_1, v_2 \rangle$$

Estudios empíricos reportados en el artículo muestran cómo esta medida obtiene su mejor eficiencia cuando los parámetros son los siguientes: $c_1=c_2=1$ y $c=1.2$; obteniéndose con ello unos resultados teóricamente mejores que los obtenidos con cualquier otra métrica (para la tarea de comparar definiciones con sus entradas del diccionario).

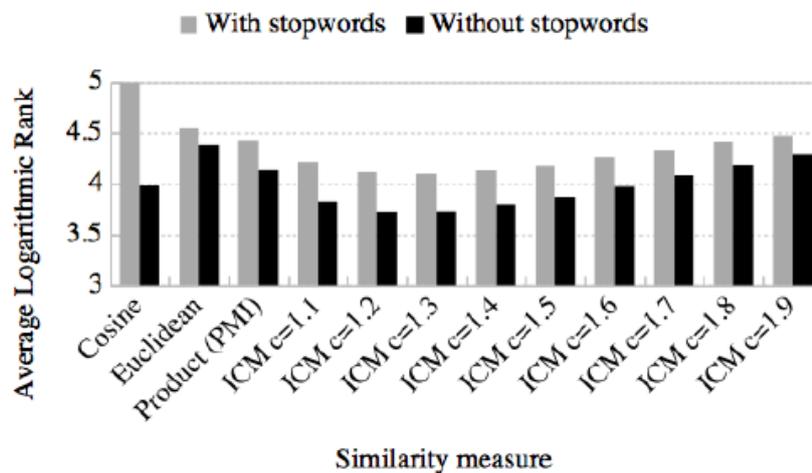


Ilustración 56 Mejores resultados para el caso de icm 1.2 sin stopwords (Amigó Cabrera & Fresno Fernández, 2018)

Los estudios que ha presentado este trabajo se han centrado en textos en inglés, definidos en prosa y usando word2vec como base para el análisis, de manera que este TFM plantea una aportación a la investigación en este punto al continuar dicho estudio sobre textos poéticos, en castellano, y usando además otras técnicas de *embeddings* como BERT.

2.4. Planteamiento de la solución

En este apartado se muestra el planteamiento de una solución inicial de un RecSys sensible al contexto afectivo para entornos de e-learning en el que se implemente la pedagogía poética utilizando un corpus de sonetos en castellano de los últimos cinco siglos.

Al comienzo de este capítulo se plantearon los dos grandes objetivos que se buscaban cubrir con este TFM.

Por un lado se planteó un análisis preliminar de la arquitectura de los RecSys sensibles al contexto afectivo y aplicados a los entornos de e-learning para la integración de la pedagogía poética junto con las ontologías necesarias para desarrollarlos de manera que sean interoperables. En el primer subapartado de esta Sección se propone una solución técnica para este primer objetivo.

Por otro lado, el segundo objetivo es desarrollar un prototipo de un nuevo componente de dichos sistemas con el que se cumpla un propósito doble: 1) verificar determinadas hipótesis de investigación sobre la aplicación de computación afectiva y NLP a textos poéticos, y 2) que se pueda integrar, a modo de API, tanto en sistemas RecSys como en otro tipo de sistemas (tales como herramientas de autor para la creación de cursos y contenidos en e-learning, o aplicaciones para el análisis de poesía desarrolladas en el contexto del proyecto POSTDATA con propósitos de investigación en lingüística de la poesía). La propuesta para este objetivo se detallará en el segundo subapartado de esta Sección.

2.4.1 Estudio tecnológico preliminar

A continuación se plantea una primera aproximación al problema tecnológico de integrar la pedagogía poética en un RecSys afectivo en entornos de e-learning, identificando algunas cuestiones clave, señalando cómo afectaría esta integración a los componentes habituales de la arquitectura de un RecSys, y haciendo algunas propuestas iniciales.

En cualquier proyecto de software, se consideran tres capas:

- Datos
- Lógica/aplicación
- Interfaz

A. Datos

El uso de ontologías permite tener representación y taxonomía común de la información contenida en las distintas bases de datos (BD) y relacionar dicha información con las reglas

que se definan en la capa de la lógica para, por ejemplo, articular las recomendaciones del sistema.

Dicho uso de ontologías permite modelar contextos con mayor riqueza (mucho más que con meras tablas en una BD) a la vez que proporcionan una semántica formal para acceder a la información y compartirla con otras herramientas.

A continuación se va a particularizar la propuesta de ontologías para los distintos módulos de un RecSys sensible al contexto afectivo en entornos de e-learning que incorpore la pedagogía poética. Para ello primero se analiza una a una las ontologías necesarias para cada modelo, y posteriormente se proponen extensiones a sus conceptos así como conceptos nuevos que las enriquezcan de cara a servir para el propósito de la investigación presente.

Modelo de contenidos

En primer lugar, se debe elegir una ontología para los poemas (de manera que estos estén representados de manera sencilla como LO, desde la instanciación de una ontología común que permita su reutilización en otras potenciales aplicaciones).

El punto de partida será la propia ontología en la que se basa DISCO (Díez Platas & Bermúdez Sabel). En el caso de DISCO aparecen, por un lado, los metadatos de los autores, expresados utilizando de base distintas ontologías (indicadas en su repositorio) mediante un esquema RDFa (sujeto-predicado-objeto) incorporado dentro de unos documentos generados con un formato XML-TEI.

A modo de ejemplo para entender esta estructura, en la Ilustración 57 aparece la ontología referida a un autor en concreto. En ella se ve cómo el identificador (id) de un autor en concreto, en este caso *disco_100n*. Para este sujeto se tiene una frase “*El autor disco_100n nació en 1837*”. En esta frase el sujeto es “[*El autor*] *disco_100n*”, el predicado “*nació en*” y el objeto “*1837*”. Esto queda reflejado en la relación “*birthDate*” y su valor “*1837*”.

Los metadatos de esta ontología acerca de los autores hacen referencia al lugar y fecha de nacimiento, fecha de fallecimiento, nombre y las obras que han compuesto.

```

<person xml:id="disco_100n" about="disco:100n" typeof="foaf:Person">
  <idno cert="high"
    property="rdfs:seeAlso"
    resource="https://viaf.org/viaf/29108480"/>
  <persName type="full">
    <forename property="foaf:givenName">Antonia</forename>
    <surname property="foaf:familyName">Díaz de Lamarque</surname>
  </persName>
  <sex property="foaf:gender" content="F"/>
  <birth>
    <location>
      <placeName>
        <settlement property="schema:birthPlace">Marchena (Sevilla)</settlement>
      </placeName>
    </location>
    <date property="schema:birthDate" content="1837" cert="high"/>
  </birth>
  <death>
    <date property="schema:deathDate" content="1892" cert="high"/>
  </death>
  <listBibl rel="blterms:hasCreated">
    <bibl resource="disco:s100n_0335" typeof="schema:CreativeWork">
      <title property="dc:title">A Dios en la Eucaristía</title>
      <title type="incipit" property="dc:alternative">Tu infinito poder en la armonía</title>
    </bibl>
  </listBibl>
</person>

```

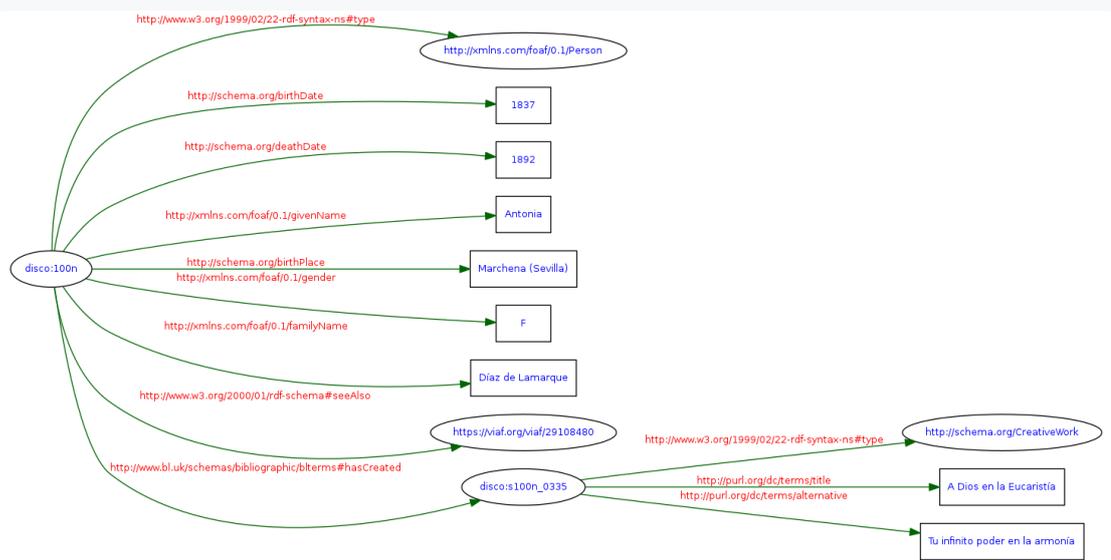


Ilustración 57 Ejemplo de definición del XML-TEI con su ampliación con RDF y su correspondiente visualización en modo RDF (Ruiz Fabo et al., 2017)

Además de los metadatos de autores, DISCO incluye una ontología (de nuevo, ya poblada con datos) para los sonetos. De ella se derivan diferentes rasgos literarios, junto con atributos como título, fecha, y textos de cada poema, divididos en sus cuartetos y tercetos.

En primer lugar aparece la *Escansión métrica*, que se define mediante el atributo @met y que indica las sílabas tónicas (las acentuadas) con un +, y las átonas (sobre las que no recae el acento) con un -, para cada frase del texto.

Aparece, además, el *Esquema de rima* para los cuartetos de los poemas, ya que estos pueden ser de tipo ABBA o de rima alterna, ABAB.

En último lugar se incluye el *Encabalgamiento*, identificado mediante una herramienta propia de POSTDATA. El encabalgamiento es una figura retórica que consiste en no terminar las frases al final del verso sino en el verso siguiente. Este se especifica en el atributo @enjamb.

```

<text>
  <front>
    <head>A un arroyo</head>
  </front>
  <body>
    <lg n="1" type="cuarteto">
      <l n="1" met="--+++----">Cuando todo era flores tu camino,</l>
      <l n="2" met="--+++----">cuando todo era pájaros tu ambiente,</l>
      <l n="3" met="--+++----">cediendo de tu curso a la pendiente</l>
      <l n="4" met="++---+--">todo era en ti fugaz y repentino.</l>
    </lg>
    <lg n="2" type="cuarteto">
      <l n="5" met="++---+--" enjamb="B-ex_subj_verb" cert="low">Vino el invierno con sus nieblas vino</l>
      <l n="6" met="++---+--" enjamb="I-ex_subj_verb" cert="low">el hielo que hoy estanca tu corriente,</l>
      <l n="7" met="++---+--">y en situación tan triste y diferente</l>
      <l n="8" met="++---+--">ni aún un pálido sol te da el destino.</l>
    </lg>
    <lg n="3" type="terceto">
      <l n="9" met="--+++----" enjamb="B-ex_subj_verb" cert="low">Y así en la vida el incesante vuelo</l>
      <l n="10"
        met="--+++----"
        enjamb="B-ex_dobj_verb I-ex_subj_verb"
        cert="low low">mientras que todo es ilusión, avanza</l>
      <l n="11" met="--+++----" enjamb="I-ex_dobj_verb" cert="low">en sólo una hora cuanto mide el cielo.</l>
    </lg>
    <lg n="4" type="terceto">
      <l n="12" met="--+++----">Y cuando el duelo asoma en lontananza</l>
      <l n="13" met="--+++----" enjamb="B-ex_subj_verb" cert="low">entonces como tú cambiada en hielo</l>
      <l n="14" met="++---+--" enjamb="I-ex_subj_verb" cert="low">no puedes reflejar ni al esperanza.</l>
    </lg>
  </body>
</text>

```

Ilustración 58 Ejemplo de metadatos para el XML de un soneto de DISCO (Ruiz Fabo et al., 2017)

Cabe recordar que la ontología de base utilizada en DISCO (Díez Platas & Bermúdez Sabel) es más extensa y contiene otros campos relacionados con el ámbito poético que pueden ser de interés para esta investigación.⁶¹ No obstante, estos campos no se contemplan en la presente memoria por no aparecer poblados para los sonetos del corpus.

Se propone ampliar el conjunto de atributos mencionado y anotar los archivos de sonetos de POSTDATA con rasgos afectivos extraídos usando las fuentes previamente descritas, de manera que, a nivel de estrofa, se tengan determinados rasgos afectivos que puedan ser útiles para la ser reutilizados posteriormente por aplicaciones de e-learning que implementen la pedagogía poética.

En consecuencia, se amplían los XML-TEI de los documentos de DISCO con una anotación que siga el esquema TEI definiendo un nuevo atributo dentro de *lg* (*line group*) de cada soneto y del *lg* de cada estrofa que incluya un *@param* denominado *AffectiveFeatures* en el cual se especifiquen unos atributos mediante la especificación *@attRef* en los que el campo *name* haga referencia al nombre del atributo afectivo en cuestión y que tengan un contenido (*content*) donde aparezca el valor de dicho atributo.⁶²

⁶¹ POSTDATA. “Domain model for European Poetry”. Web. 11/05/2019. <http://postdata-prototype.linhd.uned.es/html-model.php>

⁶² “Text Encoding Initiative”. Web. 11/05/2019. <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-attRef.html>

```

<param name="AffectiveFeatures">
  <attRef name="Hap_Mean">
    <content>62.52</content>
  </attRef>
  <attRef name="Hap_SD">
    <content>19.91</content>
  </attRef>
  <attRef name="Ang_Mean">
    <content>28.96</content>
  </attRef>
  <attRef name="Ang_SD">
    <content>14.6</content>
  </attRef>
  <attRef name="Sad_Mean">
    <content>32.59</content>
  </attRef>
  <attRef name="Sad_SD">
    <content>18.58</content>
  </attRef>

```

Ilustración 59 Ejemplo de algunos atributos anotados siguiendo la especificación de TEI sobre el XML de un soneto

Además de los atributos ya mencionados se ampliará la ontología de los poemas con otro parámetro (definido en @param) denominado *SemanticFeatures* que sirva para que los autores puedan definir tags relevantes para los distintos sonetos. Así, cuando quisiesen recomendar poemas de una determinada temática, un autor podría optar porque el sistema recomendase automáticamente en función de sus técnicas de IR, o podría usar ellos mismos el buscador, explorar el repositorio y definir los tags que les interesen, los cuales se anotarían en dicho campo. Adicionalmente a las etiquetas semánticas de *SemanticFeatures* se debe incluir la información respecto a los vectores de *embeddings* de cada poema (sonetos en la presente memoria). Se especifica otro campo denominado *SemanticEmbeddingId*, donde se incluye el identificador del archivo csv (*comma separated value*) que contenga la matriz de *embeddings* para ese soneto (en el caso de la presente investigación tendrá un vector para cada estrofa).

Análogamente, se define otro atributo denominado *PsychologicalFeatures* dentro de @params que contenga las anotaciones relacionadas con los conceptos psicológicos, mencionados en el Capítulo anterior (con los valores reales para los sonetos anotados y los valores inferidos para el resto de ellos).

Por último, esta ontología se utiliza para especificar todo el conjunto de poemas, tanto a los existentes en DISCO como a los adicionales añadidos (sonetos del S. XX). Por este motivo el proceso de poblado de información sobre autores e información lingüística de los sonetos que se lleva a cabo en DISCO se debe aplicar también sobre dicho conjunto del S. XX. Así, para la información sobre los autores se usa la información de la fuente de origen de dichos sonetos (si existe), junto con VIAF (*Virtual International Authority File*)⁶³ y completándose con una revisión manual. Para la *Escansión métrica* se usa ADSO (Navarro-Colorado, 2018), para el *Esquema de rima* se hace un análisis manual, y para el *Encabalgamiento* se usa la herramienta ANJA (Ruiz Fabo, Martínez Cantón, Poibeau, & Gonzalez-Blanco, Aug 2017), tal como indica DISCO.

Así, esta extensión de la ontología propuesta en (Díez Platas & Bermúdez Sabel) define una ontología válida para integrar la pedagogía poética dentro de un RecSys sensible al contexto afectivo en entornos de e-learning. Gracias a ello, este sistema podría, por ejemplo,

⁶³ “Virtual International Authority File”. Web. 09/08/2019. <https://www.oclc.org/research/activities/viaf.html>

complementar la formación de alumnos de psicología sobre trastornos de personalidad gracias a que estos leerían poesía que serviría para comprender esos trastornos desde una perspectiva distinta.

No obstante, la extensión no es suficiente para definir por completo un LO. Efectivamente, otros campos importantes para un LO, como por ejemplo los relacionados con la accesibilidad, están ausentes. Por este motivo, se plantea en esta propuesta inicial combinar la ontología mencionada con una segunda ontología que incluya algunos campos importantes que están ausentes. Para ello se utiliza IEEE LOM⁶⁴, ya que permite definir detalladamente el contenido los LO a nivel de dificultad, recurso con el que se relaciona (en este caso, el soneto), contexto en el que tiene lugar o tipo de interactividad.

El estándar IEEE LOM se enriquece con la propuesta de (Bustos, Broisin, Muñoz, & Guzman, 2016) en la que se extienden los campos relacionados con la accesibilidad y la navegación, tratando con ello de facilitar el acceso a los recursos educativos desde distintos dispositivos, y sobre todo de fomentar una educación más inclusiva al añadirse campos como *Discapacidad*, con el que se define el tipo de *diversidad funcional* que puede tener un estudiante (de manera que los recursos y la presentación de los mismos se pueda adecuar a ello), o *Función cognitiva*, que define las funciones cognitivas que pueden ser apoyadas con el LOM (percepción, memoria...).

⁶⁴ IEEE Standards Association. Web. 11/05/2019. https://standards.ieee.org/standard/1484_12_1-2002.html

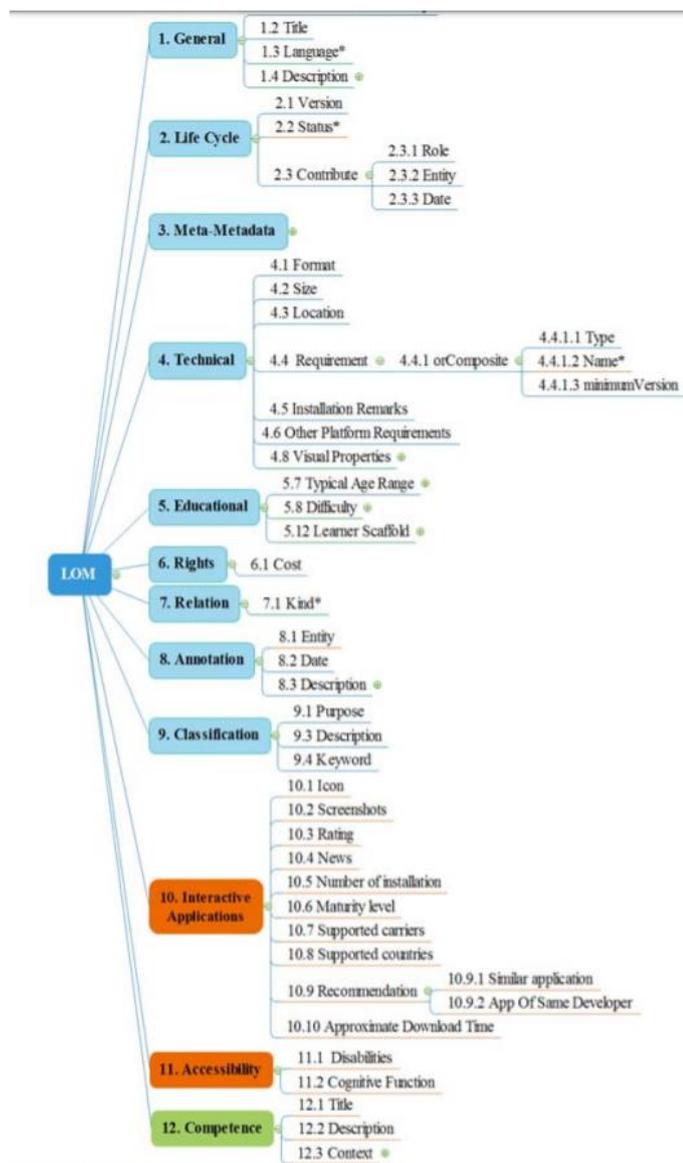


Ilustración 60 LOM extendido para incluir información sobre accesibilidad y función cognitiva (Bustos et al., 2016).

De esta manera, se propone usar IEEE LOM, ampliado con (Bustos et al., 2016), para definir los LO de los sonetos. Esta estructura de IEEE LOM será el LO principal en el que se integrarán los parámetros pertenecientes a cada uno de los sonetos. Los parámetros propios de la estructura lingüística del soneto y la información sobre los autores (es decir, las anotaciones que aparecen en DISCO), así como el texto en sí, serían atributos de la clase *General*. Por otro lado, las anotaciones afectivas, semánticas y psicológicas, al ser información propia del LO para su uso pedagógico, aparecerán como atributos de la clase *Educational*.

Para la presente investigación, el uso de ontologías se reducirá a la ampliación de DISCO para los rasgos afectivos y sin incorporar los sonetos del S. XX. El corpus entero (junto con sonetos del S. XX) con todas las anotaciones (afectivas, semánticas y psicológicas) se definirá en una estructura de datos en formato JSON en la que aparecerán sólo los campos necesarios para la presente investigación, es decir *autor*, *fecha de la obra*, *nombre de la obra*, *texto de la obra*, y todos los rasgos afectivos, semánticos y psicológicos.

El resto de campos que aparecen en la ontología IEEE LOM extendida no se han pero serán necesarios para el diseño completo de un RecSys para pedagogía poética (por ejemplo, información sobre el nivel de dificultad de los LO, contextos en que presentarlos...).

En la Tabla 2 aparecen los campos principales para incluir en el modelo de contenidos, especificando su ontología de origen.

| Campo | Fuente | Descripción |
|------------------------------|--|--|
| <i>AffectiveFeatures</i> | Ampliación sobre POSTDATA (Díez Platas & Bermúdez Sabel) | Conjunto de rasgos afectivos que caracterizan al poema |
| <i>SemanticFeatures</i> | Ampliación sobre POSTDATA (Díez Platas & Bermúdez Sabel) | Conjunto de etiquetas (anotadas) que un autor asocia al poema |
| <i>SemanticEmbeddingId</i> | Ampliación sobre POSTDATA (Díez Platas & Bermúdez Sabel) | Identificador para vincular el poema con el csv correspondiente de su <i>embedding</i> |
| <i>PsychologicalFeatures</i> | Ampliación sobre POSTDATA (Díez Platas & Bermúdez Sabel) | Conjunto de etiquetas con los rasgos psicológicos asociados al poema |
| Datos de autoría | POSTDATA (Díez Platas & Bermúdez Sabel) | Información sobre el autor del poema (nombre, fecha nacimiento...) |
| Datos lingüísticos | POSTDATA (Díez Platas & Bermúdez Sabel) | Información lingüística del poema (en el caso de que sean sonetos: escansión métrica, esquema de rima y encabalgamiento) |
| Accesibilidad | IEEE LOM ampliado con (Bustos et al., 2016) | Información para vincular determinados poemas con usuarios con una determinada diversidad funcional o para una función cognitiva específica. |
| Typical Age Range | IEEE LOM | Información sobre la edad recomendada para la lectura del poema. |
| Difficulty | IEEE LOM | Información sobre el nivel de dificultad del poema. |

Tabla 2 Campos principales para el modelo de contenidos

Modelo de usuario

El modelo de usuario debe incluir, como se ha mencionado, información sobre su perfil al margen de su uso de la aplicación, e información sobre su aprendizaje y uso de la aplicación. Todo ello ha de estar particularizado en un RecSys sensible al contexto afectivo en entornos de e-learning que integre en él la pedagogía poética.

La recomendación de un poema personalizado como LO requiere disponer de determinada información en la ontología del modelo de usuario, para comprobar que los atributos correspondientes de la ontología IEEE LOM extendida se ajustan a sus características. Dicha información incluye los intereses generales de ese usuario, ya que así se le puede recomendar un poema en función de la similitud entre dichos intereses y las etiquetas del poema (*SemanticFeatures*) o en función de su contenido semántico (a través de la similitud semántica entre el *embedding* de los intereses del usuario y de los vectores de *embeddings* de los poemas). Además, se deben considerar los rasgos afectivos del estudiante (*AffectiveFeatures*). En el curso de un RecSys para la enseñanza de los trastornos de personalidad, los conceptos psicológicos podrían vincularse a los *intereses* para una aplicación de la pedagogía poética dirigida a estudiantes de psicología. Otros parámetros relevantes a considerar en el RecSys, como son los objetivos de aprendizaje del usuario, la información sobre la accesibilidad o los estilos de aprendizaje de los estudiantes.

Otra primera ontología que procede mencionar en este apartado es **FOAF** (*Friend of a Friend*)⁶⁵. FOAF permite definir información genérica de los usuarios siguiendo un esquema RDF (*sujeto-objeto-predicado*). Es relevante destacar que en FOAF aparecen campos para los intereses del usuario, además de campos generales sobre el propio usuario (nombre y apellidos, edad, correo electrónico...).

La ontología MIRLO (Raffenne, 2013) propone un modelo abierto de estudiante que incluye aspectos como sus competencias o progreso académico, al tiempo que se vinculan con las actividades y cursos que realiza. En la ontología se especifican los objetivos del alumno y en función de ellos se calcula su progreso. Además, incluye el estilo de aprendizaje (reflexivo, visual, verbal...). MIRLO también se define en una estructura RDF en la que un *Alumno* (sujeto) tiene determinadas propiedades (objetos).

En la Ilustración 61 se muestra el modelo de usuario de MIRLO. Para la presente investigación son relevantes tanto el modelo en sí como las relaciones con *Competencia* (donde se definen los objetivos de aprendizaje) y *Actividad* (donde se especifican las relaciones del alumno con los LO a través de tareas como tests y cuestionarios). Esos modelos de *Competencias* y de *Actividad*, tal como aparecen en MIRLO, son útiles para el propósito de esta investigación. Si bien conceptos como *Curso* no son relevantes ya que los LO son poemas, cuya estructura se especifica en el apartado anterior.

⁶⁵ FOAF. Web. 11/05/2019. <http://www.foaf-project.org/>

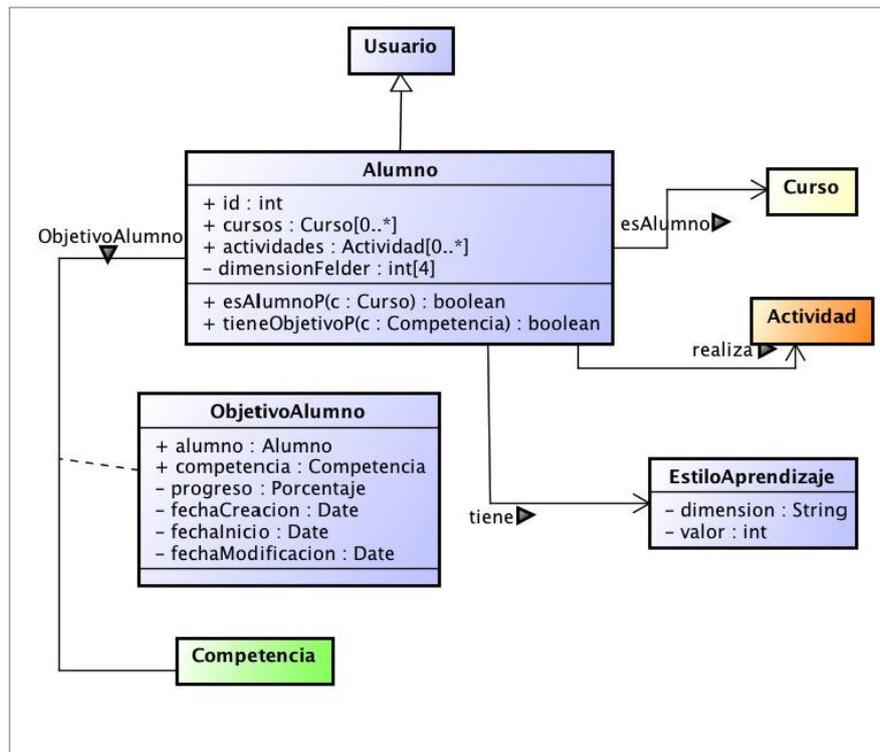


Ilustración 61 Modelo de usuario de MIRLO (Raffenne, 2013)

Como ya se mencionó, existen diversos estudios que remarcan la influencia positiva que la poesía puede tener en personas con trastornos psicológicos (Czernianin, 2016; Parastoo et al., 2016; Wassiliwizky et al., 2017). Por este motivo es pertinente considerar una ontología que incluya dentro de sus especificaciones de accesibilidad al modelado de conceptos relacionados con trastornos psicológicos. Esto se aborda en la ontología ADOOLES (Nganji, Brayshaw, & Tompsett, 2011). En ella aparecen diversos trastornos, tanto psicológicos (ansiedad, autismo, depresión...) como físicos, ofreciendo así un espectro de caracterización muy amplio para cuestiones de accesibilidad. Esta ontología incluye también el tipo de dispositivo adecuado para cada usuario.

Por último, la propuesta ontológica para el modelo de usuario debe incluir su modelado afectivo de cara a perfilar las recomendaciones afectivas de poemas. En relación a ello, el artículo (Abaalkhail et al., 2018) hace una revisión de las distintas ontologías existentes para incluir la información afectiva de los usuarios, donde aparecen propuestas basadas en *Emotion Markup Language 1.0*⁶⁶ (EmotionML).

Un ejemplo de anotaciones con EmotionML es el siguiente, donde se anota con parámetros afectivos un pasaje de un texto⁶⁷:

```
<emotionml xmlns="http://www.w3.org/2009/10/emotionml"
  xmlns:meta="http://www.example.com/metadata"
  category-set="http://www.w3.org/TR/emotion-voc/xml#everyday-categories">
  <info>
    <meta:doc>Example adapted from (Zhang, Black & Sproat 2003)
    http://www.cs.cmu.edu/~awb/papers/eurospeech2003/esper.pdf
    </meta:doc>
  </info>
```

⁶⁶ Emotion Markup Language. Web. 01/06/2019. <https://www.w3.org/TR/emotionml/>

⁶⁷ Emotion Markup Language. Web. 10/08/2019. <https://www.w3.org/TR/emotionml/#s4.1>

```

<emotion>
  <category name="Disgust" value="0.82"/>
  'Come, there's no use in crying like that!'
</emotion>
said Alice to herself rather sharply;
<emotion>
  <category name="Anger" value="0.57"/>
  'I advise you to leave off this minute!'
</emotion>
</emotionml>

```

La ontología que se propone de base es MIRLO, ya que ella se contiene el mayor número de conceptos necesarios para definir el modelo de usuario de un RecSys para pedagogía poética. MIRLO no contiene datos personales de usuario por ser una propuesta prototipo, por lo que dichos campos personales (como la edad) se extienden utilizando la ontología FOAF. De FOAF se recoge también atributo para definir los intereses del usuario, así como su personalidad. Además, como es necesario modelar la accesibilidad y la afectividad, se usan respectivamente ADOOLES y EmotionML tal como se indica en (Abaalkhail et al., 2018). Para el caso de EmotionML se usan atributos coherentes con los rasgos especificados para los LO (poemas).

En la Tabla 3 se indican algunos de los campos principales para incluir en el modelo de usuario, especificando su ontología de origen.

| Campo | Fuente | Descripción |
|---------------------------------|-----------|---|
| Información personal | FOAF | Atributos como edad, nombre, correo... |
| Intereses | FOAF | Lista de intereses particulares de ese usuario |
| Objetivos de Aprendizaje | MIRLO | Objetivos a alcanzar, competencias asociadas y progreso alcanzado hasta el momento |
| Estilo de Aprendizaje | MIRLO | Estilos de aprendizaje: Activo/Reflexivo, Intuitivo/Juicioso, Verbal/Visual, Global/Secuencial. |
| Progreso | MIRLO | Progreso del estudiante en relación a determinados objetivos de aprendizaje. |
| Personalidad | FOAF | Personalidad en base al esquema de Myer Briggs |
| Afectividad | EmotionML | Rasgos afectivos de la categoría <i>emotion</i> que incluyen Valencia, Nivel de Excitación, Alegría, Tristeza, Miedo, Ira, Asco, Nivel de Concreción, Disponibilidad de Contexto e Imaginabilidad. |
| Accesibilidad | ADOOLES | De ADOOLES se toman los campos de <i>Discapacidad</i> (especialmente <i>MentalDisability</i>) y de <i>Dispositivo Activo</i> (<i>AssistiveMechanism</i>) para especificar el dispositivo recomendado (en caso de que aplique). |

Tabla 3 Campos principales para el modelo de usuario

Modelo pedagógico

El modelo pedagógico se define vinculando contextos y características de usuario con contenidos apropiados (los LO, en este caso sonetos) mediante reglas que comparen los valores relevantes de ambas ontologías.

Aunque no se contempla en el prototipo desarrollado como parte de esta investigación, el modelo pedagógico debe ofrecer la posibilidad de definir competencias relacionadas con los distintos LO, así como el consiguiente progreso. Además, tiene que ofrecer la posibilidad de evaluar al usuario en función de los LO con que se relacione para cuantificar dicho progreso. Por este motivo, la propuesta de ontología del modelo pedagógico de un RecSys para pedagogía poética recoge así mismo la especificación de MIRLO (Raffenne, 2013) para el *modelo de actividad* (que es el que permite vincular los distintos LO con actividades de evaluación concretas) y el *modelo de competencias*. De nuevo, dichos modelos no se implementan en el presente proyecto al restringirlo a un MVP (*mínimum viable product*) requerido para comprobar las hipótesis de investigación, sin atender a la evaluación o a la especificación de reglas que ponderen el nivel del estudiante para ajustar las recomendaciones de distintos LO.

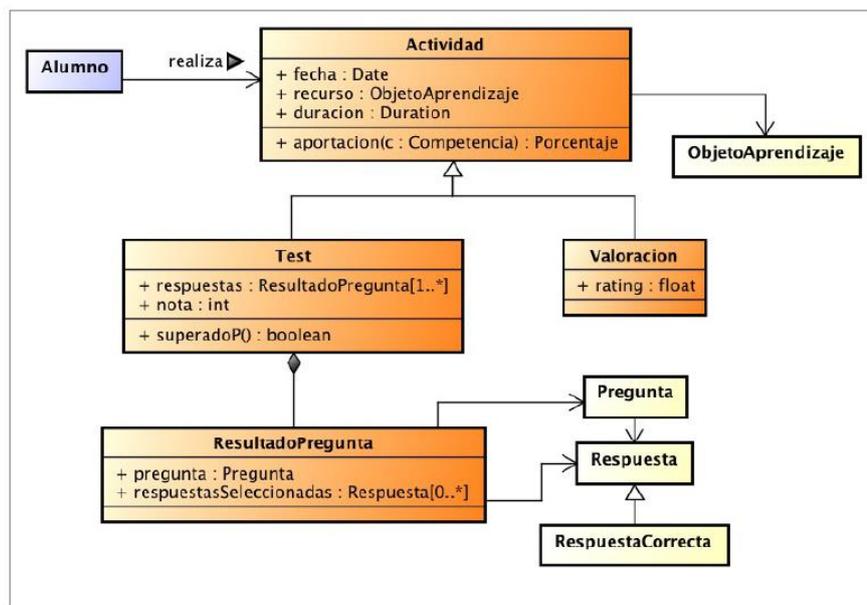


Ilustración 62 Modelo de Actividad (Raffenne, 2013)

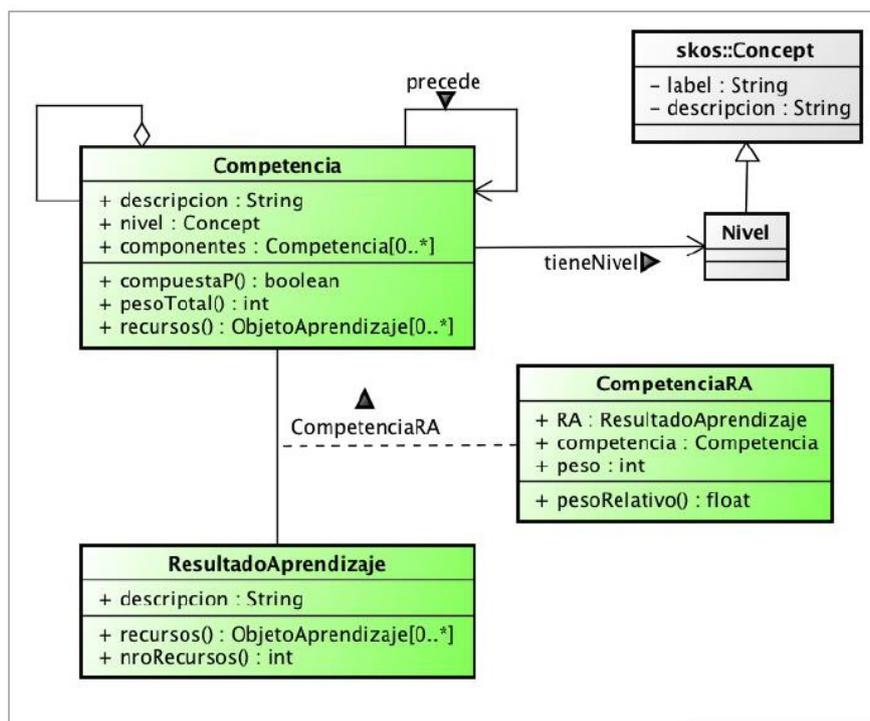


Ilustración 63 Modelo de Competencias (Raffenne, 2013)

B. Lógica

En el modelo pedagógico se ha especificado que para el RecSys para pedagogía poética se definen unas reglas que relacionen determinados atributos del modelo de usuario y del modelo de contenidos para recomendar un LO específico al estudiante.

En la presente memoria se proponen las siguientes reglas:

- Comparación entre *AffectiveFeatures* (modelo de contenidos) y *emotion* (modelo de usuario).
- Comparación entre *SemanticFeatures* (modelo de contenidos) frente a *Intereses*, *Competencias* u *Objetivos de Aprendizaje* (modelo usuario).
- Comparación entre el vector semántico de un soneto, *SemanticId*, (modelo de contenidos) frente a *Intereses*, *Competencia* u *Objetivos de Aprendizaje* (modelo de usuario).
- Comparación entre *PsychologicalFeatures* (modelo de contenidos) e *Intereses*, *Competencia* u *Objetivos de Aprendizaje* (modelo usuario).

Así, por ejemplo, si un alumno estudiante de psicología tiene entre sus *Intereses* el concepto *Grandiosidad*, o si dicho concepto aparece dentro de sus *Objetivos de Aprendizaje* (Raffenne, 2013), recibirá como recomendación un poema que entre sus *PsychologicalFeatures* contenga dicho rasgo. De esta manera se recomendarán poemas que ayuden a los estudiantes de psicología a comprender dichos conceptos psicológicos a través de la poesía en el marco de la pedagogía poética.

Además de este modelo pedagógico, la lógica incorpora un motor de IR para asistir en la búsqueda y consulta de sonetos (de manera que puedan ver que sonetos hay, definir que sonetos

aparecerían en cada LO, e incluso usarlo como buscador para anotar manualmente los sonetos con determinadas etiquetas), permitiendo una exploración del corpus por parte de los usuarios. La búsqueda de sonetos se podría hacer por palabras clave, por frases... de manera que la recuperación se hiciese en función de la similitud semántica de la consulta con respecto al valor semántico de los sonetos, comparando ambos en función del *embedding*. Este proceso es análogo al de recomendaciones semánticas para intereses de un usuario cuando no se dispone de etiquetas de temáticas anotadas en el campo *SemanticFeatures* de los sonetos, descrito previamente. La única diferencia es que se aplican técnicas de composición tras el preprocesado de las consultas de entrada (si son frases) para obtener un único vector semántico que comparar con respecto a los sonetos. Este motor de IR, diseñado atendiendo al SOTA descrito previamente, es la herramienta usada para la comprobación de las hipótesis de la investigación reportada en esta memoria.

Por último se consideran las reglas que comparan el nivel de dificultad de un poema para la comprensión de un concepto determinado con el nivel alcanzado por el usuario en dicho concepto, utilizando la información de los modelos de actividad y de competencias.

C. Interfaz

A priori, la interfaz utilizada puede ser de cualquier tipo, incluso un sistema multiplataforma. Ahora bien, para los propósitos de este proyecto se desarrolla una interfaz web, alojada en un servidor en la nube, de manera que su acceso, tanto para el diseño de cursos como para su uso, se pueda realizar desde cualquier dispositivo, pero estando todo centralizado en un mismo servidor.

3. Descripción de la experimentación

El SOTA descrito en el Capítulo 2 ha servido para detectar diversas oportunidades de investigación en el ámbito del modelado semántico y afectivo aplicado a poesía poética. Esta investigación es necesaria como punto de partida para abordar la integración de la pedagogía poética dentro de RecSys sensibles al contexto afectivo en entornos de e-learning.

La investigación realizada se ha especificado en un conjunto de hipótesis, definidas en el Capítulo 1, y que van a ser comprobadas mediante el corpus de sonetos en Castellano DISCO, ampliado con sonetos del S. XX. Este corpus contará con anotaciones de rasgos afectivos y psicológicos por parte de expertos del proyecto de POSTDATA.

Las hipótesis de investigación han definido los objetivos, experimentos y metodología que se ha llevado a cabo en este TFM. Todos estos puntos fueron introducidos en el Capítulo 1, pero se van a describir en más detalle en el presente Capítulo.

3.1. Detalle de los objetivos

En este subapartado se retoman los objetivos a cubrir en este TFM, introducidos en el Capítulo 1, con los que se buscan comprobar las distintas hipótesis de investigación.

Objetivo primario: Realizar una primera aproximación al soporte de la pedagogía poética en los SAE y, más concretamente, en los RecSys sensibles al contexto afectivo aplicados a entornos de e-learning.

- **Objetivo 1:** Delimitación de los retos tecnológicos planteados en el objetivo primario.
 - 1.1. Estudiar la arquitectura de los sistemas RecSys sensibles al contexto afectivo (especialmente para e-learning, y ver qué componentes deberían extenderse o adaptarse para dar soporte a la pedagogía poética y, en concreto, con qué funcionalidades.
 - 1.2. Ídem para las ontologías que facilitan la interoperabilidad de estos sistemas y herramientas.
- **Objetivo 2:** Aproximación a la cuestión principal: la utilización de poemas como objetos de aprendizaje personalizados, atendiendo a vincularlos con el contexto afectivo, y a capturar su significado semántico gracias a la aplicación NLP, en concreto aplicando aproximaciones basadas en semántica distribucional.
 - 2.1. Obtener un corpus de poesía en castellano de una estructura determinada (sonetos).
 - El corpus de partida será el conjunto de sonetos DISCO. Analizar y procesar dicho conjunto.
 - Ampliar el corpus de DISCO con sonetos en castellano del siglo XX
 - 2.2. Obtener y anotar el corpus de poesía con etiquetas que sean relevantes para el filtrado y personalización en contextos afectivos. Estas anotaciones deben estar en coherencia con las ontologías y lenguajes de marcado

establecidos en el campo de los RecSys sensibles al contexto afectivo aplicados a entornos de e-learning.

- Obtener anotaciones de rasgos afectivos por expertos de dominio (POSTDATA) para un subconjunto de sonetos.
 - Estudiar el nivel de acuerdo entre anotadores, y si este es suficientemente bueno, componer las anotaciones en un único *dataset* para esos sonetos. Este *dataset* es el que se usará en los pasos siguientes.
 - Obtener el valor afectivo a nivel de palabras individuales y ver si hay correlación entre las anotaciones a nivel de soneto y la composición de anotaciones a nivel de palabra.
 - Extender las anotaciones afectivas a todo el conjunto de sonetos del corpus.
 - Ampliar las anotaciones XML-TEI de DISCO con diferentes rasgos afectivos
- 2.3. Modelar semánticamente el corpus de poesía mediante técnicas basadas en *word embeddings* y composición de *embeddings*. Valorar empíricamente para el caso de uso de poesía en castellano con una estructura determinada (sonetos) cuáles de las técnicas antes mencionadas dan mejores resultados.
- Obtener distintos tipos de *word embeddings*, en concreto word2vec y BERT) para castellano.
 - Procesar el conjunto de sonetos y asignar los *embeddings* a nivel de palabras.
 - Aplicar técnicas de composición para componer los *embeddings* individuales en *embeddings* a nivel de soneto o de estrofa.
 - Desarrollar una API que permita recuperar poemas a partir de sus contenidos semánticos para las técnicas basadas en *word embeddings* y composición de *embeddings*. Esta aplicación servirá tanto para facilitar la evaluación de las distintas alternativas como para tener una primera versión de una API que se integraría en herramientas de autor.
 - Usando dicha API, crear una aplicación que muestre los poemas recuperados con dos funciones de composición distintas y una misma métrica de comparación, para poder así estudiar si la efectividad de las distintas funciones de composición depende del tipo de *embedding* usado de fondo. Para ello se mostrará dicha aplicación en la feria de la educación de AULA⁶⁸ para recoger los datos de los usuarios y estudiar posteriormente los resultados. Se va a aprovechar también para pedir información afectiva a los usuarios (sobre valencia y excitación) para hacer una analítica sobre la información afectiva declarada, la propia de la palabra introducida y la posible relación entre la que tiene el soneto seleccionado y la puntuación emitida posteriormente.
 - Los datos de puntuación emitida en los sonetos mostrados en AULA servirán para realizar un primer análisis sobre si los rasgos que se mencionan en (Ulrich et al., 2016) para poder predecir el gusto estético

⁶⁸ IFEMA. “AULA, Salón internacional del estudiante y de la oferta educativa con la presentación del espacio *Te regalo un soneto*”. Web. 15/05/2019. http://www.ifema.es/aula_01

de un poema en alemán también sirven para predecirlo con poesía en castellano.

- Comprobar si BERT aporta más que word2vec dentro de un sistema de IR para poesía. Para ello se desarrollará una aplicación que presente los poemas que se recuperarían con distintas frases de entrada, utilizando para un caso BERT y para otro word2vec, usando una misma función de composición y una misma métrica de comparación.
 - Como alternativa al punto anterior, se realizará el análisis comparativo de BERT frente a word2vec analizando la efectividad en la recuperación de sonetos según el término psicológico anotado por POSTDATA.
- 2.4. Plantear un estudio preliminar entre rasgos psicológicos vinculados a determinados trastornos psicológicos y el valor afectivo en la poesía.
- Obtener las anotaciones de rasgos psicológicos de POSTDATA.
 - Estudiar el nivel de acuerdo entre anotadores y si este es suficientemente bueno componer las anotaciones en un único *dataset* para esos sonetos. Este *dataset* es el que se usará en los pasos siguientes.
 - Analizar la correlación entre los rasgos afectivos (objetivo 2.2.) y los rasgos psicológicos.
- 2.5. Comprobar si usando como entrada los términos de trastornos psicológicos el sistema es capaz de recuperar del subconjunto de sonetos anotados aquellos que están etiquetados con dichos términos. Como referencia para los rasgos psicológicos a usar se parte de la memoria de (García Franco & Manjarrés Riesco, 2016).
- Obtener las anotaciones de rasgos psicológicos de POSTDATA.
 - Estudiar el nivel de acuerdo entre anotadores y si este es suficientemente bueno componer las anotaciones en un único *dataset* para esos sonetos. Este *dataset* es el que se usará en los pasos siguientes.
 - Usar el sistema definido en el subobjetivo 2.2 pero sólo usando como corpus de sonetos el subconjunto anotado por POSDATA.
 - Introducir los términos de trastornos psicológicos y recuperar una lista ordenada de tamaño M de los sonetos recomendados, siendo M el número de sonetos totales de este corpus.
 - Obtener los valores de precisión comparando la lista final de sonetos recuperados ordenados de forma creciente por su métrica respecto a una lista teórica ideal. Esta lista ideal contiene los M_1 primeros sonetos con la etiqueta del término psicológico (donde M_1 es el número de sonetos para un término psicológico concreto) y el resto de valores a 0. También se calculará la precisión media (MAP) sobre una lista de tamaño M con M el número de sonetos totales de este corpus.
 - Se probarán varias de las funciones de composición y se darán los resultados comparados.

Los subobjetivos 2.1 y 2.2 sirven como marco de evaluación de las hipótesis 1 y 2, mientras que el 2.3 sirve para la 3. El subobjetivo 2.4 sirve para un primer análisis de la hipótesis 4. El subobjetivo 2.5 sirve para la hipótesis 5. Esta evaluación se realiza principalmente mediante

los experimentos gracias a la colaboración con el laboratorio de POSTDATA y a la feria de AULA. En el apartado siguiente de *Metodología* se describirán los pasos a realizar en dichos experimentos para evaluar las distintas hipótesis.

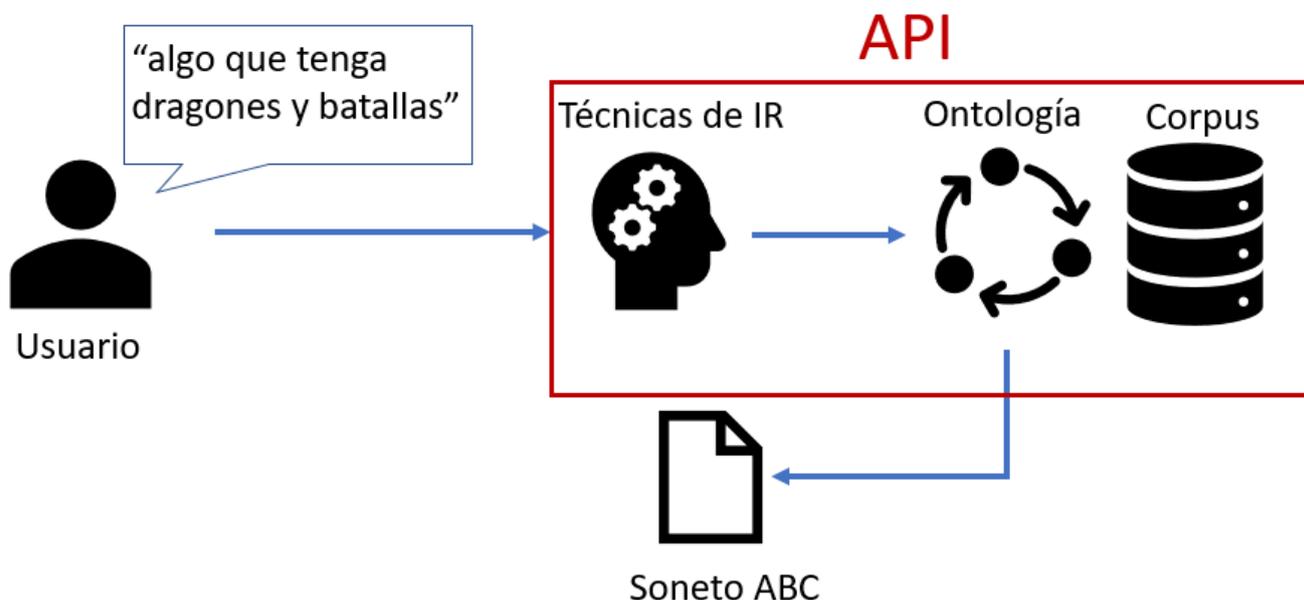


Ilustración 64 Esquema de alto nivel de la API desarrollada para la experiencia de AULA. Recibe las consultas del usuario y que devuelva el soneto más similar semánticamente.

3.2. Detalle de los experimentos

En este apartado se describen en detalle los experimentos llevados a cabo para comprobar las distintas hipótesis de investigación, tras su introducción en el Capítulo 1.

POSTDATA

Con la colaboración de POSTDATA se abordan 3 experimentos para analizar tres de las hipótesis planteadas en el apartado anterior.

- Experimento 1:
 - Para comprobar la hipótesis 2, en la que se recogía que se pueden componer el valor afectivo de palabras individuales para obtener rasgos afectivos a nivel de soneto, se comparan las anotaciones afectivas del subconjunto de sonetos etiquetados con las derivadas de agregar la información individual, aplicando los pasos del objetivo 2.2.
 - El primer paso consiste en analizar el nivel de acuerdo entre anotadores para esas anotaciones. Para ello se realiza un análisis con la *alfa de Krippendorff*. Si el valor de alfa es suficientemente bueno (> 0.21) se validan las anotaciones y se crea un único *dataset* en el que el valor asignado a cada soneto es la mediana de los valores anotados por los distintos expertos de POSTDATA.
 - Si se cumple el punto anterior se prosigue con el experimento. Con ello, la hipótesis se aceptará para el caso de que apareciese entre el vector de rasgos

inferido y correspondiente anotado una correlación superior a 0.21 (*correlación de Spearman* debido a que las variables son ordinales).

- Si no existe esa correlación no se puede asegurar que la composición a nivel individual sirva para obtener los rasgos globales. Se abrirían aquí líneas de investigación posterior para analizar otras funciones de composición, ver si hay relaciones entre ambas que no son tan triviales, ...

- Experimento 2:

- Para analizar la hipótesis 3.1, en la que se va a estudiar si BERT como *word embedding* contextual da mejores resultados que word2vec, se dan dos alternativas experimentales.

- **Opción 1)**

- Se va a desarrollar una aplicación prototipo que muestre ante una consulta de entrada dos poemas recuperados usando una misma función de composición y una misma métrica pero distinto *word embedding*.
- La composición de textos se va a hacer a nivel de estrofas, de cara a tener un equilibrio entre tamaño de texto y representación semántica de las palabras que lo componen en el vector final. Esta aproximación es adecuada para plantear este experimento. De esta manera, un soneto quedará representado por 4 vectores de composición (uno por estrofa).
- Los expertos de POSTDATA elegirán el soneto recuperado que consideren que es más próximo semánticamente a la consulta introducida. Junto con ello puntuarán el nivel de satisfacción con el soneto elegido y recuperado (escala del 1 al 5).
- Se analizará la selección de poemas que usan BERT y su puntuación frente a las de word2vec. Para ello se hará un contraste de hipótesis con *t de Student* para ver que los vectores de selección de sonetos son significativamente distintos (es decir, que se han seleccionado muchos más sonetos con un *embedding* que con otro). En caso de haber diferencia (se rechazaría H_0 que determina que las medias son iguales) y que la media de selección con BERT fuese mayor que la de word2vec, se aceptaría la hipótesis 3.1.
- En caso de que no se pueda rechazar H_0 no se tendría suficiente información para concluir nada sobre la hipótesis 3.1. Se necesitaría una investigación más exhaustiva.

- **Opción 2)**

- De forma alternativa al caso anterior, se pueden comparar los resultados de efectividad en la recuperación de sonetos usando como referencia los rasgos psicológicos anotados por POSTDATA.
- De esta manera se compararán las métricas de precisión y de precisión media (*mean average precision*, MAP) para cada *embedding* para ver si hay diferencias significativas, tal y como se ha indicado en la Opción 1.
- Aun planteándose estas dos alternativas y desarrollándose la aplicación para abordar la Opción 1, en esta memoria se incluirá el análisis para esta Opción 2, remitiéndose el uso de la herramienta mencionada en la Opción 1 para líneas futuras.

- Experimento 3:

- Para comprobar la hipótesis 4 se obtienen los datos anotados con los rasgos psicológicos relevantes (0 o 1 dependiendo de si el soneto los representa o no) procedentes de (García Franco & Manjarrés Riesco, 2016).
- El primer paso consiste en analizar el nivel de acuerdo entre anotadores para esas anotaciones. Para ello se realizará un análisis con la *alfa de Krippendorff* (Krippendorff, 2011). Si el valor de alfa es suficientemente bueno (> 0.21) se validan las anotaciones y se crea un único *dataset* en el que el valor asignado a cada soneto será la mediana de los valores anotados por los distintos expertos de POSTDATA.
- Si se cumple la condición anterior se prosigue con el experimento. Con ello, la hipótesis se acepta en el caso de que haya entre el vector de rasgos psicológicos y el vector de rasgos afectivos anotados una correlación superior a 0.21 (correlación de Spearman debido a que las variables son ordinales).
- En el caso de que no haya correlación no se tiene suficiente información para asegurar nada sobre la hipótesis 4. Se necesitarían investigaciones posteriores para analizar posibles relaciones de dependencia más complejas.

- Experimento 4:

- Para comprobar la hipótesis 5 se obtienen los datos anotados con los rasgos psicológicos relevantes (0 o 1 dependiendo de si el soneto los represente o no).
- El primer paso consiste en analizar el nivel de acuerdo entre anotadores para esas anotaciones. Para ello se realiza un análisis con la *alfa de Krippendorff*. Si el valor de alfa es suficientemente bueno (> 0.21) se validan las anotaciones y se crea un único *dataset* en el que el valor asignado a cada soneto será la mediana de los valores anotados por los distintos expertos de POSTDATA.
- Si se cumple la condición anterior se prosigue con el experimento. Se irán introduciendo los distintos términos y se obtendrá como salida una lista ordenada de los sonetos semánticamente más similares. El tamaño de la lista, N , corresponde al número de términos etiquetados.
- Se prueba la función de composición RL_SEQ (joint). Como métricas de comparación se usa el coseno y e ICM 1.2. Como *embeddings* se prueban BERT y word2vec.
- La composición de textos se va a hacer a nivel de estrofas de cara a tener un equilibrio entre tamaño de texto y representación semántica de las palabras que lo componen en el vector final. Esta aproximación es adecuada para plantear este experimento. De esta manera, un soneto quedará representado por 4 vectores de composición (uno por estrofa).
- Se obtiene la métrica de precisión dividiendo el número de sonetos en esa categoría con respecto al tamaño de la lista recuperada. También se calculará la MAP sobre una lista de tamaño M con M el número de sonetos totales de este corpus.
- Se verá que solución da los mejores resultados (mediana de MAP más alta).
- Los resultados serán aceptables si su MAP es mayor que 0.25, que corresponde con una asignación aleatoria de etiquetas con un 25% de probabilidades para una lista de 230 términos, coincidente con los 230 sonetos etiquetados. Esta lista

aleatoria tendría, por lo tanto, unos y ceros alternos siendo el primero de ellos un 1.

AULA

Con la colaboración de AULA se abordan 2 experimentos para analizar dos de las hipótesis planteadas en el apartado anterior.

- Experimento 5:

- Para comprobar la hipótesis 1, que menciona que los rasgos afectivos sirven para predecir lo que a un usuario va a gustar un soneto se van a utilizar los rasgos afectivos derivados de las palabras individuales en caso de que la hipótesis 2 sea cierta. En caso contrario se podrán utilizar las anotaciones sólo si el conjunto de sonetos elegidos por los usuarios coincide con el conjunto de los sonetos anotados (al menos 30 muestras).
- Se desarrolla una aplicación para AULA mediante la cual los usuarios escriben una consulta y el sistema les recomiende sonetos semánticamente similares.
- Además de ello, los usuarios declaran su estado afectivo en lo que respecta a nivel de excitación y valencia. En el Capítulo 3 se describe en más detalle dicha aplicación.
- Se analiza la correlación (correlación de Spearman debido a que las variables son ordinales) entre el valor afectivo de los sonetos seleccionados y la puntuación emitida. La hipótesis 1 sería cierta si la correlación es superior a 0.21 (correlación de Spearman debido a que las variables son ordinales). En ese caso será posible utilizar los rasgos afectivos como predictores.
- En caso de que no haya correlación, no se sabrá si la hipótesis 1 es cierta (ya que podría haber dependencias no triviales entre los rasgos y la salida). Sería necesaria una investigación en mayor profundidad.
- A su vez, se usa la información afectiva de los usuarios para ver si hay correlación entre ella y la puntuación emitida. Con ello se quiere ver si el estado afectivo de los usuarios ha influido en las elecciones de un soneto frente a otro. Influirá si la correlación es superior a 0.21 (correlación de Spearman debido a que las variables son ordinales).

- Experimento 6:

- Para comprobar la hipótesis 3 se usa la misma aplicación indicada en el apartado anterior.
- Se muestran 2 sonetos recuperados según la proximidad semántica ante una consulta del usuario, acotada a una única palabra.
- La función de composición usada para uno de los sonetos es RL_SEQ y para otro F sum. La métrica para ambos casos será el coseno. Los *word embedding* usados serán BERT.
- Para este caso, en (Amigó Cabrera & Fresno Fernández, 2018) se menciona que los resultados son prácticamente los mismos para el caso de definiciones de diccionario y de word2vec.
- La composición de textos se hace a nivel de estrofas de cara a tener un equilibrio entre tamaño de texto y representación semántica de las palabras que lo

componen en el vector final. Esta aproximación es adecuada para plantear este experimento. De esta manera, un soneto quedará representado por 4 vectores de composición (uno por estrofa).

- Así, la idea es comparar el vector de elecciones de los sonetos de un tipo frente a los del otro, así como las puntuaciones dadas a los mismos. Si no son significativamente distintos (no se puede rechazar H_0 del contraste de *t de Student*) no se puede rechazar la hipótesis 3 y por tanto podría ser cierto que los resultados sean extrapolables al escenario de esta memoria.
- Si por el contrario se rechaza H_0 en el caso anterior se puede concluir que las conclusiones que menciona el artículo no son extrapolables al escenario de esta memoria. Esto abriría una importante investigación posterior para entender las causas de la discrepancia, ya que el artículo propone una definición axiomática de funciones de composición.

En la Tabla 4 se muestra una tabla resumen para los experimentos mencionados.

| Experimento | Hipótesis a la que aplica | Criterio de aceptación | Resultado positivo | Resultado negativo |
|-------------|---------------------------|---|----------------------------------|------------------------|
| 1 | 2 | Alpha > 0.21. Corr (Spaeman) > 0.21 | Aceptar hipótesis 1 | No concluyente |
| 2 | 3.1 | Rechazar H_0 en t Student. Media BERT > Media word2vec | Aceptar hipótesis 3.1 | No concluyente |
| 3 | 4 | Alpha > 0.21. Corr (Spaeman) > 0.21 | Aceptar hipótesis 4 | No concluyente |
| 4 | 5 | Alpha > 0.21. MAP > 0.25 | Aceptar hipótesis 5 | Rechazar hipótesis 5* |
| 5 | 1 | Hipótesis 2 aceptada o 30 usuarios que hayan elegido sonetos anotados. Corr (Spaeman) > 0.21 | Aceptar hipótesis 1 | No concluyente |
| 6 | 3 | No rechazar H_0 en t Student. | No se puede rechazar hipótesis 3 | Se rechaza hipótesis 3 |

Tabla 4 Solo se rechaza la hipótesis 5 si el MAP medio para todos los términos es menor que 0.25 y de manera individual dicho MAP no es consistentemente menor que 0.25

3.3. Detalle de la metodología aplicada: CRISP-DM

En el Capítulo 1 se introdujo que la metodología aplicada a este proyecto es CRIPS-DM. Siguiendo con ello, en este apartado se retoman los puntos ya introducidos detallando los pasos que se llevan a cabo dentro de ellos vinculados a los experimentos realizados.

1. **Comprensión de Negocio:** Comprensión del valor de la pedagogía poética y su valor para recomendar objetos de poesía personalizada desde un IR.
2. **Comprensión de Datos:** Análisis del corpus de sonetos (DISCO) y de los sonetos del S. XX. Análisis de los *datasets* para asociar información afectiva a las palabras que aparecen en dichos sonetos.
3. **Preparación de Datos:**
 - Preprocesado de los sonetos.
 - Contexto afectivo: Para los rasgos afectivos se han eliminado *stopwords* y se ha llevado al lema (lo que se conoce como proceso de lematizado o *lemmatization*) las palabras de cara a maximizar la correspondencia con las palabras de los *datasets* afectivos, que también se han lematizado. Con ello se generan los rasgos afectivos y se anotan sobre la estructura TEI XML (para los sonetos de DISCO) y como un diccionario global (para DISCO más los sonetos del S. XX). En la fase de preparación de datos también se analizarán los sonetos anotados por los distintos expertos de POSDATA para ver el nivel de acuerdo que hay entre ellos y definir un único *dataset* final de anotaciones con las anotaciones comunes, y para el caso de que haya discrepancia, con la mediana de los valores dados por todos los anotadores.
 - Contexto semántico: Para la parte semántica se han obtenido los *word embeddings* de las palabras de los sonetos en bruto usando BERT. Después, para la composición a nivel de estrofas y de soneto de dichos *embeddings* individuales se han eliminado las *stopwords* para minimizar el ruido de dicha composición. Los vectores semánticos se han persistido en los diccionarios mencionados en el punto anterior.
4. **Modelado:** En el modelado se especifica una función que relaciona entradas con una salida concreta y con la que se busca maximizar una determinada función de utilidad.
 - Contexto afectivo: En este caso la entrada son los rasgos afectivos obtenidos partiendo de las palabras individuales y la salida es la anotación de los expertos de dominio. La evaluación se hará analizando la correlación entre ambas.
 - Contexto semántico: Para este caso, se dispone de unas entradas (consultas del usuario o los términos psicológicos anotados) y una salida (el soneto recomendado) y se elige el soneto (o lista de sonetos) que tenga una estrofa cuyo vector semántico sea lo más cercano posible al vector semántico de las entradas (función de utilidad). En concreto para este trabajo se ha experimentado con la función coseno y con ICM1.2 e ICM1.4 (Amigó Cabrera & Fresno Fernández, 2018).
5. **Evaluación:**

- Contexto afectivo: Se evalúan los resultados de la correlación entre rasgos afectivos y anotaciones, y en caso de que haya similitud se valida la anotación de todo el corpus a través de sus palabras individuales.
- Contexto semántico: Se evalúan a nivel semántico los resultados de usar unas funciones de composición (RL_SEQ y SUM) con una métrica de comparación (coseno) y usando BERT en el entorno de AULA para ver si los resultados son consistentes con (Amigó Cabrera & Fresno Fernández, 2018). A su vez se evalúa RL_SEQ y coseno usando BERT frente a word2vec mediante la selección de resultados en la aplicación por parte de POSTDATA. Por último se evalúan los resultados de las listas de sonetos recuperados según los términos psicológicos frente a las anotaciones que tienen.

6. *Iteraciones*:

- Contexto afectivo: En caso de que haya una correlación adecuada (> 0.21) entre rasgos por palabras individuales y anotaciones se valida la anotación de todo el corpus a través de sus palabras individuales. Si no la hay se plantean otras alternativas para la anotación de rasgos en el corpus usando como punto de partida las anotaciones de POSTDATA. Una opción sería usar un algoritmo de aprendizaje supervisado con los vectores semánticos del texto como entrada y las anotaciones afectivas como salida.
- Contexto semántico: Con los resultados obtenidos se ve, para la experimentación semántica, si probar con otras funciones de composición y *embeddings* o proponer ya una combinación de ellas para una primera versión de la API final.

7. *Despliegue*: Se considera como despliegue la publicación final de la primera versión de la API con el IR para poesía y con el corpus de sonetos anotados de manera que se pueda integrar con un SAE posteriormente.

4. Descripción de la aplicación

En este apartado se describe el código desarrollado para la aplicación, explicando las funciones principales que componen los distintos scripts del sistema, detallando las entradas de datos más relevantes, resumiendo las operaciones que realizan e indicando la salida esperada.

4.1. Introducción

En este apartado se va a describir el código desarrollado para la aplicación. La manera en que se va a hacer será describir las funciones principales que componen los distintos scripts del sistema, detallando las entradas de datos más relevantes, resumiendo que operaciones realizan e indicando cual es la salida esperada.

El desarrollo realizado se divide principalmente en la implementación de los siguientes componentes:

- Un componente de procesado de sonetos, que incluye la extracción de información desde el repositorio de sonetos de DISCO⁶⁹ (Ruiz Fabo et al., 2017), la construcción de *embeddings* por soneto y por estrofa usando distintas técnicas de *embeddings* para palabras y de composición para textos, y la construcción de rasgos afectivos que incluir dentro del XML TEI de DISCO (Usando los *corpora* detallados en el capítulo anterior), junto con la ampliación del repositorio con sonetos del S. XX.
- El diseño de una API con la que interactuar y que sirva para devolver el soneto más adecuado ante un texto de consulta.
- El diseño de una aplicación de escritorio para evaluar los resultados con distintas funciones de composición, para ser utilizada en la feria de AULA⁷⁰.
- El diseño de una interfaz web para acceder directamente a la API antes mencionada.
- El desarrollo de una aplicación para asistir a los anotadores de POSTDATA en la anotación afectiva y de rasgos psicológicos de un subconjunto de poemas.
- El desarrollo de una aplicación para que los anotadores de POSDATA seleccionen los sonetos recuperados con mayor acierto ante un mismo texto de entrada usando BERT y word2vec con una misma función de composición y métrica de comparación.

Todos estos puntos se explicarán y desarrollarán en los siguientes apartados.

Cabe mencionar que el código se ha desarrollado desde una perspectiva funcional, y por este motivo la manera antes citada es la mejor forma de explicarlo (ya que no se han definido prácticamente clases ni objetos).

4.2. Descripción del código

A continuación se procede a describir los distintos módulos que se han desarrollado para abordar las tareas mencionadas previamente.

⁷⁰ “AULA”. IFEMA.Web. 15/05/2019. http://www.ifema.es/aula_01

Procesado de sonetos⁷¹

La primera tarea que abordar ha consistido en, partiendo de los sonetos en XML TEI disponibles en el repositorio de DISCO y utilizando las bases de datos afectivos por palabras descritas en el Capítulo anterior, calcular para cada poema los rasgos afectivos identificados como potencialmente relevantes y añadir dicha información al XML TEI. Para ello se calculan los rasgos a nivel de estrofa y de soneto usando las palabras lematizadas, sin *stopwords* y extrayendo su valor afectivo de la palabra correspondiente desde las tablas indicadas en el Capítulo 2. Para los casos en los que haya palabras que aun lematizadas no aparezcan en las tablas de referencia no se considerarán para el cálculo de dichos rasgos.

Esta tarea, además de aplicarla sobre DISCO se quiere aplicar sobre poemas más recientes, de manera que se han obtenido desde internet algunos sonetos para autores en castellano del siglo XX. Se han extraído para ello **166 sonetos** manualmente de una web⁷², dando lugar a un total de **4251 sonetos** en el corpus. Como esos sonetos adicionales no tienen representación en XML TEI se ha construido un JSON con todos los campos relevantes y sobre dicho JSON se ha trabajado en el resto del programa. En el anexo 9.3 se detalla la información por soneto que contiene el documento.

Junto con ello se han extraído los embeddings por palabra usando BERT y se han combinado dichos embeddings a nivel de estrofa y de soneto entero usando las dos funciones de composición ya mencionadas (F sum y F joint).

La obtención de embeddings con BERT se ha llevado a cabo usando de base la librería de Deep Learning para NLP **GluonNLP**⁷³, que es una interfaz que engloba de fondo la librería de Deep Learning de Amazon de uso genérico Apache MXNet⁷⁴. Esto se ha hecho tras evaluar otras posibles librerías que incluyen la implementación de dicho modelo de *embeddings*, como ocurre con TensorFlow y PyTorch, al verse que GluonNLP acababa ofreciendo la interfaz de uso más sencilla, y aún más gracias al *wrapper* que ofrece sobre ella la librería *bert_embedding*⁷⁵.

BERT ofrece distintos datasets en los que está preentrenado con distinta dimensionalidad del espacio vectorial generado. Para este caso se ha usado, por ser una buena combinación de tamaño en memoria y efectividad, el corpus multilingüe bert_12_768_12 (12 capas, 12 *heads*, 768 parámetros)⁷⁶.

Antes de obtener los *embeddings* con BERT se usan distintas librerías de NLP para llevar a cabo un preprocesado de los textos de origen, donde aparecerán algunos procesos como los siguientes:

- Tokenización de los strings originales para tener listas de palabras

⁷¹ Para más información se puede consultar el repositorio publicado con el correspondiente código: “sonnet-recommender”. Alberto Barbado. Web. 25/05/2019.: <https://github.com/AlbertoBarbado/sonnet-recommender>

⁷² A media voz. Web. 25/05/2019. <http://amediavoz.com/>

⁷³ “gluon-nlp”. MXNet. Web. 23/05/2019. <https://gluon-nlp.mxnet.io/>

⁷⁴ “Apache MXNet”. Wikipedia. Web. 23/05/2019. https://en.wikipedia.org/wiki/Apache_MXNet

⁷⁵ “BERT embedding”. Google. Web. 23/05/2019. <https://pypi.org/project/bert-embedding/>

⁷⁶ “gluon-nlp”. MXNet. Web. 23/05/2019. <https://github.com/dmlc/gluon-nlp/releases>

- Eliminación de mayúsculas
- Eliminación de caracteres numéricos y de otros no alfanuméricos
- Eliminación de stopwords
- Lematización de las palabras del texto

Con ello se obtendrían distintas listas de palabras por soneto y por estrofa para su uso posterior (principalmente la lista de las palabras originales sin mayúsculas, la lista de las palabras originales sin stopwords y la lista de las palabras originales lematizadas y sin stopwords).

De esta manera, en primer lugar se usa BERT para obtener los *embeddings* de los textos de origen. BERT usa como base el conjunto de todas las palabras originales sin mayúsculas (para que los vectores numéricos sean lo más realistas posibles al considerar toda la realidad semántica y de relación de palabras de las frases originales).

Ahora bien, para obtener la composición de los textos en un único vector sí que se eliminan, de los vectores obtenidos previamente para cada texto, las *stopwords*, porque introducen ruido en la agregación de los vectores con las funciones de composición. Esto es así debido a que dichos vectores quieren representar el valor semántico de los textos y, las *stopwords*, aunque son importantes para ver relaciones sintácticas entre palabras, desambiguar o conocer el género de una determinada palabra y no introducir así sesgos (p.e. el/la médico), no tienen de por sí un valor semántico que aportar después de ello y por este motivo son eliminadas en este punto.

Este proceso es el que se considerará para las consultas que entren al IR cuando se usa BERT: obtener los *embeddings* por palabra usando toda la frase sin mayúsculas, y después componer la consulta usando únicamente las palabras sin *stopwords*.

Por otro lado, otra de las funciones del procesado de sonetos es la obtención de *embeddings* por palabra y la agregación de estos por texto usando word2vec. Como matriz de partida para word2vec se usa como corpus **spanish-word-embeddings**⁷⁷, concretamente el esquema construido sobre el corpus *spanish-billion-corpora*, construyendo el espacio vectorial usando Skip-Gram desde su implementación con GenSim, que contiene más de 1M de palabras en un espacio vectorial de 300 dimensiones.

Junto con ello se preprocesan también (aunque no se usará en los experimentos actuales) los vectores word2vec usando como base los vectores de FastText⁷⁸ (así incluyendo los n-gramas de las palabras que no aparezcan en el corpus).

Para todos los procesados de NLP mencionados previamente se usan las librerías:

- spaCy⁷⁹ para la lematización de las palabras en word2vec usando el corpus que incluye para el castellano *es_core_news_sm*
- NLTK⁸⁰ para eliminar las stopwords

Se menciona finalmente que para todas las comparaciones el nivel de granularidad que se ha escogido es el de **estrofa**. Es decir, cada poema estará representado por cuatro vectores (uno por estrofa) compuestos usando las funciones de composición previamente mencionadas, de manera que la comparativa con respecto al mensaje de entrada se hará estrofa a estrofa y se devolverá el soneto cuya estrofa sea más similar semánticamente al mensaje de entrada

⁷⁷ “Spanish Word Embeddings”. Web. 25/05/2019. <https://github.com/dccuchile/spanish-word-embeddings>

⁷⁸ “fastText”. Facebook. Web. 25/05/2019. <https://github.com/facebookresearch/fastText>

⁷⁹ “spaCy”. Web. 25/05/2019. <https://spacy.io/>

⁸⁰ “NLTK”. Web. 25/05/2019. <https://www.nltk.org/>

recibido. Este método podría refinarse aún más en el futuro considerando en la comparativa también las métricas del resto de estrofas de dicho soneto.

Aplicación para AULA

Como se ha indicado previamente, la aplicación para AULA buscó comprobar si el estado afectivo de los usuarios (declarado por ellos) guarda relación con las palabras que introdujeron en la búsqueda y con su elección de soneto, buscando con ello analizar principalmente si las conclusiones sobre composición afectiva reportadas en (Amigó Cabrera & Fresno Fernández, 2018) para word2vec, utilizando definiciones de diccionario y sus palabras de entrada correspondientes en inglés, podrían extrapolarse al castellano, usando como referencia palabras de consulta y sonetos recuperados, y como *embedding* el que ofrece BERT.

Así, se ha desarrollado una aplicación denominada SAREK (*Sonnet Affective Recommender using Embedding Knowledge*) usando la librería Tkinter⁸¹ desarrollando una interfaz con la que a modo de Asistente Virtual el sistema fuese solicitando información y pudiese presentar los sonetos recuperados al usuario.

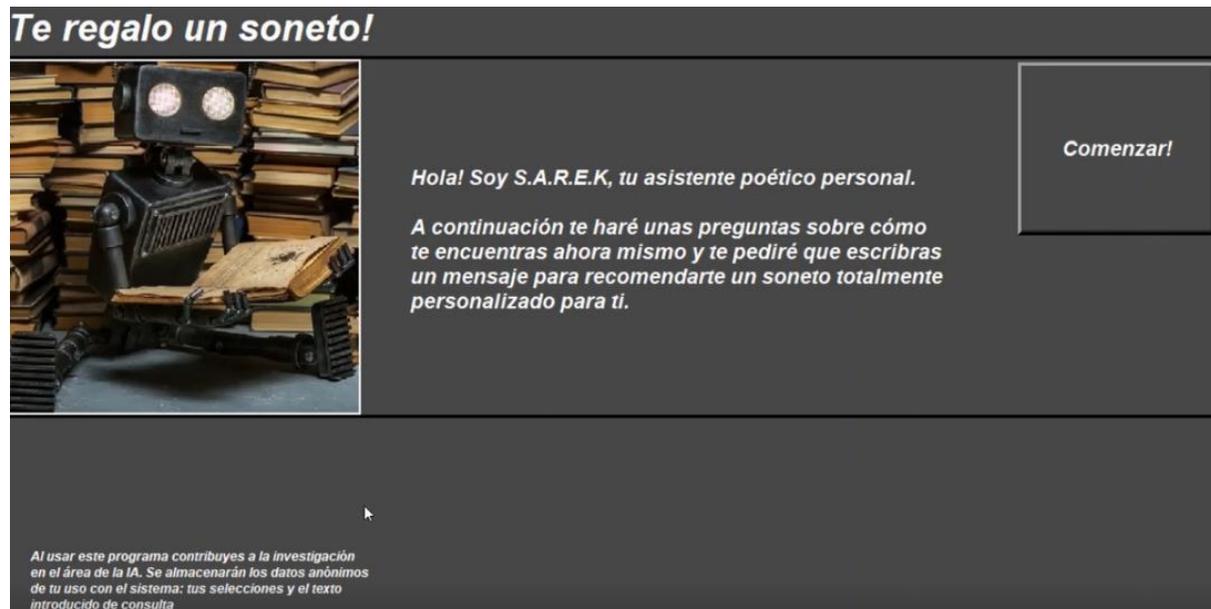


Ilustración 65 La primera pantalla consiste en una presentación del sistema, una descripción de en qué va a consistir la interacción con él, y un breve declaración de responsabilidad sobre los datos que se van a guardar de la interacción del usuario

⁸¹ “Tkinter”. Web. 25/05/2019. <https://wiki.python.org/moin/TkInter>

Te regalo un soneto!



¿Cómo te sientes ahora?

- Muy tranquilo/a
- Irritado/a
- Tranquilo/a
- Frustrado/a
- Activo/a
- Triste
- Muy activo/a
- Seguro/a
- Entusiasmado/a
- Alegre

Reiniciar

Siguiente página

Ilustración 66 Se muestran dos columnas con categorías de nivel de excitación/valencia para que el usuario eligiese una de cada columna

Te regalo un soneto!



Dime una palabra y te sugeriré un par de sonetos: guerra

consultar!

¿Oís...? Es el silbar de los cañones;
el bote de la lanza en la cimera;
el grito atronador de la guerrera
trompa y el relinchar de los bridones.

Mirad en las murallas los pendones;
lagos de sangre son monte y pradera,
y extendidas están por la ancha esfera,
cual campo de batalla, las naciones.

La guerra es nuestro dios; sacrificamos
en el altar de la ambición la vida;
la victoria de lauros coronamos,

sin ver nuestra razón oscurecida
que el paladín, que hoy héroe contemplamos,
mañana ha de pasar por homicida!

La guerra por un caso inevitable
invadió la Española Monarquía,
juzgando que aceptada, acabaría
de una vez con la gente miserable,

y rehusada, al Monarca respetable
la gloria militar rebajaría.
El pueblo frece a Carlos a porfía
dones mil del tesoro inagotable

de su amor; y por Carlos negociada,
viene la paz con palma de victoria.
La guerra cruel corriendo apresurada

tantos despojos deja en nuestra tierra,
que Carlos de la paz saca la gloria
y el pueblo la abundancia de la guerra.

Click aquí si prefieres este

Click aquí si prefieres este

Reiniciar

Ilustración 67 Después se presenta el IR. El usuario introduce una palabra de consulta y el sistema muestra dos sonetos recuperados usando la función coseno como métrica y F_{joint} y F_{sum} como funciones de composición. El orden en que se muestran es aleatorio para no introducir sesgos. El usuario elegiría uno de los dos sonetos y pasaría a la pantalla final.

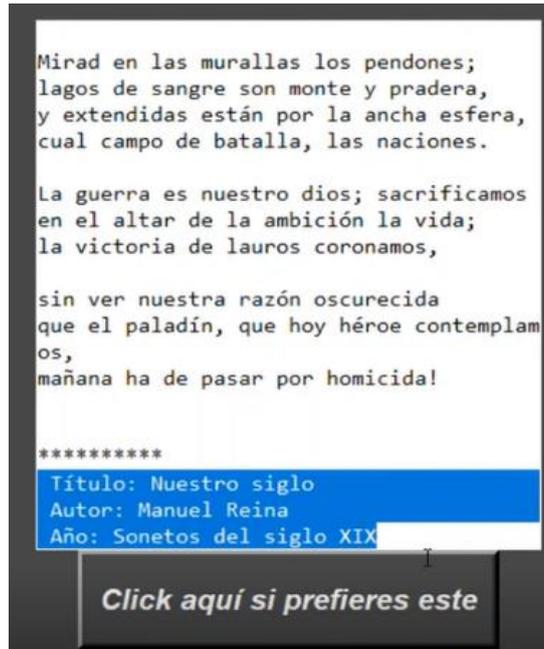


Ilustración 68 Haciendo scroll sobre el soneto se pueden ver además los datos del título, autor y siglo, tal y como aparecen en los XML TEI de DISCO o de forma análoga en el json con los sonetos adicionales del siglo XX

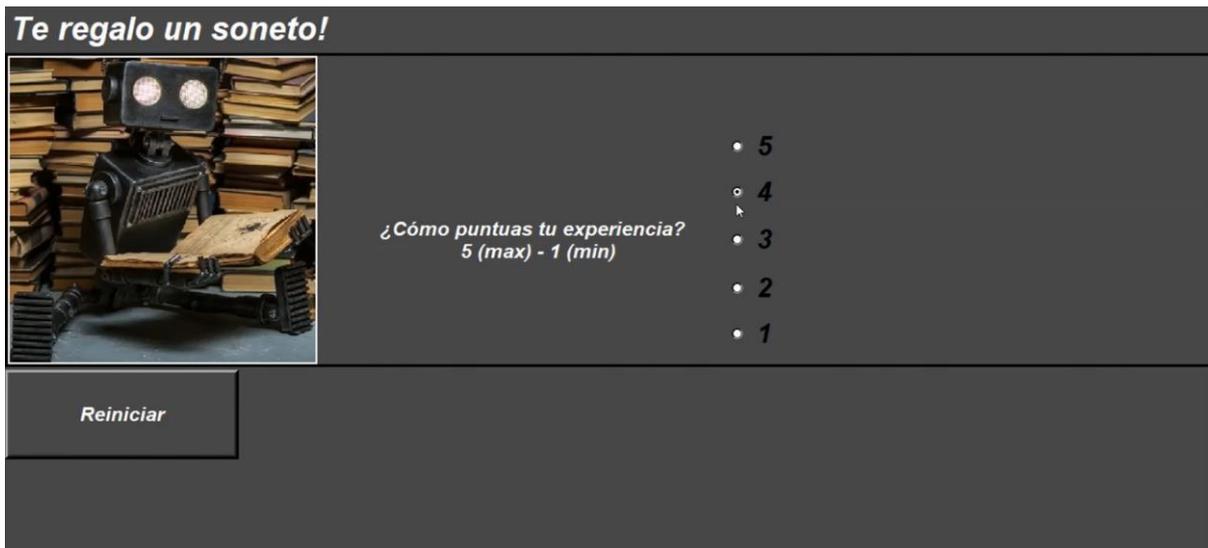


Ilustración 69 Finalmente el usuario puntúa su experiencia con el sistema

Los resultados de la interacción del usuario con el sistema se recogen en un JSON con fecha y hora (*timestamp*) de comienzo de sesión en el sistema. Un ejemplo de registro recogido para un usuario aparece en el Anexo 9.5.

Aplicación para POSTDATA

Para POSTDATA se ha presentado una aplicación análoga a la de AULA salvo en lo siguiente:

- No aparece la opción de indicar el estado afectivo en que se encuentran los usuarios (ya que serán las interacciones de unos anotadores expertos que usarán repetidas veces el sistema).
- La comparación se hace entre F joint y BERT frente a F joint y con word2vec como base de *embeddings*, usando ambas el coseno como métrica de comparación.

Excel para anotaciones de POSTDATA

Adicionalmente se genera un Excel para POSDATA de manera que los anotadores puedan etiquetar para cada soneto la información siguiente en una escala de 1 a 4 (siendo 1 el mínimo valor y 4 el máximo valor):

1. *Happiness*
2. *Anger*
3. *Sadness*
4. *Fear*
5. *Disgust*
6. *Valencia*
7. *Arousal*
8. *concreteness* (Concreteness se define como el grado de especificidad de la palabra, valiendo 1 cuando la palabra es muy abstracta y 4 cuando es muy concreta. Palabras como ‘objeto’ son más abstractas que otras como ‘mesa’).
9. *imageability* (se define como la facilidad o dificultad de construir una imagen mental asociada a dicha palabra, valiendo 1 cuando la palabra es muy difícil de imaginar y 4 cuando es muy fácil. Es más fácil imaginarse algo con palabras como ‘bandera’ que con otras como ‘caridad’).
10. *context availability* (se define como la facilidad o dificultad en asociar dicha palabra con un contexto en el que pudiese aparecer, valiendo 1 cuando la palabra es muy difícil de asociar a un contexto y 4 cuando es muy fácil. Es más fácil construir frases o buscar ejemplos de uso para palabras como ‘mesa’ que para otras como ‘ciudadanía’).

Junto con ello se anotan 21 rasgos psicológicos que se han definido como relevantes para investigaciones posteriores (García Franco & Manjarrés Riesco, 2016) que conciernen a la integración del enfoque de la pedagogía poética en RecSys educativos afectivos para la enseñanza y educación sobre conceptos psicológicos relacionados con trastornos de personalidad. Estas investigaciones están motivadas por diversos estudios del ámbito psicológico han establecido conexiones entre la poesía y el tratamiento de determinados trastornos (Parastoo et al., 2016; Wassiliwizky et al., 2017).

Dichas anotaciones son binarias, indicándose con un 1 en caso de que el soneto haga referencia a ese concepto y 0 en caso contrario.

- Soledad
- Ilusión
- Ensoñación
- Grandiosidad
- Orgullo
- Irritabilidad
- Ansiedad
- Ira
- Inestabilidad
- Idealización
- Depresión
- Desilusión
- Aversión
- Inseguridad
- Impotencia
- Vulnerabilidad
- Temor
- Obsesión
- Compulsión
- Prejuicio
- Dramatización

Estas anotaciones permitirán calcular distintas métricas (como precisión, por ejemplo) en la recuperación de poemas que servirán de criterio de validación de las hipótesis en la presente investigación.

API

Por último se diseña una API en Python usando la librería Flask⁸² (los endpoints se definen en el apartado 9.2). De cara a probar su uso se ha añadido una visualización web mediante el uso de html y css.

4.3. Diagramas de procesos

En este apartado se incluyen unas gráficas que representan la lógica del programa implementado.

En la Ilustración 70 se muestra el esquema de la lógica para el procesado de los textos. Desde todos los repositorios de información se obtienen tanto las ontologías como un json con la información de los distintos sonetos. Además, se crean varios cvs que contienen los *embeddings* finales para cada soneto tras aplicar las funciones de composición (un csv por *word*

⁸² “Flask”. Web. 25/05/2019. <http://flask.pocoo.org/>

embedding y por función de composición usada). Cada fila del cvs tiene una clave única que identifica a cada soneto y que coincide con la clave única asociada en el json para dicho soneto.

Por otro lado, en la Ilustración 72 se muestra la aplicación de escritorio utilizada para las pruebas de AULA. Con ella es con la que se recogen los estados afectivos de los usuarios y, además, se muestran los resultados obtenidos con BERT para dos funciones de composición con el fin de que los participantes elijan uno y puntúen su experiencia.

De forma análoga, la Ilustración 73 muestra la aplicación de escritorio utilizada en las pruebas con POSTDATA. Ante un mensaje de entrada se usan distintas técnicas de *word embedding*, para luego componer con una misma función de composición y presentar los dos sonetos obtenidos para que el usuario elija uno y puntúe.

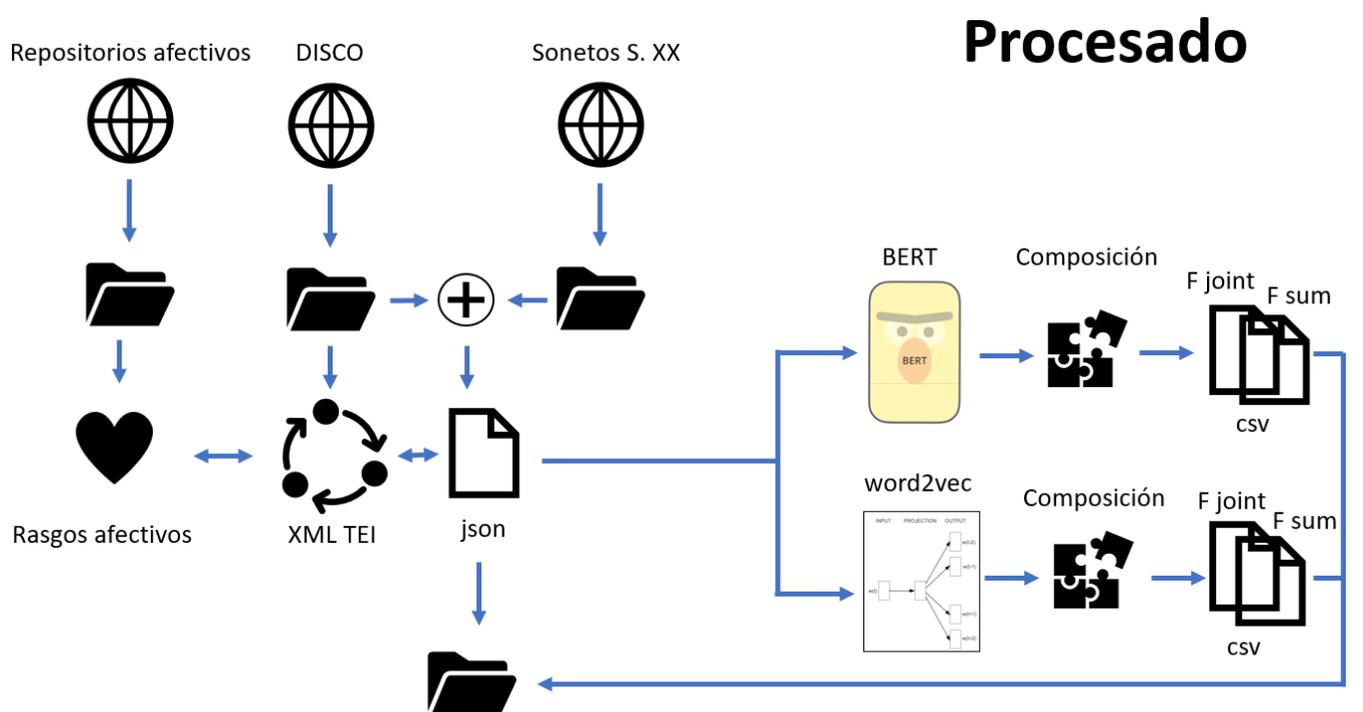


Ilustración 70 Esquema de la lógica para el procesamiento de los textos.

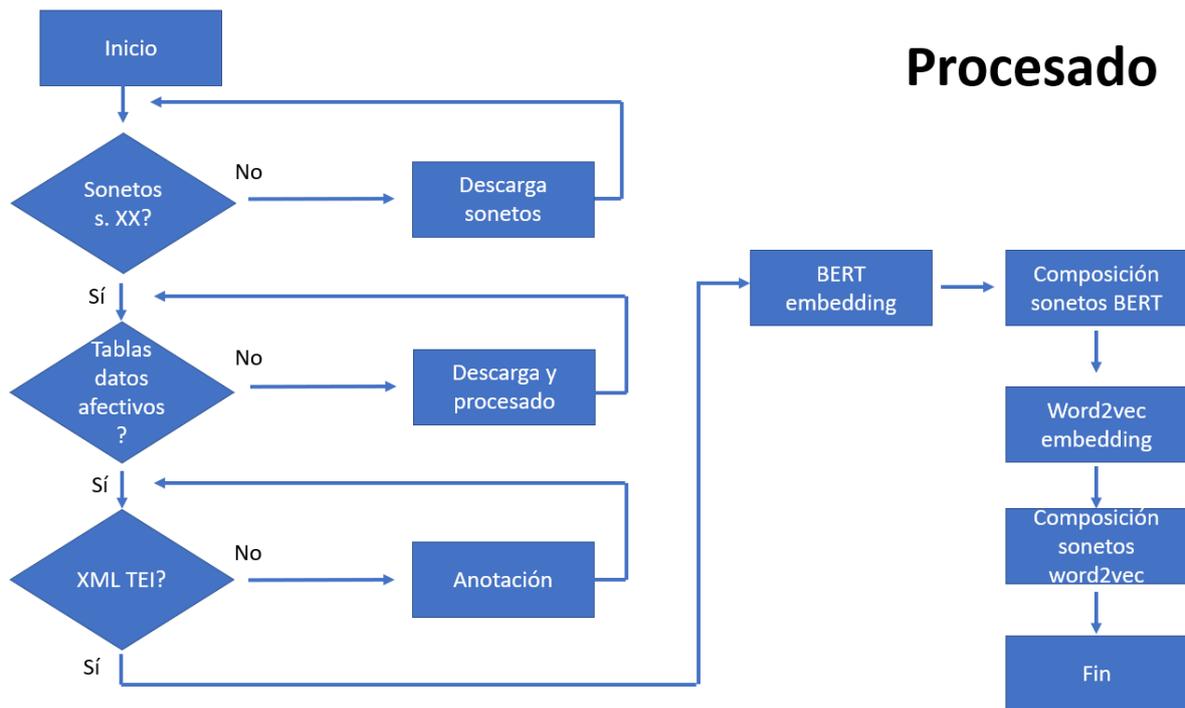


Ilustración 71 Flujo del proceso esquematizado en la Ilustración 68

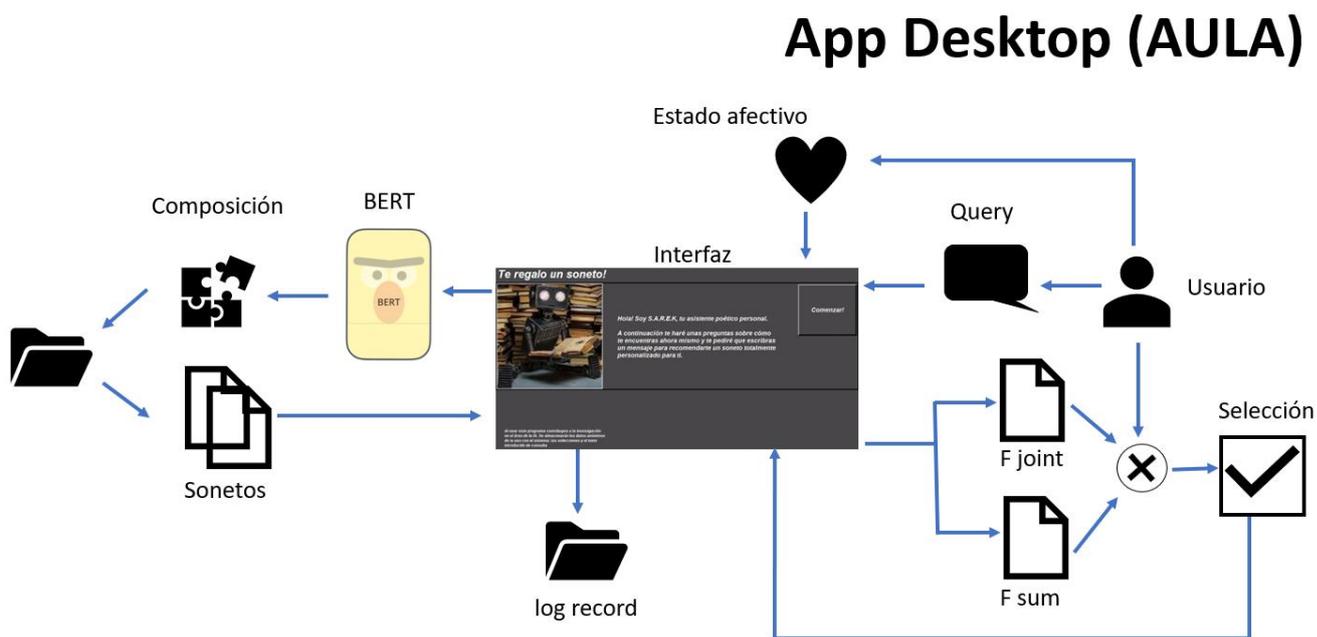


Ilustración 72 Aplicación de escritorio utilizada para las pruebas de AULA⁸³

⁸³ Imagen del robot dentro de la interfaz tomada de: Robot-child reading a book in the workshop of its creator. Vasilyev Alexandr. Web. 15/09/2019. <https://www.shutterstock.com/image-photo/robotchild-reading-book-workshop-creator-333995531?src=xJPWZuG5eIRT0y5xKB1Ctg-1-26>

App Desktop (POSTDATA)

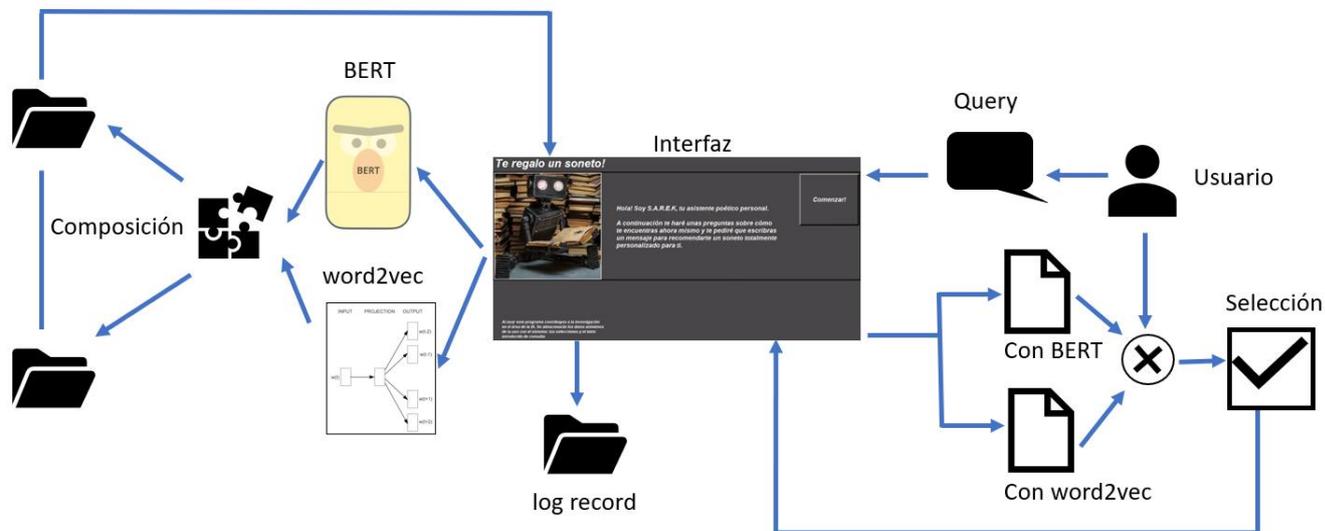


Ilustración 73 Aplicación de escritorio para las pruebas con POSTDATA.

API + web

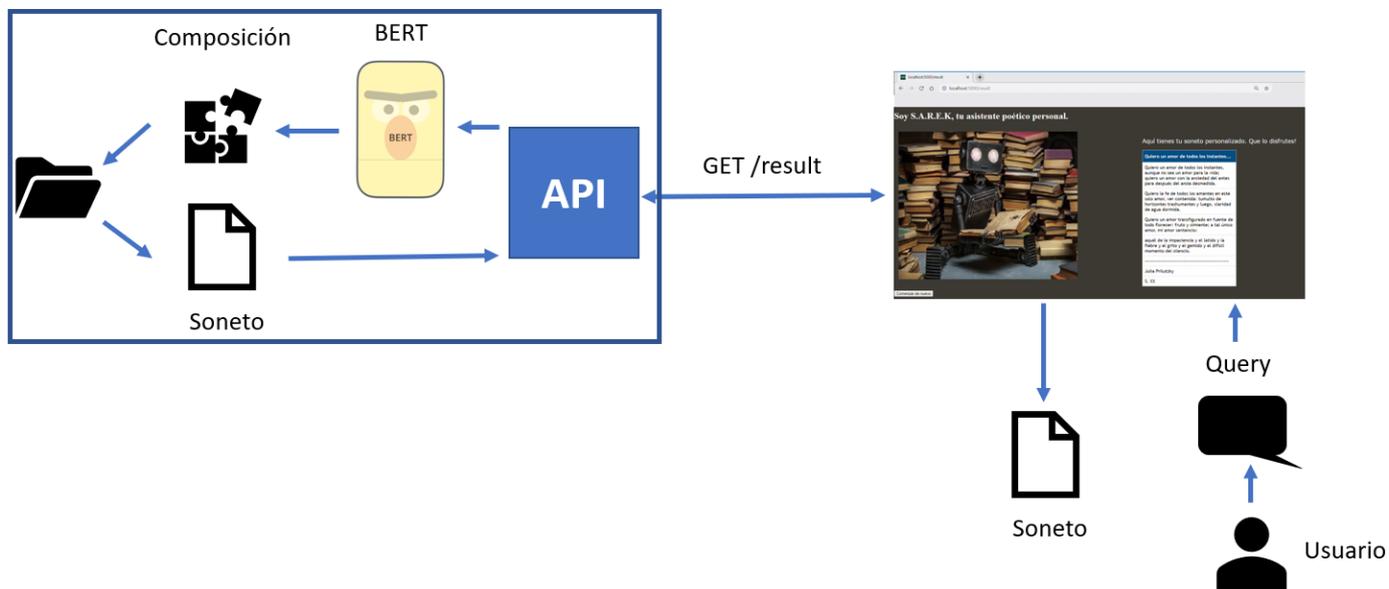


Ilustración 74 Diseño de la API que de entrada tiene un endpoint para que el usuario envíe un mensaje y reciba el soneto más similar semánticamente usando BERT y F Joint RL SEQ con métrica de comparación de coseno

Esquema general Apps

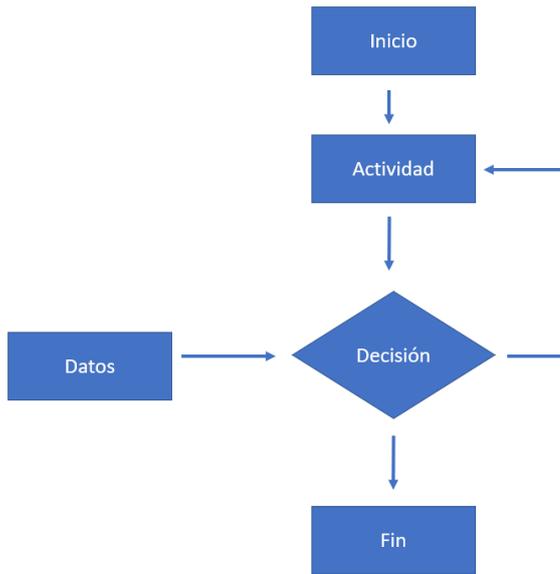


Ilustración 75 Flujograma genérico para cualquiera de las aplicaciones desarrolladas

Excel (anotaciones)

| 1 | index | text | happiness | sadness | | | |
|---|-------|---|-------------------|-----------------|--------------------|-----------------|-------------------|
| | 445 | <p>Que sentimiento cielo tan extraño, descubro en este Bulto peregrino, de aquel que el río Tabor en luz divino, en Navarra se de sangriento un baño.</p> <p>Si cada fuente vierte por su caño, licor precioso que del cielo vino, segunda vez redime su amor fino, al mundo todo, que rindió el engaño.</p> <p>Mas ay, que afán le hiere, aunque distante, que del Indio Javier por Dios padece, correspondido con dolor profundo.</p> <p>Y siendo de Javier el más amante, si por el Orbe su coral ofrece, le vierte por Javier, que pesa un mundo.</p> | | | | | |
| 2 | 97 | <p>Muere Jesús del Gólgota en la cumbre, con amor perdonando al que le hería, siente deshecho el corazón María del dolor en la inmensa pesadumbre.</p> <p>Se aleja con pavor la muchedumbre cumplida ya la Santa Profecía, tiembla la tierra, el iluminar del día cegando a tal horror, pierde su lumbré.</p> <p>Se abren las tumbas, se desgarran el velo, y a impulsos de un amor grande y fecundo parece estar la cruz, signo de duelo,</p> <p>cerrando augusta con el pie el profundo, con la excelsa cabeza abriendo el cielo y con los brazos abarcando el mundo.</p> | | | | | |
| | | | Happiness_Sadness | Valence_Arousal | Anger_Fear_Disgust | Age_Familiarity | Conc_Imag_Context |

Ilustración 76 Excel que usaron los anotadores de POSDATA para llevar a cabo las anotaciones para cada soneto de los rasgos de los que se dispone de información a nivel de palabras.

5. Planificación

En la Ilustración 77 se adjunta un cronograma de las distintas tareas, hitos y entregas realizadas en la elaboración de este TFM. Se muestran las tareas realizadas con fecha de inicio y fin, así como su estado de completitud. Se puede observar que están todas realizadas al presente de la redacción de esta memoria.

Para más detalle sobre las tareas realizadas se remite al archivo adjunto **gantt.png** en el que se puede visualizar el diagrama de Gantt con todas las tareas y sus correspondientes hitos.

Por último, hay que mencionar que el tiempo total dedicado al proyecto desde la fecha en que se empezó a trabajar (01/10/2018) hasta la entrega de la memoria (15/09/2019), con una media de dedicación diaria de 2.5 horas, ha supuesto un tiempo de trabajo de **875 horas**, cumpliéndose las 750 horas estimadas que requería la carga de créditos ECTS para el TFM. La entrega se decidió hacer en Septiembre para poder concluir algunas investigaciones y redactar el artículo científico adjunto en esta memoria.

TFM UNED

| ▼ Estudio del SOTA | 100% | | Start | Due |
|---|-------------------------------------|---------------------------------------|--------------|--------------|
| Pedagogía Poética | 100% | <div style="width: 100%;"></div> | Oct 1, 2018 | Oct 14, 2018 |
| RecSys afectivos en e-learning | 100% | <div style="width: 100%;"></div> | Oct 14, 2018 | Oct 21, 2018 |
| Ontologías e-learning | 100% | <div style="width: 100%;"></div> | Oct 21, 2018 | Nov 4, 2018 |
| Corpus de poesía | 100% | <div style="width: 100%;"></div> | Nov 4, 2018 | Nov 11, 2018 |
| Word embeddings | 100% | <div style="width: 100%;"></div> | Nov 11, 2018 | Dec 2, 2018 |
| Task Milestone Group of Tasks | | | | |
| ▼ Desarrollo de prototipos | 100% | | Start | Due |
| Script para procesado de sonetos | 100% | <div style="width: 100%;"></div> | Dec 9, 2018 | Feb 16, 2019 |
| Script extracción rasgos afectivos | 100% | <div style="width: 100%;"></div> | Jan 13, 2019 | Mar 17, 2019 |
| Obtención de nuevos sonetos de internet | 100% | <div style="width: 100%;"></div> | Jan 6, 2019 | Jan 13, 2019 |
| Desarrollo aplicación AULA | 100% | <div style="width: 100%;"></div> | Mar 10, 2019 | Mar 30, 2019 |
| Desarrollo aplicación POSTDATA | 100% | <div style="width: 100%;"></div> | Mar 24, 2019 | Apr 7, 2019 |
| Desarrollo aplicación API REST | 100% | <div style="width: 100%;"></div> | Apr 21, 2019 | Apr 27, 2019 |
| Diseño documento para anotaciones de POSTDATA | 100% | <div style="width: 100%;"></div> | May 19, 2019 | May 31, 2019 |
| Anotación de rasgos afectivos en corpus DISCO | 100% | <div style="width: 100%;"></div> | Jan 20, 2019 | Feb 17, 2019 |
| Task Milestone Group of Tasks | | | | |
| ▼ Realización de pruebas | 50% | | Start | Due |
| Pruebas en AULA | 100% | <div style="width: 100%;"></div> | Mar 30, 2019 | Mar 30, 2019 |
| Anotación afectiva por grupo expertos | 100% | <div style="width: 100%;"></div> | Jun 6, 2019 | Jun 23, 2019 |
| Pruebas POSTDATA | 100% | <div style="width: 100%;"></div> | Mar 6, 2019 | Jun 23, 2019 |
| Task Milestone Group of Tasks | | | | |
| ▼ Redacción de la memoria | 100% | | Start | Due |
| Análisis resultados obtenidos | 100% | <div style="width: 100%;"></div> | Mar 31, 2019 | Apr 6, 2019 |
| Redacción del documento final | 100% | <div style="width: 100%;"></div> | May 1, 2019 | Jun 15, 2019 |
| Task Milestone Group of Tasks | | | | |
| ▼ Hitos | 88% | | Start | Due |
| AULA 2019 | <input checked="" type="checkbox"/> | ◆ | Mar 30, 2019 | Mar 30, 2019 |
| Entrega del TFM | <input type="checkbox"/> | ◆ | Sep 15, 2019 | Sep 15, 2019 |
| Revisión con tutores 1 | <input checked="" type="checkbox"/> | ◆ | Oct 1, 2018 | Oct 1, 2018 |
| Revisión con tutores 2 | <input checked="" type="checkbox"/> | ◆ | Nov 30, 2018 | Nov 30, 2018 |
| Revisión con tutores 3 | <input checked="" type="checkbox"/> | ◆ | Mar 11, 2019 | Mar 11, 2019 |
| Revisión con tutores 4 | <input checked="" type="checkbox"/> | ◆ | Jun 16, 2019 | Jun 16, 2019 |
| Revisión con tutores 5 | <input checked="" type="checkbox"/> | ◆ | Apr 29, 2019 | Apr 29, 2019 |
| Revisión con tutores 6 | <input checked="" type="checkbox"/> | ◆ | Friday | Friday |
| Task Milestone Group of Tasks | | | | |

Ilustración 77 Diagrama de Gantt para el TFM

6. Resultados

A continuación se detallan los resultados de los diversos experimentos. El primer punto de este subapartado muestra un análisis exploratorio de los datos (EDA) presentes en el corpus de poemas de POSTDATA (ya ampliado con los poemas del S. XX). Seguidamente se analiza los experimentos realizados tanto dentro del marco de POSTDATA como los planteados para la feria de AULA.

6.1. Estudio del corpus de sonetos

El primer estudio que se realiza, antes de llevar a cabo cualquier experimento, es el del corpus de sonetos disponibles.

Como se ha mencionado previamente, el corpus consiste en 4085 sonetos en lengua castellana del S. XV al S. XIX, recogidos del corpus DISCO de POSTDATA (UNED) y completados mediante sonetos concretamente del S. XX (166 sonetos) obtenidos de internet, los cuales han sido anotados con rasgos afectivos específicos, inspirados en la literatura (Ullrich et al., 2016). Dichos rasgos se detallan en el Anexo 9.7. El total de sonetos es por lo tanto 4251.

El artículo mencionado indica cómo trabajan con el conjunto de datos BAWL que contiene 6000 palabras en alemán, y cómo, para asociar el valor de los distintos rasgos a las palabras individuales, se consultan las distintas palabras del conjunto de poemas en dicho conjunto de datos. Para aumentar el número de palabras que coincidan con las entradas de dichas tablas, las palabras de los poemas se lematizan al tiempo que también se eliminan *stopwords*. De esta manera se consigue encontrar el valor afectivo para el 90% de las palabras que aparecen en los poemas, siendo el 10% restante un conjunto de palabras que no aparecen en dichas tablas por ser, principalmente, nombres propios. De esta manera, el primer punto a considerar será analizar cuántas palabras del conjunto de sonetos en castellano disponibles aparecen en las tablas utilizadas.

En el caso del corpus disponible de sonetos, se plantea una comparativa en la que se analice que porcentaje de palabras del corpus están presentes en cada uno de los conjuntos de datos propuestos para obtener los rasgos afectivos.

En la Tabla 5 aparecen las palabras del corpus presentes en los distintos conjuntos de datos. Se plantean diversos escenarios, en los que se trabaja con las palabras lematizadas, con o sin verbos y haciendo *stemming*, además de mostrar una comparativa usando únicamente los sonetos del siglo XX.

| | Emociones (<i>Happiness, Fear...</i>) (Ferré et al., 2017) | Valencia y Excitación (Stadthagen-Gonzalez et al., 2017) | Valencia, Excitación y conceptos (FAM, ...) (Guasch et al., 2016) | Valencia y Excitación (ambos datasets) | <i>Age of Acquisition</i> (Alonso et al., 2015) |
|---|--|--|---|---|--|
| Número de palabras | 2267 | 14031 | 1400 | 14133 | 7040 |
| Corpus lematizado (n°) | 1414 | 6416 | 905 | 6449 | 3568 |
| Corpus lematizado (%) | 0.075 | 0.343 | 0.048 | 0.344 | 0.191 |
| Corpus lematizado sin verbos (n°) | 1245 | 4700 | 811 | 4727 | 2937 |
| Corpus lematizado sin verbos (%) | 0.091 | 0.343 | 0.059 | 0.345 | 0.091 |
| Corpus <i>stemmed</i> sin verbos (n°) | 1424 | 5017 | 939 | 5036 | 3187 |
| Corpus <i>stemmed</i> sin verbos (%) | 0.136 | 0.480 | 0.090 | 0.482 | 0.305 |
| Corpus <i>stemmed</i> sin verbos S. XX (n°) | 597 | 1762 | 398 | 1766 | 1307 |
| Corpus <i>stemmed</i> sin verbos S. XX (%) | 0.265 | 0.781 | 0.177 | 0.783 | 0.580 |

Tabla 5 Palabras presentes del corpus en los distintos conjuntos de datos.

El análisis planteado muestra como los ratios de presencia de las palabras del corpus en los conjuntos de datos son muy bajos, más bajos que el 90% de coincidencia que menciona el artículo (Ullrich et al., 2016) con el que se plantea la comparativa de los valores afectivos para el castellano.

Como se puede apreciar, eso sí, dos de los motivos principales de esta problemática son los verbos (que no suelen estar recogidos en los conjuntos de datos) y, principalmente, el léxico presente, perteneciente al castellano antiguo. El análisis muestra como para todos los conjuntos de datos los ratios de presencia se duplican prácticamente si se utilizan para la comparativa exclusivamente los sonetos del S. XX. Así pues, gran parte del contenido afectivo, especialmente de los sonetos más antiguos, no se va a poder agregar con demasiada precisión. En cualquier punto, con los datos disponibles se procedió igualmente al análisis como se había planteado, en donde para todas las palabras y análisis de los parámetros afectivos se trabajará con las palabras lematizadas y sin *stopwords*.

Tras el punto anterior, se plantea el análisis de la matriz de correlaciones de todos los rasgos afectivos por soneto para ver si hay relación o no entre ellos. Debido al gran tamaño de dicha matriz, se remite al Anexo 9.8 para ver su detalle; a continuación se recogen las conclusiones principales de ésta.

En el caso del análisis sobre todos los sonetos se puede ver lo siguiente:

- Uno de los mayores índices de correlación de los rasgos es el de **ValenceSpan** con el de **Min_Valence** (-0.9) remarcándose como los sonetos cuanto menor suele ser el valor de valencia que tienen menor suele ser el *intervalo* (*span*). Es decir, que en los sonetos el *intervalo* suele ser menor cuanto más alto sea el valor mínimo de valencia que aparezca en el texto. Así, un soneto puede contener una palabra de valencia muy negativa y esto suele causar un mayor *intervalo* ya que el resto de palabras del texto suelen mantenerse en unos niveles de valencia más o menos constantes (de hecho, la correlación con **Max_Valence** es de 0.5, lo que remarca que cuando hay un *span* muy significativo de valencia principalmente es por la contribución de **Min_Valence** y no tanto por la del **Max_Valence** que haya).
- Ídem para el **ArousalSpan** con respecto al **Min_Arousal** (-0.8). Aquí la contribución del nivel de excitación de las palabras de alto valor de excitación, **Max_Arousal** es un poco más cercana, eso sí (0.6).
- Se observa una alta correlación positiva entre el nivel de excitación, **ArousalMean** y la valencia, **ValenceMean** (del orden de 0.7), remarcándose como el nivel de excitación tiende a aumentar a medida que los sentimientos son más positivos. Como dato complementario hay que indicar que el nivel de *ruido* también está relacionado; es decir, que en los casos en los que hay mayor desviación en el valor del nivel de excitación (porque había más discrepancia entre los usuarios que valoraban el nivel de excitación de ese conjunto de palabras) suele darse también acompañado de un mayor nivel de ruido en la valencia.
- Se observa un alto nivel de correlación entre el nivel de felicidad, **Hap_Mean** y el de **ValenceMean** (0.6) indicándose como los sonetos con alto contenido de felicidad suelen estar asociados con un mayor nivel de sentimiento positivo (valencia). No hay demasiada correlación entre el nivel de valencia y el resto de sentimientos.
- Se aprecia una mayor correlación positiva entre el nivel de excitación y el de todos los demás sentimientos (Hap, Ang, Dis, Fear, Sad), mayor del 0.5 o 0.6, lo que remarca como dicho nivel aumenta a medida que haya un mayor nivel de excitación provocado por las palabras de los textos, y viceversa (si hay una gran presencia de dichos sentimientos el nivel de excitación aumenta); evidentemente, cuanto mayor es el valor del mínimo de valencia que hay en un texto, mayor valor de valencia tiene dicho texto y por lo tanto menor presencia de sentimientos negativos.
- **Max_Val** afecta sobre todo de forma positiva a **Hap_Mean** (y de forma menos positiva para el resto de sentimientos), mientras que **Min_Val** no afecta casi a **Hap_Mean** pero si tiene una influencia negativa en el resto de sentimientos.
- El nivel de complejidad de los textos se recoge con la información del **AoA**. Se puede ver como de forma general, el nivel de cualquiera de las emociones que transmiten los textos aumentan con el nivel de complejidad de estos (es decir, con el uso de vocabulario más avanzado). Es interesante remarcar que esto ocurre cuando el conjunto de palabras es de nivel avanzado (**MaxAoA**, que representa la suma de los máximos por palabra) pero no cuando es una única palabra aislada la que contribuye (**Max_AoA**, que es el máximo de AoA de todas las palabras del texto). De igual forma se aprecia como dicho AoA tiene una correlación positiva con los valores de media del nivel de excitación y valencia (**ValenceMean** y **ArousalMean**), en cambio

no se aprecia dicha relación con otros parámetros como el máximo o mínimo de dichos rasgos, su posición en el texto...

- Se ve además la gran correlación positiva entre las variables de los mismos grupos (valencia con el nivel de excitación, los distintos sentimientos entre ellos...)

Además de todo el análisis anterior, se muestra una visualización de los distintos rasgos en aras a encontrar valores atípicos.

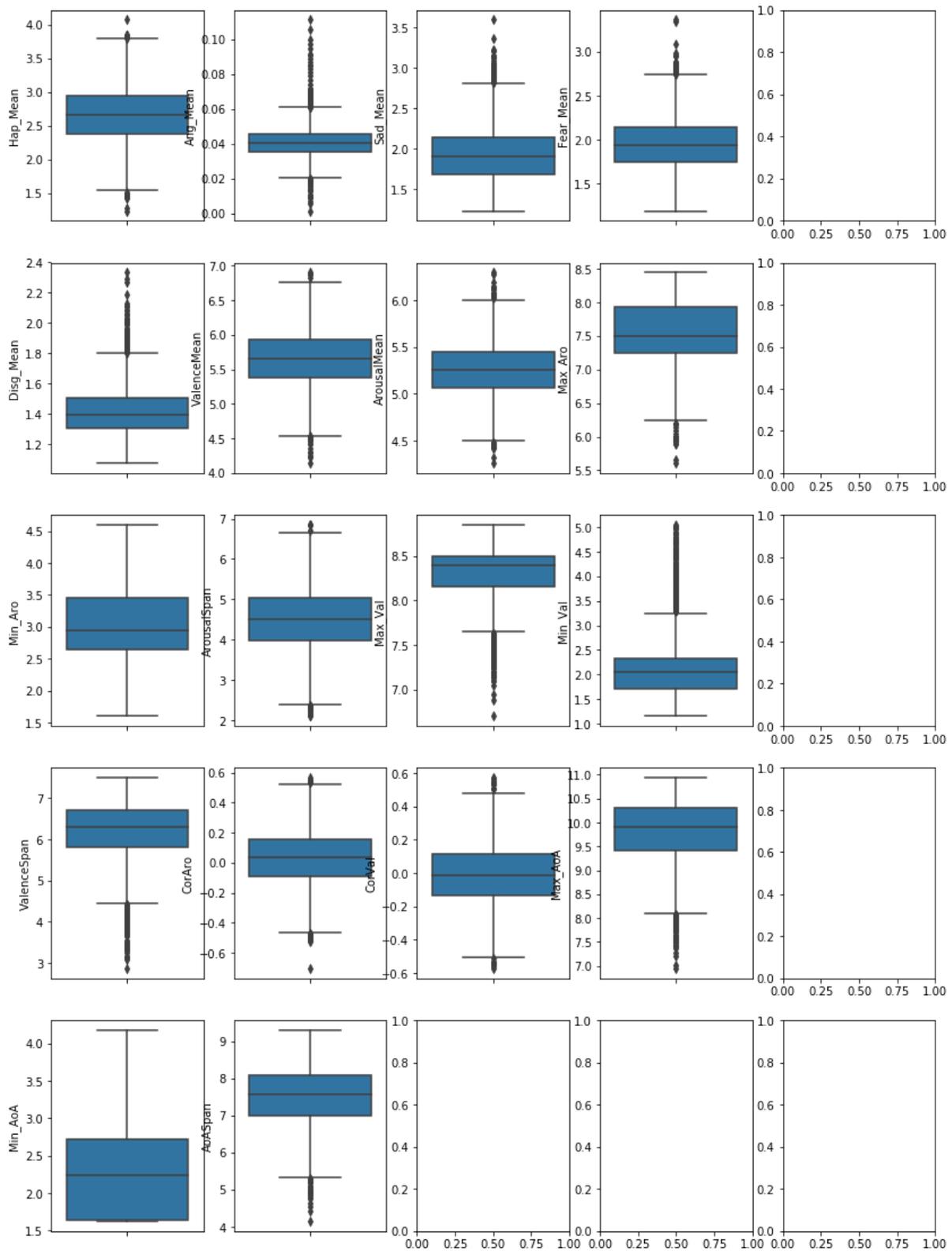


Ilustración 78 Box-plot de los distintos rasgos

En los diagramas box-plot se puede ver que para algunos parámetros existe mucha cantidad de atípicos, algo que contribuirá a las medidas obtenidas. Especialmente se ve en **ValenceSpan**, **Max_AoA** (máximo de los máximos de AoA), **Min_Val**, **AoASpan**...

Por último, se muestra el top 10 de las palabras más habituales sin lematizar y tras lematizar.

| count | | count | |
|-------|------|-------|------|
| words | | words | |
| si | 1833 | si | 1833 |
| amor | 1499 | amor | 1667 |
| vida | 1096 | ver | 1180 |
| alma | 903 | vida | 1127 |
| tan | 879 | ser | 1098 |
| cielo | 817 | alma | 1000 |
| pues | 815 | cielo | 912 |
| sol | 811 | tan | 879 |
| luz | 801 | luz | 871 |
| mas | 780 | dar | 833 |

Ilustración 79 A la derecha se muestra el top de palabras sin lematizar. A la izquierda el de las lematizadas

6.2 Experimentos con POSTDATA

Como ya se mencionó en el capítulo anterior, POSTDATA colabora con el proyecto principalmente en dos tareas específicas, por un lado en la anotación de los sonetos con categorías afectivas con conceptos que aparecen representados en los poemas (especialmente relacionados con el ámbito psicológico) para evaluar con ello distintas métricas de recuperación de sonetos. Con estos puntos, además de permitir evaluaciones posteriores, se pretende comprobar:

- La primera hipótesis de investigación planteada, en la que se mencionaba si los rasgos afectivos, que proceden de los que se han visto relevantes en diversos estudios sobre poesía en otros idiomas, son igual de relevantes para el castellano, y que la estimación de dichos rasgos, en base a la agregación del contenido afectivo individual de cada palabra del poema, responde al contenido afectivo global. (**Experimento 1**).
- La hipótesis de investigación 5, en la que se mencionaba que existe una correlación entre determinados trastornos psicológicos y el valor afectivo de la poesía. (**Experimento 3**).
- La hipótesis de investigación 6, en la que se indicaba que es posible etiquetar poemas utilizando metadatos coherentes con las ontologías de uso común en los RecSys educativos afectivos, de modo que sean recuperables como objetos de aprendizaje personalizados, haciendo uso de técnicas de NLP. (**Experimento 4**).

El otro aspecto con el que está colaborando es en el uso de una herramienta para analizar con los resultados de uso, y con los que se pretende corroborar la segunda hipótesis de investigación, en la que se plantaba el interés por ver si BERT aportaría más que word2vec para un sistema de IR para poesía, dado que el SOTA de *embeddings* habla de los grandes avances que supone BERT (desambiguaciones, capturar relaciones a largo plazo en los textos, multi idioma...). Como ya se mencionó en el apartado de objetivos de investigación, este experimento se ha planteado pero no se ha llegado a realizar dentro de la planificación temporal estipulada para la presente memoria. Sin embargo, como también se mencionó en el apartado de Objetivos, existe una alternativa para analizarlo, comparando las métricas de recuperación de sonetos para los términos psicológicos usando un *embedding* u otro. De esta manera el experimento 2 se realiza de forma conjunta con el experimento 4 (**experimento 2**).

Así, este apartado habla de los resultados de los tres experimentos reportados en la presente memoria (experimentos 1, 3 y 4), tras realizar previamente un EDA de los datos anotados por los expertos de POSTDATA.

Análisis exploratorio de datos sobre las anotaciones

A modo de recordatorio, los 3 expertos de POSTDATA han anotado sobre un subconjunto de sonetos con valores para distintos rasgos. La Tabla 6 contiene los rasgos psicológicos anotados, cuyo valor es un valor binario (1 si el soneto guarda relación con dicho concepto, 0 en caso contrario). En la Tabla 7 aparecen los rasgos afectivos, cuyos valores van de 1 (mínimo) a 4 (máximo) según sea la intensidad de dicho rasgo en el poema.

| | | | | | |
|------------|---------------|---------------|---------------|-----------|----------------|
| Ansiedad | Desilusión | Idealización | Inseguridad | Orgullo | Vulnerabilidad |
| Aversión | Dramatización | Ilusión | Ira | Prejuicio | |
| Compulsión | Ensoñación | Impotencia | Irritabilidad | Soledad | |
| Depresión | Grandiosidad | Inestabilidad | Obsesión | Temor | |

Tabla 6 Rasgos psicológicos

| | | |
|--------------|----------------------|--------------|
| Anger | Context availability | Imageability |
| Annotator | Disgust | Sadness |
| Arousal | Fear | Valence |
| Concreteness | Happiness | |

Tabla 7 Rasgos afectivos

El estudio parte de un subconjunto de **230 sonetos anotados** completamente para esos rasgos especificados. Los sonetos anotados pertenecen tanto al intervalo del S. XIX (167 sonetos), como al intervalo del S. XV al XVII (72), y al del S. XVIII (9). Este es un dato relevante a tener en cuenta en todos los estudios posteriores, ya que no sólo se está trabajando con poesía en castellano sino que, además, se trabaja con castellano antiguo (gran parte de los sonetos son además del periodo más antiguo), algo que puede afectar notablemente a los *embeddings* de las palabras de los sonetos y la composición de los mismos para un fragmento de texto.

Como 230 es mucho mayor que 30, hay suficientes sonetos para plantear un análisis estadístico de valor significativo. Con ello se procede a calcular la Alpha de Krippendorff (k-alpha) para las anotaciones que han hecho los 3 expertos por cada una de las variables. Es decir, se calcula k-alpha para 3 listas (una por anotador) para cada variable.

K-alpha es una métrica que generaliza otras métricas que se encargan de cuantificar la fiabilidad entre anotadores (*inter-rater reliability*), pudiendo trabajar para ello con anotaciones tanto ordinales como nominales, así como con cualquier número de anotadores.

K-alpha tendrá un valor menor que 1, donde 1 representa el total acuerdo entre anotadores. Sin embargo, existen diferentes criterios respecto a cuándo considerar que hay acuerdo entre anotadores. En algunas ocasiones se usan criterios estrictos, en los que sólo se aceptan como verdaderamente válidas las anotaciones de expertos si existe un k-alpha de al menos 0.8. Otros criterios más laxos establecen el mínimo en 0.21, definiendo los siguientes umbrales⁸⁴:

- $K < 0$: Muy bajo
- $0 < K < 2$: Ligero
- $0.21 < K < 0.4$: Aceptable
- $0.41 < K < 0.60$: Moderado
- $0.61 < K < 0.80$: Substancial
- $0.81 < K < 1$: Perfecto

Con ello, los resultados obtenidos para los distintos anotadores se muestran en la Tabla 8.

| Variable | K-alpha | Variable | K-alpha | Variable | K-alpha |
|----------------------|---------|-----------------------------|---------|---------------------|---------|
| Ansiedad | 0.2475 | Orgullo | 0.2722 | Soledad | 0.1943 |
| Aversión | 0.2948 | Prejuicio | 0.2745 | anger | 0.2010 |
| Depresión | 0.2260 | Temor | 0.2738 | arousal | 0.0912 |
| Desilusión | 0.2291 | Vulnerabilidad | 0.2826 | disgust | 0.1836 |
| Dramatización | 0.2300 | concreteness | 0.2774 | fear | 0.1642 |
| Ilusión | 0.3063 | context availability | 0.2269 | happiness | 0.0279 |
| Impotencia | 0.2671 | Compulsión | 0.1932 | imageability | 0.2087 |
| Inestabilidad | 0.2540 | Ensoñación | 0.1852 | sadness | 0.1916 |
| Inseguridad | 0.3123 | Grandiosidad | 0.1690 | valence | 0.0051 |
| Ira | 0.2501 | Idealización | 0.1791 | | |
| Obsesión | 0.2603 | Irritabilidad | 0.2058 | | |

Tabla 8 Valores de k-alpha para todos los anotadores por cada variable

De la Tabla 8 se puede concluir lo siguiente:

- Hay coincidencia aceptable ($K \geq 0.21$) para las siguientes variables teniendo en cuenta las anotaciones de los 3 expertos; 'Ansiedad', 'Aversión', 'Depresión', 'Desilusión', 'Dramatización', 'Ilusión', 'Impotencia', 'Inestabilidad', 'Inseguridad', 'Ira', 'Obsesión', 'Orgullo', 'Prejuicio', 'Temor', 'Vulnerabilidad', 'concreteness', 'context availability'. Esto corresponde a un 54.84 % del total de variables existentes. Hay que mencionar

⁸⁴ "Interpreting reliability results". Jean Carletta. Web. 15/08/2019
<http://homepages.inf.ed.ac.uk/jeanc/maptask-coding-html/node23.html>

que en todos los términos que pasan esta comprobación se incluye el conjunto de términos psicológicos.

- Hay una ligera coincidencia ($0 > K < 0.21$) para las demás variables teniendo en cuenta las anotaciones de los 3 expertos. Para algunos casos la K es alta (cercana al 0.21 aunque algo inferior), algo que ocurre especialmente en los términos psicológicos.
- Las mayores discrepancias se ven en las variables: arousal 0.0912, happiness 0.0279, valence 0.0051

Además del análisis conjunto de los anotadores, en las Tablas 9, 10 y 11 se incluyen las comparativas entre cada par de anotadores para comprobar entre cuáles se da la mayor discrepancia.

| Variable | K-alpha | Variable | K-alpha | Variable | K-alpha |
|----------------------|---------|-----------------------------|---------|---------------------|---------|
| Ansiedad | 0.2363 | Orgullo | 0.1095 | Soledad | 0.1156 |
| Aversión | 0.2178 | Prejuicio | 0.1111 | anger | 0.0956 |
| Depresión | 0.1411 | Temor | 0.2402 | arousal | 0.1082 |
| Desilusión | 0.1065 | Vulnerabilidad | 0.245 | disgust | 0.0832 |
| Dramatización | 0.2749 | concreteness | 0.167 | fear | 0.0993 |
| Ilusión | 0.1986 | context availability | 0.1779 | happiness | 0.0339 |
| Impotencia | 0.156 | Compulsión | 0.0568 | imageability | 0.1918 |
| Inestabilidad | 0.2195 | Ensoñación | 0.0866 | sadness | 0.0757 |
| Inseguridad | 0.2734 | Grandiosidad | 0.009 | valence | 0.1075 |
| Ira | 0.1886 | Idealización | 0.009 | Soledad | 0.1156 |
| Obsesión | 0.3102 | Irritabilidad | 0.1323 | | |

Tabla 9 Comparativa entre los anotadores 1 y 2

| Variable | K-alpha | Variable | K-alpha | Variable | K-alpha |
|----------------------|---------|-----------------------------|---------|---------------------|---------|
| Ansiedad | 0.0842 | Orgullo | 0.1156 | Soledad | 0.1943 |
| Aversión | 0.1806 | Prejuicio | 0.158 | anger | 0.0447 |
| Depresión | 0.0115 | Temor | 0.1117 | arousal | -0.026 |
| Desilusión | 0.0867 | Vulnerabilidad | 0.1513 | disgust | 0.0528 |
| Dramatización | 0.1222 | concreteness | 0.0508 | fear | 0.0517 |
| Ilusión | 0.2382 | context availability | 0.029 | happiness | -0.0575 |
| Impotencia | 0.1583 | Compulsión | 0.2079 | imageability | 0.034 |
| Inestabilidad | 0.1639 | Ensoñación | 0.0759 | sadness | 0.0393 |
| Inseguridad | 0.188 | Grandiosidad | -0.0467 | valence | -0.0447 |
| Ira | 0.0506 | Idealización | 0.0722 | Soledad | 0.0295 |
| Obsesión | 0.1669 | Irritabilidad | 0.0913 | | |

Tabla 10 Comparativa entre los anotadores 1 y 3

| Variable | K-alpha | Variable | K-alpha | Variable | K-alpha |
|----------------------|---------|-----------------------------|---------|---------------------|---------|
| Ansiedad | 0.3884 | Orgullo | 0.595 | Soledad | 0.4663 |
| Aversión | 0.46 | Prejuicio | 0.518 | anger | 0.2491 |
| Depresión | 0.5536 | Temor | 0.4545 | arousal | 0.1144 |
| Desilusión | 0.4762 | Vulnerabilidad | 0.4337 | disgust | 0.2782 |
| Dramatización | 0.2819 | concreteness | 0.1361 | fear | 0.2858 |
| Ilusión | 0.49 | context availability | 0.0257 | happiness | -0.0723 |
| Impotencia | 0.491 | Compulsión | 0.2997 | imageability | 0.0363 |
| Inestabilidad | 0.3659 | Ensoñación | 0.4284 | sadness | 0.1648 |
| Inseguridad | 0.4564 | Grandiosidad | 0.5543 | valence | 0.0347 |
| Ira | 0.4848 | Idealización | 0.4295 | Soledad | 0.1943 |
| Obsesión | 0.267 | Irritabilidad | 0.3743 | | |

Tabla 11 Comparativa entre los anotadores 2 y 3

Las Tablas 9, 10 y 11 muestran cómo los valores de k-alpha cambian significativamente según qué par de anotadores se compare. De nuevo, la validación de variables se hace para aquellas con k-alpha mayor que 0.21. En particular se comprueba lo siguiente:

- Para los anotadores 1 y 2 (Tabla 9) sólo se validan las variables 'Ansiedad', 'Aversión', 'Dramatización', 'Inestabilidad', 'Inseguridad', 'Obsesión', 'Temor', 'Vulnerabilidad'. Esto representa un 25.81 % respecto al total de variables.
- Para los anotadores 1 y 3 (Tabla 10) sólo se valida la variable 'Ilusión', lo que representa un 3.23 % respecto al total de variables.
- Para los anotadores 2 y 3 (Tabla 11) las variables validadas son: 'Ansiedad', 'Aversión', 'Compulsión', 'Depresión', 'Desilusión', 'Dramatización', 'Ensoñación', 'Grandiosidad', 'Idealización', 'Ilusión', 'Impotencia', 'Inestabilidad', 'Inseguridad', 'Ira', 'Irritabilidad', 'Obsesión', 'Orgullo', 'Prejuicio', 'Soledad', 'Temor', 'Vulnerabilidad', 'anger', 'disgust', 'fear'. Esto representa el 77.42 % de las variables (e incluye todas las variables psicológicas).
- Para 2 y 3 sólo aumenta mucho la k-alpha de algunas variables, especialmente las de los términos psicológicos. Sin embargo baja mucho para las variables de menor k-alpha (happiness, valence).
- Las mayores discrepancias se dan con el anotador 1, especialmente entre el 1 y el 3 donde únicamente una variable tiene k-alpha mayor que 0.21: Ilusión

Así, la mayor discrepancia se presenta respecto al anotador 1. Este estudio de acuerdo entre anotadores se puede enriquecer comparando la media del vector de anotaciones por variable

entre cada par de anotadores para esas variables en las que k-alpha es mayor que 0.21. Con ello se busca estudiar si hay diferencias significativas entre las medias por anotador, y ver que variables son las que tienen un mayor nivel de acuerdo. Por ello, se lleva a cabo un análisis ANOVA (*One-way Anova*) para comparar la media para las distintas variables por cada par de anotadores para ver formalmente si existe entre ellos una diferencia significativa. La hipótesis nula en el contraste (H0) es que no hay diferencia. Si el p-valor del contraste es menor que un umbral (0.05 como referencia) se descarta H0 y se dice que para esa variable en concreto hay diferencias significativas en la media de valores asignados según qué anotador. Si el p-valor está por encima de 0.05 no se puede descartar H0 y por tanto puede que no haya diferencias significativas.

En la Tabla 12 se muestran las variables con k-alpha superior a 0.21 para cada par de anotadores junto con el p-valor obtenido comparando los resultados de dichas variables.

| pair | variable | k_alpha | p_value_oneway |
|--------|----------------|---------|----------------|
| [1, 2] | Ansiedad | 0.2363 | 0.1251 |
| [1, 2] | Aversión | 0.2178 | 0.0017 |
| [1, 2] | Dramatización | 0.2749 | 0.7795 |
| [1, 2] | Inestabilidad | 0.2195 | 0.4572 |
| [1, 2] | Inseguridad | 0.2734 | 0.0915 |
| [1, 2] | Obsesión | 0.3102 | 0.0874 |
| [1, 2] | Temor | 0.2402 | 0.1236 |
| [1, 2] | Vulnerabilidad | 0.2450 | 0.0767 |
| [1, 3] | Ilusión | 0.2382 | 0.0115 |
| [2, 3] | Ansiedad | 0.3884 | 0.0000 |
| [2, 3] | Aversión | 0.4600 | 0.0000 |
| [2, 3] | Compulsión | 0.2997 | 0.0000 |
| [2, 3] | Depresión | 0.5536 | 0.0289 |
| [2, 3] | Desilusión | 0.4762 | 0.3482 |
| [2, 3] | Dramatización | 0.2819 | 0.0016 |
| [2, 3] | Ensoñación | 0.4284 | 0.1198 |
| [2, 3] | Grandiosidad | 0.5543 | 0.0010 |
| [2, 3] | Idealización | 0.4295 | 0.0000 |
| [2, 3] | Ilusión | 0.4900 | 0.0032 |
| [2, 3] | Impotencia | 0.4910 | 0.4572 |
| [2, 3] | Inestabilidad | 0.3659 | 0.0000 |
| [2, 3] | Inseguridad | 0.4564 | 0.0268 |
| [2, 3] | Ira | 0.4848 | 0.0024 |
| [2, 3] | Irritabilidad | 0.3743 | 0.0229 |
| [2, 3] | Obsesión | 0.2670 | 0.0012 |
| [2, 3] | Orgullo | 0.5950 | 0.0562 |
| [2, 3] | Prejuicio | 0.5180 | 0.3522 |
| [2, 3] | Soledad | 0.4663 | 0.0044 |
| [2, 3] | Temor | 0.4545 | 0.0007 |
| [2, 3] | Vulnerabilidad | 0.4337 | 0.0000 |

| | | | |
|--------|---------|--------|--------|
| [2, 3] | anger | 0.2491 | 0.8963 |
| [2, 3] | disgust | 0.2782 | 0.0003 |
| [2, 3] | fear | 0.2858 | 0.0781 |

Tabla 12 One-way ANOVA para cada variable por par de anotadores

Una consideración adicional a tener en cuenta es que en las variables binarias existe un gran desbalanceo de datos, debido a que hay muchos más sonetos sin etiquetas de 1 que etiquetados con 1. Así, si por ejemplo se comparan las anotaciones de un anotador frente a otro, para una categoría concreta se tendría una matriz de confusión sobre la que calcular unas métricas de precisión, *recall*, y F1 para evaluar si coinciden con etiquetas de 1 un número adecuado de sonetos.

Así, por ejemplo, para la variable *Prejuicio* analizando las etiquetas del anotador 2 frente al 1 se tiene la siguiente matriz de confusión.

$$\begin{pmatrix} 192 & 22 \\ 12 & 4 \end{pmatrix}$$

De manera que sólo coinciden 4 anotaciones con un 1, frente a 22 que el anotador 2 ha etiquetado como 0 cuando el anotador 1 ha etiquetado como 1. Esto representa un 0.15 aproximadamente de tasa de acierto, valor que coincide con el de *precisión*. También interesa analizar el valor de *recall* con el que se ve que los 16 sonetos que el anotador 1 ha etiquetado con 1 (12+4) sólo coinciden los anotadores en 4 de ellos.

En la Tabla 13 se muestran los resultados de las métricas para todas las parejas de anotadores, incluyendo una métrica F1 para ponderar de forma conjunta tanto la *precisión* como el *recall*.

| feature_col | Precision (1, 2) | Precision (1, 3) | Precision (2, 3) | Recall (1, 2) | Recall (1, 3) | Recall (2, 3) | F1 (1, 2) | F1 (1, 3) | F1 (2, 3) |
|---------------|------------------|------------------|------------------|---------------|---------------|---------------|-----------|-----------|-----------|
| Ansiedad | 0.41 | 0.39 | 0.90 | 0.52 | 0.20 | 0.37 | 0.46 | 0.26 | 0.53 |
| Aversión | 0.45 | 0.48 | 0.93 | 0.65 | 0.38 | 0.51 | 0.53 | 0.42 | 0.66 |
| Compulsión | 0.22 | 0.37 | 0.83 | 0.45 | 0.29 | 0.32 | 0.29 | 0.32 | 0.46 |
| Depresión | 0.31 | 0.22 | 0.83 | 0.28 | 0.12 | 0.49 | 0.29 | 0.15 | 0.61 |
| Desilusión | 0.24 | 0.22 | 0.53 | 0.27 | 0.30 | 0.63 | 0.26 | 0.26 | 0.58 |
| Dramatización | 0.60 | 0.55 | 0.69 | 0.62 | 0.39 | 0.48 | 0.61 | 0.46 | 0.56 |
| Ensoñación | 0.32 | 0.34 | 0.62 | 0.24 | 0.19 | 0.44 | 0.27 | 0.24 | 0.51 |
| Grandiosidad | 0.47 | 0.47 | 0.90 | 0.40 | 0.25 | 0.57 | 0.43 | 0.33 | 0.70 |
| Idealización | 0.37 | 0.41 | 0.91 | 0.53 | 0.33 | 0.51 | 0.43 | 0.37 | 0.66 |
| Ilusión | 0.40 | 0.51 | 0.81 | 0.42 | 0.32 | 0.48 | 0.41 | 0.40 | 0.60 |
| Impotencia | 0.38 | 0.39 | 0.66 | 0.39 | 0.35 | 0.59 | 0.38 | 0.37 | 0.62 |
| Inestabilidad | 0.40 | 0.48 | 0.81 | 0.45 | 0.23 | 0.35 | 0.42 | 0.31 | 0.49 |
| Inseguridad | 0.35 | 0.32 | 0.71 | 0.49 | 0.29 | 0.45 | 0.40 | 0.30 | 0.55 |
| Ira | 0.32 | 0.24 | 0.82 | 0.43 | 0.18 | 0.46 | 0.37 | 0.21 | 0.59 |
| Irritabilidad | 0.22 | 0.21 | 0.62 | 0.32 | 0.18 | 0.37 | 0.26 | 0.19 | 0.46 |
| Obsesión | 0.32 | 0.31 | 0.62 | 0.50 | 0.18 | 0.24 | 0.39 | 0.23 | 0.34 |
| Orgullo | 0.35 | 0.37 | 0.82 | 0.38 | 0.30 | 0.61 | 0.36 | 0.33 | 0.70 |

| | | | | | | | | | |
|-----------------------|------|------|------|------|------|------|------|------|------|
| Prejuicio | 0.15 | 0.20 | 0.65 | 0.25 | 0.25 | 0.50 | 0.19 | 0.22 | 0.57 |
| Soledad | 0.37 | 0.36 | 0.79 | 0.30 | 0.16 | 0.43 | 0.33 | 0.22 | 0.56 |
| Temor | 0.48 | 0.43 | 0.82 | 0.58 | 0.33 | 0.52 | 0.52 | 0.38 | 0.64 |
| Vulnerabilidad | 0.57 | 0.57 | 0.89 | 0.67 | 0.41 | 0.54 | 0.62 | 0.48 | 0.68 |

Tabla 13 Métricas para comparar las anotaciones de variables binarias entre los distintos anotadores

Con los resultados mostrados en la Tabla 13 se reafirman las discrepancias respecto al anotador 1, con el cual se tienen de forma sistemática peores métricas para todas las variables.

Para poder plantear los experimentos posteriores es necesario componer las anotaciones de los tres expertos en un único vector por variable. La propuesta inicial para construir este vector es utilizar la mediana del valor de los 3 anotadores en cada variable por soneto, de manera que se le asigne a cada soneto dicho valor mediano.

Así, si por ejemplo dos anotadores dan un 1 en una variable para un soneto y el tercer anotador da un 0, la mediana es 1 y ese es el valor final que se usa en esa variable para ese soneto.

Las Tablas 14, 15 y 16 muestran la comparativa tanto con el contraste ANOVA como con k-alpha de ese vector, con los valores medianos de los anotadores por soneto y variable.

| pair | variable | k_alpha | p_value | variable | k_alpha | p_value |
|--------|---------------|---------|---------|----------------------|---------|---------|
| [1, 4] | Ansiedad | 0.4532 | 0.1520 | Obsesión | 0.5807 | 0.4037 |
| [1, 4] | Aversión | 0.4740 | NaN | Orgullo | 0.2699 | 0.1288 |
| [1, 4] | Compulsión | 0.4313 | 0.7026 | Prejuicio | 0.2288 | 0.5659 |
| [1, 4] | Depresión | 0.2344 | 0.0264 | Soledad | 0.2755 | 0.0014 |
| [1, 4] | Desilusión | 0.2938 | 0.6045 | Temor | 0.4433 | 0.4257 |
| [1, 4] | Dramatización | 0.5576 | 0.3474 | Vulnerabilidad | 0.4892 | 0.5739 |
| [1, 4] | Ensoñación | 0.2812 | 0.0024 | anger | 0.3848 | 0.0280 |
| [1, 4] | Grandiosidad | 0.1845 | 0.0004 | arousal | 0.5049 | 0.2961 |
| [1, 4] | Idealización | 0.3191 | 0.7693 | concreteness | 0.6449 | 0.6679 |
| [1, 4] | Ilusión | 0.4221 | 0.0311 | context availability | 0.6392 | 0.3615 |
| [1, 4] | Impotencia | 0.3796 | 0.2344 | disgust | 0.3724 | 0.0062 |
| [1, 4] | Inestabilidad | 0.4486 | 0.0097 | fear | 0.3292 | 0.0001 |
| [1, 4] | Inseguridad | 0.4705 | 0.5957 | happiness | 0.6300 | 0.8008 |
| [1, 4] | Ira | 0.3606 | 0.6302 | imageability | 0.6733 | 0.8794 |
| [1, 4] | Irritabilidad | 0.3829 | 0.6622 | sadness | 0.4416 | 0.1265 |
| [1, 4] | valence | 0.6753 | 0.5922 | | | |

Tabla 14 Comparativa del soneto con los valores medianos frente al anotador 1

| pair | variable | k_alpha | p_value | variable | k_alpha | p_value |
|--------|---------------|---------|---------|-------------------------|---------|---------|
| [2, 4] | Ansiedad | 0.6883 | 0.0030 | Orgullo | 0.7498 | 0.0328 |
| [2, 4] | Aversión | 0.7033 | 0.0017 | Prejuicio | 0.6366 | 0.0296 |
| [2, 4] | Compulsión | 0.4498 | 0.0000 | Soledad | 0.7146 | 0.0559 |
| [2, 4] | Depresión | 0.7676 | 0.0822 | Temor | 0.7542 | 0.0194 |
| [2, 4] | Desilusión | 0.7429 | 0.3110 | Vulnerabilidad | 0.7480 | 0.0196 |
| [2, 4] | Dramatización | 0.7090 | 0.2226 | anger | 0.7137 | 0.1764 |
| [2, 4] | Ensoñación | 0.6637 | 0.1198 | arousal | 0.5817 | 0.0005 |
| [2, 4] | Grandiosidad | 0.7807 | 0.0402 | concreteness | 0.7336 | 0.0861 |
| [2, 4] | Idealización | 0.6358 | 0.0002 | context availability | 0.7152 | 0.4491 |
| [2, 4] | Ilusión | 0.6630 | 0.0100 | disgust | 0.7054 | 0.0044 |
| [2, 4] | Impotencia | 0.7207 | 0.1958 | fear | 0.7010 | 0.6839 |
| [2, 4] | Inestabilidad | 0.6220 | 0.0009 | happiness | 0.5542 | 0.0006 |
| [2, 4] | Inseguridad | 0.6980 | 0.0268 | imageability | 0.6901 | 0.9242 |
| [2, 4] | Ira | 0.7303 | 0.0311 | sadness | 0.7585 | 0.1826 |
| [2, 4] | Irritabilidad | 0.6117 | 0.0334 | valence | 0.5344 | 0.4233 |
| [2, 4] | Obsesión | 0.5819 | 0.0115 | | | |

Tabla 15 Comparativa del soneto con los valores medianos frente al anotador 2

| pair | variable | k_alpha | p_value | variable | k_alpha | p_value |
|--------|---------------|---------|---------|-------------------------|---------|---------|
| [3, 4] | Ansiedad | 0.6611 | 0.0469 | Obsesión | 0.5730 | 0.4511 |
| [3, 4] | Aversión | 0.7407 | 0.0834 | Orgullo | 0.8215 | 0.8216 |
| [3, 4] | Compulsión | 0.8033 | 0.5044 | Prejuicio | 0.7720 | 0.2068 |
| [3, 4] | Depresión | 0.7493 | 0.6511 | Soledad | 0.7247 | 0.3435 |
| [3, 4] | Desilusión | 0.6742 | 0.0514 | Temor | 0.6790 | 0.3015 |
| [3, 4] | Dramatización | 0.5696 | 0.0529 | Vulnerabilidad | 0.6764 | 0.0268 |
| [3, 4] | Ensoñación | 0.7244 | NaN | anger | 0.6609 | 0.2011 |
| [3, 4] | Grandiosidad | 0.7624 | 0.2127 | arousal | 0.4696 | 0.0000 |
| [3, 4] | Idealización | 0.7781 | 0.2285 | concreteness | 0.5774 | 0.1481 |
| [3, 4] | Ilusión | 0.7977 | 0.7086 | context availability | 0.4876 | 0.0002 |
| [3, 4] | Impotencia | 0.7448 | 0.5821 | disgust | 0.6968 | 0.3077 |
| [3, 4] | Inestabilidad | 0.6794 | 0.3370 | fear | 0.6535 | 0.0260 |
| [3, 4] | Inseguridad | 0.7024 | NaN | happiness | 0.2872 | 0.0000 |
| [3, 4] | Ira | 0.7238 | 0.3728 | imageability | 0.4576 | 0.0000 |
| [3, 4] | Irritabilidad | 0.7037 | 0.8802 | sadness | 0.6459 | 0.0110 |
| [3, 4] | valence | 0.2694 | 0.0618 | | | |

Tabla 16 Comparativa del soneto con los valores medianos frente al anotador 3

Como se puede apreciar, en todos los casos salvo para el término *Grandiosidad* con el anotador 1 (k-alpha = 0.1845) se obtiene un k-alpha superior a 0.21. Esto muestra como el vector con los valores medianos tiene un alto nivel de acuerdo con todos los anotadores para prácticamente

todas las variables (a pesar de que en algunos casos la media sea distinta, como muestran los p valores del contraste ANOVA).

En los análisis que se han ido exponiendo se puede observar como para dos de las variables afectivas, en particular, *arousal* y *valence*, hay una especial discrepancia entre los anotadores. Como estas son dos variables afectivas de especial relevancia para los experimentos, se analizan con más detalle sus anotaciones por parte de los expertos.

En las Ilustraciones 80 y 81 se muestran los histogramas para las anotaciones de las variables *arousal* y *valence*, respectivamente. En dichas gráficas se puede apreciar cómo hay una diferencia entre las etiquetas que usan los anotadores. El anotador 2, por ejemplo, tiende más a utilizar los valores bajos para la variable *arousal*, frente al anotador 3 que es quien suele dar los valores más altos. El anotador 2, a su vez, tiende a dar puntuaciones más bajas también para la variable *valence*. Como comparativa respecto a uno de los términos de más acuerdo, *Ansiedad*, se ve que en el histograma (Ilustración 82) coinciden mucho más el número de sonetos que los expertos anotan con 0 o con 1. El anotador 2, por lo tanto, parece que tiende a dar valores más bajos en las métricas de *arousal* y *valence*. Esto es un complemento al análisis de k-alpha ya que aunque haya acuerdo en los valores de los sonetos anotados, la discrepancia para los demás se debe posiblemente a esa tendencia a puntuar más bajo.

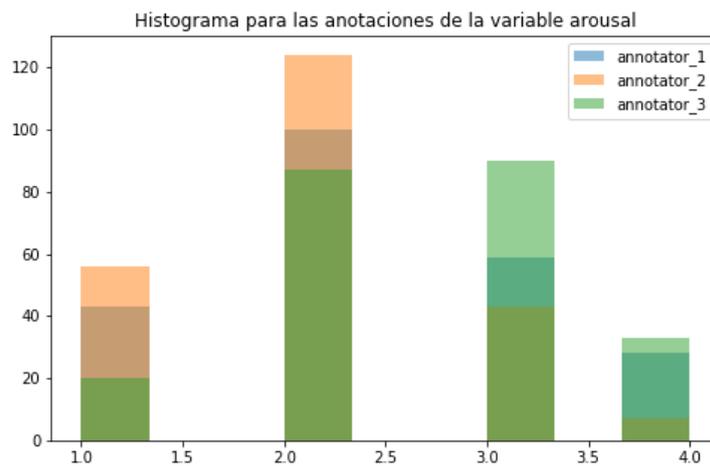


Ilustración 80 Histograma para las anotaciones de la variable *arousal*

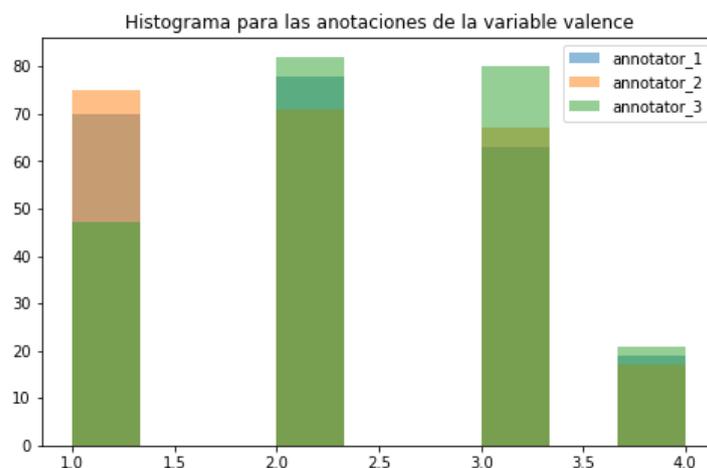


Ilustración 81 Histograma para las anotaciones de la variable *valence*

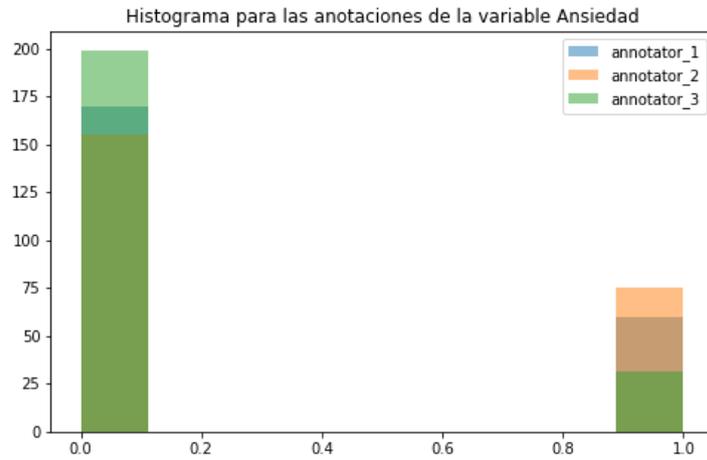


Ilustración 82 Histograma para las anotaciones de la variable Ansiedad

En las Ilustraciones 83, 84 y 85 se muestra la diferencia de puntuación por soneto y para cada anotador para la variable *arousal*. Como se puede apreciar, no hay tantos sonetos cuya diferencia sea 0 (es decir, misma puntuación) y en cambio hay muchos sonetos en los que el anotador 1 o el anotador 3 puntúan con más valor que el anotador 2 (diferencia mayor que 0).

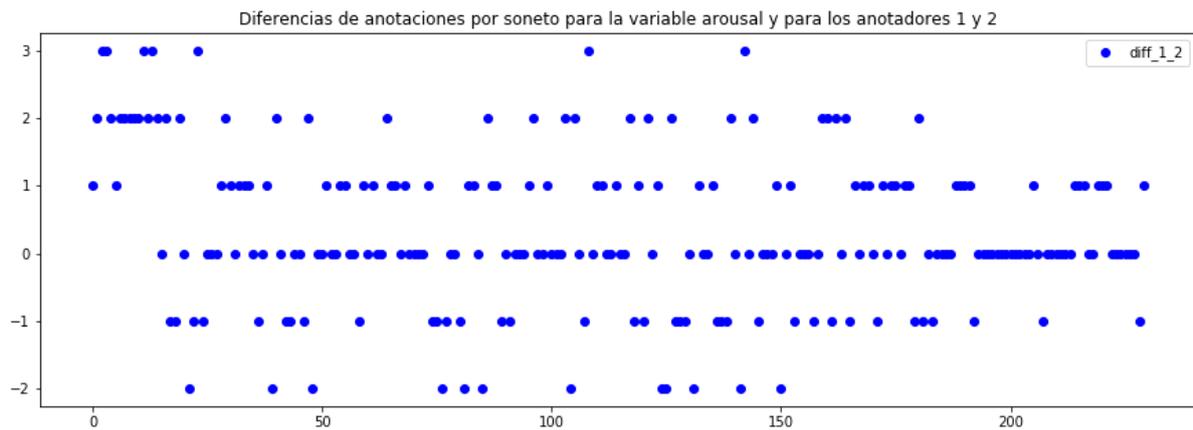


Ilustración 83 Diferencias de anotaciones por soneto para la variable arousal y los anotadores 1 y 2

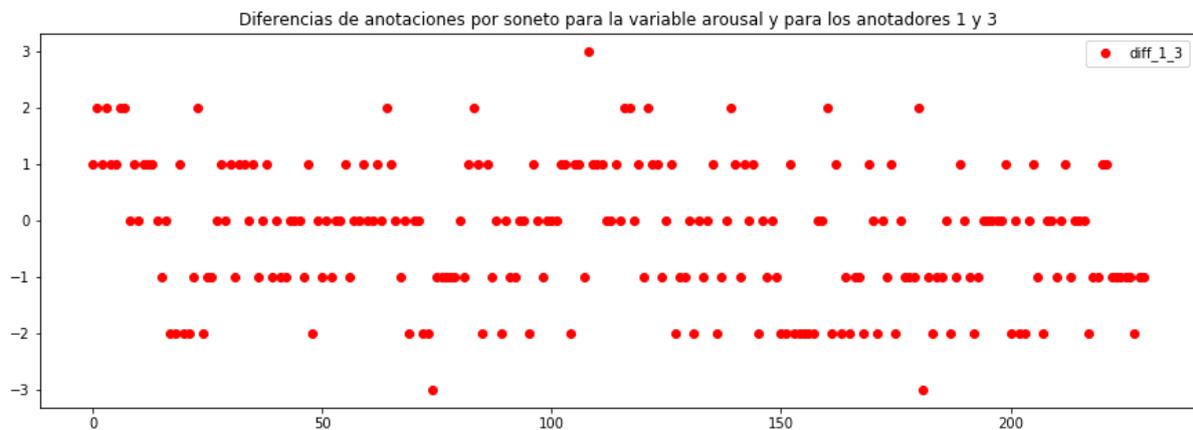


Ilustración 84 Diferencias de anotaciones por soneto para la variable arousal y los anotadores 1 y 3

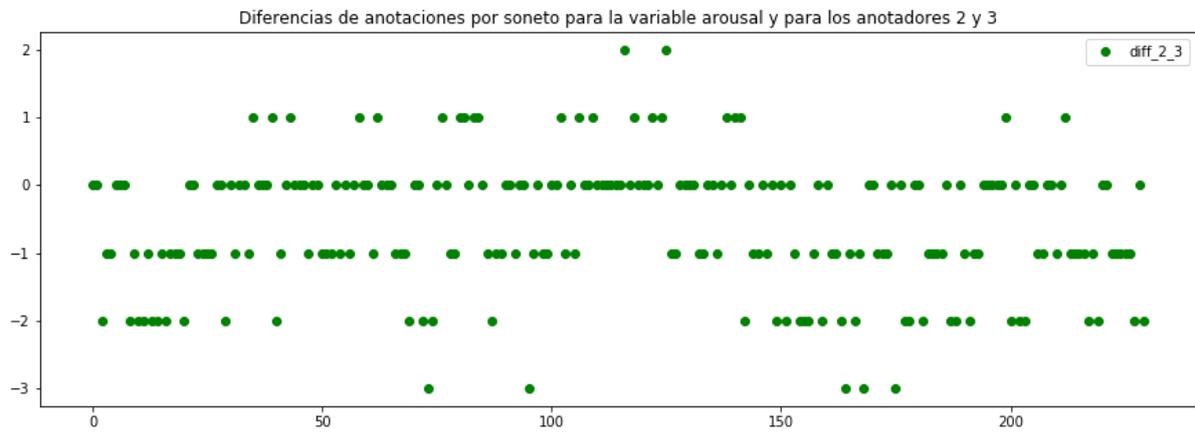


Ilustración 85 Diferencias de anotaciones por soneto para la variable arousal y los anotadores 2 y 3

En las Ilustraciones 86, 87 y 88 se muestra el mismo análisis para la variable *valence*, donde de nuevo se ve una discrepancia análoga por soneto.

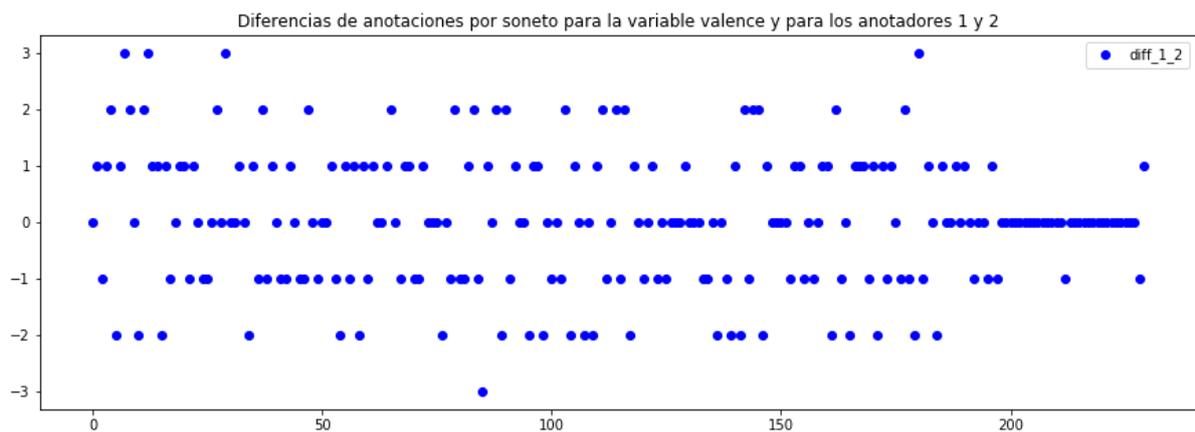


Ilustración 86 Diferencias de anotaciones por soneto para la variable valence y los anotadores 1 y 2

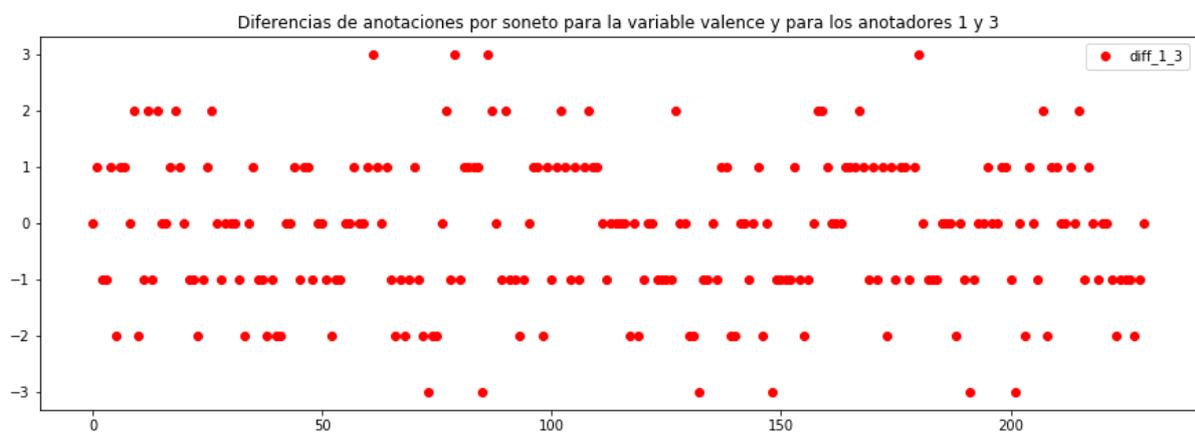


Ilustración 87 Diferencias de anotaciones por soneto para la variable valence y los anotadores 1 y 3

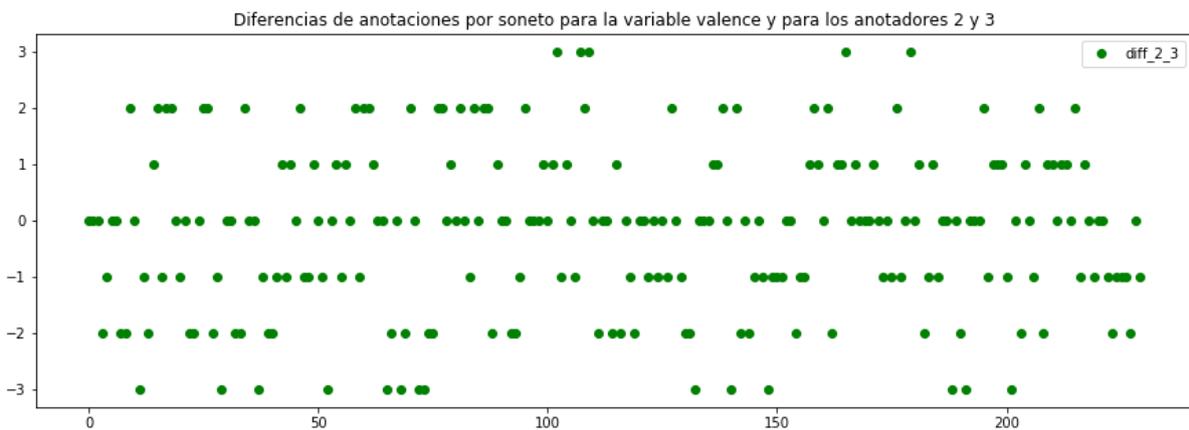


Ilustración 88 Diferencias de anotaciones por soneto para la variable valence y los anotadores 1 y 2

Las conclusiones de este subapartado se muestran a continuación:

- Las variables que tienen un nivel de acuerdo aceptable entre los tres anotadores son: 'Ansiedad', 'Aversión', 'Depresión', 'Desilusión', 'Dramatización', 'Ilusión', 'Impotencia', 'Inestabilidad', 'Inseguridad', 'Ira', 'Obsesión', 'Orgullo', 'Prejuicio', 'Temor', 'Vulnerabilidad', 'concreteness', 'context availability'. Esto corresponde a un 54.84 % del total de variables existentes.
- Las variables indicadas serán las que se utilicen para el estudio de la recuperación de sonetos según los términos psicológicos.
- Sólo 2 de las variables indicadas en el punto anterior pertenecen al ámbito de las variables afectivas ('concreteness', 'context availability') siendo el resto variables pertenecientes al subconjunto de las variables del dominio psicológico (variables que además son binarias).
- Por ello, aunque se plantee un análisis para todas las variables, las que verdaderamente servirán para el estudio de la composición de rasgos afectivos son 'concreteness' y 'context availability'.
- Para todos los estudios se usará el vector con el valor mediano de las anotaciones de los tres expertos por cada variable.
- Cabe recordar aun así la complejidad inherente a trabajar con textos en castellano antiguo. Igualmente cabe remarcar que los niveles de acuerdo entre anotadores no son excesivamente altos aun cuando han servido para validar mínimamente las anotaciones realizadas.

Experimento 1

Como se mencionó en el Capítulo 1, el experimento 1 busca comprobar la hipótesis de investigación 2, en la que se recoge que el valor de los rasgos afectivos de los poemas coincide con el valor afectivo individual derivado de las palabras que lo componen, como se hace en (Ullrich et al., 2016) para poesía en alemán.

Para ello se utilizan los corpus afectivos mencionados en el Capítulo 2 para construir, en base a las palabras del soneto que aparezcan en ellos, determinados rasgos afectivos para el soneto en su totalidad. Ciertamente, algunas de las palabras de los sonetos no estarán presentes en los *corpora* utilizados, aun cuando se van a lematizar y eliminar *stopwords* de cara a maximizar la representatividad. Esto se debe, principalmente, a que ciertas palabras del castellano antiguo no aparecen recogidas en esos *corpora*. A pesar de ello, se procede con el estudio para analizar si con las palabras que están presentes es suficiente para obtener conclusiones relevantes con este experimento.

Para todo este análisis se recuerda que se está utilizando el vector con los valores medianos de los anotadores. Aunque el acuerdo entre ellos se da principalmente para los rasgos psicológicos (ya que en los afectivos sólo se da en “*concreteness*” y “*context availability*”), se va a estudiar respecto a todas las variables debido a que el vector mediano manifiesta un acuerdo relevante respecto a cada anotador de manera individual.

La Tabla 17 contiene los vectores de correlación de las variables anotadas por POSTDATA “*anger*”, “*fear*”, “*happiness*”, “*sadness*” y “*disgust*” con los valores inferidos para esas mismas variables. De (Ferré, Guasch, Martínez-García, Fraga, & Hinojosa, 2017) se usa su valor medio y desviación típica. Como se puede observar, hay una correlación relevante (>0.21) para las variables:

- *Anger* [valor medio] con 0.35
- *Anger* [desviación típica] con 0.33
- *Sadness* [valor medio] con 0.48
- *Sadness* [desviación típica] con 0.27
- *Happiness* [valor medio] con 0.29
- *Happiness* [desviación típica] con 0.21
- *Fear* [valor medio] con 0.27

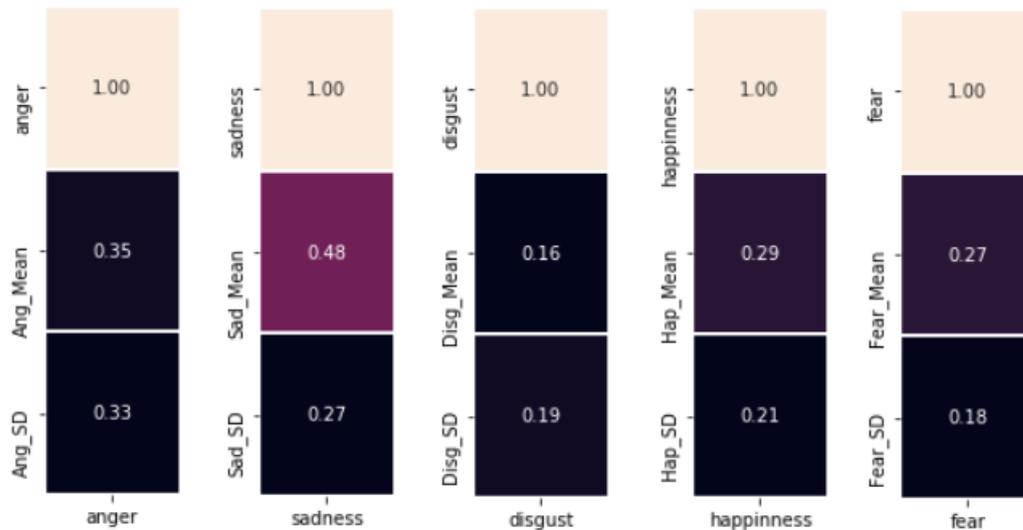


Tabla 17 Correlación de los valores de media y desviación típica de “anger”, “sadness”, “disgust”, “happiness” y “fear” (eje Y) inferidas del texto usando (Ferré, Guasch, Martínez-García, Fraga, & Hinojosa, 2017) con los anotados por POSTDATA.

En la Tabla 18 aparecen los vectores de correlación para las variables “arousal” y “valence” anotados por POSTDATA con respecto a determinados rasgos para ambas variables. Estos rasgos son: valor máximo, valor mínimo, media, desviación, diferencia entre valor máximo y mínimo, y correlación de los valores con la posición de cada palabra. Los valores afectivos individuales por palabra se han obtenido del corpus (Stadthagen-Gonzalez, Imbault, Pérez Sánchez, & Brysbaert, 2017). En la Tabla 18 se ve cómo para ninguna de las variables mencionadas se alcanza el mínimo de correlación especificada para el experimento (>0.21). De esta manera, no se consigue validar que los rasgos afectivos compuestos a partir de las palabras individuales usados en (Ullrich et al., 2016) modelen afectivamente todo el conjunto del soneto, usando los datos recogidos en la presente memoria.

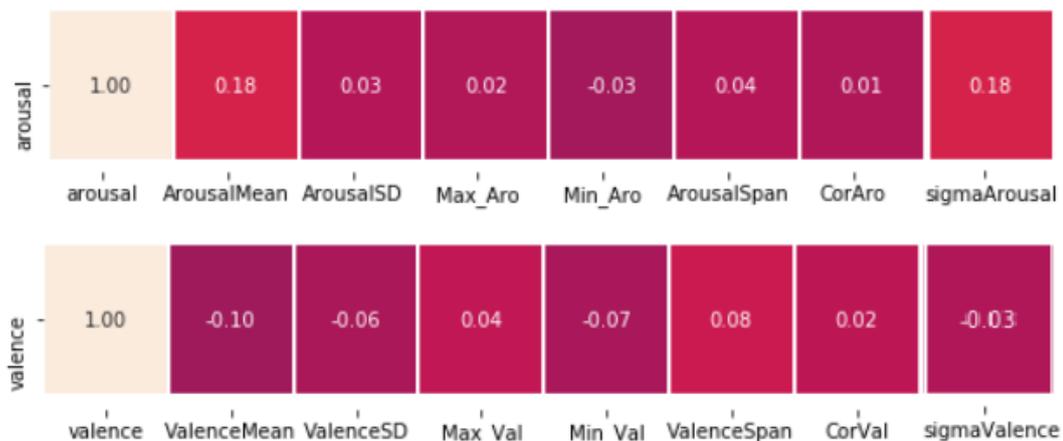


Tabla 18 Correlación de algunos rasgos usados en (Ullrich et al., 2016) (eje X) inferida del texto usando (Stadthagen-Gonzalez, Imbault, Pérez Sánchez, & Brysbaert, 2017) con los anotados por POSTDATA (eje Y)

Por último, la Tabla 19 recoge la correlación entre los rasgos afectivos de "concreteness", "context availability" e "imageability" (las anotaciones de POSTDATA) con los rasgos afectivos inferidos desde (Guasch, Ferré, & Fraga, 2016). Como el corpus mencionado contiene también valores para "arousal" y "valence" se comprueban también los resultados usando estos valores. Para ninguno de los casos se consigue superar el mínimo de correlación estipulado para este experimento (>0.21) de manera que no se consigue validar la hipótesis 2 para estas variables.

| | | | | | | | | | | | | | | | | |
|----------------------|------|-------|--------|-------|-----------|-----------|------------|--------------|------------|------------|-------------|---------------|-----------|-----------|------------|--------------|
| concreteness | 1.00 | 0.02 | -0.10 | 0.03 | -0.01 | -0.06 | 0.01 | -0.00 | -0.03 | -0.03 | 0.01 | -0.12 | 0.06 | 0.03 | 0.05 | 0.01 |
| concreteness | | CON_M | CON_SD | CON_N | max_CON_M | min_CON_M | span_CON_M | median_CON_M | max_CON_SD | min_CON_SD | span_CON_SD | median_CON_SD | max_CON_N | min_CON_N | span_CON_N | median_CON_N |
| context availability | 1.00 | 0.00 | 0.12 | -0.03 | -0.08 | 0.14 | -0.15 | -0.07 | 0.10 | 0.14 | -0.07 | 0.08 | -0.10 | -0.03 | -0.08 | -0.04 |
| context availability | | AVA_M | AVA_SD | AVA_N | max_AVA_M | min_AVA_M | span_AVA_M | median_AVA_M | max_AVA_SD | min_AVA_SD | span_AVA_SD | median_AVA_SD | max_AVA_N | min_AVA_N | span_AVA_N | median_AVA_N |
| imageability | 1.00 | 0.09 | -0.07 | -0.02 | -0.03 | 0.12 | -0.13 | 0.07 | 0.04 | -0.02 | 0.02 | -0.07 | -0.08 | 0.10 | -0.11 | -0.02 |
| imageability | | IMA_M | IMA_SD | IMA_N | max_IMA_M | min_IMA_M | span_IMA_M | median_IMA_M | max_IMA_SD | min_IMA_SD | span_IMA_SD | median_IMA_SD | max_IMA_N | min_IMA_N | span_IMA_N | median_IMA_N |
| valence | 1.00 | -0.04 | -0.00 | -0.01 | 0.08 | -0.09 | 0.15 | -0.04 | 0.04 | -0.10 | 0.08 | -0.00 | -0.01 | -0.08 | 0.00 | -0.02 |
| valence | | VAL_M | VAL_SD | VAL_N | max_VAL_M | min_VAL_M | span_VAL_M | median_VAL_M | max_VAL_SD | min_VAL_SD | span_VAL_SD | median_VAL_SD | max_VAL_N | min_VAL_N | span_VAL_N | median_VAL_N |
| arousal | 1.00 | 0.11 | 0.06 | -0.03 | 0.05 | 0.01 | 0.02 | 0.11 | 0.01 | -0.06 | 0.03 | 0.06 | -0.01 | -0.07 | 0.02 | -0.02 |
| arousal | | ARO_M | ARO_SD | ARO_N | max_ARO_M | min_ARO_M | span_ARO_M | median_ARO_M | max_ARO_SD | min_ARO_SD | span_ARO_SD | median_ARO_SD | max_ARO_N | min_ARO_N | span_ARO_N | median_ARO_N |

Tabla 19 Correlaciones de los rasgos afectivos anotados (eje Y) con los rasgos afectivos inferidos (eje X) para "concreteness", "context availability", "imageability", "valence" y "arousal" con el conjunto de datos (Guasch, Ferré, & Fraga, 2016)

De esta manera, este primer punto del experimento sirve para reconocer que la composición de los rasgos afectivos desde las palabras individuales de los sonetos es una aproximación válida para modelar afectivamente el soneto en su conjunto para las variables "anger", "sadness", "happiness" y "fear", usando el corpus de (Ferré, Guasch, Martínez-García, Fraga, & Hinojosa,

2017). Sobre el resto de las variables que no han alcanzado el mínimo de correlación no se puede concluir nada, siendo así necesarias investigaciones adicionales al respecto.

Experimento 3

En el experimento 3 se comprueba la hipótesis 4, que sostiene que existe una correlación entre determinados trastornos psicológicos y el valor afectivo de la poesía. Para ello se analiza en primer lugar el nivel de correlación entre las anotaciones de POSTDATA de los rasgos psicológicos y sus propias anotaciones afectivas. A continuación se realiza este mismo estudio pero con los valores afectivos inferidos, especialmente los rasgos validados en el experimento 1.

Para todo este análisis se recuerda que se está utilizando el vector con los valores medianos de los anotadores. Aunque el acuerdo entre ellos se da principalmente para los rasgos psicológicos (ya que en los afectivos sólo se da en “*concreteness*” y “*context availability*”), se estudian todas las variables dado que el vector mediano manifiesta un acuerdo relevante respecto a cada anotador de manera individual.

En la Tabla 20 se muestra la correlación de los 22 rasgos psicológicos con los 11 rasgos afectivos anotados por los anotadores de POSTDATA. A continuación se destacan las correlaciones relevantes (>0.21) y se incluyen las correlaciones significativamente negativas (<-0.21). Las correlaciones positivas indican que un aumento de los rasgos afectivos conlleva una mayor presencia de los rasgos psicológicos (y viceversa) mientras que las correlaciones negativas implican lo contrario: un aumento de los rasgos afectivos conlleva una menor presencia de los rasgos psicológicos (y viceversa). Simplemente se indicarán las correlaciones significativas detectadas, sin entrar a profundizar en el análisis de sus posibles motivos.

- **Ansiedad:** La variable *Ansiedad* tiene una correlación positiva con *Anger* (0.22), *Disgust* (0.21), *Sadness* (0.36) y *Valence* (0.22).
- **Aversión:** La variable *Aversión* tiene una correlación positiva con *Anger* (0.31) y *Disgust* (0.40). En cambio, tiene una correlación negativa con *Happiness* (-0.34).
- **Compulsión:** La variable *Compulsión* no tiene ninguna correlación significativa.
- **Depresión:** La variable *Depresión* tiene una correlación positiva con *Disgust* (0.22) y con *Sadness* (0.25).
- **Desilusión:** La variable *Desilusión* tiene una correlación positiva con *Sadness* (0.25) y con *Disgust* (0.26).
- **Dramatización:** La variable *Dramatización* tiene una correlación negativa con *Happiness* (-0.31).
- **Ensoñación:** La variable *Ensoñación* no tiene ninguna correlación significativa.

- **Grandiosidad:** La variable Grandiosidad tiene una correlación negativa con Anger (-0.24), Arousal (-0.21), Disgust (-0.39), Sadness (-0.38) y Valence (-0.21).
- **Idealización:** La variable Idealización tiene una correlación negativa con Anger (-0.25), Disgust (-0.24) y Sadness (-0.28).
- **Ilusión:** La variable Ilusión tiene una correlación positiva con Happiness (0.41) y negativa con Disgust (-0.26) y Sadness (-0.30).
- **Impotencia:** La variable Impotencia tiene una correlación positiva con Disgust (0.24), Fear (0.24) y Sadness (0.33), así como negativa con Imageability (-0.21).
- **Inestabilidad:** La variable Inestabilidad no tiene ninguna correlación significativa.
- **Inseguridad:** La variable Inseguridad no tiene ninguna correlación significativa.
- **Ira:** La variable Ira tiene correlación positiva, como cabría esperar, con Anger (0.43). También tiene correlación positiva con Disgust (0.35). Tiene correlación negativa con Happiness (-0.21).
- **Irritabilidad:** La variable Irritabilidad tiene correlación positiva con Anger (0.37), Disgust (0.25) y Sadness (0.21).
- **Obsesión:** La variable Obsesión tiene correlación positiva con Disgust (0.22) y Sadness (0.22).
- **Orgullo:** La variable Orgullo tiene correlación negativa con Disgust (-0.25) y Happiness (-0.29).
- **Prejuicio:** La variable Prejuicio no tiene ninguna correlación significativa.
- **Soledad:** La variable Soledad no tiene ninguna correlación significativa.
- **Temor:** La variable Temor tiene una correlación positiva con Sadness (0.21) y negativa con Happiness (-0.25). Curiosamente con Fear solo alcanza un 0.20.
- **Vulnerabilidad:** La variable Vulnerabilidad tiene una correlación positiva con Sadness (0.33) y negativa con Imageability (-0.22).

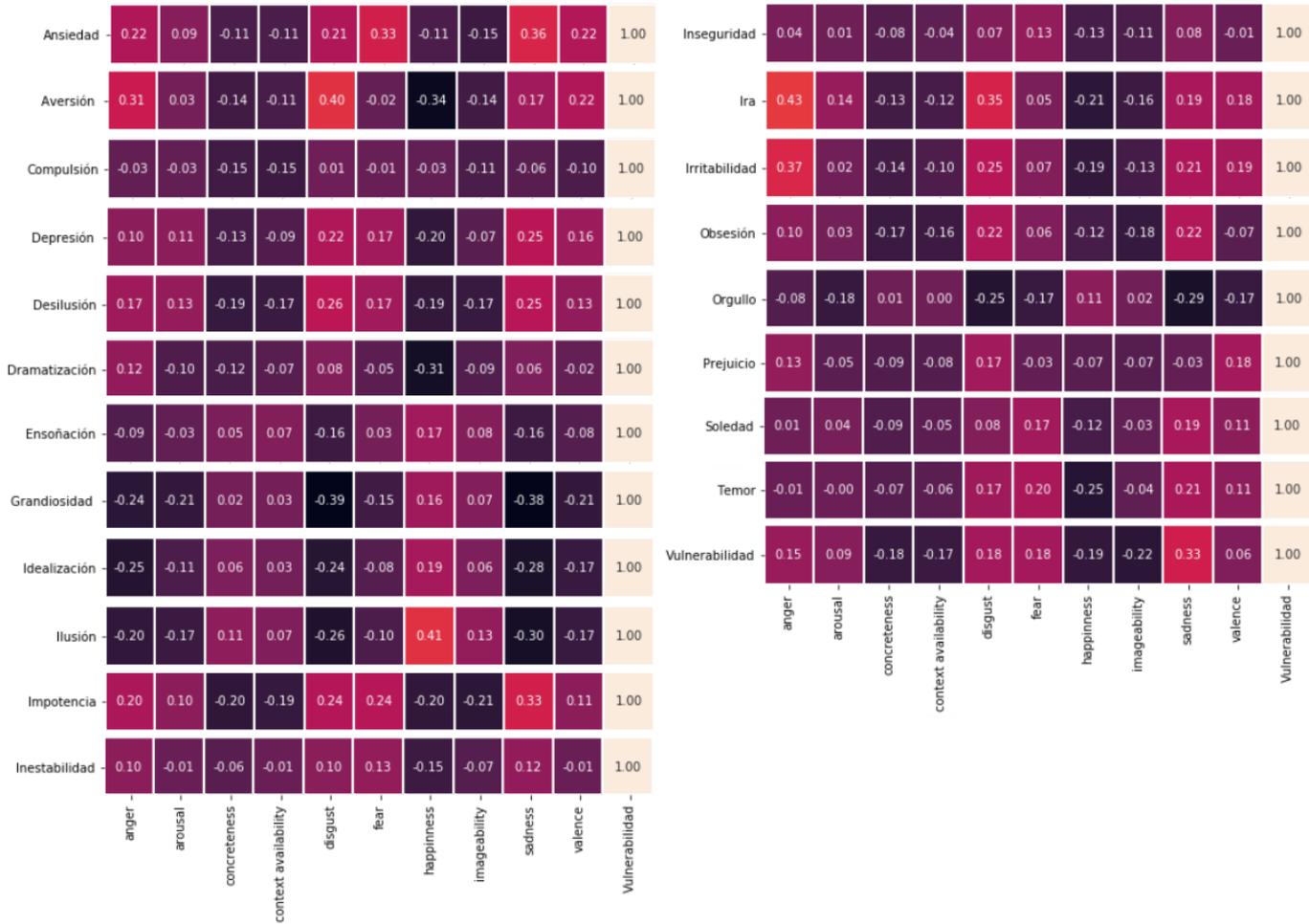


Tabla 20 Correlaciones de los términos psicológicos con los rasgos afectivos anotados por POSTDATA

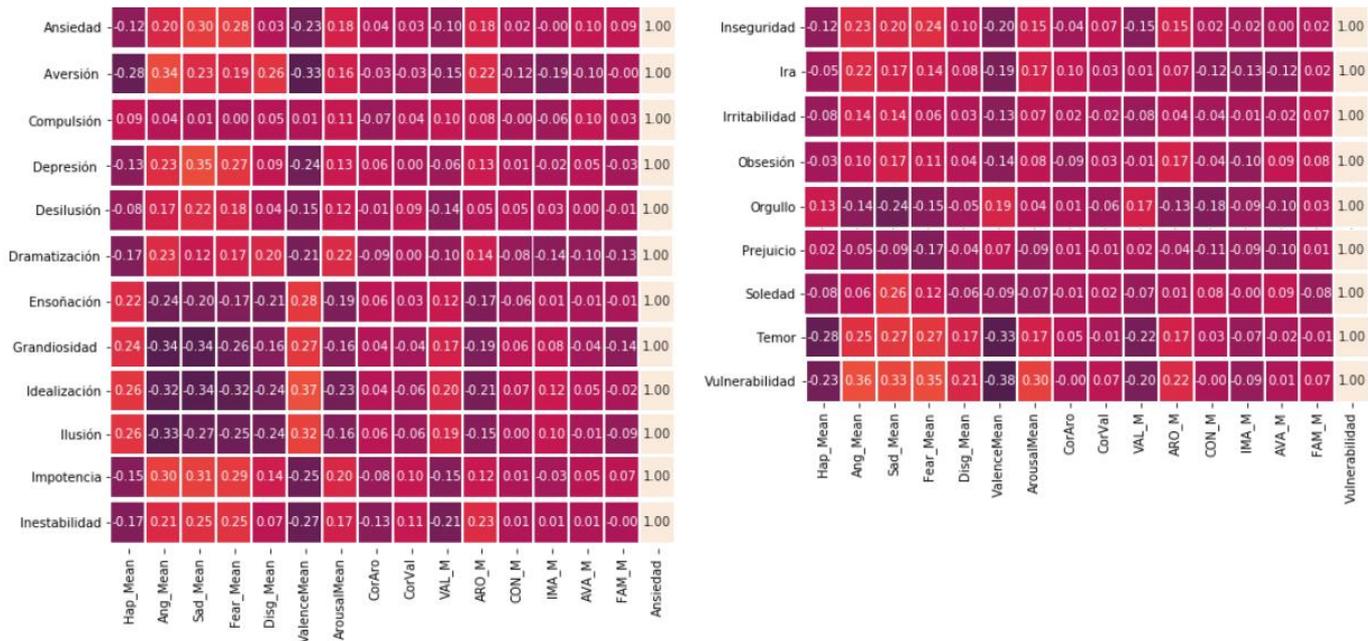


Tabla 21 Correlaciones de los términos psicológicos con los rasgos afectivos inferidos desde las palabras individuales

Experimento 2 y 4

Como se comentó en la introducción de este capítulo, en el experimento 2 y 4 se realizan de manera conjunta debido a que ambos versan sobre la recuperación de sonetos en base a los términos psicológicos anotados por POSTDATA. Se comienza analizando los resultados del experimento 4 para comentar posteriormente los del experimento 2.

El experimento 4 se centra en comprobar la hipótesis de investigación 5, que planteaba que es posible etiquetar poemas utilizando metadatos coherentes con las ontologías de uso común en los RecSys educativos afectivos, de modo que sean recuperables como LO personalizados, haciendo uso de técnicas de NLP.

Para ello se toman los siguientes rasgos psicológicos, validados en la exploración del conjunto de anotaciones proporcionadas por los expertos de POSTDATA:

- 'Ansiedad', 'Aversión', 'Depresión', 'Desilusión', 'Dramatización', 'Ilusión', 'Impotencia', 'Inestabilidad', 'Inseguridad', 'Ira', 'Obsesión', 'Orgullo', 'Prejuicio', 'Temor', 'Vulnerabilidad', 'concreteness', 'context availability'

Estos términos se utilizan como consultas de entrada al RecSys de poesía descrito en esta memoria. Como respuesta se obtiene una lista ordenada del subconjunto de 230 sonetos anotados por POSTDATA, ordenados de forma creciente por el valor de relevancia a la consulta. Como función de ranking que establezca dicha relevancia se emplea la similitud semántica que tenga el vector de *embeddings* compuesto del texto con el de la palabra de entrada.

Cada soneto está formado por 4 vectores de *embeddings* compuestos, uno por estrofa. Para este primer análisis, de cada soneto se elige el vector de la estrofa de mayor similitud semántica con el de la palabra de entrada, de manera que todo el soneto quedará representado por dicha estrofa. De esta manera la lista ordenada final tendrá los sonetos ordenados según la estrofa que ha sido más representativa semánticamente hablando.

Como métrica para evaluar la calidad del proceso de recuperación se utiliza la *Precisión* (P), y para establecer el valor de P habrá que comparar la lista de sonetos recuperada frente a una lista de referencia con las anotaciones (lo que se conoce como juicios de relevancia). El objetivo del proceso es recuperar en las primeras posiciones del ranking sonetos que hayan sido anotados como relevantes para la consulta (aquellos que fueran anotados con un 1 en relación al término que sirve como consulta).

Los sonetos recuperados, ordenados por medio de la función de ranking (que será una métrica de similitud semántica entre el soneto y el término de entrada al sistema), pueden contener o no la etiqueta del término a 1. Así, si en un conjunto de N sonetos recuperados hay N1 anotados con la etiqueta 1, una recuperación perfecta sería aquella que devolviese una lista de tamaño N donde los N1 primeros términos contuviesen el término en cuestión. Esa lista perfecta es la que se usa para calcular la precisión con respecto a la lista recuperada de sonetos.

No tiene sentido calcular otras métricas como *Recall*, ya que el número de sonetos identificados con la etiqueta 1 en la lista final va a coincidir siempre con el número real de sonetos, por lo

que la matriz de confusión será simétrica, dando el mismo valor de precisión que de *Recall*. Si se plantea el análisis con respecto sólo a una lista de valores 1 (esos N1 primeros términos) de cara a comprobar si en la lista recuperada todo son valores 1 o no, se tendría una matriz de confusión en la que no habría falsos positivos (FP) y todo serían verdaderos positivos (TP) o falsos negativos (FN, etiquetas de 0 en la lista recuperada donde debería haber un 1). Por este motivo en este escenario sólo se podría calcular el *Recall* y este además daría el mismo resultado que daba en el caso anterior.

Además de la *Precisión* se calcula la métrica *MAP* (*mean average precision*) que pondera los documentos relevantes obtenidos en función de su posición en la lista final. De esta manera, si la lista final tuviese todos los N1 primeros términos con un documento con la etiqueta del término a 1, el valor de MAP sería 1 (caso perfecto). MAP toma valores entre 0 y 1.

$$Precision = \frac{K1}{N1}$$

Siendo K1 el número de documentos con etiqueta 1 de los N1 primeros documentos.

$$MAP = \frac{\sum_{i=1}^N \frac{N_acc(i)}{J(i)}}{N}$$

Con J(i) la posición del documento actual, N el número total de documentos con etiqueta 1 y N_acc(i) el número de sonetos con etiqueta 1 (valor acumulado) hasta ese documento inclusive. Si en la posición *i* el soneto tiene la etiqueta a 0, N_acc(i) vale 0.

Los experimentos se plantean para los siguientes escenarios:

- *Embedding*: BERT | Función de composición: Joint (RL_SEQ) | Métrica: Coseno
- *Embedding*: Word2Vec | Función de composición: Joint (RL_SEQ) | Métrica: Coseno
- *Embedding*: BERT | Función de composición: Joint (RL_SEQ) | Métrica: ICM 1.2
- *Embedding*: Word2Vec | Función de composición: Joint (RL_SEQ) | Métrica: ICM 1.2

El motivo de usar ICM 1.2. es que se trata de un valor para el parámetro de la función ICM que ha funcionado razonablemente bien en diferentes escenarios (Amigó Cabrera & Fresno Fernández, 2018).

De esta forma se analiza cuáles de los casos dan unas métricas buenas (MAP > 0.25) así como si hay diferencias significativas o no para la recuperación de información en función de si se usa una técnica de *embedding* u otra.

A modo de ejemplo, para el término *Vulnerabilidad*, el primer poema que se recupera (y que coincide con una anotación de dicho término a 1) para el escenario 3 (word2vec, joint, coseno) es:

A tu divina frente, ¡oh poderoso
Niño!, una venda con trabajo y arte
tejí de oro y colores, donde parte
retraté de tu triunfo más glorioso.

Allí se muestra atado al victorioso
carro el gran Febo, que la luz reparte,
preso Mercurio, encadenado Marte,
y Vulcano con muestras de celoso.

Ni se pudo librar con los reales
insignias Jove; mal pudiera Psique
resistir, si a estos rinde la fiereza.

Agravan mi prisión mayores males,
siendo fuerza que a un niño sacrifique
mi firme amor, y a un ciego mi belleza.

En la Tabla 22 se muestran los resultados para los 4 escenarios mencionados con las métricas de evaluación Precisión y MAP para las variables que superan el umbral de 0.25. En ella se ven 7 del total de 15 variables utilizadas, lo que representa un total del 0.47%. Cabe mencionar que las variables que han pasado el umbral establecido son variables que aparecen de forma significativa en los sonetos. Así, *Vulnerabilidad* aparece en 98 de los 230 sonetos; sin embargo, otras variables como *Prejuicio* aparecen sólo en 13 de ellos. En la Tabla 20 se incluye como dato adicional si para ese escenario y esa variable el primer soneto recuperado contiene o no la etiqueta de dicha variable (lo que supone la métrica P@1). Este dato es útil ya que un RecSys para pedagogía poética podría recomendar un único poema, y no una lista, de manera que es importante que para el primer soneto haya una buena tasa de acierto.

En el Anexo 9.8 se incluyen los resultados completos para los experimentos aquí incluidos. Cabe mencionar que para algunos términos las tasas de precisión y MAP son muy bajas, principalmente porque los términos anotados sobre el total son muy reducidos. Esto hace que la lista sobre la que se evalúa la precisión sea muy corta y que sea más fácil que no haya términos anotados en ella, dando por ello lugar a valores cercanos a 0.

| variable | n_labels | embeddings | type | metric | precision | P@1 | map |
|-----------------------|----------|------------|-------|--------|-----------|-----|--------|
| Vulnerabilidad | 98 | word2vec | joint | icm | 0.449 | 1 | 0.4649 |
| Vulnerabilidad | 98 | word2vec | joint | cosine | 0.449 | 1 | 0.461 |
| Vulnerabilidad | 98 | bert | joint | cosine | 0.4388 | 1 | 0.4547 |
| Dramatización | 94 | bert | joint | cosine | 0.383 | 1 | 0.4498 |
| Dramatización | 94 | bert | joint | icm | 0.383 | 0 | 0.424 |
| Vulnerabilidad | 98 | bert | joint | icm | 0.4388 | 0 | 0.4209 |
| Dramatización | 94 | word2vec | joint | icm | 0.4043 | 0 | 0.4065 |
| Dramatización | 94 | word2vec | joint | cosine | 0.3723 | 0 | 0.4037 |
| Aversión | 77 | word2vec | joint | icm | 0.4026 | 1 | 0.394 |
| Aversión | 77 | bert | joint | icm | 0.3247 | 1 | 0.3912 |
| Aversión | 77 | word2vec | joint | cosine | 0.3506 | 1 | 0.389 |
| Aversión | 77 | bert | joint | cosine | 0.3636 | 0 | 0.3671 |
| Temor | 70 | bert | joint | icm | 0.3429 | 1 | 0.3652 |
| Temor | 70 | word2vec | joint | icm | 0.3143 | 1 | 0.3433 |
| Temor | 70 | bert | joint | cosine | 0.3429 | 0 | 0.3432 |
| Temor | 70 | word2vec | joint | cosine | 0.3143 | 1 | 0.343 |
| Orgullo | 49 | bert | joint | cosine | 0.2449 | 1 | 0.2628 |
| Ira | 40 | word2vec | joint | cosine | 0.225 | 1 | 0.2476 |
| Orgullo | 49 | bert | joint | icm | 0.2041 | 0 | 0.2469 |
| Ansiedad | 47 | bert | joint | icm | 0.2553 | 0 | 0.241 |
| Ira | 40 | word2vec | joint | icm | 0.2 | 1 | 0.2385 |

Tabla 22 Resultados de los Experimentos 2 y 4: métricas MAP para los distintos escenarios

Así, las variables que pasan las pruebas planteadas y sobre las que se comprueba la validez del sistema de IR para poesía con las anotaciones proporcionadas por POSTDATA son:

- 'Vulnerabilidad', 'Dramatización', 'Aversión', 'Temor'

Con ello, el experimento 4 sirve para corroborar parcialmente la hipótesis 5. Por tanto, se puede afirmar que, para los términos psicológicos mencionados previamente, es posible etiquetar poemas utilizando metadatos coherentes con las ontologías de uso común en los RecSys educativos afectivos, de modo que sean recuperables como objetos de aprendizaje personalizados, haciendo uso de técnicas de NLP.

El siguiente paso es analizar, para el experimento 2, si hay diferencias entre los distintos escenarios para ver si el tipo de *embedding* utilizado y la métrica afectan a la eficiencia del sistema de IR. Este análisis se muestra en las Tablas 23 y 24.

| Word2Vec | BERT | |
|---------------------------|--------------------------|---|
| Precision (median) 0.1765 | Precision (median) 0.129 | Comparison - One-Way (Precision) 0.8954 |
| MAP (median) 0.2045 | MAP (median) 0.1875 | Comparison - One-Way (MAP) 0.8918 |

Tabla 23 Comparativa de escenarios para la métrica coseno

| Word2Vec | BERT | |
|---------------------------|---------------------------|---|
| Precision (median) 0.1765 | Precision (median) 0.2041 | Comparison - One-Way (Precision) 0.8954 |
| MAP (median) 0.2049 | MAP (median) 0.2008 | Comparison - One-Way (MAP) 0.8918 |

Tabla 24 Comparativa de escenarios para la métrica icm

Las métricas MAP y Precisión no son estadísticamente diferentes entre los resultados de BERT y de word2vec (p-valor muy superior a alpha, con alpha 0.05) para ninguna de las métricas (ni con coseno ni con ICM). Así, no se puede rechazar la hipótesis de que las medias de métricas para cada uno de esos dos casos sean iguales. Por este motivo, los *embeddings* contextuales no aparecen como un elemento sensiblemente diferenciador de cara a la construcción de *embeddings* en sonetos de poesía en castellano antiguo.

De esta manera el experimento 2 no sirve para confirmar la hipótesis 3.1 que recogía que debido a las ventajas de los *embeddings* contextuales como BERT (desambiguaciones, capturar relaciones a largo plazo en los textos, multi idioma...), estos aportan más que word2vec para un sistema IR para poesía. Este experimento, como se detalló en el apartado de objetivos, no es suficiente para validarla, pero sería necesario un estudio en más profundidad para ver si descartarla o no.

El estudio se complementa con el análisis de métricas de recuperación por anotador, de cara a comprobar si varían significativamente respecto a la del vector con la mediana. A continuación se muestran los contrastes de hipótesis del análisis ANOVA para ver si hay diferencias significativas en las métricas obtenidas usando las anotaciones individuales de los anotadores con respecto a las del vector de los datos conjuntos.

Como se ve en la Tabla 25, en ningún escenario de combinación de *embeddings* y métricas se obtienen métricas significativamente distintas a las obtenidas con el vector de los valores medianos, salvo para el caso de BERT con coseno y usando los datos del anotador 2. Esto es así porque las métricas de recuperación usando únicamente la información del anotador 2 son algo superiores respecto a usar las del vector mediano, tal y como se ve en la Tabla 26, donde se incluye el Top 10 de las mejores métricas, en las que 8 de ellas pertenecen al anotador 2 (*annotator* = 2) y las otras 2 al vector mediano (*annotator* = 4).

| Scenario | Anotador 1 | Anotador 2 | Anotador 3 |
|------------------|---|---|---|
| Word2Vec, Cosine | Comparison - One-Way 1 vs median Precision 0.4141 | Comparison - One-Way 2 vs median Precision 0.0670 | Comparison - One-Way 3 vs median Precision 0.8814 |
| Word2Vec, Cosine | Comparison - One-Way 1 vs median MAP 0.3810 | Comparison - One-Way 2 vs median MAP 0.0812 | Comparison - One-Way 3 vs median MAP 0.6086 |
| Word2Vec, ICM | Comparison - One-Way 1 vs median Precision 0.4103 | Comparison - One-Way 2 vs median Precision 0.0940 | Comparison - One-Way 3 vs median Precision 0.9227 |
| Word2Vec, ICM | Comparison - One-Way 1 vs median MAP 0.3720 | Comparison - One-Way 2 vs median MAP 0.0657 | Comparison - One-Way 3 vs median MAP 0.6590 |
| BERT, Cosine | Comparison - One-Way 1 vs median Precision 0.5222 | Comparison - One-Way 2 vs median Precision 0.0260 | Comparison - One-Way 3 vs median Precision 0.5873 |
| BERT, Cosine | Comparison - One-Way 1 vs median MAP 0.4646 | Comparison - One-Way 2 vs median MAP 0.0290 | Comparison - One-Way 3 vs median MAP 0.4979 |
| BERT, ICM | Comparison - One-Way 1 vs median Precision 0.5785 | Comparison - One-Way 2 vs median Precision 0.0810 | Comparison - One-Way 3 vs median Precision 0.7183 |
| BERT, ICM | Comparison - One-Way 1 vs median MAP 0.4680 | Comparison - One-Way 2 vs median MAP 0.0521 | Comparison - One-Way 3 vs median MAP 0.5727 |

Tabla 25 Contrastes de hipótesis comparando las métricas de recuperación por anotador frente a las obtenidas con los valores medianos de los 3

| annotator | feature_col | embedding | composition | metric | best_index | n_labels | precision | map |
|-----------|----------------|-----------|-------------|--------|------------|----------|-----------|--------|
| 2 | Vulnerabilidad | word2vec | joint | cosine | 73 | 123 | 0.5935 | 0.6217 |
| 2 | Vulnerabilidad | word2vec | joint | icm | 76 | 123 | 0.5935 | 0.6203 |
| 2 | Vulnerabilidad | bert | joint | cosine | 137 | 123 | 0.5691 | 0.5666 |
| 2 | Dramatización | bert | joint | cosine | 137 | 107 | 0.4953 | 0.5450 |
| 2 | Aversión | bert | joint | cosine | 137 | 110 | 0.4909 | 0.5260 |
| 2 | Aversión | word2vec | joint | cosine | 73 | 110 | 0.4818 | 0.5172 |
| 2 | Aversión | word2vec | joint | icm | 76 | 110 | 0.5000 | 0.5146 |
| 4 | Vulnerabilidad | word2vec | joint | cosine | 73 | 98 | 0.5000 | 0.4968 |

| | | | | | | | | |
|---|----------------|----------|-------|-----|-----|-----|--------|--------|
| 2 | Vulnerabilidad | bert | joint | icm | 193 | 123 | 0.5041 | 0.4961 |
| 4 | Vulnerabilidad | word2vec | joint | icm | 76 | 98 | 0.4796 | 0.4881 |

Tabla 26 Top 10 de las mejores métricas de recuperación usando los datos de los anotadores individuales

El análisis de este apartado se termina con un primer estudio comparativo de los resultados obtenidos con las técnicas de IR frente a otra alternativa para la recuperación de poemas por términos concretos, incluyendo los de dominio psicológico aquí utilizados. Esta alternativa es el uso de modelos de ML supervisado, que se entrenen en unos rasgos representativos de los sonetos y que a través de ellos infieran relaciones con los términos anotados de cara a predecirlos sobre poemas futuros no anotados.

La pretensión de la presente memoria en este punto no es analizar de forma exhaustiva las posibles alternativas de modelos de ML para predecir las etiquetas del dominio psicológico, sino introducir esta posibilidad y realizar una primera comparativa con los resultados obtenidos hasta ahora con técnicas de IR para apuntar a posibles líneas de investigación futuras.

Para ello se utiliza un modelo sencillo de *Random Forest* (RF). Este modelo recibe un vector de rasgos de entrada (predictores) y con esa información tratará de predecir las categorías asociadas al soneto. Para este análisis base se construye un modelo por categoría psicológica, prediciendo así una etiqueta binaria (1 pertenece a esa categoría, 0 no). Como predictores se utilizan los datos de los *embeddings* de los sonetos, tratando de obtener así relaciones entre la información semántica de los poemas y los rasgos psicológicos. Como cada soneto está caracterizado por 4 vectores de *embedding* tras aplicar las funciones de composición (uno por estrofa) se agregan en un único vector por soneto utilizando el valor mediano por componente para los 4 vectores. El *embedding* de base usado es el de BERT. Así, se tiene un único vector por soneto (768 componentes) y las etiquetas psicológicas asociadas.

Para evitar problemas de balanceo de datos debido a la poca proporción de sonetos anotados con 1 frente a los anotados con 0, se aplica un *undersampling* utilizando un subconjunto por modelo en el que el 50% sean sonetos anotados con 1 y el otro 50% sonetos anotados con 0. Los modelos son entrenados usando un 66% de los datos del subconjunto (aleatorios) para predecir el 33% restante (test). Aun definiéndose una semilla, para asegurar unos resultados más estables se ejecutan 100 iteraciones de modelos entrenados con muestras de datos obtenidas de forma aleatoria con la misma semilla de cara a garantizar una estabilidad en los resultados. En cada una de esas iteraciones se coge un subconjunto distinto (aleatoriamente) de train y de test, con el fin de tener unos datos generalizables y estables.

Los parámetros de todos los modelos de RF son los siguientes:

- `n_estimators = 100, max_depth=2` con `random_state=0` de semilla de cara a la reproducibilidad.

En la Tabla 27 se muestran los resultados por variable utilizando como métricas la precisión, el *recall* y *f1* para tener una visión conjunta de ambas por modelo.

| feature_col | precision | recall | f1 |
|----------------|-----------|--------|--------|
| Inseguridad | 0.6667 | 0.6000 | 0.6316 |
| Prejuicio | 0.6000 | 0.6000 | 0.6000 |
| Ilusión | 0.7273 | 0.5000 | 0.5926 |
| Depresión | 0.7143 | 0.5000 | 0.5882 |
| Inestabilidad | 0.4667 | 0.6363 | 0.5385 |
| Ira | 0.5000 | 0.5714 | 0.5333 |
| Temor | 0.4643 | 0.5909 | 0.5200 |
| Impotencia | 0.4348 | 0.6250 | 0.5128 |
| Dramatización | 0.5000 | 0.4688 | 0.4839 |
| Orgullo | 0.5833 | 0.4118 | 0.4828 |
| Ansiedad | 0.3214 | 0.8182 | 0.4616 |
| Obsesión | 0.5000 | 0.4286 | 0.4615 |
| Vulnerabilidad | 0.3429 | 0.4000 | 0.3692 |
| Aversión | 0.5000 | 0.2667 | 0.3479 |
| Desilusión | 0.3333 | 0.0714 | 0.1177 |

Tabla 27 Métricas de los distintos modelos de ML supervisado por categoría psicológica

Una comparativa directa de las métricas de precisión obtenidas por estos modelos respecto a las técnicas de IR utilizadas previamente muestra una clara ventaja de los modelos de ML, obteniendo datos significativamente mejores tras aplicar analizar los resultados del contraste ANOVA, tal y como muestra la Tabla 28. Como comparativa se analiza para el caso de BERT (dado que no había diferencias significativas con word2vec) tanto para ICM como para coseno.

| IR (BERT, coseno) | IR (BERT, icm) | Modelos ML |
|--|--|----------------------------|
| Precision (mean) 0.1978 | Precision (mean) 0.2235 | Precision (mean) 0.5103 |
| Precision (median) 0.129 | Precision (median) 0.2041 | Precision (median) 0.5 |
| Comparison - One-Way (Precision) 8.8169127164739e-07 | Comparison - One-Way (Precision) 1.6774015171150e-06 | |

Tabla 28 Comparativas entre los modelos de ML y los de IR

Hay que tener en cuenta en esta comparativa que los datos sobre los que trabajan las técnicas de IR son el conjunto entero de sonetos (debido a que son técnicas no supervisadas) frente a unos datos balanceados con los que trabajan los de ML. Esto penaliza los resultados que se obtienen con las técnicas de IR, si bien es lo que ocurriría en un escenario real.

A modo ilustrativo se muestra el análisis de un escenario en el que se aplican las técnicas de IR sobre el mismo subconjunto balanceado usado en los modelos de ML. Las nuevas métricas que se obtienen en este escenario ficticio se comparan con las de los modelos de ML de cara a comprobar si hay diferencias significativas. El contraste de estas métricas con los modelos de ML aparece en la Tabla 29.

| IR (BERT, coseno) | IR (BERT, icm) | Modelos ML |
|--|--|----------------------------|
| Precision (mean) 0.50409 | Precision (mean) 0.4980 | Precision (mean) 0.5103 |
| Precision (median) 0.5 | Precision (median) 0.5098 | Precision (median) 0.5 |
| Comparison - One-Way (Precision) 0.8631 | Comparison - One-Way (Precision) 0.7331 | |

Tabla 29 Comparativas con un escenario ficticio de sonetos balanceados por etiquetas

Con los resultados sobre el mismo subconjunto de sonetos, la medida de evaluación (Precisión) para ambas métricas de IR (BERT coseno, BERT icm) no da unos resultados significativamente distintos que los modelos simples (*vanilla*) que se han entrenado. De esta manera se puede deducir que uno de los principales motivos que causaba el bajo valor de Precisión indicada en los apartados previos es el desbalanceo de datos existente (ya que los sonetos etiquetados son muchos menos que los no etiquetados, de manera que esto actúa en detrimento de las técnicas de IR).

Con todo, cabe la posibilidad de que con un modelo ajustado (usando un subconjunto de datos de validación) se puedan llegar a obtener unos resultados significativamente mejores que los de los métodos de IR aun en el escenario teórico de no desbalanceo de clases.

El problema de las técnicas de IR es que son técnicas no supervisadas que no aprovechan la información de la que se dispone (las anotaciones) para hacer el etiquetado, y que necesitan utilizar todo el corpus de datos para dar sus resultados. Por este motivo la comparativa adecuada requiere que las técnicas de IR usen toda la cantidad de sonetos disponibles, y en este escenario los modelos simples de ML parece que aportan resultados consistentemente mejores.

En cualquier caso, convendría explorar en más detalle el uso de modelos de ML supervisado, ajustando adecuadamente los hiper parámetros con datos de validación. También se deberían analizar los resultados no de entrenar modelos independientes sino de entrenar un modelo multiclase en un conjunto de datos balanceado, para ver si este puede capturar relaciones más complejas y aportar con ello mejores predicciones.

Aun así, ambas técnicas pueden darse de manera conjunta. Esto es así ya que se puede etiquetar el corpus entero de sonetos en distintas categorías con los modelos de ML con las etiquetas anotadas de las que se dispone (y ante una consulta de usuario devolver el que tenga el porcentaje de clasificación más alto). Por otro lado, cuando un usuario introduzca una consulta que no responde a una etiqueta, aplicar sobre ello las técnicas de IR para devolver el vector semánticamente más próximo.

6.3. Resultados en AULA

El siguiente estudio presenta los resultados obtenidos durante la feria de AULA 2019⁸⁵ en Madrid en el rincón donde se presentó el proyecto, titulado “*Te receto un soneto*”. Se han analizado los datos de 35 usuarios que voluntariamente accedieron a utilizar la aplicación la mañana del sábado 30 de marzo (desde las 11:00 hasta la 13:00).

⁸⁵ Web. 05/06/2019. http://www.ifema.es/aula_01

El sistema presentaba una sencilla interfaz (indicada en el Capítulo 3) a través de la cual se preguntaba a los usuarios por su estado afectivo y por su estado anímico ofreciendo diferentes opciones de respuesta (podían elegir únicamente una de las respuestas posibles de cada categoría).

Pregunta: "¿Cómo te sientes ahora?"

Respuestas posibles - Grupo 1 [*Valence*]

- "Triste"
- "Irritado/a"
- "Entusiasmado/a"
- "Alegre"
- "Frustrado/a"
- "Seguro/a"

Respuestas posibles - Grupo 2 [*Arousal*]

- "Muy tranquilo/a"
- "Tranquilo/a"
- "Activo/a"
- "Muy activo/a"

Posteriormente, se pedía a los usuarios que escribiesen una palabra para presentarles dos posibles opciones de sonetos, recuperados de un corpus de 4251 sonetos en lengua castellana del S. XV al S. XX, recogidos desde el corpus DISCO de POSTDATA y completados mediante sonetos del S. XX (166 sonetos) obtenidos de Internet.

Los sonetos de la base de datos tenían asociadas, para cada una de las palabras, un *embedding* obtenido gracias al corpus multilingüe de BERT con mayúsculas⁸⁶; dichos *embeddings* se componían a nivel de estrofa utilizando dos técnicas distintas basadas en lo que se describe en el artículo (Amigó Cabrera & Fresno Fernández, 2018). Tras obtener los *embeddings* con el texto en bruto (sin ningún tipo de preprocesamiento), se eliminaron las *stopwords* antes de agregar los valores de los vectores a nivel de estrofas. La agregación se hizo con dos funciones de composición: por un lado usando una función suma (F Sum) que simplemente sumaba los vectores [1x768] del *embedding* de cada una de las palabras. Por otro lado se usó una tipo *joint* con la que se agregaban las palabras usando el método RL_SEQ, agregando secuencialmente por el orden que aparecen en la frase, de derecha a izquierda y de dos en dos con la fórmula que se describe en el artículo citado.

Con dichas funciones se obtienen dos sonetos de entre todos los posibles usando una misma métrica de comparación (similitud basada en el coseno entre el vector de la consulta y el vector de una estrofa) y se devuelve el soneto cuya estrofa sea más cercana a la consulta de entrada (que será el vector correspondiente a una única palabra que introduzca el usuario). Los dos sonetos recuperados se muestran a izquierda y derecha en la interfaz (orden aleatorio cada vez y en orden desconocido para el usuario) para que este elija el que prefiera.

⁸⁶ Fuente: bert-base-multilingual-cased | <https://github.com/google-research/bert/blob/master/multilingual.md>
Última visita: 05/06/2019

Tras la selección el usuario puntúa cómo de bien se ha aproximado el poema seleccionado con respecto a la palabra que ha introducido (de 0 a 5, 5 muy bien y 0 muy mal).

| | query_text | selection_arousal | selection_rating | selection_valence | sonnet_selected |
|---|------------|-------------------|------------------|-------------------|-----------------|
| 0 | guerra | activo | 3 | Frustrado | sum |
| 1 | esfuerzo | muy_activo | 5 | alegre | joint |
| 2 | pabellón | activo | 5 | entusiasmado | joint |
| 3 | cuna\n | muy_activo | 5 | alegre | sum |
| 4 | pasión | activo | 4 | entusiasmado | sum |

Ilustración 89 Ejemplo de datos recogidos por usuario

El foco de este experimento en AULA era doble. Por un lado se pretende analizar la relación que tienen los rasgos afectivos en los sonetos elegidos con respecto a la puntuación dada a dichos sonetos, de cara a poder ver si hay una relación entre ello para poesía en castellano, como parece que la hay para poesía en alemán, tal como se menciona en (Ulrich et al., 2016). Este punto se aborda simplemente analizando la matriz de correlaciones entre los rasgos ya calculados de dichos sonetos con respecto a la puntuación emitida. En caso de haber una correlación, se podrá plantear un sencillo modelo de regresión logística para ver la varianza explicada por cada uno de los rasgos, cuáles han sido más relevantes y si serían capaces de servir para predecir la puntuación final emitida. Teniendo los datos declarados de los usuarios se puede ver también si, en caso de que haya correlación entre el estado declarado y el valor afectivo del soneto, se podría usar dicha información para predecir mejor la calificación de salida. Así, esta primera parte del análisis busca llevar a cabo el experimento 5.

Por otro lado, se busca analizar la propensión de los usuarios a elegir un poema u otro, y con ello ver qué resultados se obtienen para cada función de composición utilizada, y averiguar si los resultados con BERT son similares para poesía en castellano a los que aparecen en (Amigó Cabrera & Fresno Fernández, 2018). Esto se podrá comprobar analizando qué sonetos han sido elegidos por cada usuario, y la tendencia en la puntuación de los sonetos de un grupo respecto a otro. Además, analizando el estado afectivo previo declarado por los usuarios se podrá investigar si dicho estado declarado tiene relación con el contenido afectivo del soneto, en aras a plantear si este podría o no haber influido su elección o si, por el contrario, la elección posiblemente se haya hecho más bien por el contenido semántico (y por lo tanto, por la similitud del contenido del poema con el de la palabra de la consulta de entrada). De esta manera la segunda parte del análisis será llevar a cabo el experimento 6 indicado en el Capítulo 3.

Análisis inicial de datos

Antes de proceder con los experimentos se muestra un análisis descriptivo inicial del conjunto de datos obtenidos en AULA.

En primer lugar, se visualiza qué proporción de usuarios ha elegido sonetos con un tipo de función de composición o con otra. Esto se muestra en la Ilustración 90:

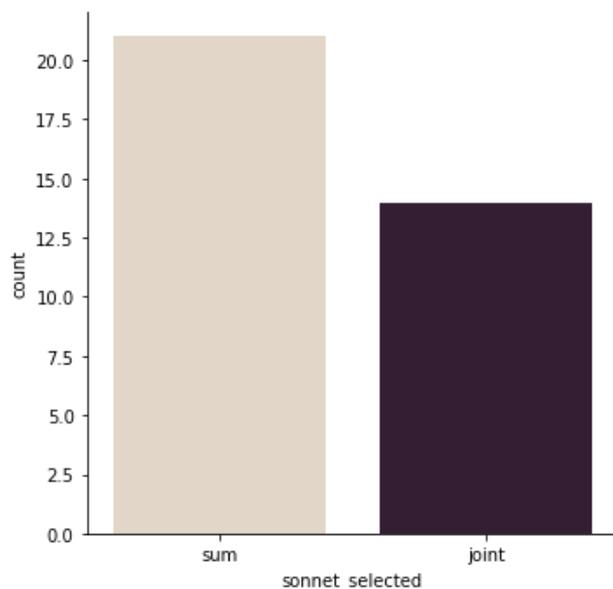


Ilustración 90 Selección según la función de composición

Como se puede apreciar, un mayor número de usuarios ha elegido los sonetos que estaban seleccionados según la función de composición F sum, aunque con una diferencia muy pequeña respecto a los que han elegido los de F joint (21 F sum y 13 F joint).

Tras esto, se prosigue mostrando en la Ilustración 91 un histograma de las selecciones de nivel de excitación:

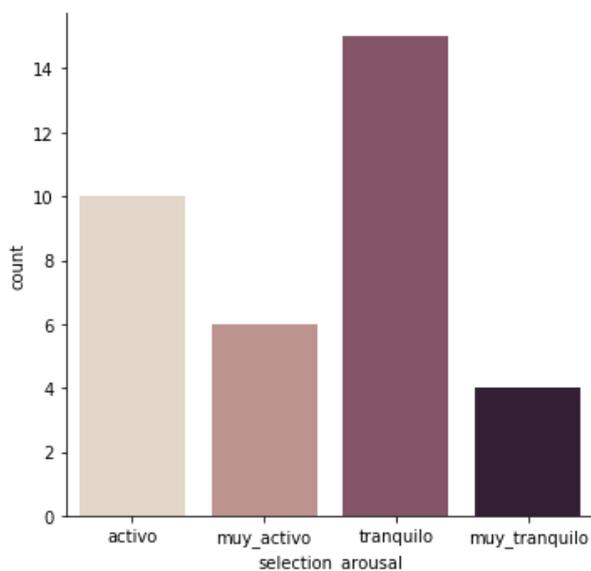


Ilustración 91 Histograma con las selecciones de nivel de excitación; se puede observar cómo los estados intermedios son los más habituales

De igual forma, un histograma para las selecciones de niveles de valencia aparece en la Ilustración 92:

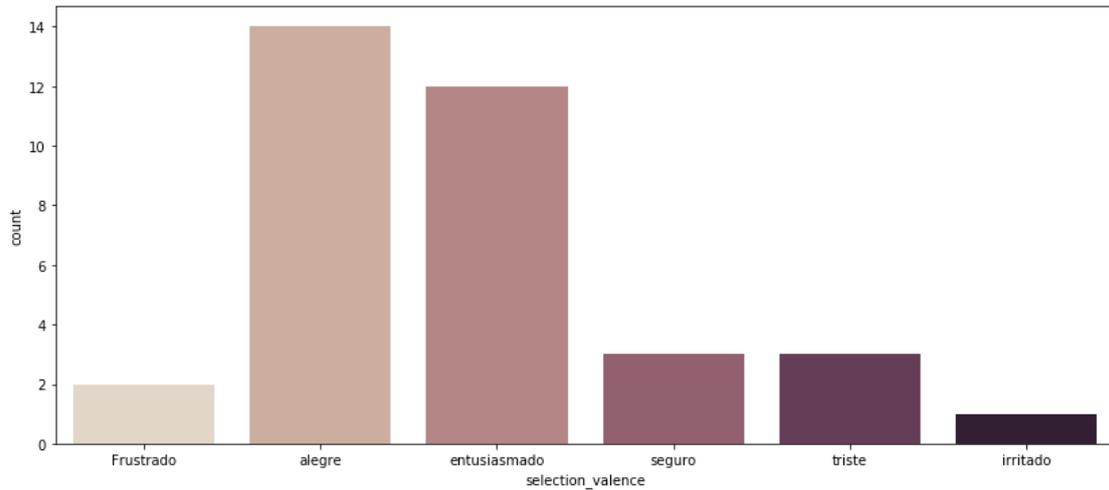


Ilustración 92 Se puede apreciar en las selecciones de valencia cómo los estados claramente más habituales son los de alegría y entusiasmo, habiendo datos muy minoritarios en comparación para los otros casos

La puntuación emitida por los usuarios se muestra en la Ilustración 93:

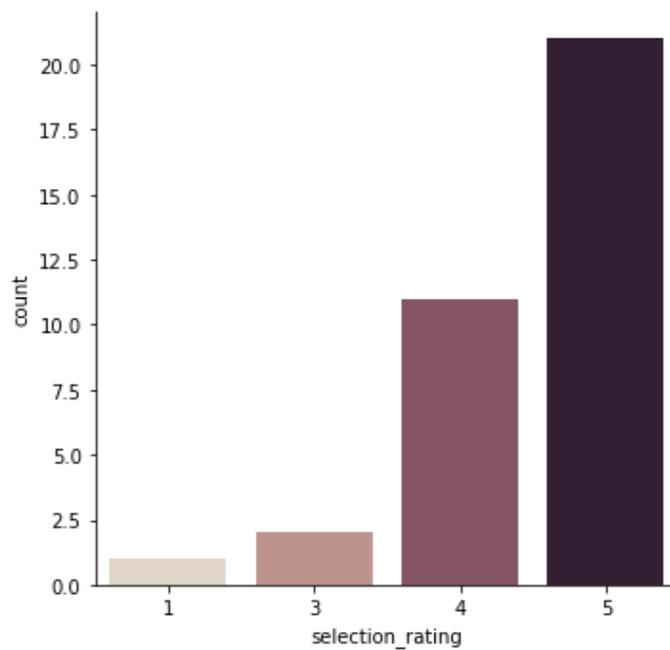


Ilustración 93 Para la puntuación se aprecia en el histograma cómo la mayoría de los usuarios han puntuado a su experiencia con el uso de la aplicación con un 5 o un 4, siendo así la experiencia de uso extremadamente positiva

Previo a comenzar el estudio, se debe asociar un valor numérico a las variables afectivas declaradas por los usuarios (categóricas), partiendo de datos con el formato tabular que aparece en la Ilustración 94.

| | query_text | selection_arousal | selection_rating | selection_valence | sonnet_selected | sonnet_id |
|---|-------------|-------------------|------------------|-------------------|-----------------|-----------|
| 0 | guerra | activo | 3 | Frustrado | sum | 642 |
| 1 | esfuerzo | muy_activo | 5 | alegre | joint | 2475 |
| 2 | pabellón | activo | 5 | entusiasmado | joint | 3463 |
| 3 | cuna\n | muy_activo | 5 | alegre | sum | 3281 |
| 4 | pasión | activo | 4 | entusiasmado | sum | 68 |
| 5 | estrellas\n | tranquilo | 5 | alegre | sum | 1541 |
| 6 | futuro | activo | 5 | entusiasmado | sum | 4125 |
| 7 | libros | tranquilo | 5 | alegre | joint | 1662 |
| 8 | magia | tranquilo | 5 | alegre | sum | 3539 |

Ilustración 94 Ejemplo de los datos introducidos por cada usuario junto con el id del soneto seleccionado correspondiente. Los datos completos aparecen como archivos adjuntos y además hay una referencia a ellos en el Anexo 7.5

Partiendo de los datos con el formato recogido en la Ilustración 94, se vinculan las palabras preprocesadas con los rasgos afectivos que corresponderían a dichas palabras individualmente y se comparan con la media de los valores afectivos de los mismos rasgos de los sonetos seleccionados.

Experimento 5

Con el experimento 5 se busca investigar la hipótesis 1, que plantea que los rasgos afectivos incluidos, y que proceden de diversos estudios sobre poesía en otros idiomas como (Ullrich et al., 2016), son igual de relevantes para el castellano y sirven para predecir su gusto (*liking*) estético.

Esto se comprobará analizando la matriz de correlación entre los rasgos afectivos de los sonetos seleccionados y la puntuación emitida por los usuarios. El análisis se plantea para las variables *happiness*, *fear*, *disgust*, *anger*, *sadness*, para *valence* y *arousal*, y para otros términos como *concreteness*, *context availability* e *imaginability*.

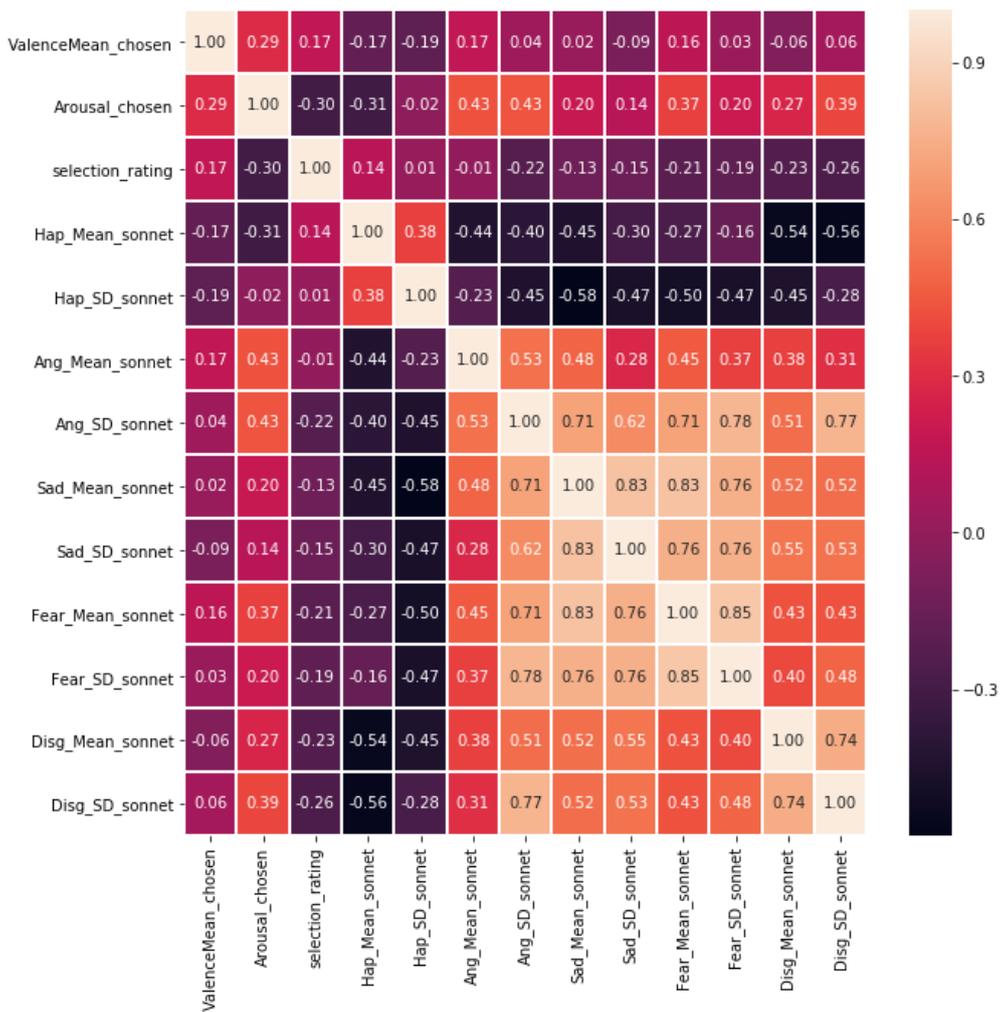


Ilustración 95 Correlación entre los valores declarados, la puntuación emitida y las variables afectivas (I)

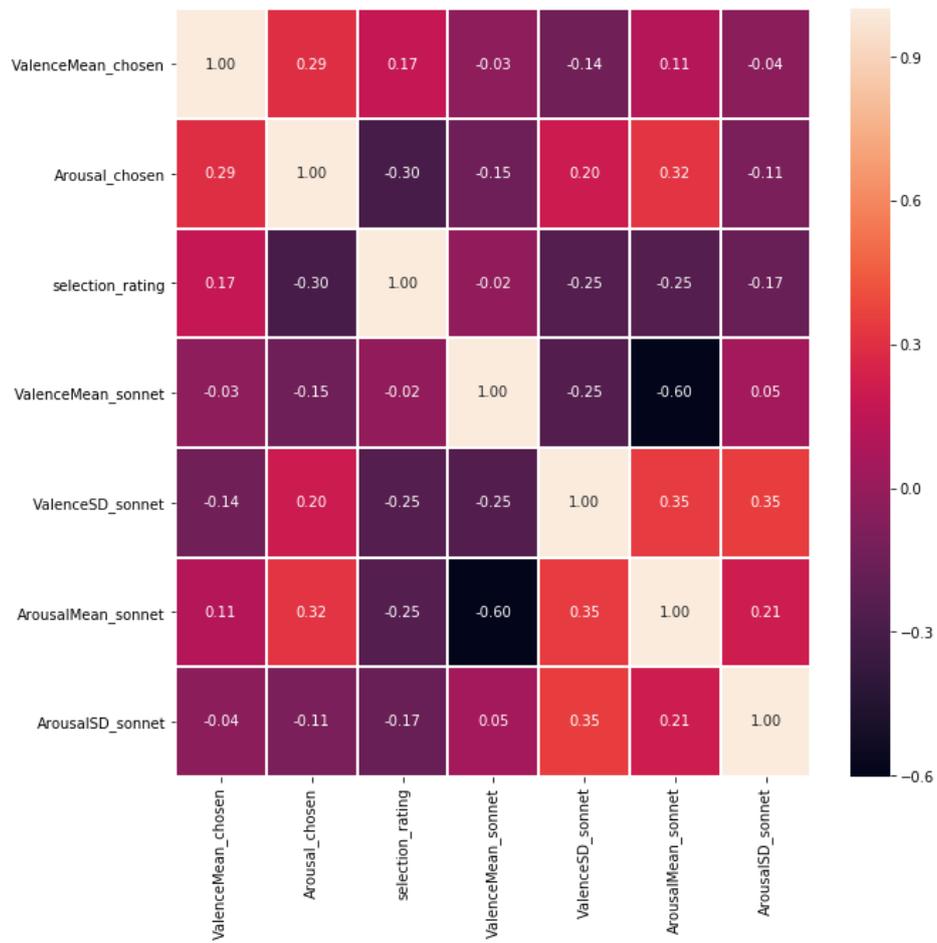


Ilustración 96 Correlación entre los valores declarados, la puntuación emitida y las variables afectivas (II)

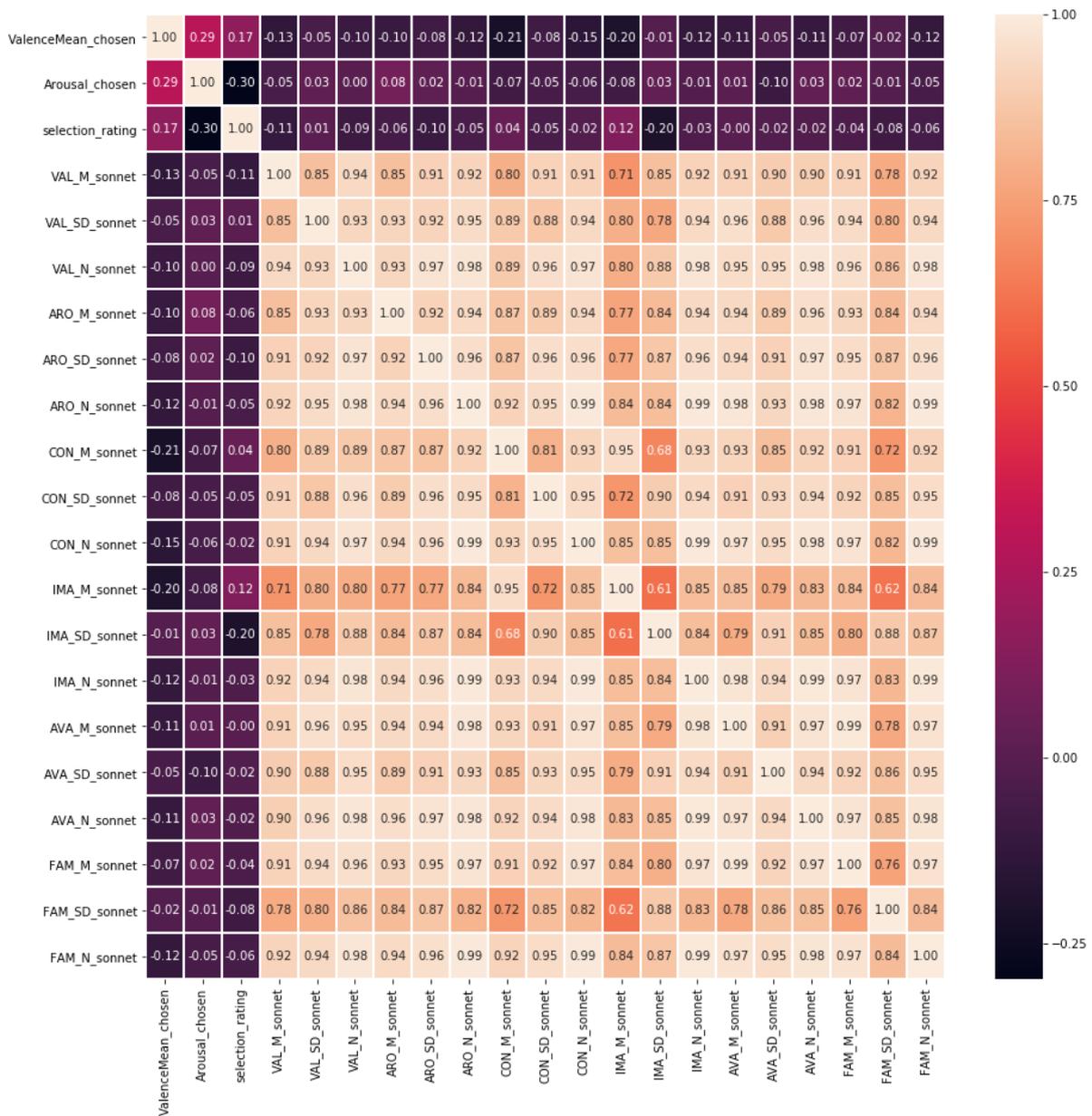


Ilustración 97 Correlación entre los valores declarados, la puntuación emitida y las variables afectivas (III)

En primer lugar, interesa analizar si el estado afectivo declarado por los usuarios ha podido influir significativamente en la puntuación emitida, de cara a ver si la influencia se podría deber puramente al valor del soneto o si hay una predisposición inicial significativa. En cualquiera de las Ilustraciones 95, 96, o 97 se aprecia como la variable de *valence* del estado de los usuarios no influye en la puntuación (0.17), mientras que el estado de *arousal* sí que parece influir (-0.30) condicionado parcialmente los resultados de los usuarios. Aparentemente los usuarios más tranquilos (menor *arousal*) tienden a dar puntuaciones más altas.

Pese a existir este sesgo, se procede al análisis de la correlación respecto a los rasgos del soneto para ver su posible influencia en la puntuación emitida. Como se ve en las Ilustraciones 95, 96 y 97, las siguientes variables tienen una correlación mayor de 0.21 y podrían servir para predecir la puntuación que se emitiría para el poema:

- Anger [desviación]: -0.22
- Fear [media]: -0.21
- Disgust [media]: -0.23
- Disgust [desviación]: -0.26
- Valence [desviación]: -0.25
- Arousal [media]: -0.25

Como se puede observar, todas las correlaciones son negativas, con lo que se ve que en el conjunto de datos disponibles se dan mayores puntuaciones para los sonetos cuanto menor sea el valor de dichas variables.

Lo interesante en este punto es que, dado que hay una correlación entre el estado afectivo declarado para el nivel de excitación (*arousal*) y la puntuación emitida, si se puede plantear un sistema de predicción que en función del contenido afectivo del soneto y del estado del usuario pudiese predecir cuánto le iba a gustar. Para ello se analizan las variables de los sonetos previamente identificadas con correlación con la puntuación de salida para ver en cuáles de ellas hay también correlación con el estado afectivo declarado. Las correlaciones de dichas variables con el nivel de excitación declarado son:

- Anger [desviación]: 0.43
- Fear [media]: 0.37
- Disgust [media]: 0.27
- Disgust [desviación]: 0.39
- Arousal [media]: 0.32

De esta manera, los usuarios con mayor nivel de excitación han introducido palabras de consulta y han recibido sonetos como resultado de ello de más valor para esas variables. Con toda esta información se podría plantear un modelo de predicción que intentase estimar cuánto fuese a gustar a un usuario un soneto en concreto en función de los rasgos afectivos del soneto y su estado afectivo. Así, la información afectiva del usuario si parece que es relevante para ser considerada en las recomendaciones.

Experimento 6

Con el experimento 6 se busca comprobar la hipótesis 3, que sostiene que las conclusiones derivadas del uso de funciones de composición para *word embeddings* que se mencionan en la literatura (Amigó Cabrera & Fresno Fernández, 2018) para word2vec y recuperación de definiciones en inglés son extrapolables a otros *embeddings*, como BERT, y a otro tipo de textos, como es el caso de sonetos, con respecto a una consulta de entrada, debido a que son definiciones axiomáticas.

Se comienza el estudio analizando las correlaciones de las distintas variables involucradas. Debido a que gran parte de ellas son variables categóricas se lleva a cabo un estudio de correlación basado en Spearman. Este análisis aparece en la Ilustración 98.

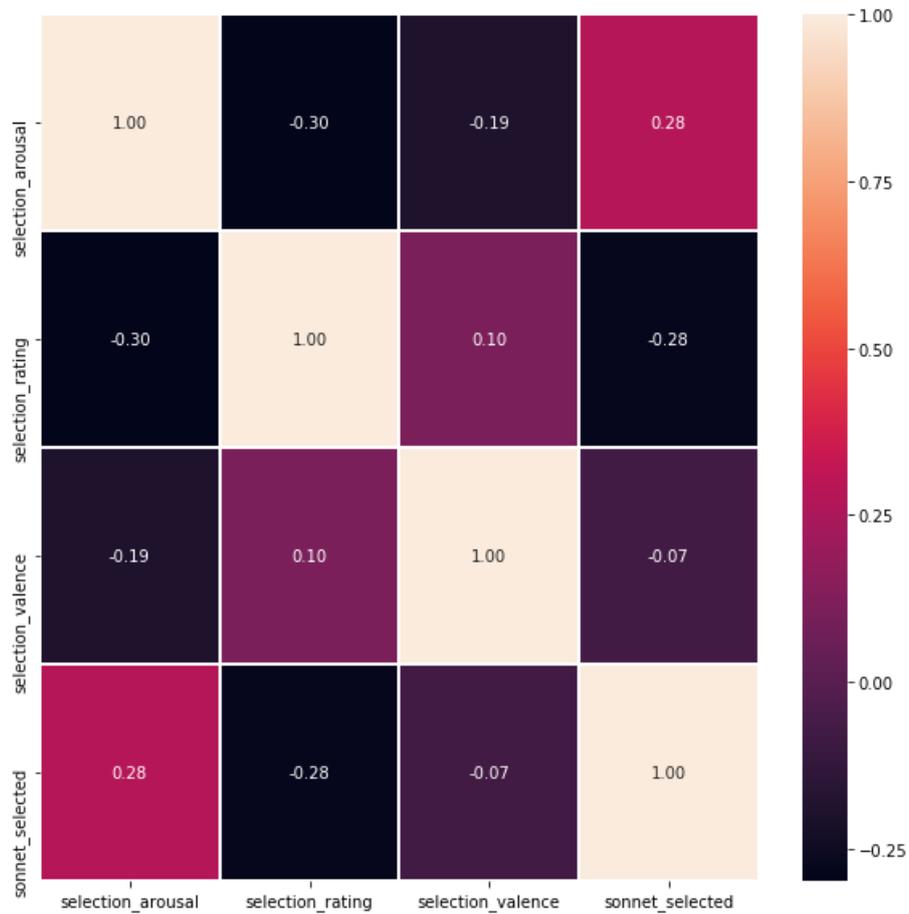


Ilustración 98 Correlación entre las variables seleccionadas por los usuarios

Se puede apreciar que la correlación más significativa se presenta entre *selection_arousal* y *selection_rating* (-0.30), como ya se comentó en el experimento anterior; al parecer cuanto menos excitación (más tranquila está la persona) mejores ratings se suelen dar. Por otro lado, el nivel de valencia de la persona (*selection_valence*) no parece afectar a la puntuación emitida. Aparece una correlación significativa entre la puntuación emitida y la función de composición utilizada (-0.28). Como F joint está representada con una etiqueta 0 y F sum con 1, al parecer puntuaciones más altas van asociadas a F joint.

La última afirmación se puede completar mostrando las distribuciones de puntuación, dependiendo de la función de composición utilizada para la recuperación del soneto elegido, tal y como muestra la Ilustración 99.

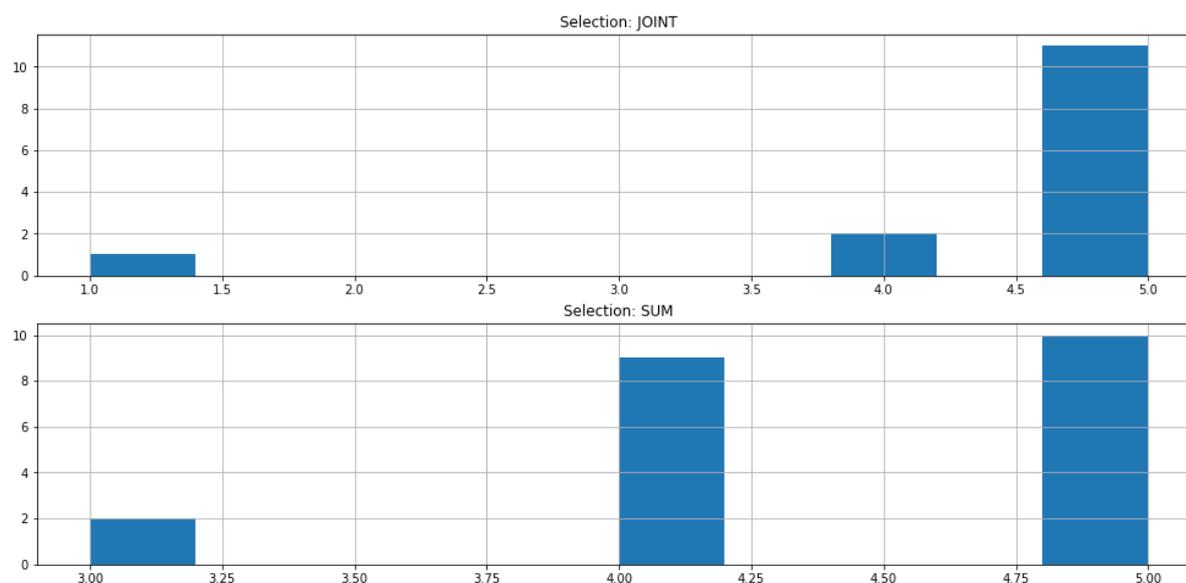


Ilustración 99 Histogramas con las puntuaciones según las funciones de composición.

El reparto de ratings por grupo es más o menos el mismo, si bien es cierto que para el caso de la función tipo *joint* los ratings atribuidos suelen ser mayores. Es decir, aunque se elijan más veces los tipo *sum* cuando se eligen los tipo *joint* suelen darse mejores puntuaciones (esto podría ser porque cuando el usuario no está satisfecho con ninguna de las dos posibilidades quizá tienda más a escoger los tipo *sum*, y luego, eso sí, darles peor puntuación).

Aun así, el análisis exacto se hace mediante un contraste de hipótesis con la prueba t de Student⁸⁷ entre las puntuaciones emitidas por tipo de función de composición. El resultado es el siguiente:

- P-valor = 0.5251

La diferencia entre ratings para los dos grupos no es significativa, ya que 0.53 es mucho mayor p-valor de 0.05 con lo que no se puede descartar la H0 que diría que las medias son iguales. Así, se puede apreciar cómo utilizando una función de composición tipo suma frente a una tipo *joint* se han obtenido resultados muy similares. Esto es a priori coherente con los resultados expuestos en (Amigó Cabrera & Fresno Fernández, 2018) donde los autores muestran cómo la efectividad de un sistema de recuperación de información es prácticamente idéntica cuando se usan funciones de composición tipo suma frente a usar unas tipo *joint* con una agrupación secuencial de derecha a izquierda (RL_SEQ).

⁸⁷ Prueba T de Student. Web. 15/09/2019. https://ast.wikipedia.org/wiki/Prueba_t_de_Student

7. Conclusiones y Líneas Futuras

En este Capítulo se recoge una reflexión final en base a los resultados de los experimentos realizados, para posteriormente indicar posibles líneas de investigación desde las que continuar el trabajo presentado en esta memoria.

7.1. Conclusiones

En la presente memoria se ha reportado una primera investigación en el ámbito de la pedagogía poética y su aplicación a través de RecSys sensibles al contexto afectivo para entornos de e-learning. La motivación principal es incorporar en e-learning un enfoque formativo integral en el que la dimensión afectiva del estudiante sea una parte activa de su proceso formativo, usando para ello la poesía.

Al ser un ámbito novedoso y sobre el que no hay prácticamente investigación publicada, hasta donde se ha podido comprobar, la presente memoria ha detectado las áreas principales de investigación desde las que comenzar a abordar la pedagogía poética en e-learning. Estas incluyen un análisis inicial de qué componentes debe incluir un RecSys afectivo para entornos de e-learning donde se incorporen poemas como LO. El análisis recoge una propuesta inicial de estándares y ontologías para utilizar en cada uno de los componentes de cara a facilitar su interoperabilidad.

Este RecSys afectivo recomienda poesía personalizada a los estudiantes según el contexto afectivo de un corpus de poesía. Para la investigación reportada en esta memoria, se ha usado el corpus de sonetos en castellano de DISCO ampliado con sonetos del S. XX. Este corpus ha sido parcialmente anotado por expertos de POSTDATA con términos psicológicos y rasgos afectivos, con el fin de ilustrar la propuesta con un prototipo de RecSys educativo afectivo para el aprendizaje de trastornos de personalidad.

La recomendación de un poema, un soneto en este caso, para un estudiante se realiza según el contenido semántico y afectivo de dichos textos. Por ello, se plantea el uso de técnicas de NLP y minería de texto para modelar semánticamente poesía, así como el uso de fuentes de información adicionales, para modelar afectivamente los sonetos.

La búsqueda del modelado semántico de los sonetos ha implicado investigar el uso de *word embeddings* y funciones de composición de textos en el dominio de la poesía en castellano.

1. En lo que respecta a la investigación de *word embeddings* se han comparado técnicas clásicas como word2vec con técnicas más novedosas como los *embeddings* contextuales de BERT, analizando cómo influyen en la recuperación de poemas ante una consulta de entrada al sistema y concluyendo que no hay diferencias significativas entre ambos métodos. Esto se ha comprobado tras analizar para ambos *embeddings* la recuperación de los poemas anotados con las categorías psicológicas de POSTDATA, usando como entrada el propio término psicológico y comprobando si los sonetos recomendados están o no etiquetados con dichas categorías.
2. Para la investigación y análisis de distintas funciones de composición se ha desarrollado una aplicación que ha sido usada con usuarios reales en la feria de AULA de IFEMA.

La literatura que se ha tomado como referencia para las funciones de composición propone técnicas axiomáticas que deben dar resultados similares independientemente del dominio en que se utilice. Por ello se han comparado los resultados para dos funciones de composición, usando distintos *embeddings* de los experimentados en la literatura de referencia (BERT en el caso de esta memoria), y aplicándose sobre el dominio de los sonetos en castellano. Los resultados obtenidos son coherentes con los presentados en la literatura, por lo que el análisis experimental corrobora las afirmaciones de partida.

En lo que respecta al modelado afectivo ha servido para investigar sobre la obtención de rasgos afectivos que caractericen un soneto, tanto usando información supervisada (las anotaciones de POSTDATA) como usando diversos *corpora* públicos que ofrecen información afectiva de palabras individuales.

1. En este ámbito, el primer punto que se ha investigado es si es posible construir rasgos afectivos para un soneto en castellano en base al contenido afectivo de las palabras individuales que lo componen. Para ello se han inferido rasgos generales en base al peso afectivo de las palabras individuales gracias a los *corpora* disponibles y se han comparado los resultados con las anotaciones proporcionadas por POSTDATA. El resultado ha servido para validar parcialmente la construcción de rasgos afectivos con esta aproximación: para determinados rasgos si se ha podido caracterizar un soneto desde sus palabras individuales, mientras que para otros no se ha podido asegurar con los experimentos llevados a cabo en este TFM.
2. Además, mediante la aplicación desarrollada para AULA mencionada previamente, se ha investigado si es posible predecir el impacto que un poema va a tener en un usuario en base a los rasgos afectivos que modelan el poema. Esto se ha llevado a cabo en la literatura para otros idiomas, y en este TFM se ha experimentado en el dominio de los sonetos en castellano. El resultado de los experimentos ha servido para ver que determinados rasgos afectivos si sirven para predecir la puntuación que los usuarios van a dar a un soneto. Es más, el estado afectivo del usuario también está relacionado con los rasgos afectivos del soneto que elige además de con la puntuación que va a emitir, de manera que se puede considerar la información afectiva declarada por un usuario para ajustar la poesía que se le recomienda.
3. Por último, se ha investigado la potencial influencia de la poesía desde su contenido afectivo en determinados rasgos psicológicos, en línea con la literatura psicológica que afirma la influencia de la poesía en la psicología. Se ha analizado para ello la correlación entre rasgos afectivos y rasgos psicológicos anotados por POSTDATA, lo que ha servido para dar una primera confirmación de que, efectivamente, aparece una relación entre contenido afectivo y rasgos psicológicos. Esto sugiere la idoneidad del enfoque de la pedagogía poética para la enseñanza de la psicología clínica, dada la potencialidad de la poesía para comunicar tanto el significado como la afectividad características de los diferentes trastornos mentales.

Así, la presente memoria presenta unas primeras investigaciones sobre la pedagogía poética y su integración en RecSys afectivos para entornos de e-learning, identificando algunas de las principales áreas de investigación y proporcionando algunos resultados empíricos utilizando tanto datos anotados como usuarios reales. Como resultado de este TFM se proporciona

también un conjunto de datos anotado por expertos en el dominio poético que abre la posibilidad de líneas de investigación futuras.

7.2. Líneas futuras

En este subapartado se detallan las posibles líneas de investigación con las que continuar a partir del trabajo expuesto en la presente memoria. Estas líneas de investigación se dividen en tres grupos. Los dos primeros grupos guardan relación directa con lo expuesto en esta memoria. El primero de ellos recoge las líneas derivadas de la investigación realizada en base a la minería de textos y NLP sobre textos poéticos. El segundo grupo engloba las líneas de investigación derivadas del estudio realizado para los RecSys afectivos, especialmente en su aplicación a entornos de e-learning. Por último, se incluye una línea adicional en la que se apunta a la investigación que se puede realizar en el ámbito de los sistemas éticos de IA, especialmente en lo relacionado con las áreas de IA cubiertas en esta memoria: *embeddings* y RecSys aplicados a entornos de e-learning.

7.2.1. Minería de textos y NLP en poesía

La presente memoria cubre diversas áreas relacionadas con la minería de textos y la aplicación de técnicas de NLP al ámbito poético. A este respecto sugiere líneas de investigación relacionadas con el uso de *embeddings* y funciones de composición, la estandarización de resultados de mediante el uso de ontologías apropiadas, la investigación más detallada de la aplicación de la minería de textos para castellano antiguo, y el estudio multidisciplinar de los resultados obtenidos en los diversos experimentos, en particular, con la colaboración de expertos de los dominios psicológico y lingüísticos.

Investigación adicional en embeddings y funciones de composición

En primer lugar, recordando el artículo (Amigó Cabrera & Fresno Fernández, 2018), existen otras funciones de composición que se pueden utilizar para abordar los análisis recogidos en la presente memoria. Así, sería interesante repetir ciertos experimentos pero utilizando funciones de composición más complejas y que recojan la estructura sintáctica de la frase en la construcción del vector de composición. Principalmente esto servirá para ver si los resultados mejoran a los expuestos en este trabajo, y también para seguir explorando si los resultados que aparecen en el artículo citado para el ámbito de la recuperación de definiciones en inglés son plenamente extrapolables al ámbito de recuperación de sonetos en Castellano.

Junto con ello, se abre también la continuación en la investigación del uso de modelos de ML supervisados para la predicción de las categorías anotadas por POSTDATA, en base a la información que contienen los textos poéticos. En el ejemplo sencillo que se ha mostrado en esta memoria se han utilizado los rasgos semánticos derivados del vector de composición de los sonetos, pero los predictores podrían ampliarse y considerarse también la información

afectiva y sintáctica de los sonetos para predecir las etiquetas, más aún tras haberse comprobado en los experimentos realizados que algunos rasgos afectivos tienen correlación con los rasgos psicológicos anotados por POSTDATA. Con ello, se dispondría de un sistema que etiquetase nuevos sonetos con las categorías anotadas por POSTDATA, permitiendo que para categorías no etiquetadas se pudiesen usar las técnicas indicadas de IR.

Entrenar un modelo de ML supervisado en el subconjunto de sonetos para etiquetar el resto del corpus es una opción, pero también se pueden explorar técnicas de *semi-supervised learning* que fomenten el uso de todo el corpus para el entrenamiento del modelo. Un ejemplo de ello sería la aplicación previa de algoritmos de *clustering* para agrupar todos los sonetos dentro de las categorías identificadas infiriendo, para cada clúster, los rasgos anotados por POSTDATA en función de los sonetos usados que estén en los *clústers* y aplicando esas etiquetas al resto de sonetos no anotados.

Además de todo ello, como se indicó en el Capítulo 3, la comparativa se hace a nivel de estrofas de cada soneto compuestas en un único vector con respecto a los mensajes de entrada (compuestos siempre en un único vector), devolviéndose el soneto con la estrofa semánticamente más similar al mensaje de entrada. Se podría investigar más en esta línea y probar si la recuperación mejora si se considera también la información relativa semántica del resto de estrofas del poema que se va a seleccionar (aplicando por ejemplo unos pesos de corrección sobre el valor de la métrica final en función de la similitud semántica con respecto al resto de estrofas).

Estandarización de poemas

En el presente TFM se han utilizado poemas más recientes para enriquecer el corpus de sonetos disponibles. Ahora bien, estos últimos se han incorporado al repositorio como ficheros JSON con los campos relevantes para la aplicación, pero no se han recopilado con el esquema de ontología en TEI XML que tienen los demás sonetos, procedentes de DISCO. Una línea de trabajo a desarrollar en este punto es generalizar e integrar en la herramienta el trabajo de etiquetado llevado a cabo en DISCO para que, de manera automática, se pudiesen expresar en dicho formato TEI XML (enriquecido ya con los campos afectivos relevantes) cualquier nuevo soneto que se incorporase al sistema. Esto serviría para que si este prototipo fuese parte de una herramienta de autor, los autores pudiesen subir nuevos sonetos al sistema y que estos pasasen a estar expresados con el mismo esquema ontológico, generado de manera automática.

Junto con ello se puede estudiar la posibilidad de anotar los rasgos afectivos de los sonetos usando ontologías específicamente pensadas para dicha tarea, como la ya mencionada EmotionML, tratando de conseguir con ello una mayor interoperabilidad de los LO con otros sistemas de e-learning.

Anotación de palabras en castellano antiguo

Otra línea futura con la que trabajar es la anotación de las palabras en castellano antiguo, de cara a aumentar el porcentaje de palabras de los textos presentes en los corpus afectivos y refinar la composición de rasgos afectivos a nivel del texto completo. Esto se podría realizar

de dos maneras. Una manera sería mediante la anotación manual de dichas palabras en castellano antiguo con el valor afectivo correspondiente. Otra manera, más sencilla, sería mediante el uso de diccionarios de sinónimos que mapeen las palabras no presentes en el corpus a sus sinónimos que sí están presentes.

Relación entre afectividad y psicología

Como han mostrado los experimentos realizados, existe una correlación clara entre determinados rasgos afectivos y etiquetas psicológicas que caracterizan el texto de un soneto en castellano. Es interesante seguir profundizando en estas correlaciones en el contexto de una investigación multidisciplinar que implicase a expertos de psicología y de lingüística.

7.2.2. RecSys afectivos

Junto con la minería de textos, el otro ámbito principal de esta memoria es el uso de RecSys de poesía en particular para entornos de e-learning, teniendo en cuenta el contexto afectivo. Por este motivo en este apartado se detallan posibles continuaciones de investigación y desarrollo relacionadas con el ámbito de los RecSys afectivos para la pedagogía poética. Así, se mencionarán líneas de investigación para la aplicación de la pedagogía poética en el dominio psicológico mediante su uso en cursos, el desarrollo de sistemas interactivos que respondan mediante sonetos, y de manera general el desarrollo de herramientas de autor para la creación de cursos que incluyan la pedagogía poética.

Desarrollo de un Sistema de Recomendación sensible al contexto afectivo para aplicaciones de e-learning

Una línea de trabajo adicional es el desarrollo de un RecSys para entornos afectivos de e-learning que integre la pedagogía poética y que esté aplicado a un dominio concreto. Un dominio inicial, coherente con la investigación expuesta en esta memoria, es el dominio psicológico, estudiándose el impacto que la poesía tiene en el aprendizaje de determinados conceptos de esta área, en particular para los anotados por POSTDATA, derivados de (García Franco & Manjarrés Riesco, 2016).

Para ello, además de desarrollar el sistema completo, partiendo de los resultados de este TFM se analizaría su repercusión sobre estudiantes de psicología. El desarrollo implicaría las siguientes fases:

1. Implementación de las ontologías del resto de modelos. En este TFM se han planteado propuestas para las ontologías de los modelos de usuario y de dominio en aras de facilitar la interoperabilidad con sistemas de e-learning. Por este motivo, el siguiente paso sería crear dichas ontologías y automatizar su instanciación con los datos pertinentes de usuarios, LO...
2. Tras disponer de un repositorio de información de usuarios, LO... así como de técnicas para incorporar en ellos nuevas instancias, implementación del modelo pedagógico y

las técnicas de recomendación que relacionasen usuarios con objetivos y con una potencial meta u objetivo, en este caso pedagógico.

3. Desarrollo de una interfaz adaptativa que pudiese adecuarse a las particularidades del contexto y del usuario.

Talk to Sonnets

Como se mencionó en la introducción del TFM, una línea de trabajo muy interesante que se ha comenzado con el desarrollo del prototipo de este TFM es el diseño de una aplicación similar a *Google Talk To Books*⁸⁸ pero en la que el diálogo diese como resultado un determinado soneto (e incluso mostrase qué estrofa es la más relevante semánticamente hablando). Esto se podría implementar con el sistema actual, ya que la similitud semántica se evalúa con las estrofas individualmente (por ser ese el nivel de granularidad escogido), de manera que se podría fácilmente indicar cuál de ellas ha sido la más significativa.

Según los resultados de los experimentos con las anotaciones afectivas de POSTDATA, la recomendación de sonetos podría ser distinta según el contexto afectivo detectado en el usuario (en su mensaje de entrada) y el contenido afectivo del soneto.

Con los experimentos llevados a cabo en AULA no es suficiente para investigar en la potencial utilidad de esta herramienta. Habría que desarrollar experimentos adicionales que estudiaran si se tiende a puntuar mejor un poema que tuviese en cuenta determinados criterios de similitud afectiva frente a uno que no los tuviese. Para esto sería necesario incluir un conocimiento de dominio experto que indicase que relaciones afectivas se deberían incluir en el experimento (por ejemplo, ¿recomendar más poesía alegre cuando se detecta que el usuario está triste por el mensaje introducido, o recomendar contenido que tenga valores afectivos similares?).

Herramienta de autor

Otro desarrollo a plantear en relación con el anterior es abstraer el RecSys planteado para definir una herramienta de autor que pueda ser utilizada por personal docente para instanciar un SAE particular basado en pedagogía poética.

Dicho SAE se definiría en base a estándares de e-learning que garantizarían la interoperabilidad con otros sistemas, de manera que se podrían definir tanto cursos en los que el enfoque pedagógico estuviese determinado íntegramente por la pedagogía poética, como módulos complementarios de SAE que posibiliten otros usos de la poesía (por ejemplo, refuerzos motivacionales) en función del contexto afectivo del estudiante.

Investigaciones adicionales

Por último, se puede continuar el estudio de la influencia de los estados afectivos declarados por un usuario de cara a ajustar con ellos la recomendación personalizada de poesía. Para ello

⁸⁸ “Talk to Books”. Google. Web. <https://books.google.com/talktobooks/> 22/08/2019

se podría realizar un estudio análogo al de AULA con un grupo de control más amplio para ver si los resultados son coherentes con los mostrados en la presente memoria.

7.2.3. Trusted AI

Otra posibilidad interesante, siguiendo lo planteado por (Kerly, Hall, & Bull, 2007) que en línea con lo que plantea la nueva ley de protección de datos europea (LOPD) que entró en vigor el 25 de mayo de 2018⁸⁹ dentro del Artículo 22 (“*Article 22. Automated individual decision making, including profiling*”) es incluir técnicas relacionadas con el ámbito del Trusted AI para garantizar que los SAE respondan a las necesidades éticas planteadas. Para ello, se debe investigar y plantear soluciones al menos en dos áreas especialmente importantes: la eliminación de sesgos en los algoritmos de IA (principalmente en los algoritmos de aprendizaje automático que pueden estar trabajando con datos sesgados) y el diseño de modelos de IA transparentes que permitan a sus distintos interlocutores comprender el proceso de toma de decisiones del sistema, y qué datos suyos están siendo utilizados y cómo.

En lo que respecta al primer punto, cuando se trabaja con datos y en base a ellos se infieren decisiones, relaciones o conclusiones, existe la posibilidad de que los datos de origen posean algún potencial sesgo que se arrastre hacia las conclusiones obtenidas y que conlleve una toma de decisiones discriminatoria hacia algún determinado colectivo. Aunque esto es habitual especialmente en el uso de algoritmos de aprendizaje automático (clásicamente en los algoritmos supervisados donde, por ejemplo, la clasificación en una determinada categoría va a estar muy influida por dicho sesgo en los datos) también aparece en escenarios como, por ejemplo el de este TFM. Esto es así ya que, en la literatura, se ha detectado ampliamente cómo el uso de *word embeddings* puede conllevar presencia de determinados sesgos dependiendo del sesgo existente en el corpus de textos utilizado de base (Garg, Schiebinger, Jurafsky, & Zou, 2018; Park, Shin, & Fung, 2018). Es más, aunque los artículos mencionados hablen de sesgos especialmente presentes con *embeddings word2vec*, esta problemática parece que se ha detectado también recientemente en el uso de *embeddings* contextualizados, como en el caso de ELMo (Basta, Costa-jussà, & Casas, 2019; Zhao et al., 2019).

Por este motivo, es crucial detectar y tratar esta problemática. De lo contrario podrían ocurrir situaciones como, por ejemplo, que se quisiese recuperar poesía relacionada con el *trabajo* y se estuviesen recuperando al hacerlo más habitualmente poemas donde se hablase del trabajo de hombres frente a poemas que hablasen del trabajo de mujeres. Esto podría ocurrir si en el corpus de origen la relación semántica entre, por ejemplo, *ingeniero* y *trabajo* fuese mayor que la de *ingeniera* y *trabajo*, incurriéndose así en un sesgo de género.

Para el caso del aprendizaje supervisado existen diversas técnicas que, de manera general, sirven para eliminar sesgos en el conjunto de datos antes de entrenar los modelos, durante el entrenamiento del mismo o para mitigar los sesgos en modelos entrenados ya en datos sesgados, como se detalla ampliamente en (Bellamy et al., 2018). Además de esto, la comunidad de ML también ha investigado tanto en la eliminación de sesgos existentes en el caso concreto de *word embeddings* (Bolukbasi, Chang, Zou, Saligrama, & Kalai, 2016; Park et

⁸⁹ Fuente: <https://www.protecciondatos-lopd.com/empresas/reglamento-europeo-proteccion-datos-2018/> Última visita: 02/06/2019

al., 2018) como en las técnicas adecuadas para asegurar que el entrenamiento de los *embeddings* no va a acarrear sesgos (Zhao, Zhou, Li, Wang, & Chang, 2018), de manera que aparecen soluciones tanto para mejorar el entrenamiento de los *embeddings* como para mitigar el sesgo ya existente en *embeddings* entrenados.

Se abre así una interesante propuesta de investigación para comprobar si existen sesgos en el uso de un *embedding* contextualizado como BERT para el castellano y comprobar específicamente si se propagan los sesgos para los sonetos utilizados en este corpus y si es posible mitigar dichos sesgos en caso de ocurrir.

Con respecto al segundo punto, se debe identificar para cualquier sistema que use técnicas de IA quiénes serán sus interlocutores para, en primera instancia, analizar los niveles de transparencia que será necesario especificar y abordar (Theodorou, Wortham, & Bryson, 2017). Ello permitirá aplicar las técnicas de IA explicable (xAI) adecuadas para cada caso, que pueden pasar por explicaciones globales sobre cómo funciona el sistema y qué criterios y qué relaciones ha inferido desde los datos disponibles, como se detalla en el estudio realizado en (Guidotti et al., 2019), o pueden implicar explicar determinadas decisiones en particular e incluso detallar qué tendría que haber ocurrido para que la decisión tomada por el sistema hubiese sido otra (explicaciones contrafactuales), algo que se puede conseguir con técnicas de explicabilidad local como LIME, *Local Interpretable Model-agnostic Explanations*, (Ribeiro, Singh, & Guestrin, Aug 13, 2016). Este último caso sería relevante para que, por ejemplo, el RecSys afectivo explicase al estudiante el porqué de la recomendación de un determinado soneto en particular. Esta justificación podría ser, bien por el contexto afectivo detectado del estudiante (por ejemplo, se detecta un estado de cansancio y se quiere reforzar el aprendizaje con un soneto de nivel de excitación elevado) o bien justificando, para un escenario como el del presente TFM, qué información en concreto ha sido relevante de la consulta de entrada para elegir un determinado soneto y no otro como el candidato idóneo ante ella. Soluciones como la mencionada de LIME han servido para que en aplicaciones de NLP se pudiese ver qué palabras han sido las más relevantes en una determinada toma de decisión, pero para aplicaciones como los IR (donde no se usa un algoritmo de aprendizaje supervisado) o aplicaciones de aprendizaje no supervisado, el campo de investigación tiene por delante mucho más por descubrir. De igual manera, para sistemas IR como el del presente TFM, donde se usan funciones de composición para todos los sonetos y mensajes de entrada, incorporar explicaciones para por qué recomendar un soneto ante una determinada consulta de manera relevante y comprensible para el usuario, es algo aún más complejo. Por este motivo también este trabajo abre una vía de investigación potencialmente interesante en relación con la aplicación de técnicas de xAI en RecSys afectivos para entornos de e-learning donde se trabaja con *embeddings* y técnicas de composición de textos.

Para finalizar este apartado se menciona que el diseño de soluciones que respondan a algunos de los retos planteados de *Trusted AI* en entornos de e-learning no es algo del todo nuevo. Un ejemplo de ello son los casos donde se implementa un sistema en el que el modelo del usuario sea accesible al propio estudiante, como también hace ELM-ART (Weber & Brusilovsky, 2016), y que incluso puede entrar en debate con el propio estudiante sobre las cuestiones que el programa ha inferido para que el estudiante le intente corregir si no está de acuerdo con alguno de los aspectos inferidos. Sería interesante por ello plantear una vía de diálogo y de debate para que el sistema no sólo le explique qué estados afectivos ha inferido, sino que el estudiante pueda intentar modificarlos con los estados afectivos que él mismo siente. Así, no sólo se conseguirá perfilar mejor al estudiante con ese feedback que se obtiene explícitamente desde él sino que también se fomentará que el estudiante haga introspección y descubra de

forma reflexiva cómo se siente en ese momento gracias a esa mayéutica que se lleva a cabo con el sistema. Junto con ello se podría incluir dentro del sistema la posibilidad de que explique cómo ha inferido concretamente el modelo de usuario con técnicas como las que se plantea en (Datta, Sen, & Zick, 2016) usando métricas como la QII (*Quantitative Input Influence*), con las que se busca capturar el grado de influencia de los inputs en los outputs del sistema, y con ello diseñar informes de transparencia sobre cómo funcionan los sistemas de decisión.

8. Referencias

1. Abaalkhail, R., Guthier, B., Alharthi, R., & El Saddik, A. (2018). Survey on ontologies for affective states and their influences. *Semantic Web*, 9(4), 441-458. doi:10.3233/SW-170270
2. Aggarwal, C. C. (2016). *Recommender systems*. Cham ; Heidelberg ; New York ; Dordrecht ; London: Springer.
3. Alonso, M., Fernandez, A., & Díez, E. (2015). Subjective age-of-acquisition norms for 7,039 spanish words. *Behavior Research Methods*, 47(1), 268-274. doi:10.3758/s13428-014-0454-2
4. Al-Yahya, M., George, R., & Alfares, A. (2015). Ontologies in E-learning : Review of the literature. *International Journal of Software Engineering and its Applications*, 9(2), 67-84. Retrieved from <http://www.earticle.net/Article.aspx?sn=242009>
5. Amigó Cabrera, E., & Fresno Fernández, V. (2018). An axiomatic approach to unsupervised distributional semantic compositionality.(Manuscript in preparation)
6. Andújar, J. C. (2016). *Saber docente desde la pedagogía poética*
7. Arevalillo-Herraez, M., Arnau, D., & Marco-Gimenez, L. (2013). Domain-specific knowledge representation and inference engine for an intelligent tutoring system. *Knowledge-Based Systems*, 49, 97-105.
8. Barros, L., Rodriguez, P., & Ortigosa, A. (Sep 2013). (Sep 2013). Automatic classification of literature pieces by emotion detection: A study on quevedo's poetry. Paper presented at the pp. 141-146. doi:10.1109/ACII.2013.30
9. Basta, C., Costa-jussà, M. R., & Casas, N. (2019). Evaluating the underlying gender bias in contextualized word embeddings. Retrieved from <https://arxiv.org/abs/1904.08783>
10. Beck, J., Stern, M., & Haugsjaa, E. (2016). Applications of AI in education. *Crossroads*, 3(1), 11-15. doi:10.1145/332148.332153
11. Bellamy, R. K. E., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., et al. (2018). AI fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias. Retrieved from <https://arxiv.org/abs/1810.01943>
12. Bernd J. Krämer. (2001). *Personalized curriculum builder in the federated virtual university of the europe of regions* Unpublished.
13. Bolukbasi, T., Chang, K., Zou, J., Saligrama, V., & Kalai, A. (2016). *Man is to computer programmer as woman is to homemaker? debiasing word embeddings* Retrieved from https://www.openaire.eu/search/publication?articleId=od_18::65b0af7fdca3b78f60f6538967faeb9f
14. Boticario, J. G., & Gaudioso Vázquez, E. (2003). *Sistemas interactivos de enseñanza aprendizaje*. Madrid: Sanz y Torres.
15. Boticario, J., Santos, O., Cabestrero, R., Quirós, P., Salmerón-Majadas, S., Uria-Rivas, R., et al. BIG-AFF: Exploring low cost and low intrusive infrastructures for affective computing in secondary schools. pp. 287-292.

16. Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), 49-59.
17. Bustos, V., Broisin, J., Muñoz, J., & Guzman, J. E. (2016). Extension of IEEE LOM standard for describing educational interactive application: An accessibility approach. *IEEE Latin America Transactions*, 14(8), 3847-3855. doi:10.1109/TLA.2016.7786372
18. Carletta, J. (1996). Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics*, 22(2), 249. Retrieved from http://gateway.proquest.com/openurl?ctx_ver=Z39.88-2003&xri:pqi:res_ver=0.2&res_id=xri:ilcs-us&rft_id=xri:ilcs:rec:abell:R03037566
19. Cechinel, C., da Silva Camargo, S., Sicilia, M., & Sánchez-Alonso, S. (2016, Jun 23,). Mining models for automated quality assessment of learning objects. *Computer Business Week*, pp. 334.
20. Cobos, C., Rodriguez, O., Rivera, J., Betancourt, J., Mendoza, M., León, E., et al. (2013). A hybrid system of pedagogical pattern recommendations based on singular value decomposition and variable data attributes. *Information Processing and Management*, 49(3), 607-625. doi:10.1016/j.ipm.2012.12.002
21. Czernianin, W. (2016). Poetry as a therapeutic medium in shaping mood. *Journal of Poetry Therapy*, 29(3), 135-145. doi:10.1080/08893675.2016.1199513
22. Datta, A., Sen, S., & Zick, Y. Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems. pp. 598-617.
23. Descartes, R. (2013). *René descartes: Meditations on first philosophy: With selections from the objections and replies* Cambridge University Press.
24. Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. Retrieved from <https://arxiv.org/abs/1810.04805>
25. Díez Platas, M. L., & Bermúdez Sabel, H. *Postdata-core ontology* (V 1.0 ed.)
26. Eberl, J. T. (2004). Aquinas on the nature of human beings. *The Review of Metaphysics*, , 333-365.
27. Fernández-Breis, J. T., Cañavate-Cañavate, D., Castellanos-Nieves, D., Valencia-García, R., & Vivancos-Vicente, P. J. (2005). OeLE: Applying ontologies to support the evaluation of open questions-based tests. *SW-EL '05: Applications of Semantic Web Technologies for E-Learning*, , 37.
28. Ferré, P., Guasch, M., Martínez-García, N., Fraga, I., & Hinojosa, J. (2017). Moved by words: Affective ratings for a set of 2,266 spanish words in five discrete emotion categories. *Behavior Research Methods*, 49(3), 1082-1094. doi:10.3758/s13428-016-0768-3
29. García Franco, J. D., & Manjarrés Riesco, Á. (2016). Modelado de los trastornos de la personalidad en OWL - lógica descriptiva y aplicación al desarrollo de un sistema experto de diagnóstico clínico y educativo ; Universidad Nacional de Educación a Distancia (UNED)). , 157.
30. Garg, N., Schiebinger, L., Jurafsky, D. & Zou, J. (2018). *Word embeddings quantify 100 years of gender and ethnic stereotypes.*<http://scholar.aci.info/view/1555a4c688800130002/162d5de6c9400014ec0b850>

31. Gloria Schultz Eastman. (2015). Making metaphor visible: The common core, poetry, and visual literacy. *The English Journal*, 104(6), 40-46. Retrieved from Research Library Prep database. Retrieved from <https://www.jstor.org/stable/24484429>
32. González Fernández, E. (2016). La vida como realidad radical en el pensamiento de Julián Marías. *Scio*, (12), 97-123.
33. Guasch, M., Ferré, P., & Fraga, I. (2016). Spanish norms for affective and lexico-semantic variables for 1,400 words. *Behavior Research Methods*, 48(4), 1358-1369. doi:10.3758/s13428-015-0684-y
34. Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2019). A survey of methods for explaining black box models. *ACM Computing Surveys (CSUR)*, 51(5), 1-42. doi:10.1145/3236009
35. J. Richard Landis, & Gary G. Koch. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174. doi:10.2307/2529310
36. Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jégou, H., & Mikolov, T. (2016). *FastText.zip: Compressing text classification models*. Retrieved from https://www.openaire.eu/search/publication?articleId=od_____18::02a044da2e2bc2b8a047f3507939f28a
37. Kant, I. (1998). *Critique of pure reason* Cambridge University Press.
38. Kant, I. (1999). Fundamentación de la metafísica de las costumbres. (M. García Morente Trans.). (pp. 21-22). Alicante : Biblioteca Virtual Miguel de Cervantes, 1999:
39. Kaur, J., & Saini, J. (Feb 24, 2017). (Feb 24, 2017). Punjabi poetry classification. Paper presented at the pp. 1-5. doi:10.1145/3055635.3056589
40. Krippendorff, K. (2011). *Computing krippendorff's alpha-reliability*ScholarlyCommons.
41. Lapedra Gutiérrez, G. (1997). Una comparación entre Razón Vital y Razón Poética. María Zambrano y la Filosofía de la Religión. Retrieved from <http://revistas.ucm.es/index.php/ILUR/article/view/ILUR9797110063A>
42. Liu, B. (2012). *Sentiment analysis and opinion mining* Morgan & Claypool Publishers. doi:10.2200/S00416ED1V01Y201204HLT016
43. Manjarrés Riesco, Á, Santos, O. C., & Boticario, J. G. (2013). Eliciting affective recommendations to support distance learning students. doi:10.1007/978-3-642-38844-6_35
44. Manjarrés-Riesco, Á, Santos, O. C., Boticario, J. G., & Saneiro, M. (2013). Open issues in educational affective recommendations for distance learning scenarios. *CEUR Workshop Proceedings*, 997, 26-33.
45. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient estimation of word representations in vector space* Retrieved from https://www.openaire.eu/search/publication?articleId=od_____18::7d58576a90b92bef1fcc9e579134295f
46. Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). *Distributed representations of words and phrases and their compositionality* Retrieved from

https://www.openaire.eu/search/publication?articleId=od_18::583007a114816fb993488e2b3687de99

47. Ming-Che, L., Ding, Y. Y., & Tzone, I. W. Java learning object ontology. pp. 538-542.
48. Navarro-Colorado, B. (2018). A metrical scansion system for fixed-metre spanish poetry. *Digital Scholarship in the Humanities*, 33(1), 112-127. doi:10.1093/llc/fqx009
49. Nurjanah, D. (2016). Good and similar learners' recommendation in adaptive learning systems.42(1) doi:10.5220/0005864304340440
50. Nussbaum, M. C., & Rorty, A. O. (1992). *Essays on aristotle's de anima* Clarendon Press.
51. Obermeier, C., Menninghaus, W., von Koppenfels, M., Raettig, T., Schmidt-Kassow, M., Otterbein, S., et al. (2013). Aesthetic and emotional effects of meter and rhyme in poetry. *Frontiers in Psychology*, 4, 10. doi:10.3389/fpsyg.2013.00010
52. Parastoo, A., Amenehsadat, K., & Shahla, M. (2016). *The effect of group poetry therapy on cognitive emotion regulation and moods of psychotic patients: The case of razi psychiatric hospital, tehran*
53. Park, J. H., Shin, J., & Fung, P. (2018). *Reducing gender bias in abusive language detection*
Retrieved from https://www.openaire.eu/search/publication?articleId=od_18::a92d58648f7e02ab83e72cddec8a722c
54. Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global vectors forWord representation. *Schedae Informaticae*, 25, 103. doi:10.4467/20838476SI.16.008.6189
55. Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., et al. (2018). *Deep contextualized word representations*
Retrieved from https://www.openaire.eu/search/publication?articleId=od_18::6b3dff98b9e2ca05e9870289f788ecfa
56. Platón. (1997). Fedón en diálogos, tomo III. *Gredos, Madrid*,
57. POSTDATA. (2015). *POSTDATA: Poetry standardization and linked open data. Ref. ERC-2015-STG-679528 project Starting Grant from European Research Council at H2020.*
58. Raffenne, E. (2013). *MIRLO: Una ontología para dar soporte a un modelo de estudiante abierto*
59. Ribeiro, M. T., Singh, S., & Guestrin, C. (Aug 13, 2016). (Aug 13, 2016). "Why should I trust you?". Paper presented at the pp. 1135-1144. doi:10.1145/2939672.2939778
60. Ricci, F., Rokach, L., & Shapira, B. (2011). Introduction to recommender systems handbook. *Recommender systems handbook* (pp. 1-35) Springer.
61. Ricci, F., Rokach, L., & Shapira, B. (2015). *Recommender systems handbook* (Second edition. ed.). New York, NY: Springer Verlag. doi:10.1007/978-1-4899-7637-6
62. Ruiz Fabo, Pablo, Martínez Cantón, Clara Isabel y Calvo Tello, José (2018) .DISCO: Diachronic Spanish Sonnet Corpus. DHd Konferenz 2018: Kritik der digitalen Vernunft (2018. Köln) At: Colonia (Germany). (2018-02-26)

63. Ruiz Fabo, P., Martínez Cantón, C., Poibeau, T., & Gonzalez-Blanco, E. Enjambment detection in a large diachronic corpus of spanish sonnets. pp. 27-32.
64. Sabitha, S., Mehrotra, D., & Bansal, A. (2015). *Knowledge enriched learning by converging knowledge object & learning object*. Reading: Academic Conferences International Limited.
65. Salmeron-Majadas, S., Santos, O. C., & Boticario, J. G. (2014). Exploring indicators from keyboard and mouse interactions to predict the user affective state. *Proceedings of the 7th International Conference on Educational Data Mining*, 35, 691-700. doi://doi.org/10.1016/j.procs.2014.08.151
66. Santos, O. C. (2016). *Emotions and personality in adaptive e-learning systems: An affective computing perspective* (1st ed. 2016 ed.). DE: Springer Verlag.
67. Santos, O. C., & Boticario, J. G. (2011). (2011). TORMES methodology to elicit educational oriented recommendations. Paper presented at the *Artificial Intelligence in Education*, pp. 541-543.
68. Santos, O. C., Boticario, J. G., Arevalillo-Herráez, M., Saneiro, M., Cabestrero, R., del Campo, E., et al. (2012). MAMIPEC - affective modeling in inclusive personalized educational scenarios. *Bulletin of the IEEE Technical Committee on Learning Technology*, 14(4), 35-38.
69. Santos, O. C., Saneiro, M., Boticario, J. G., & Rodriguez-Sanchez, M. C. (2016). Toward interactive context-aware affective educational recommendations in computer-assisted language learning. *New Review of Hypermedia and Multimedia*, 22(1-2), 27-57. doi:10.1080/13614568.2015.1058428
70. Saquicela, V., Baculima, F., Orellana, G., Piedra, N., Orellana, M., & Espinoza, M. (2018). *Similarity detection among academic contents through semantic technologies and text mining*
71. Shapiro, J., & Rucker, L. (2003). Can poetry make better doctors? teaching the humanities and arts to medical students and residents at the university of california, irvine, college of medicine. *Academic Medicine : Journal of the Association of American Medical Colleges*, 78(10), 953-957. doi:10.1097/00001888-200310000-00002
72. Stadthagen-Gonzalez, H., Imbault, C., Pérez Sánchez, M., & Brysbaert, M. (2017). Norms of valence and arousal for 14,031 spanish words. *Behavior Research Methods*, 49(1), 111-123. doi:10.3758/s13428-015-0700-2
73. Theodorou, A., Wortham, R. H., & Bryson, J. J. (2017). *Designing and implementing transparency for real time inspection of autonomous robots*
74. Ullrich, S., Aryani, A., Kraxenberger, M., Jacobs, A. M., & Conrad, M. (2016). On the relation between the general affective meaning and the basic sublexical, lexical, and inter-lexical features of poetic Texts—A case study using 57 poems of H. M. enzensberger. *Frontiers in Psychology*, 7, 2073. doi:10.3389/fpsyg.2016.02073
75. Vö, M., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M., & Jacobs, A. (2009). The berlin affective word list reloaded (BAWL-R). *Behavior Research Methods*, 41(2), 534-538. doi:10.3758/BRM.41.2.534
76. Walshe, N. (2017). An interdisciplinary approach to environmental and sustainability education: Developing geography students' understandings of sustainable development using poetry. *Environmental Education Research*, 23(8), 1130-1149. doi:10.1080/13504622.2016.1221887

77. Wassiliwizky, E., Koelsch, S., Wagner, V., Jacobsen, T., & Menninghaus, W. (2017). The emotional power of poetry: Neural circuitry, psychophysiology and compositional principles. *Social Cognitive and Affective Neuroscience*, 12(8), 1229-1240. doi:10.1093/scan/nsx069
78. Weber, G., & Brusilovsky, P. (2016). ELM-ART – an interactive and intelligent web-based electronic textbook. *International Journal of Artificial Intelligence in Education*, 26(1), 72-81. doi:10.1007/s40593-015-0066-8
79. Zhang, X., Zhao, J., & LeCun, Y. (2015). Character-level convolutional networks for text classification. Retrieved from <http://arxiv.org/abs/1509.01626>
80. Zhang, X., Zhao, J., & Lecun, Y. (2015). *Character-level convolutional networks for text classification*. Retrieved from https://www.openaire.eu/search/publication?articleId=od_____18::cb1878e4f9caacc0ef20e47db7f18f3f
81. Zhang, Y., Jin, R., & Zhou, Z. (2010). Understanding bag-of-words model: A statistical framework. *International Journal of Machine Learning and Cybernetics*, 1(1), 43-52. doi:10.1007/s13042-010-0001-0
82. Zhao, J., Wang, T., Yatskar, M., Cotterell, R., Ordonez, V., & Chang, K. (2019). Gender bias in contextualized word embeddings. Retrieved from <https://arxiv.org/abs/1904.03310>
83. Zhao, J., Zhou, Y., Li, Z., Wang, W., & Chang, K. (2018). *Learning gender-neutral word embeddings*. Retrieved from https://www.openaire.eu/search/publication?articleId=od_____18::f476526606a15368fbaa30489bfb5d18

9. Anexos

9.1. Campos de las ontologías

FOAF Core⁹⁰:

| | | | | |
|---------------|-------------------------------------|-------------------|------------------------------|---|
| <u>Agent</u> | <u>title</u> | <u>familyName</u> | <u>based_near</u> | <u>primaryTopic</u> (<u>primaryTopicOf</u>) |
| <u>Person</u> | <u>img</u> | <u>givenName</u> | <u>age</u> | <u>Project</u> |
| <u>name</u> | <u>depiction</u> (<u>depicts</u>) | <u>knows</u> | <u>made</u> (<u>maker</u>) | <u>Organization</u> |
| <u>Group</u> | <u>member</u> | <u>Document</u> | <u>Image</u> | |

Tabla 30 Lista de campos en FOAF Core

DISCO:

- Metadatos de autores⁹¹

| field name | value |
|------------------|---|
| author | author's last and first name |
| aid | alphanumeric ID for the author |
| normdate | "normalized date" |
| birth | year or century of birth |
| death | year or century of death |
| raw | raw author information (available as unstructured text for each author) in the corpus sources |
| hasbirth | 1 or 0 whether an exact year is given in the source or not |
| hasdeath | 1 or 0 whether an exact year is given in the source or not |
| hasdates | 1 when both <i>hasbirth</i> and <i>hasdeath</i> are 1, else 0 |
| estimate | 0 unless the date was indicated as an estimation (with a question mark) in the sources |
| place-birth | place of birth |
| place-death | place of death |
| country-birth | country of birth |
| continent-birth | continent of birth |
| first-name | first name |
| preposition-name | particle in last name |
| second-name | last name |
| gender | author's gender |

⁹⁰ FOAF. Web. 11/05/2019. <http://xmlns.com/foaf/spec/>

| | |
|---------------|---|
| viafid | ID we selected programmatically from VIAF's API responses |
| vf_bi | birthdate for our selected VIAF candidate |
| vf_de | date of death for our selected VIAF candidate |
| vf_bi_esti | 1 if the birthdate for our VIAF candidate has question marks in VIAF, else 0 |
| vf_de_esti | 1 if the date of death for our VIAF candidate has question marks in VIAF, else 0 |
| birth_diff | difference between birthdate in our sources and the one for our VIAF candidate |
| death_diff | difference between date of death in our sources and the one for our VIAF candidate |
| vf_cand_n | number of candidates returned by VIAF's API for our queries |
| auname_viaf | author name for the VIAF candidate we selected |
| redo | project-internal information |
| match_type | indicates how the author's name in our corpus compares to the VIAF candidate we selected: <i>exact</i> match, <i>contained</i> in the VIAF name using full-word matching or non-full-word match (<i>wordpart</i>) |
| vf_validation | this is rendered in the TEI headers as the @certvalue for <i>idno</i> [@type="viaf"] |

Tabla 31 Metadatos de autores en DISCO

xAPI⁹²:

| Property | Type | Description | Required |
|------------------------|--------|---|-------------|
| id | UUID | UUID assigned by LRS if not set by the Learning Record Provider. | Recommended |
| actor | Object | Whom the Statement is about, as an Agent or Group Object. | Required |
| verb | Object | Action taken by the Actor. | Required |
| object | Object | Activity, Agent, or another Statement that is the Object of the Statement. | Required |

⁹² Experience API. Web. 11/05/2019. <https://github.com/adlnet/xAPI-Spec/blob/master/xAPI-Data.md#parttwo>

| | | | |
|-----------------------------|-------------------------------------|--|-----------------|
| result | Object | Result Object, further details representing a measured outcome. | Optional |
| context | Object | Context that gives the Statement more meaning. Examples: a team the Actor is working with, altitude at which a scenario was attempted in a flight simulator. | Optional |
| timestamp | Timestamp | Timestamp of when the events described within this Statement occurred. Set by the LRS if not provided. | Optional |
| stored | Timestamp | Timestamp of when this Statement was recorded. Set by LRS. | Set by LRS |
| authority | Object | Agent or Group who is asserting this Statement is true. Verified by the LRS based on authentication. Set by LRS if not provided or if a strong trust relationship between the Learning Record Provider and LRS has not been established. | Optional |
| version | Version | The Statement's associated xAPI version, formatted according to Semantic Versioning 1.0.0 . | Not Recommended |
| attachments | Ordered array of Attachment Objects | Headers for Attachments to the Statement | Optional |

Tabla 32 Campos en xAPI

9.2. API endpoints

A continuación se describen sucintamente los endpoints de la API, remitiendo al repositorio de esta para profundizar más en la documentación técnica⁹³.

Estos endpoints corresponden a la primera versión del prototipo, la 0.1.

- **[GET] /result** | Endpoint para llevar a cabo la tarea de IR sobre los sonetos del corpus en base a una consulta de entrada. La llamada a este endpoint debe tener los siguientes argumentos.

| parameter | values | description |
|------------------|----------------------|--|
| query_text | | Input query sent |
| composition_type | "joint" or "sum" | |
| metric | "cosine" or "icm" | |
| log | "True" or "False" | To indicate if the user wants to see the scoring obtained for the last top of retrieved sonnets. |

Tabla 33 Campos para el método GET de la API desarrollada

- **[POST] /create_corpus** | Endpoint para generar el corpus de sonetos en base a los datos de entrada que se han mencionado previamente.
- **[POST] /create_ontology** | Endpoint para modificar los XML TEI de POSTDATA – DISCO y añadir los parámetros afectivos.

⁹³ Barbado González, Alberto. "Sonnet Recommender System". Web. 17/05/2019. <https://github.com/AlbertoBarbado/sonnet-recommender>

9.3. JSON con la información para cada soneto

```
▼ object {11}
  title : Al plan de Matanzas
  text : ¿Quién eres tú, gigante, en cuya frente\nse detienen las nieblas apiñadas,\nen tanto que
a tus plantas, humilladas\nrugen las tempestades sordamente?...\n\nTu fantástica forma
sorprendente,\ntus crestas a los cielos levantadas,\ntus abismos, tus rocas
despeñadas,\n¿qué misterios encubren a la mente?...\n\n¿Y pretendo tu origen
misterioso\npenetrar, al través del tiempo inmenso\nque miraste pasar?... De luz un
rayo\n\nilumina mi espíritu; y lloroso,\nque eres la tumba perdurable pienso\ndel pueblo
antiguo que habitó en Yucayo.\n\n
  id_doc : disco:po_099n0334
  ▶ dct_stanzas {4}
  ▶ affective_features {38}
  ▶ words_lem [45]
  ▶ words_lem_ngrams [90]
  ▶ words_lem_complete [80]
  file_path : poems/per-sonnet\\Acosta,_Ignacio_María__099n~Al_plan_de_Matanzas__0334.xml
  ngrams2str : ment espíritu inmenso apiñar encubri través travé frent habita planta lloros
pretende rugí antigu puebl yucay ray quí humillar tiempo plantar form yucayo
tumba penetra perdurabl niebla mira perdurable sorprendent encubrir gigant antiguo
mirar tempestad niebl fantástic despeñar tiemp iluminar cuyo rayo misterio
fantástica ciel lloroso detene levantar gigante inmens orige lu pensa levanta
detener luz sordamente abism humilla cuy habitar cielo rugir quién origen pueblo
abismo pensar misterí pasa tumb penetrar pasar pretender apiña misterios
sorprendente crest frente espíritu roca forma mente roc sordament despeña ilumina
cresta tempesta misterioso
  lem2str : quién gigante cuyo frente detener niebla apiñar plantar humillar rugir tempestad
sordamente fantástica forma sorprendente cresta cielo levantar abismo roca despeñar
misterio encubrir mente pretender origen misterioso penetrar través tiempo inmenso
mirar pasar luz rayo iluminar espíritu lloroso tumba perdurable pensar pueblo antiguo
habitar yucayo
```

Ilustración 100 JSON ejemplo de respuesta del RecSys desarrollado

9.4. JSON con la información de uso de AULA

```
{"selection_arousal":
```

```
  "tranquilo", "selection_valence": "triste",
```

```
  "query_text": "pena",
```

```
  "sonnet_selected": "joint",
```

```
  "selection_rating": "5"
```

```
}
```

```
▼ object {11}
```

```
  title : Al plan de Matanzas
```

```
  text : ¿Quién eres tú, gigante, en cuya frente\nse detienen las nieblas apiñadas,\nen tanto que  
a tus plantas, humilladas\nrugen las tempestades sordamente?... \n\nTu fantástica forma  
sorprendente, \ntus crestas a los cielos levantadas, \ntus abismos, tus rocas  
despeñadas, \n¿qué misterios encubren a la mente?... \n\n¿Y pretendo tu origen  
misterioso\npenetrar, al través del tiempo inmenso\nque miraste pasar?... De luz un  
rayo\nnilumina mi espíritu; y lloroso, \nque eres la tumba perdurable pienso\n\ndel pueblo  
antiguo que habitó en Yucayo. \n\n
```

```
  id_doc : disco:po_099n0334
```

```
▶ dct_stanzas {4}
```

```
▼ affective_features {38}
```

```
  Hap_Mean : 45.4
```

```
  Hap_SD : 21.18
```

```
  Ang_Mean : 35.48
```

```
  Ang_SD : 15.8
```

```
  Sad_Mean : 45.25
```

```
  Sad_SD : 21.45
```

```
  Fear_Mean : 49.41
```

```
  Fear_SD : 23.7
```

```
  Disg_Mean : 31.21
```

```
  Disg_SD : 14.29
```

```
  ValenceMean : 418.24
```

```
  ValenceSD : 96.24
```

```
  ArousalMean : 405.28
```

```
  ArousalSD : 119.36
```

```
  Max_Aro : 7.4
```

```
  Min_Aro : 2.95
```

Ilustración 101 JSON con los campos del TEI XML de los sonetos de DISCO incluyendo la extensión para los campos afectivos

9.5. Resultados de Aula 2019

Datos de uso de los usuarios de la aplicación

| | query_text | selection_arousal | selection_rating | selection_valence | sonnet_selected | sonnet_id |
|----|-------------|-------------------|------------------|-------------------|-----------------|-----------|
| 0 | guerra | activo | 3 | Frustrado | sum | 642 |
| 1 | esfuerzo | muy_activo | 5 | alegre | joint | 2475 |
| 2 | pabellón | activo | 5 | entusiasmado | joint | 3463 |
| 3 | cuna\n | muy_activo | 5 | alegre | sum | 3281 |
| 4 | pasión | activo | 4 | entusiasmado | sum | 68 |
| 5 | estrellas\n | tranquilo | 5 | alegre | sum | 1541 |
| 6 | futuro | activo | 5 | entusiasmado | sum | 4125 |
| 7 | libros | tranquilo | 5 | alegre | joint | 1662 |
| 8 | magia | tranquilo | 5 | alegre | sum | 3539 |
| 9 | futuro\n | tranquilo | 5 | seguro | sum | 4125 |
| 10 | primavera | tranquilo | 5 | alegre | joint | 1750 |
| 11 | luna | tranquilo | 5 | entusiasmado | joint | 1898 |
| 12 | luna\n | tranquilo | 5 | entusiasmado | sum | 4117 |
| 13 | pena | tranquilo | 5 | triste | joint | 980 |
| 14 | hola | muy_activo | 4 | triste | sum | 1399 |
| 15 | felicidad | tranquilo | 5 | triste | sum | 3311 |
| 16 | morado | activo | 5 | alegre | sum | 1784 |
| 17 | luminosidad | activo | 1 | entusiasmado | joint | 1074 |
| 18 | terminar | muy_tranquilo | 4 | Frustrado | sum | 302 |
| 19 | periodico | tranquilo | 5 | seguro | sum | 1983 |
| 20 | feliz | tranquilo | 5 | entusiasmado | joint | 2019 |
| 21 | sensualidad | activo | 3 | alegre | sum | 153 |
| 22 | motivada | muy_activo | 4 | entusiasmado | sum | 700 |
| 23 | relajada | tranquilo | 5 | entusiasmado | joint | 2333 |
| 24 | feria | activo | 4 | alegre | sum | 1892 |
| 25 | jirafa | muy_activo | 5 | alegre | sum | 3113 |
| 26 | buscar | tranquilo | 4 | entusiasmado | sum | 4028 |
| 27 | tiempo | activo | 4 | entusiasmado | sum | 2272 |
| 28 | voluntad | muy_activo | 4 | alegre | joint | 3828 |
| 29 | comer | muy_tranquilo | 5 | irritado | joint | 2014 |
| 30 | esperanza | muy_tranquilo | 5 | alegre | joint | 4231 |
| 31 | alegria | tranquilo | 5 | alegre | joint | 3424 |
| 32 | emocion | tranquilo | 4 | entusiasmado | sum | 2462 |
| 33 | cariño | activo | 4 | alegre | sum | 3358 |
| 34 | agotado | muy_tranquilo | 4 | seguro | joint | 27 |

Tabla 34 Resultados de los usuarios de AULA

Número de veces que se han elegido las distintas palabras

| query_text | count |
|-------------|-------|
| luna | 2 |
| futuro | 2 |
| pena | 1 |
| magia | 1 |
| morado | 1 |
| motivación | 1 |
| pabellón | 1 |
| pasión | 1 |
| alegría | 1 |
| luminosidad | 1 |
| primavera | 1 |
| relajado | 1 |
| saludo | 1 |
| sensualidad | 1 |
| terminar | 1 |
| tiempo | 1 |
| periódico | 1 |
| libro | 1 |
| buscar | 1 |
| jirafa | 1 |
| guerra | 1 |
| feria | 1 |
| feliz | 1 |
| felicidad | 1 |
| estrella | 1 |
| esperanza | 1 |
| esfuerzo | 1 |
| emoción | 1 |
| cuna | 1 |
| comer | 1 |
| cariño | 1 |
| cansado | 1 |
| voluntad | 1 |

Tabla 35 Frecuencia de las palabras usadas por los usuarios de AULA

9.6. Diccionario de siglas

SAREK: Sonnet Affective Recommender using Embedding Knowledge
SAE: Sistema Adaptativo en la Educación
API: Application program interface
FOAF: Friend of a Friend
SCORM: Sharable Content Object Reference Model
NLP: Natural Language Processing
NLU: Natural Language Understanding
NLG: Natural Language Generation
OWL: Web Ontology Language
DISCO: Diachronic Spanish Sonnet Corpus
POSTDATA: Poetry Standardization and Linked Open Data
RecSys: Recommender System
IMS: Integrated Management System
LIP: Learner Information Package
ECTS: European Credit Transfer and Accumulation System
TFM: Trabajo Fin de Máster
CRISP-DM: Cross Industry Standard Process for Data Mining
LIME: Local Interpretable Model-agnostic Explanations
TORMES: Tutor-Oriented Recommendations Modelling for Educational Systems
ELMo: Embeddings from language model design
BERT: Bidirectional Encoder Representations from Transformers
ML: Machine Learning
DL: Deep Learning
IA/AI: Inteligencia Artificial/Artificial Intelligence
LRS: Learner Record Store
LO: Learning Object
IEEE: Institute of Electrical and Electronics Engineers
LSTM: Long Short-Term Memory
ANN: Artificial Neural Network
CNN: Convolutional Neural Network
OOV: Out of vocabulary
UNK: Unknown
BoW: Bag of Words
SOTA: State of the Art
RAE: Real Academia de la lengua española
IR: Information Retrieval
GAM: General Affective Meaning
TEL: Technology Enhanced Learning
NAcc: Núcleo *accumbens*
IR: Information Recovery
SAM: *Self-Assessment Manikin*
LU: *Learning Unit*

9.7. Rasgos afectivos utilizados

Los rasgos afectivos utilizados para anotar los sonetos están principalmente recogidos en (Ullrich et al., 2016), aunque se proponen algunos otros que amplíen éstos. Son los siguientes:

- **ValenceMean:** Media de la valencia de las palabras individuales
- **ValenceSD:** Desviación típica de la valencia de las palabras individuales
- **ArousalMean:** Media del nivel de excitación de las palabras individuales
- **ArousalSD:** Desviación típica del nivel de excitación de las palabras individuales
- **ArousalSpan:** Diferencia entre el valor máximo y mínimo del nivel de excitación de las palabras individuales.
- **MaxAro:** Valor máximo del nivel de excitación
- **MinAro:** Valor mínimo del nivel de excitación.
- **ValenceSpan:** Diferencia entre el valor máximo y mínimo de la valencia de las palabras individuales.
- **MaxVal:** Valor máximo de valencia
- **MinVal:** Valor mínimo de valencia
- **CorAro:** Coeficiente de correlación (Spearman) entre los valores de nivel de excitación de las palabras individuales y el índice de su posición dentro del texto.
- **CorVal:** Coeficiente de correlación (Spearman) entre los valores de valencia de las palabras individuales y el índice de su posición dentro del texto.
- **SigmaAro:**

$$\text{SigmaAro} = \frac{M(\text{arousal})}{1/\sqrt{N}}$$

Con $M(\text{arousal})$ la media de nivel de excitación de las palabras del texto
y N el número total de palabras que aparecen en ese texto

- **SigmaVal:** Ídem para valencia
- **Hap_Mean:** Ídem para *happiness*
- **Hap_SD:** Ídem para *happiness*
- **Ang_Mean:** Ídem para *anger*
- **Ang_SD:** Ídem para *anger*
- **Sad_Mean:** Ídem para *sadness*
- **Sad_SD:** Ídem para *sadness*
- **Fear_Mean:** Ídem para *fear*
- **Fear_SD:** Ídem para *fear*
- **Disg_Mean:** Ídem para *disgust*
- **Disg_SD:** Ídem para *disgust*
- **averageAoA:** Ídem para *Age of Acquisition*
- **sdAOA:** Ídem para *Age of Acquisition*
- **MaxAoA:** Ídem para *Age of Acquisition*
- **MinAoA:** Ídem para *Age of Acquisition*
- **AoASpan:** Ídem para *Age of Acquisition*
- **CorAoA:** Ídem para *Age of Acquisition*

9.8. Resultados completos del Experimento 2 y 4

| feature_col | n_labels | type_embeddings | type | metric | index | precision | 1st_correct | map |
|----------------|----------|-----------------|-------|--------|-------|-----------|-------------|--------|
| Vulnerabilidad | 98 | word2vec | joint | icm | 230 | 0.449 | yes | 0.4649 |
| Vulnerabilidad | 98 | word2vec | joint | cosine | 230 | 0.449 | yes | 0.461 |
| Vulnerabilidad | 98 | bert | joint | cosine | 220 | 0.4388 | yes | 0.4547 |
| Dramatización | 94 | bert | joint | cosine | 220 | 0.383 | yes | 0.4498 |
| Dramatización | 94 | bert | joint | icm | 60 | 0.383 | no | 0.424 |
| Vulnerabilidad | 98 | bert | joint | icm | 60 | 0.4388 | no | 0.4209 |
| Dramatización | 94 | word2vec | joint | icm | 230 | 0.4043 | no | 0.4065 |
| Dramatización | 94 | word2vec | joint | cosine | 230 | 0.3723 | no | 0.4037 |
| Aversión | 77 | word2vec | joint | icm | 230 | 0.4026 | yes | 0.394 |
| Aversión | 77 | bert | joint | icm | 60 | 0.3247 | yes | 0.3912 |
| Aversión | 77 | word2vec | joint | cosine | 230 | 0.3506 | yes | 0.389 |
| Aversión | 77 | bert | joint | cosine | 220 | 0.3636 | no | 0.3671 |
| Temor | 70 | bert | joint | icm | 60 | 0.3429 | yes | 0.3652 |
| Temor | 70 | word2vec | joint | icm | 230 | 0.3143 | yes | 0.3433 |
| Temor | 70 | bert | joint | cosine | 220 | 0.3429 | no | 0.3432 |
| Temor | 70 | word2vec | joint | cosine | 230 | 0.3143 | yes | 0.343 |
| Orgullo | 49 | bert | joint | cosine | 220 | 0.2449 | yes | 0.2628 |
| Ira | 40 | word2vec | joint | cosine | 230 | 0.225 | yes | 0.2476 |
| Orgullo | 49 | bert | joint | icm | 60 | 0.2041 | no | 0.2469 |
| Ansiedad | 47 | bert | joint | icm | 60 | 0.2553 | no | 0.241 |
| Ira | 40 | word2vec | joint | icm | 230 | 0.2 | yes | 0.2385 |
| Impotencia | 51 | bert | joint | icm | 60 | 0.2157 | no | 0.2196 |
| Impotencia | 51 | word2vec | joint | cosine | 230 | 0.1765 | no | 0.2172 |
| Impotencia | 51 | word2vec | joint | icm | 230 | 0.1765 | no | 0.2143 |
| Ansiedad | 47 | word2vec | joint | icm | 230 | 0.1702 | no | 0.2112 |
| Ansiedad | 47 | word2vec | joint | cosine | 230 | 0.1702 | no | 0.2096 |
| Orgullo | 49 | word2vec | joint | icm | 230 | 0.1633 | no | 0.2049 |
| Orgullo | 49 | word2vec | joint | cosine | 230 | 0.1633 | no | 0.2045 |
| Ira | 40 | bert | joint | icm | 60 | 0.2 | no | 0.2008 |
| Inestabilidad | 34 | bert | joint | icm | 60 | 0.2059 | no | 0.1919 |
| Ansiedad | 47 | bert | joint | cosine | 220 | 0.1277 | no | 0.1909 |
| Ira | 40 | bert | joint | cosine | 220 | 0.15 | no | 0.1884 |
| Impotencia | 51 | bert | joint | cosine | 220 | 0.1176 | no | 0.1875 |
| Inestabilidad | 34 | word2vec | joint | icm | 230 | 0.1765 | no | 0.184 |
| Inestabilidad | 34 | bert | joint | cosine | 220 | 0.1765 | no | 0.1777 |
| Inestabilidad | 34 | word2vec | joint | cosine | 230 | 0.2059 | no | 0.1743 |
| Ilusión | 40 | bert | joint | icm | 60 | 0.15 | no | 0.171 |
| Desilusión | 33 | word2vec | joint | icm | 230 | 0.2121 | no | 0.1684 |
| Desilusión | 33 | word2vec | joint | cosine | 230 | 0.1818 | no | 0.1678 |
| Desilusión | 33 | bert | joint | icm | 60 | 0.1818 | no | 0.1657 |
| Inseguridad | 31 | bert | joint | icm | 60 | 0.1613 | no | 0.1657 |
| Ilusión | 40 | word2vec | joint | icm | 230 | 0.175 | no | 0.1638 |
| Ilusión | 40 | word2vec | joint | cosine | 230 | 0.15 | no | 0.1628 |

| | | | | | | | | |
|--------------------|----|----------|-------|--------|-----|--------|----|--------|
| Inseguridad | 31 | word2vec | joint | cosine | 230 | 0.1613 | no | 0.1528 |
| Inseguridad | 31 | word2vec | joint | icm | 230 | 0.129 | no | 0.1513 |
| Desilusión | 33 | bert | joint | cosine | 220 | 0.1212 | no | 0.1491 |
| Ilusión | 40 | bert | joint | cosine | 220 | 0.1 | no | 0.1423 |
| Depresión | 26 | word2vec | joint | cosine | 230 | 0.0769 | no | 0.1367 |
| Depresión | 26 | word2vec | joint | icm | 230 | 0.0769 | no | 0.1313 |
| Inseguridad | 31 | bert | joint | cosine | 220 | 0.129 | no | 0.1304 |
| Depresión | 26 | bert | joint | icm | 60 | 0.1538 | no | 0.1196 |
| Depresión | 26 | bert | joint | cosine | 220 | 0.0769 | no | 0.1132 |
| Obsesión | 17 | bert | joint | icm | 60 | 0.0588 | no | 0.1045 |
| Obsesión | 17 | bert | joint | cosine | 220 | 0.1176 | no | 0.0985 |
| Prejuicio | 13 | word2vec | joint | icm | 230 | 0.0 | no | 0.0753 |
| Obsesión | 17 | word2vec | joint | icm | 230 | 0.0588 | no | 0.0746 |
| Prejuicio | 13 | word2vec | joint | cosine | 230 | 0.0 | no | 0.0732 |
| Obsesión | 17 | word2vec | joint | cosine | 230 | 0.0588 | no | 0.0713 |
| Prejuicio | 13 | bert | joint | icm | 60 | 0.0769 | no | 0.0706 |
| Prejuicio | 13 | bert | joint | cosine | 220 | 0.0769 | no | 0.0669 |

Tabla 36 Tabla completa de resultados para los Experimentos 2 y 4

DISCO PAL: Diachronic Spanish Sonnet Corpus with Psychological and Affective Labels

Alberto Barbado González⁹⁴, Víctor Fresno Fernández⁹⁵, Ángeles Manjarrés Riesco⁹⁶

Abstract

Nowadays there are many public repositories that contain texts in prose with different kinds of meta information that can help the research of text mining. However, there are few resources for text analytics applied to poetry, and much less for Spanish poetry.

This article presents a dataset that contains 230 Spanish sonnets from authors for different centuries, from 15th to 19th. This dataset was annotated by different domain experts. The experts labelled the poems with affective features, as well as with domain concepts that belong to psychology. Thanks to that, a rich dataset is available, so it can contribute to the text mining research.

The dataset is available for downloading from the public repository where it was published.

Keywords: Poetry. Spanish Sonnets. Affective. Semantic. Psychology.

Introduction

Text mining techniques aim to extract insights from a text and discover patterns within it using different kinds of information from that text. As an example, the information contained in a text could be related to its syntactical structure, to its semantical meaning or it can even consider information sources such as the affective value of the text like, for instance, if a text inspires a certain emotion when read.

This is the base for many researches, such as sentiment analysis. Sentiment analysis, also called opinion mining, is the field of study that analyses people's opinions, feelings, assessments, attitudes, and emotions towards entities such as products, services, organizations, individuals, problems, events, topics and its attributes (Liu, 2012).

Thus, this is applicable to the field of text mining where these feelings can be extracted, for example, through the use of supervised learning algorithms in which the labels are the different categories of feelings (for example, joy) and input features are the semantic information of a specific text for which that feeling has been annotated.

However, this is not the only way to focus the automatic treatment of affective content in texts. In general, the aim is to be able to obtain a general

affective meaning (GAM) of the text with its direct information, which can be semantic information, affective information of the individual words that compose it, the type of text used and its syntactic characteristics... Similarly, the relationship between the information in the text and the GAM can be carried out through the use of different functions, which can range from simple composition functions (such as directly adding the individual information of the words in the text) to more sophisticated solutions, such as supervised ML algorithms when there are labels annotated with the GAM values.

In general, a set of features are defined in order to represent a text and those features serve as input for a function that outputs the corresponding GAM tags. The process can use the value of manual GAM tags (supervised relationship) or not (unsupervised relationship). As mentioned before, supervised ML algorithms can be used when there are labels annotated for the texts. However, the approaches are different for scenarios when those labels are not available. An example of an unsupervised approach is the use of information retrieval (IR) techniques such as comparing the GAM of a label and rank the texts according to the similarity of their GAM values to the labelled value.

GAM are not the only type of labels that can be used to model the global meaning of a text. Any kind of labels can be considered, including words related to the semantic meaning of the text, such as the relation between definitions and their associated words in a dictionary (Amigó Cabrera & Fresno Fernández, 2018). All the different kinds of information contained within a text (semantic, syntactic, affective...) will depend according to the type of texts considered. Because of this, the approach is different depending on whether the text is, for example, prose or verse. It also depends on the language of the texts used.

That said, there are not many datasets available to perform text mining tasks on poetry texts, and much less for the Spanish language. It is true that there are available corpora for Spanish poetry, such as the corpus DISCO (Ruiz Fabo, Calvo Tello, Bermúdez Sabel, & Martínez Cantón, 2018), but these are only corpora with semantic and syntactic information that

⁹⁴ National Distance Education University (UNED). Madrid. Spain

⁹⁵ Department of Languages and Computer Systems. National Distance Education University (UNED). Madrid. Spain

⁹⁶ Department of Informatics and Automation. National Distance Education University (UNED). Madrid. Spain

cannot help directly for other text modelling tasks such as obtaining the GAM mentioned before. There is research published for the GAM of poetry in Spanish, such as (Barros, Rodriguez, & Ortigosa, Sep 2013) but that proposal obtains the GAM in an unsupervised manner by counting how many instances of words such as Fear or Joy appear within a set of Quevedo's poems. In fact, more recent researches of the topic for text mining with poetry, such as (Kaur & Saini, Feb 24, 2017) only list that dataset of Quevedo's poems annotated with sentiment labels according to the presence of certain words as Spanish corpora sources for text mining and GAM modelling.

This is the reason why the present research aimed to increase the available datasets for text mining tasks with Spanish poetry by presenting a dataset labelled of Spanish sonnets from different time periods labelled with both affective and semantic labels in order to contribute to the research of text mining in both areas. The article presents DISCO PAL, Diachronic Spanish Sonnet Corpus with Psychological and Affective Labels (together with this paper), a labelled dataset annotated by POSTDATA⁹⁷ experts in humanities. This dataset includes binary labels for a group of concepts depending on whether that concept appears within the text or not. The concepts used all belong to the psychological domain, since the literature have identified a potential relationship between poetry and psychology (Czernianin, 2016; Parastoo et al., 2016; Shapiro & Rucker, 2003). The labelled dataset provided also includes other ordinal features related to the affective modelling of the sonnets. These annotated features could help the research of other areas, such as the GAM obtention of a text.

The structure of this article is as follows: after the Introduction presented in this first chapter, the second chapter summarizes the state of the art (SOTA) for the areas relevant for this article. First, the SOTA related to text mining of affective information from poems. Then, the SOTA related to affective modelling of Spanish language by using public corpora for modelling individual words.

After that, the third chapter presents the dataset used annotated by POSTDATA experts in humanities, analyzing the agreement between the annotators and the reliability of the dataset. It also mentions possible way to effectively use it.

The last chapter mentions the potential lines of research that could be carried out thanks to this dataset. It also includes a summary with the conclusions of this article.

Overall, the main contributions of this article are:

- Provide an annotated dataset of a corpus of Spanish sonnets by different domain experts. This dataset has labels for psychological concepts that are expressed through that sonnet. The dataset also includes ordinal labels for affective features with values that depend on the intensity of that affection within each sonnet.
- Analyze the dataset provided, studying the level of agreement between the different annotators in order to check the reliability of the dataset.
- Highlight possible ways to use the dataset for text mining of poetry. mainly through the affective and semantic modelling of texts.

Related Work

This chapter presents a brief review for the related work relevant for this article. As it was mentioned in the Introduction, a text contains information related to different areas such as semantics, syntactics or affections. This is applicable to any kind of text, including poetic ones. Since this article provides a research related to affective and semantic modelling for Spanish poetry, the main area covered in this chapter is related to text mining of affective information in poetry, followed by a section describing some public corpora for the affective modelling of individual Spanish words.

Text mining of affective information in poetry

As previously indicated, texts in general, and particularly poetic ones, contain affective information that can be extracted using different techniques, like for instance aggregating the individual affective values of the individual words present in the text. It is important to quantify this affective contribution of poetic texts in order to operationalize how to work with them. Thus, the task consists in detecting which poetic elements are especially relevant in order to calculate through them the affective contribution of the whole poem. The articles shown below analyze precisely different ways of extracting and quantifying affective aspects from poetic texts.

An example of text mining of affective content in poetic texts appears in (Ullrich et al., 2016) for German language. This article explores how the features of a poetic text (at sub-lexical, lexical and

⁹⁷ *Poetry Standardization and Linked Open Data*, Ref. ERC-2015-STG-679528 proyecto Starting Grant from European Research Council within the horizon H2020.

inter-lexical level) influence in GAM that is perceived. Thus, this article serves as an example to see which affective features are relevant to a text based on how related they are to the GAM as well as how to calculate them. To calculate those features they use the BAWL database for German words (Berlin Affective Word List), described in (Võ et al., 2009) in its revision. This database contains affective values for individual words that belong to German, and they aggregate these individual values into a global value that models the whole poem.

As texts they use a corpus of poems is composed by 57 poems from the German author H.M. Enzenberg. These poems are annotated by a group of readers with the following features:

1. Score on a scale of 7 for the valence (valence or level of positive or negative affect of the text), where -3 would be very negative, 0 neutral and 3 very positive.
2. Score on a scale of 5 for the arousal (level of excitement of the text of the poem, which goes from texts that inspire peacefulness or calm to others who seek to motivate or are more exciting), where 1 is very quiet and 5 very exciting.
3. Score on a scale of 1 to 5 for the level of friendliness, where 1 indicates that the text is not friendly and 5 that it is very friendly.
4. Score on a scale of 1 to 5 for the level of sadness, where 1 would be that the text is nothing sad and 5 that it is very much.
5. Score from 1 to 5 for the level of malevolence, with 5 being much the level.
6. Score from 1 to 5 indicating if they liked the poem a lot or a little (5 a lot)
7. Score from 1 to 5 for the level of poeticity, where 5 would indicate that the poem is very poetic and 1 that it is little.
8. Score from 1 to 5 for the level of onomatopoeia (level that quantifies the use of this literary resource). 5 would indicate a lot of their presence

These annotations by users at a global level serve to analyze the correlation of them against different features derived from the individual value of the words that appear within the text, not considering *stopwords*. The purpose of these study is to check if the features could serve to predict a GAM for the poem.

The features are from three different levels: sub-lexical, lexical and inter-lexical. The lexical level captures the valence and arousal average values from the words present in the text, the inter-lexical level quantifies peaks, ranges and changes within the lexical affective content, and the sub-lexical level considers sources such as phonological information of the poems.

All these specifications are considered to define 55 affective features (using the 3 levels described above). Approximately the 50% of the explained variance is reached using only the lexical features, and together with the inter-lexical ones, the explained variance reaches 75%. This indicate that the best predictors would be the ones related to these two levels, particularly the average of valence and the average of arousal derived from the individual words.

Of course, considering only these two features would indicate that the order of the words in the text is irrelevant for the affective impact, and that is not the case; the order matters, and experiencing crescendos or affective decrescendos is something fundamental, so the span of the level of excitation is another key aspect to consider.

Together with that, the article also considers how the valence and arousal level evolve during the poem. This is important because, for example, poems are generally perceived as sadder when the valence of words is becoming less and less (more negative) and when the arousal at the end is lower, and poems are perceived as friendlier when the valence of the last words of the text is more positive. In this way it is important to consider the correlation coefficient between the vector of affectivity (arousal / valence) of the individual words with the vector of their positions in the text.

An article that also sheds light on the poetic sphere and that can help to propose a solution for affective computing is (Gloria Schultz Eastman, 2015). This article addresses two relevant issues in this regard. On the one hand, it reminds how poetry widely uses metaphors and figurative language (words open to many meanings and interpretations). This makes the extraction of affective information not always as obvious as simply assigning to each word a value contained in a repository and then composing all the individual values. Metaphors are interpreted in large part also from the subjectivity of the reader and from his personal experience, so it is not trivial and immediate to incorporate all the possible information. This is also applicable to the semantic scope (that is, to represent a single word with a vector based on fixed information from a repository). On the other hand it also mentions that the understanding of the words of a poetic text should not be done only based on the text itself but that a poem by a given author can be understood in greater depth if compared with other poems by that author or with poems of other authors. Two types of poetic learning transfers are mentioned in this regard:

- Near transfer: learn to understand a type of poem by an author in order to understand other texts by that author

- Far transfer: learn to understand poems by an author and that learning helps to understand any other type of document

As what would be sought is far transfer to avoid falling into an understanding only of a very specific niche, it is necessary to include different contexts for the affective modelling. For this reason the relevant features that can be automated to extract should cover as many texts as possible and not be done on subsets that are too particular.

As a last comment, the work of (Obermeier et al., 2013) is worth mentioning, where a study of the influence of poetry on affections is presented thanks to certain aesthetic and emotional elements such as the metric of the poem and its rhyme. Thus, the starting hypothesis is that metrics and rhyme have an impact on aesthetic perception, emotional involvement and valence. To verify this the authors analyze the influence of metrics and rhyme in the aesthetic and emotional perception of poetry, as well as their interaction with the lexicon, using the stanzas of the poems as references for the study.

It is true that features such as the lexical form and structural variables (for example, poetic form) influence in the emotional response provoked. However, they mention that there is abundant literature on how the valence of words influences the perception of the text, and that there is less information on how other formal features can affect. The article reminds that Aristotle already talks about how the sweetness of the metric and the harmony of the sound in a tragedy make the readers experience pleasure (*hedoné*) even considering that the text is drama. This happens because while contemplating the tragedy, the reader grasp that the beauty in the structural harmony of the text. Thus, metrics and rhyme are different features but both serve to structure poetry: they represent patterns of recurrence or similarity (for example, both serve to structure poetry over time).

Thus, with the starting hypothesis that metrics and rhyme influence the emotional response caused in the reader, the article conducts a research to empirically analyze it. To do that, they work with a group of 60 adults that listened to audios of German poems (100 poems from the 19th and 20th centuries). The poems had stanzas of 4 verses in which there were sets of poems with lexical differences (for instance, real words vs pseudowords; pseudowords were modified original words that kept the vowels but changed some consonant ensuring that they were still pronounceable). Poems also were divided depending on whether they had rhymes or not, or if they had accent or not.

With this, the users scored four metrics for the poems that they were listening: liking (aesthetic appreciation), intensity (power of emotional

response), perceived emotion (emotion that was expressed within the stanza) and felt emotion (emotion experienced by the users).

The results are as follows:

- **Liking:** results had better aesthetic ratings for poetry with metric as well as for stanzas with rhyme compared to those without it.
- **Intensity rating:** for all kind of poems the results were better with the stanzas that contain real words and not pseudowords.
- **Perceived emotion:** influence of lexicon, metric and rhyme (especially the last two); best score for stanzas with pseudowords if they don't have metric versus those that do. This last difference does not appear for poems with only real words
- **Felt emotion:** the main influence is the rhyme. There is also a triple interaction between lexical-metric-rhyme. When there's rhyme the emotion felt is stronger.

To conclude, the articles mentioned in this section show how there are different features that could be used to model the global GAM of a poetic text. These features can be affective, such as the valence or arousal. They can also be related to other areas, such as to the structure of the poem (rhyme, metric...). And they can be obtained by modelling the whole text or by modelling the individual stanzas of the poem. For the case of affective features, they can be inferred using as input the individual affective values of the words that appear in the text as long as there is an available corpus that contain those individual affective values, such as BAWL in (Ullrich et al., 2016).

However, it is important to note that the understanding of a text, and hence the context for the individual words, it is best done if the words are understood not only within the context of a specific poem or a specific author but in a bigger context that includes poems from other authors. This is something important in any text comprehension task, but it is even more critical for poems where the language used is sometimes full of metaphors and other stylistic figures not so easily understood. The proper comprehension of the text is important for both the semantic modelling of the poem but also for the affective one.

Corpora for affective word modelling in Spanish

Just as BAWL, as mentioned in (Ullrich et al., 2016) is a dataset used as source information for the affective modelling of individual words, there are similar corpora for Spanish vocabulary. Some of these corpora are described below.

In (Ferré et al., 2017) 2267 words are written in Spanish (along with their English translation) with the following fields⁹⁸:

- *Spanish_Word*: word in Spanish
- *English_Translation*: translation of that word into English
- *Hap_Mean*: average value associated with this feeling (happiness) thanks to the set of users
- *Hap_SD*: typical deviation associated with this feeling (happiness) thanks to the set of users
- *Ang_Mean*: idem for this feeling (anger)
- *Ang_SD*: idem for this feeling (anger)
- *Sad_Mean*: idem for this feeling (sadness)
- *Sad_SD*: idem for this feeling (sadness)
- *Fear_Mean*: idem for this feeling (fear)
- *Fear_SD*: idem for this feeling (fear)
- *Disg_Mean*: idem for this feeling (disgust)
- *Disg_SD*: idem for this feeling (disgust)
- *N*: number of subjects used in the sample

In (Guasch et al., 2016) 1400 words are written in Spanish with the following fields:

- *ID*: mere auto incremental field
- *Word*: word in Spanish
- *English Trans.*: translation of the words into English
- *POS*: Part of Speech tag for that word
- *VAL_M*: average value of the valence for the subjects that there are
- *VAL_SD*: standard deviation of the valence for the subjects that there are
- *VAL_N*: number of subjects used to obtain valence values
- *ARO_M*: idem for excitation level
- *ARO_SD*: idem for excitation level
- *ARO_N*: idem for excitation level
- *CON_N*: idem for concreteness
- *CON_SD*: idem for concreteness
- *CON_N*: idem for concreteness
- *IMA_M*: idem for imageability
- *IMA_SD*: idem for imageability
- *IMA_N*: idem for imageability
- *AVA_M*: idem for context availability
- *AVA_SD*: idem for context availability
- *AVA_N*: idem for context availability
- *FAM_M*: idem for familiarity⁹⁹
- *FAM_SD*: idem for familiarity
- *FAM_N*: idem for familiarity

Regarding the concepts used, **Concreteness** is defined as the degree of specificity of the word, being 1 when the word is very abstract and 7 when

it is very concrete. Words like ‘object’ are more abstract than others like ‘table’.

Imageability is defined as the easiness or difficulty of constructing a mental image associated with that word, being 1 when the word is very difficult to imagine and 7 when it is very easy. It is easier to imagine something with words like ‘flag’ than with others like ‘charity’.

Context availability is defined as the easiness or difficulty in associating that word with a context in which it could appear, being 1 when the word is very difficult to associate with a context and 7 when it is very easy. It is easier to construct sentences or search for examples of usage for words like ‘table’ than for others like ‘citizenship’

Familiarity is defined as the degree of familiarity, being 1 when the word is not very familiar and 7 when it is a lot. A word like ‘fish’ is more familiar than another like ‘quark’.

In (Stadthagen-Gonzalez et al., 2017) the following fields are collected for 14031 words¹⁰⁰:

- *Word*: dictionary word
- *ValencieMean*: average value of the inferred valence of the different subjects of the analysis
- *ArousalMean*: average value of the level of excitation inferred from the different subjects of the analysis
- *ValenceSD*: standard deviation of the valence values given by the different subjects
- *ArousalSD*: standard deviation of the excitation level values given by the different subjects
- *% ValenceRaters*: percentage of total subjects that have given a value to the valence
- *% ArousalRaters*: percentage of the total of subjects that has given a value to the level of excitation

Finally (Alonso et al., 2015) describes for 7040 words other characteristics such as the average age at which a word is usually learned (averageAoA), the minimum (Min) and maximum (Max) age and the deviation in these age data (SD), as well as the literary frequency with which it is usually found.

The Dataset

This chapter introduces the dataset included in this paper, Diachronic Spanish Sonnet Corpus with Psychological and Affective Labels, DISCO PAL. The first part of the chapter introduces the

⁹⁸ All the fields have ranges from 1 (minimum) to 5 (maximum)

⁹⁹ Familiarity is defined as the degree of familiarity, being 1 when the word is not very familiar and 7 when it is a lot. A word like ‘fish’ is more familiar than another like ‘quark’

¹⁰⁰ All the fields have ranges between 1 (minimum) and 9 (maximum)

participants that labelled the dataset, and after that the second part describes the dataset provided. The last part of the chapter briefly mentions how the procedure to label the sonnets took place.

Participants

As mentioned before, the labels were annotated by three experts in humanistic studies that belong to POSTDATA project.

Materials

DISCO PAL is a subset of a larger corpus, DISCO (Ruiz Fabo et al., 2018) DISCO that consists of 4085 sonnets in Spanish language from the 15th to 19th century. From that corpus, in order to create DISCO PAL, the experts of POSTDATA have annotated a subset of 230 sonnets, with 167 belonging to the 19th century, 9 belonging to the 18th century and the other 72 belonging to the interval of 15th to 17th century. This is a relevant fact to consider because some sonnets are written in old Spanish, something that can significantly affect all the text mining analysis applied to the poems. Also, the number of authors used is 47. Additionally, since the number of sonnets is 230, much bigger than 30, there are enough sonnets to propose a statistical analysis of significant value. With that, the corpus provided is very rich, with many different authors belonging to different centuries, in line with the proposals of the scientific literature (Gloria Schultz Eastman, 2015).

There are two types of features annotated: affective and psychological. Affective features are detailed in Table 1 and have a range of 1 to 4, being 1 the minimum value (the sonnet does not inspire that affection very much) and 4 the maximum (the sonnet does inspire that affection very much). The scale only uses integer values. Psychological features are binary values and indicate whether the sonnet is related to that concept (1) or not (0). These features are described in Table 2.

Table 1 Affective features used with a scale of 1 to 4, using integer values

| Affective features | |
|--------------------|-------------------------------------|
| Happiness | Disgust |
| Anger | Concreteness ¹⁰¹ |
| Sadness | Imageability ¹⁰² |
| Fear | Context availability ¹⁰³ |

¹⁰¹ Same as ‘Concreteness’ defined in the SOTA

¹⁰² Same as ‘Imageability’ described in the SOTA.

Table 2 Psychological features used with a scale of 0 to 1 using binary values

| Psychological features | |
|-------------------------------|--------------------------------|
| Soledad (Solitude) | Ansiedad (Anxiety) |
| Ilusión (Illusion) | Ira (Anger/Wrath) |
| Ensoñación (Daydream) | Inestabilidad (Instability) |
| Grandiosidad (Grandeur) | Idealización (Idealization) |
| Orgullo (Pride) | Depresión (Depression) |
| Irritabilidad (Irritability) | Desilusión (Disappointment) |
| Compulsión (Compulsion) | Prejuicio (Prejudice) |
| Aversión (Aversion/Loathing) | Inseguridad (Insecurity) |
| Impotencia (Helplessness) | Vulnerabilidad (Vulnerability) |
| Temor (Fear) | Obsesión (Obsession) |
| Dramatización (Dramatization) | |

The features mentioned before were annotated by three different domain experts belonging to POSTDATA project. Each of those experts has individually annotated the same 230 sonnets for all those features.

Procedure

As mentioned in the SOTA regarding affective modelling, there are many possible features that could be used to label a sonnet. However, the features finally used to annotate the poetic corpus are chosen considering not only the relevance within the literature but also their presence within affective corpora for Spanish words.

(Ulrich et al., 2016) uses BAWL database to obtain the individual affective value, expressed through affective features, of German words in order to analyze the GAM obtained by aggregating those individual values compared to the GAM labelled by the users. In a similar way, the affective features used to label the sonnets by POSTDATA experts are the same as the ones found within the corpora described in the previous chapter for Spanish words. Some of these features are the ones used by POSTDATA to annotate the sonnets. In order to facilitate the task, a system annotator program for POSDATA is provided so that they can write down for each sonnet information on a scale of 1 to 4 (1

¹⁰³ Same as ‘Context Availability’ described in the SOTA

being the minimum value and 4 the maximum value), as was indicated in Table 1.

The affective features are based on the literature mentioned in the SOTA of this article. Because of this, all the fields are the same as the ones described in (Ferré, Guasch, Martínez-García, Fraga, & Hinojosa, 2017) and (Guasch, Ferré, & Fraga, 2016), except for Familiarity since that feature was not included. Thus, the meaning of those features is the same as the one mentioned in that section. Together with this, the experts have annotated 21 psychological features that have been defined as relevant in the literature (García Franco & Manjarrés Riesco, 2016). They are described in Table 2. These annotations allow to calculate different metrics (such as precision, for example) in the recovery of poems. The experts have annotated the sonnets independently (without knowing the annotations from the other experts) and following the same sonnet order during a time period of one month. The experts did not know nor the author nor the time period of the different sonnets; they only had access to the text itself. They used a system annotation program where they had rows containing the sonnet texts and columns with the different variables. Each of them assigned a value within the available range in the corresponding cells.

Results and Discussion

Supplementary material

The materials included in this article are three CSV files with the annotations made by the experts, as well as a csv file with metadata information about the labelled sonnets. This metadata csv is included in order to allow the reference between the DISCO PAL and the original source DISCO. The fields included in the metadata csv are:

- *author*: author of the sonnet
- *year*: year or century of publication
- *title*: title of the sonnet
- *id_sonnet*: unique id used by DISCO for that sonnet
- *file_path*: file name path to that sonnet in the per-sonnet folder in DISCO

Reliability and validity of the dataset

A first approach to study the reliability and validity of DISCO PAL is to analyze the agreement between the three annotators. This is accomplished by obtaining the Krippendorff Alpha (Krippendorff, 2011), or k-alpha, for the annotations made by the 3 experts for each of the variables.

K-alpha is a metric that generalizes other metrics that are responsible for quantifying the reliability between annotators (inter-rater reliability), being able to work for both ordinal and nominal annotations, as well as with any number of annotators. K-alpha is a value up to 1, where 1 represents a full agreement. However, there are different criteria regarding when to consider that there is agreement between scorers. Sometimes strict criteria are used, in which only expert annotations are accepted as truly valid if there is a k-alpha of at least 0.8 (Carletta, 1996). Other laxer criteria set the minimum at 0.21, defining the following thresholds (J. Richard Landis & Gary G. Koch, 1977):

- $K < 0$: Very low
- $0 < K < 2$: Light
- $0.21 < K < 0.4$: Acceptable
- $0.41 < K < 0.60$: Moderate
- $0.61 < K < 0.80$: Substantial
- $0.81 < K < 1$: Perfect

The k-alpha results considering the three annotators together are shown in Table 3. This shows the following information.

There is acceptable coincidence ($K \geq 0.21$) for the following variables considering the annotations of the 3 experts: 'Anxiety', 'Aversion', 'Depression', 'Disappointment', 'Dramatization', 'Illusion', 'Helplessness', 'Instability', 'Anxiety', 'Aversion', 'Depression', 'Disappointment', 'Dramatization', 'Illusion', 'Helplessness', 'Instability', 'Insecurity', 'Anger', 'Obsession', 'Pride', 'Prejudice', 'Fear', 'Vulnerability', 'concreteness', 'context availability'. This corresponds to 54.84% of the total existing variables. It should be mentioned that all the terms that pass this check correspond to the set of psychological terms.

There is a slight coincidence ($0 > K < 0.21$) for the other variables considering the annotations of the 3 experts. In some cases the K is high (close to 0.21 although somewhat lower), something that occurs especially in psychological terms.

The greatest discrepancies are seen in the variables: arousal 0.0912, happiness 0.0279, valence 0.0051. In addition to the joint analysis of the scorers, Tables 4, 5 and 6 include comparisons between scorers (one versus one) to check which ones gave the greatest discrepancy.

Table 3 K-alpha values comparing the agreement of all annotators for each variable

| Variable | K-alpha | Variable | K-alpha | Variable | K-alpha |
|-----------------------|---------|----------------------|---------|-----------------------|---------|
| <i>Anxietiy</i> | 0.2475 | <i>Pride</i> | 0.2722 | <i>Solitude</i> | 0.1943 |
| <i>Aversion</i> | 0.2948 | <i>Prejudice</i> | 0.2745 | <i>anger</i> | 0.2010 |
| <i>Depression</i> | 0.2260 | <i>Fear (binary)</i> | 0.2738 | <i>arousal</i> | 0.0912 |
| <i>Disappointment</i> | 0.2291 | <i>Vulnerability</i> | 0.2826 | <i>disgust</i> | 0.1836 |
| <i>Dramatisation</i> | 0.2300 | <i>concreteness</i> | 0.2774 | <i>Fear (ordinal)</i> | 0.1642 |
| <i>Illusion</i> | 0.3063 | <i>context</i> | 0.2269 | <i>happiness</i> | 0.0279 |
| | | <i>availability</i> | | | |
| <i>Helplessness</i> | 0.2671 | <i>Compulsion</i> | 0.1932 | <i>imageability</i> | 0.2087 |
| <i>Instability</i> | 0.2540 | <i>Daydream</i> | 0.1852 | <i>sadness</i> | 0.1916 |
| <i>Insecurity</i> | 0.3123 | <i>Grandeur</i> | 0.1690 | <i>valence</i> | 0.0051 |
| <i>Anger</i> | 0.2501 | <i>Idealization</i> | 0.1791 | | |
| <i>Obsession</i> | 0.2603 | <i>Irritability</i> | 0.2058 | | |

Table 4 Comparison between annotators 1 and 2

| Variable | K-alpha | Variable | K-alpha | Variable | K-alpha |
|----------------|---------|---------------|---------|----------------|---------|
| Anxiety | 0.2363 | Pride | 0.1095 | Solitude | 0.1156 |
| Aversion | 0.2178 | Prejudice | 0.1111 | anger | 0.0956 |
| Depression | 0.1411 | Fear (binary) | 0.2402 | arousal | 0.1082 |
| Disappointment | 0.1065 | Vulnerability | 0.245 | disgust | 0.0832 |
| Dramatization | 0.2749 | concreteness | 0.167 | fear (ordinal) | 0.0993 |
| Illusion | 0.1986 | context | 0.1779 | happiness | 0.0339 |
| | | availability | | | |
| Helplessness | 0.156 | Compulsion | 0.0568 | imageability | 0.1918 |
| Instability | 0.2195 | Daydream | 0.0866 | sadness | 0.0757 |
| Insecurity | 0.2734 | Grandeur | 0.009 | valence | 0.1075 |
| Anger | 0.1886 | Idealization | 0.009 | Solitude | 0.1156 |
| Obsession | 0.3102 | Irritability | 0.1323 | | |

Table 5 Comparison between annotators 1 and 3

| Variable | K-alpha | Variable | K-alpha | Variable | K-alpha |
|----------------|---------|---------------|---------|----------------|---------|
| Anxiety | 0.0842 | Pride | 0.1156 | Solitude | 0.1943 |
| Aversion | 0.1806 | Prejudice | 0.158 | anger | 0.0447 |
| Depression | 0.0115 | Fear (binary) | 0.1117 | arousal | -0.026 |
| Disappointment | 0.0867 | Vulnerability | 0.1513 | disgust | 0.0528 |
| Dramatization | 0.1222 | concreteness | 0.0508 | fear (ordinal) | 0.0517 |
| Illusion | 0.2382 | context | 0.029 | happiness | -0.0575 |
| | | availability | | | |
| Helplessness | 0.1583 | Compulsion | 0.2079 | imageability | 0.034 |
| Instability | 0.1639 | Daydream | 0.0759 | sadness | 0.0393 |
| Insecurity | 0.188 | Grandeur | -0.0467 | valence | -0.0447 |
| Anger | 0.0506 | Idealization | 0.0722 | Solitude | 0.0295 |
| Obsession | 0.1669 | Irritability | 0.0913 | | |

Table 6 Comparison between annotators 2 and 3

| Variable | K-alpha | Variable | K-alpha | Variable | K-alpha |
|----------------|---------|---------------|---------|----------------|---------|
| Anxiety | 0.3884 | Pride | 0.595 | Solitude | 0.4663 |
| Aversion | 0.46 | Prejudice | 0.518 | anger | 0.2491 |
| Depression | 0.5536 | Fear (binary) | 0.4545 | arousal | 0.1144 |
| Disappointment | 0.4762 | Vulnerability | 0.4337 | disgust | 0.2782 |
| Dramatization | 0.2819 | concreteness | 0.1361 | fear (ordinal) | 0.2858 |
| Illusion | 0.49 | context | 0.0257 | happiness | -0.0723 |
| | | availability | | | |
| Helplessness | 0.491 | Compulsion | 0.2997 | imageability | 0.0363 |
| Instability | 0.3659 | Daydream | 0.4284 | sadness | 0.1648 |
| Insecurity | 0.4564 | Grandeur | 0.5543 | valence | 0.0347 |
| Anger | 0.4848 | Idealization | 0.4295 | Solitude | 0.1943 |
| Obsession | 0.267 | Irritability | 0.3743 | | |

Tables 4, 5 and 6 help to see how k-alpha values change significantly according to which pair of annotators are compared. Again, the validation of variables is done for those with k-alpha greater than 0.21. In particular, the following is checked:

- For scorers 1 and 2 (Table 5) only the following variables are validated, representing 25.81% of the total variables: 'Anxiety', 'Aversion', 'Dramatization', 'Instability', 'Insecurity', 'Obsession', 'Fear', 'Vulnerability'
- For annotators 1 and 3 (Table 6), only one variable is validated, which represents 3.23% of the total variables: 'Illusion'
- For annotators 2 and 3 the following variables are validated, representing 77.42% of the total variables (that includes all psychological variables): 'Anxiety', 'Aversion', 'Compulsion', 'Depression', 'Disappointment', 'Dramatization', 'Daydream', 'Grandeur', 'Idealization', 'Illusion', 'Helplessness', 'Instability', 'Insecurity', 'Anger', 'Irritability', 'Obsession', 'Pride', 'Prejudice', 'Solitude', 'Fear', 'Vulnerability', 'anger', 'disgust', 'fear'.
- For annotators 2 and 3 the k-alpha is higher for most of the variables comparing it with the previous analysis (including for all the psychological terms). However, it is especially low (including one negative value) for the following variables: 'happiness', 'valence'
- The biggest discrepancies are with the annotator 1, especially between 1 and 3 where only one variable has k-alpha greater than 0.21: 'Illusion'

This study for the agreement between annotators can be further enriched by comparing the mean of the vector of annotations per variable between each pair of annotators for those variables in which k-alpha is greater than 0.21. This seeks to study whether there are significant differences between the averages per annotator, and with it seeing which variables are the ones with the highest level of agreement. Therefore, an ANOVA (one-way ANOVA) analysis is carried out to compare the mean for the different variables for each pair of annotators to formally see if there is a significant difference between them. The null hypothesis in the contrast (H0) is that there is no difference. If the p-value of the contrast is less than a threshold (0.05 as a reference), H0 is discarded and it is said that for that specific variable there are significant differences in the mean value for the assigned annotations. If the p-value is above 0.05, H0 cannot be discarded and therefore there may be no significant differences.

Table 7 shows the variables with k-alpha greater than 0.21 for each pair of annotators together with the p-value obtained by comparing the results of these variables.

An additional consideration is that in the binary variables there is a lot of imbalanced data because there are many more sonnets without 1 label than with it. As an example of this situation is illustrated comparing the labels of a category for two annotators through the confusion matrix on which to calculate precision, recall, and F1 metrics to assess whether an adequate number of sonnets coincide with labels of 1. Thus, for example, for the variable 'Prejudice' analyzing the labels of the annotator 2 versus 1, the following confusion matrix is found.

$$\binom{192 \quad 22}{12 \quad 4}$$

The value 4 represents number of sonnets that are labelled with a 1 by both annotators. However, the annotator 2 has labelled 22 sonnets with a 1 when the annotator 1 used a 0. This represents approximately 0.15 of the success rate, just like the

Table 7 One-way ANOVA of each variable for each pair of annotators

| pair | variable | K alpha | P value one-way |
|--------|----------------|---------|-----------------|
| [1, 2] | Anxiety | 0.2363 | 0.1251 |
| [1, 2] | Aversion | 0.2178 | 0.0017 |
| [1, 2] | Dramatization | 0.2749 | 0.7795 |
| [1, 2] | Instability | 0.2195 | 0.4572 |
| [1, 2] | Insecurity | 0.2734 | 0.0915 |
| [1, 2] | Obsession | 0.3102 | 0.0874 |
| [1, 2] | Fear | 0.2402 | 0.1236 |
| [1, 2] | Vulnerability | 0.2450 | 0.0767 |
| [1, 3] | Illusion | 0.2382 | 0.0115 |
| [2, 3] | Anxiety | 0.3884 | 0.0000 |
| [2, 3] | Aversion | 0.4600 | 0.0000 |
| [2, 3] | Compulsion | 0.2997 | 0.0000 |
| [2, 3] | Depression | 0.5536 | 0.0289 |
| [2, 3] | Disappointment | 0.4762 | 0.3482 |
| [2, 3] | Dramatization | 0.2819 | 0.0016 |
| [2, 3] | Daydream | 0.4284 | 0.1198 |
| [2, 3] | Grandeur | 0.5543 | 0.0010 |

accuracy. It is also interesting to analyze the recall value. The recall metric shows that 16 sonnets that the annotator 1 labelled with 1 (12 plus 4) the annotator 2 labelled with 0.

Table 8 in the Annexes shows the results of the metrics for all pairs of scorers, including an F1 metric to jointly weigh both accuracy and recall.

| pair | variable | K alpha | P value one-way |
|--------|----------------|---------|-----------------|
| [2, 3] | Idealization | 0.4295 | 0.0000 |
| [2, 3] | Illusion | 0.4900 | 0.0032 |
| [2, 3] | Helplessness | 0.4910 | 0.4572 |
| [2, 3] | Instability | 0.3659 | 0.0000 |
| [2, 3] | Insecurity | 0.4564 | 0.0268 |
| [2, 3] | Anger | 0.4848 | 0.0024 |
| [2, 3] | Irritability | 0.3743 | 0.0229 |
| [2, 3] | Obsession | 0.2670 | 0.0012 |
| [2, 3] | Pride | 0.5950 | 0.0562 |
| [2, 3] | Prejudice | 0.5180 | 0.3522 |
| [2, 3] | Solitude | 0.4663 | 0.0044 |
| [2, 3] | Fear (binary) | 0.4545 | 0.0007 |
| [2, 3] | Vulnerability | 0.4337 | 0.0000 |
| [2, 3] | anger | 0.2491 | 0.8963 |
| [2, 3] | disgust | 0.2782 | 0.0003 |
| [2, 3] | fear (ordinal) | 0.2858 | 0.0781 |

With the results of Table 8, the discrepancies with respect to annotator 1 are reaffirmed since with that annotator the metrics are systematically worse for all variables.

The different annotations made by the experts can be combined into one vector per variable in order to have only one reference for the different experiments that could be done using this dataset. The initial proposal to construct this vector is to use the median of the value of the 3 annotators for each variable per sonnet.

With that, each sonnet has one vector for each variable with the median value of the annotations of the three experts. Thus, if for example two scorers give a 1 in a variable for a sonnet and the third scorer gives a 0, the median is 1 and that is the final value used in that variable for that sonnet.

Tables 9, 10 and 11 in Annex (Annex) show the comparison with both the ANOVA and k-alpha

contrast of that vector with the median values of the annotators per sonnet and variable.

In all cases except for the term ‘Grandeur’ with the annotator 1 (k-alpha = 0.1845) a k-alpha greater than 0.21 is obtained. This shows how the vector with the median values has a high level of agreement with all the annotators for practically all the variables (although in some cases the mean is different, as shown by the p values of the ANOVA contrast). In the analysis that have been exposed, it can be observed that for two of the affective variables, arousal and valence, there is a special discrepancy between the annotators.

Conclusion and Future Work

This chapter concludes with a final reflection based on the results of the analyses carried out as well as indicating possible lines of research that can be pursued.

Conclusions

This article presents a dataset of Spanish poetry (sonnets) with 230 annotated sonnets from authors of different time periods (from 15th to 19th century). These sonnets were annotated using both affective labels that indicate the intensity level of that affection within the sonnet, and concepts that belong to the psychological domain, indicating whether a sonnet content is related to that concept or not.

The annotations were labelled by three domain experts that belong to POSTDATA project (UNED). Those experts have annotated the sonnets independently (without knowing the annotations from the other experts) and following the same sonnet order. The experts did not know nor the author nor the time period of the different sonnets; they only had access to the text itself.

This article also conducts an analysis on the level of agreement of the annotations labelled by the three experts. The result is that around the 54% of the variables have an adequate agreement, all of them belonging to the psychological labels. This analysis was made comparing the three annotators together. However, the article proposes the usage of the median value for the three annotators since it seems that two of them are similar and is the third one that is introducing more disagreement. This median vector reaches an acceptable agreement for all the variables and all the annotators except for only one variable with respect to one annotator.

Future work

This subsection details the possible lines of research that can be pursued following the results presented in this article. There are two main group of research lines that are considered at this point. One is directly related to semantic modelling and the other one is related to the research that can be carried out in the area of affective modelling thanks to the other labels annotated within DISCO PAL.

First, the psychological labels annotated in DISCO PAL can be used to analyze the semantic modelling of Spanish poetry (sonnets). Thanks to the dataset provided there are labels that indicate when a sonnet content is related to a concept, expressed through those annotations. This allows to, for instance, semantically model a sonnet with techniques such as word embeddings and use supervised machine learning algorithms to see if it is possible to infer the content labels using that semantic information.

Another future line that can be pursued is working in everything related to affective modelling of poetry. DISCO PAL includes 10 affective labels that can be used to study how to infer the GAM of a Spanish sonnet. This could be accomplished by using ML models that predict the GAM labels based

on the semantic vector of the sonnet, or it could also be done by using the corpora for the individual affective value of words, described in Section 3.3, trying to map the values of all the individual words of a poem to the global GAM labelled.

To end with, another line of work to continue is the standardization of these labels and include them within the TEI XML format used for the ontology of DISCO sonnets.

Author note: This work was possible thanks to the POSTDATA project, and particularly because of Salvador Ros Muñoz, Laura Alises, Marie Olivier and Aroa Rabdán.

References

- Alonso, M., Fernandez, A., & Díez, E. (2015). Subjective age-of-acquisition norms for 7,039 spanish words. *Behavior Research Methods*, 47(1), 268-274. doi:10.3758/s13428-014-0454-2
- Amigó Cabrera, E., & Fresno Fernández, V. (2018). An axiomatic approach to unsupervised distributional semantic compositionality.(Manuscript in preparation)
- Barros, L., Rodriguez, P., & Ortigosa, A. (Sep 2013). Automatic classification of literature pieces by emotion detection: A study on Quevedo's poetry. Paper presented at the pp. 141-146. doi:10.1109/ACII.2013.30
- Carletta, J. (1996). Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics*, 22(2), 249. Retrieved from http://gateway.proquest.com/openurl?ctx_ver=Z39.88-2003&xri:pqil:res_ver=0.2&res_id=xri:ilcs-us&rft_id=xri:ilcs:rec:abell:R03037566
- Czernianin, W. (2016). Poetry as a therapeutic medium in shaping mood. *Journal of Poetry Therapy*, 29(3), 135-145. doi:10.1080/08893675.2016.1199513
- Ferré, P., Guasch, M., Martínez-García, N., Fraga, I., & Hinojosa, J. (2017). Moved by words: Affective ratings for a set of 2,266 spanish words in five discrete emotion categories. *Behavior Research Methods*, 49(3), 1082-1094. doi:10.3758/s13428-016-0768-3
- García Franco, J. D., & Manjarrés Riesco, Á. (2016). Modelado de los trastornos de la personalidad en OWL - lógica descriptiva y aplicación al desarrollo de un sistema experto de diagnóstico clínico y educativo; Universidad Nacional de Educación a Distancia (UNED). , 157.
- Gloria Schultz Eastman. (2015). Making metaphor visible: The common core, poetry, and visual literacy. *The English Journal*, 104(6), 40-46. Retrieved from Research Library Prep database. Retrieved from <https://www.jstor.org/stable/24484429>
- Guasch, M., Ferré, P., & Fraga, I. (2016). Spanish norms for affective and lexico-semantic variables for 1,400 words. *Behavior Research Methods*, 48(4), 1358-1369. doi:10.3758/s13428-015-0684-y
- J. Richard Landis, & Gary G. Koch. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174. doi:10.2307/2529310

Kaur, J., & Saini, J. (Feb 24, 2017). (Feb 24, 2017). Punjabi poetry classification. Paper presented at the pp. 1-5. doi:10.1145/3055635.3056589

Krippendorff, K. (2011). Computing krippendorff's alpha-reliability Scholarly Commons.

Liu, B. (2012). Sentiment analysis and opinion mining Morgan & Claypool Publishers. doi:10.2200/S00416ED1V01Y201204HLT016

Obermeier, C., Menninghaus, W., von Koppenfels, M., Raettig, T., Schmidt-Kassow, M., Otterbein, S., et al. (2013). Aesthetic and emotional effects of meter and rhyme in poetry. *Frontiers in Psychology*, 4, 10. doi:10.3389/fpsyg.2013.00010

Parastoo, A., Amenehsadat, K., & Shahla, M. (2016). The effect of group poetry therapy on cognitive emotion regulation and moods of psychotic patients: The case of razi psychiatric hospital, tehran

Ruiz Fabo, Pablo, Martínez Cantón, Clara Isabel y Calvo Tello, José (2018) .DISCO: Diachronic Spanish Sonnet Corpus. DHd Konferenz 2018: Kritik der digitalen Vernunft (2018. Köln) At: Colonia (Germany). (2018-02-26)

Shapiro, J., & Rucker, L. (2003). Can poetry make better doctors? teaching the humanities and arts to medical students and residents at the university of california, irvine, college of medicine. *Academic Medicine : Journal of the Association of American Medical Colleges*, 78(10), 953-957. doi:10.1097/00001888-200310000-00002

Stadthagen-Gonzalez, H., Imbault, C., Pérez Sánchez, M., & Brysbaert, M. (2017). Norms of valence and arousal for 14,031 spanish words. *Behavior Research Methods*, 49(1), 111-123. doi:10.3758/s13428-015-0700-2

Võ, M., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M., & Jacobs, A. (2009). The berlin affective word list

reloaded (BAWL-R). *Behavior Research Methods*, 41(2), 534-538. doi:10.3758/BRM.41.2.534

Annexes

Table 8 Metrics comparing the annotations between the different domain experts

| Feature column | Precision (1, 2) | Precision (1, 3) | Precision (2, 3) | Recall (1, 2) | Recall (1, 3) | Recall (2, 3) | F1 (1, 2) | F1 (1, 3) | F1 (2, 3) |
|----------------|---------------------|---------------------|---------------------|------------------|------------------|------------------|--------------|--------------|--------------|
| Anxiety | 0.41 | 0.39 | 0.90 | 0.52 | 0.20 | 0.37 | 0.46 | 0.26 | 0.53 |
| Aversion | 0.45 | 0.48 | 0.93 | 0.65 | 0.38 | 0.51 | 0.53 | 0.42 | 0.66 |
| Compulsion | 0.22 | 0.37 | 0.83 | 0.45 | 0.29 | 0.32 | 0.29 | 0.32 | 0.46 |
| Depression | 0.31 | 0.22 | 0.83 | 0.28 | 0.12 | 0.49 | 0.29 | 0.15 | 0.61 |
| Disappointment | 0.24 | 0.22 | 0.53 | 0.27 | 0.30 | 0.63 | 0.26 | 0.26 | 0.58 |
| Dramatization | 0.60 | 0.55 | 0.69 | 0.62 | 0.39 | 0.48 | 0.61 | 0.46 | 0.56 |
| Daydream | 0.32 | 0.34 | 0.62 | 0.24 | 0.19 | 0.44 | 0.27 | 0.24 | 0.51 |
| Grandeur | 0.47 | 0.47 | 0.90 | 0.40 | 0.25 | 0.57 | 0.43 | 0.33 | 0.70 |
| Idealization | 0.37 | 0.41 | 0.91 | 0.53 | 0.33 | 0.51 | 0.43 | 0.37 | 0.66 |
| Illusion | 0.40 | 0.51 | 0.81 | 0.42 | 0.32 | 0.48 | 0.41 | 0.40 | 0.60 |
| Helplessness | 0.38 | 0.39 | 0.66 | 0.39 | 0.35 | 0.59 | 0.38 | 0.37 | 0.62 |
| Instability | 0.40 | 0.48 | 0.81 | 0.45 | 0.23 | 0.35 | 0.42 | 0.31 | 0.49 |
| Insecurity | 0.35 | 0.32 | 0.71 | 0.49 | 0.29 | 0.45 | 0.40 | 0.30 | 0.55 |
| Anger | 0.32 | 0.24 | 0.82 | 0.43 | 0.18 | 0.46 | 0.37 | 0.21 | 0.59 |
| Irritability | 0.22 | 0.21 | 0.62 | 0.32 | 0.18 | 0.37 | 0.26 | 0.19 | 0.46 |
| Obsession | 0.32 | 0.31 | 0.62 | 0.50 | 0.18 | 0.24 | 0.39 | 0.23 | 0.34 |
| Pride | 0.35 | 0.37 | 0.82 | 0.38 | 0.30 | 0.61 | 0.36 | 0.33 | 0.70 |
| Prejudice | 0.15 | 0.20 | 0.65 | 0.25 | 0.25 | 0.50 | 0.19 | 0.22 | 0.57 |
| Solitude | 0.37 | 0.36 | 0.79 | 0.30 | 0.16 | 0.43 | 0.33 | 0.22 | 0.56 |
| Fear | 0.48 | 0.43 | 0.82 | 0.58 | 0.33 | 0.52 | 0.52 | 0.38 | 0.64 |
| Vulnerability | 0.57 | 0.57 | 0.89 | 0.67 | 0.41 | 0.54 | 0.62 | 0.48 | 0.68 |

Table 9 Comparison of the vector with median values against annotator 1

| pair | variable | K alpha | P value one-way | variable | K alpha | P value one-way |
|--------|----------------|---------|-----------------|----------------------|---------|-----------------|
| [1, 4] | Anxiety | 0.4532 | 0.1520 | Obsession | 0.5807 | 0.4037 |
| [1, 4] | Aversion | 0.4740 | 0.0001 | Pride | 0.2699 | 0.1288 |
| [1, 4] | Compulsion | 0.4313 | 0.7026 | Prejudice | 0.2288 | 0.5659 |
| [1, 4] | Depression | 0.2344 | 0.0264 | Solitude | 0.2755 | 0.0014 |
| [1, 4] | Disappointment | 0.2938 | 0.6045 | Fear (binary) | 0.4433 | 0.4257 |
| [1, 4] | Dramatization | 0.5576 | 0.3474 | Vulnerability | 0.4892 | 0.5739 |
| [1, 4] | Daydream | 0.2812 | 0.0024 | anger | 0.3848 | 0.0280 |
| [1, 4] | Grandeur | 0.1845 | 0.0004 | arousal | 0.5049 | 0.2961 |
| [1, 4] | Idealization | 0.3191 | 0.7693 | concreteness | 0.6449 | 0.6679 |
| [1, 4] | Illusion | 0.4221 | 0.0311 | context availability | 0.6392 | 0.3615 |
| [1, 4] | Helplessness | 0.3796 | 0.2344 | disgust | 0.3724 | 0.0062 |
| [1, 4] | Instability | 0.4486 | 0.0097 | fear (ordinal) | 0.3292 | 0.0001 |
| [1, 4] | Insecurity | 0.4705 | 0.5957 | happiness | 0.6300 | 0.8008 |
| [1, 4] | Anger | 0.3606 | 0.6302 | imageability | 0.6733 | 0.8794 |
| [1, 4] | Irritability | 0.3829 | 0.6622 | sadness | 0.4416 | 0.1265 |
| [1, 4] | valence | 0.6753 | 0.5922 | | | |

Table 10 Comparison of the vector with median values against annotator 2

| pair | variable | K alpha | P value one-way | variable | K alpha | P value one-way |
|--------|----------------|---------|-----------------|----------------------|---------|-----------------|
| [2, 4] | Anxiety | 0.6883 | 0.0030 | Pride | 0.7498 | 0.0328 |
| [2, 4] | Aversion | 0.7033 | 0.0017 | Prejudice | 0.6366 | 0.0296 |
| [2, 4] | Compulsion | 0.4498 | 0.0000 | Solitude | 0.7146 | 0.0559 |
| [2, 4] | Depression | 0.7676 | 0.0822 | Fear (binary) | 0.7542 | 0.0194 |
| [2, 4] | Disappointment | 0.7429 | 0.3110 | Vulnerability | 0.7480 | 0.0196 |
| [2, 4] | Dramatization | 0.7090 | 0.2226 | anger | 0.7137 | 0.1764 |
| [2, 4] | Daydream | 0.6637 | 0.1198 | arousal | 0.5817 | 0.0005 |
| [2, 4] | Grandeur | 0.7807 | 0.0402 | concreteness | 0.7336 | 0.0861 |
| [2, 4] | Idealization | 0.6358 | 0.0002 | context availability | 0.7152 | 0.4491 |
| [2, 4] | Illusion | 0.6630 | 0.0100 | disgust | 0.7054 | 0.0044 |
| [2, 4] | Helplessness | 0.7207 | 0.1958 | fear (ordinal) | 0.7010 | 0.6839 |
| [2, 4] | Instability | 0.6220 | 0.0009 | happiness | 0.5542 | 0.0006 |
| [2, 4] | Insecurity | 0.6980 | 0.0268 | imageability | 0.6901 | 0.9242 |
| [2, 4] | Anger | 0.7303 | 0.0311 | sadness | 0.7585 | 0.1826 |
| [2, 4] | Irritability | 0.6117 | 0.0334 | valence | 0.5344 | 0.4233 |
| [2, 4] | Obsession | 0.5819 | 0.0115 | | | |

Table 11 Comparison of the vector with median values against annotator 3

| pair | variable | K alpha | P value one-way | variable | K alpha | P value one-way |
|--------|----------------|---------|-----------------|----------------------|---------|-----------------|
| [3, 4] | Anxiety | 0.6611 | 0.0469 | Obsession | 0.5730 | 0.4511 |
| [3, 4] | Aversion | 0.7407 | 0.0834 | Pride | 0.8215 | 0.8216 |
| [3, 4] | Compulsion | 0.8033 | 0.5044 | Prejudice | 0.7720 | 0.2068 |
| [3, 4] | Depression | 0.7493 | 0.6511 | Solitude | 0.7247 | 0.3435 |
| [3, 4] | Disappointment | 0.6742 | 0.0514 | Fear (ordinal) | 0.6790 | 0.3015 |
| [3, 4] | Dramatization | 0.5696 | 0.0529 | Vulnerability | 0.6764 | 0.0268 |
| [3, 4] | Daydream | 0.7244 | 0.0001 | anger | 0.6609 | 0.2011 |
| [3, 4] | Grandeur | 0.7624 | 0.2127 | arousal | 0.4696 | 0.0000 |
| [3, 4] | Idealization | 0.7781 | 0.2285 | concreteness | 0.5774 | 0.1481 |
| [3, 4] | Illusion | 0.7977 | 0.7086 | context availability | 0.4876 | 0.0002 |
| [3, 4] | Impotence | 0.7448 | 0.5821 | disgust | 0.6968 | 0.3077 |
| [3, 4] | Instability | 0.6794 | 0.3370 | fear (binary) | 0.6535 | 0.0260 |
| [3, 4] | Insecurity | 0.7024 | 0.0001 | happiness | 0.2872 | 0.0000 |
| [3, 4] | Anger | 0.7238 | 0.3728 | imageability | 0.4576 | 0.0000 |
| [3, 4] | Irritability | 0.7037 | 0.8802 | sadness | 0.6459 | 0.0110 |
| [3, 4] | valence | 0.2694 | 0.0618 | | | |

