

Trabajo Fin de Grado

Predicción de rasgos de personalidad a partir
de marcadores lingüísticos:
revisión histórica y estudio experimental

Personality Prediction through linguistic
Features:
a historical Review and an experimental Study

Autora

Covadonga Cuesta Fontela

Directora

María del Carmen Horno Chéliz

Facultad de Filosofía y Letras
2018

Índice

1.	Introducción	3
2.	Características lingüísticas asociadas a los rasgos de personalidad. El estudio de Big Data.....	4
2.1.	Teoría de la personalidad	4
2.1.1	La extraversión	5
2.1.2	El neuroticismo.....	6
2.1.3	El psicoticismo	7
2.2.	Relación entre lenguaje y características psicológicas: de los primeros estudios a los más recientes	7
2.2.1	Métodos de análisis.....	9
2.2.2	Resultados obtenidos.....	10
2.3.	Conclusiones	13
3.	¿Es posible replicar los resultados con muestras lingüísticas pequeñas? Un análisis experimental sobre el rasgo de extroversión.	14
3.1.	Objetivos e hipótesis de partida	14
3.2.	Participantes.....	16
3.3.	Materiales.....	16
3.4.	Procedimiento	16
3.5.	Resultados y análisis de datos.....	18
3.6.	Discusión de resultados y conclusiones.....	20
4.	Conclusiones.....	23
5.	Bibliografía	26
	Anexo I: Inventario de personalidad de Eysenck para adultos – Formato B	29
	Anexo II: Datos del análisis lingüístico obtenidos en el recuento manual	32
	Anexo III: Datos obtenidos del análisis con la página web www.linguakit.com	34

1. Introducción

El lenguaje, como capacidad de aprender y usar lenguas naturales del ser humano, es una propiedad universal (de especie). No obstante, a nivel individual, se realiza de forma distinta en cada hablante. Como dicen Pennebaker y Tausczik (2010), “el lenguaje es la más común y fiable de las maneras de traducir las emociones y pensamientos a material que los otros puedan entender. Las palabras y el lenguaje son, por lo tanto, la esencia misma de la psicología y la comunicación” (traducción propia).

A lo largo del tiempo, estas particularidades en el empleo de la lengua se han utilizado, principalmente, para identificar la autoría de textos anónimos. Hoy en día, gracias a los avances en el estudio de la psicología y la lingüística y al uso masivo de las redes sociales, se puede, no solo reconocer al autor de un texto, si no determinar ciertas características de su personalidad.

Para ello, las redes sociales y la comunicación vía Internet juegan un papel fundamental, ya que ofrecen algo que no existía hasta ahora: un corpus casi ilimitado de mensajes escritos, en todo tipo de registros y, generalmente, dirigidos a amigos y/o conocidos. Esto permite el estudio de la lengua escrita emitida de forma natural y espontánea, algo que no era posible a partir de los textos previos a la era de Internet, de naturaleza formal y rígida. El alcance teórico de estos estudios está aún por determinar y su potencial práctico todavía no ha sido del todo explorado.

En la primera parte de este trabajo se realizará una revisión crítica de los estudios sobre lenguaje y personalidad centrados en los Big Data (apartado 2). En primer lugar, se introducirá la teoría de personalidad y se incluirán las descripciones de los tres factores de personalidad que más aparecen en la bibliografía: extraversión, neuroticismo y psicoticismo (epígrafe 2.1). A continuación, se hará un repaso de los estudios experimentales más relevantes en la relación entre el lenguaje y los rasgos de personalidad desde el siglo pasado (epígrafe 2.2). En concreto, se presentarán los dos métodos de análisis – el cerrado y el abierto – a los que estos estudios han dado lugar (epígrafe 2.2.1) y los resultados obtenidos en los mismos (epígrafe 2.2.2). El apartado finalizará con una conclusión a modo de reflexión sobre la información presentada (epígrafe 2.3).

En la segunda parte del trabajo se estudiará experimentalmente la posibilidad de llegar a los mismos resultados con muestras pequeñas (apartado 3). Se dividirá en la

presentación del experimento, en la que se detallarán el procedimiento seguido (epígrafes 3.1, 3.2, 3.3 y 3.4), los resultados y el análisis de datos (epígrafe 3.5) y la discusión de resultados y las conclusiones (epígrafe 3.6).

Por último, se añadirá un apartado de conclusiones generales (apartado 4), en el que se comentarán aspectos relativos tanto a la teoría existente como al experimento realizado.

2. Características lingüísticas asociadas a los rasgos de personalidad. El estudio de Big Data.

2.1. Teoría de la personalidad

El estudio de la personalidad no cuenta, hoy en día, con una teoría completamente aceptada por todos los psicólogos. En este trabajo se hablará de uno de los modelos de aproximación de rasgos: el modelo factorial-biológico de Eysenck (PEN, cuyas siglas hacen referencia a los tres factores de personalidad: psicoticismo, extroversión y neuroticismo). La justificación de que nos centremos en este se basa en que es uno de los modelos que cuenta con mayor aceptación y se sustenta en un amplio fundamento biológico que lo respalda.

Eysenck (1992) parte de un modelo cuantitativo de las diferencias individuales y su modelo es descriptivo, explicativo y causal. Es la teoría más completa de las que existen porque incluye estudios sobre el origen, el desarrollo, la naturaleza y la consecuencia de los rasgos. Define la personalidad como:

Una organización más o menos estable y duradera del carácter, temperamento, intelecto y físico de una persona que determina su adaptación única al ambiente. El carácter denota el sistema más o menos estable y duradero de la conducta conativa (voluntad) de una persona; el temperamento, su sistema más o menos estable y duradero de la conducta afectiva (emoción); el intelecto, su sistema más o menos estable y duradero de la conducta cognitiva (inteligencia); el físico, su sistema más o menos estable y duradero de la configuración corporal y de la dotación neuroendocrina. (Eysenck y Eysenck, 1985: 9, apud Schmidt 2010: 9)

Las ventajas de este modelo son que está claramente formulado, evitando las ambigüedades, permite hacer pronósticos contrastables en muchos ámbitos del comportamiento y es posible contrastar sus resultados con los obtenidos por otros

métodos. Defiende, por último, la heredabilidad y, por lo tanto, la base biológica de la personalidad.

Eysenck (1992) propone la existencia de tres grandes factores temperamentales: la extraversión (E), el neuroticismo (N) y el psicoticismo (P). En los siguientes subapartados se procederá a la presentación de los tres y se incluirán también experimentos que demuestran su relevancia a nivel clínico y cotidiano.

2.1.1 La extraversión

La extraversión es el factor que ha recibido más atención a lo largo del tiempo y cuya base teórica y biológica está más consolidada. El típico extravertido es definido por Eysenck en su inventario de personalidad para adultos (1954:3) como:

Le gusta participar en fiestas, tiene muchos amigos, necesita tener gente hablándole, no le gusta leer o estudiar para sí mismo, es despreocupado, bromista y optimista, le gusta reír y ser divertido, prefiere más actuar que pensar, tiende a ser agresivo y pierde rápidamente el temple. Anhela situaciones y es mudable, está fijado al ambiente, actúa sobre estímulos del momento y es generalmente un individuo impulsivo. Es aficionado a los trabajos, tiene respuestas "listas". En conjunto, sus sentimientos no son mantenidos bajo ajustado control y no es siempre previsible, ni confiado.

Todos los factores están compuestos por rasgos típicos que los definen. De los individuos extravertidos, suelen mencionarse la sociabilidad, vivacidad, actividad, asertividad, búsqueda de sensaciones socializadas, despreocupación, dominancia, espontaneidad y espíritu aventurero (Eysenck, 1990). De ellos se destacan dos centrales: la sociabilidad y la actividad (Eysenck y Eysenck, 1985).

Como se puede comprobar a continuación, el interés en la extroversión está justificado. Por un lado, se trata del factor cuya evidencia biológica es más robusta. Además, varios estudios han demostrado que los rasgos de personalidad influyen en muchos aspectos del comportamiento individual relacionado con el rendimiento en el trabajo y que el éxito de las tareas interpersonales depende de la personalidad de los participantes. Se han encontrado pruebas de la influencia de la extraversión sobre la capacidad de liderazgo (Hogan, Curphy, & Hogan, 1994) y la emergencia de líderes en grupos desorganizados (Stogdill, 1948). Furnham, Jackson, & Miller (1999) estudiaron la correlación de los factores de personalidad con distintos aspectos del rendimiento

general. Encontraron una correlación significativa entre altas puntuaciones en extraversión y la dedicación al activismo. Por el contrario, las bajas puntuaciones en extraversión correlacionaban con carreras teóricas.

Por otro lado, la influencia de la personalidad en el bienestar laboral ha sido en más de una ocasión objeto de estudio (Furnham et al., 1999). Las variadas puntuaciones obtenidas en el test de Eysenck por un grupo de trabajadores mostraban que, tanto su nivel de satisfacción, como su respuesta a los estímulos, variaba en función de la personalidad. En el caso de los extrovertidos, en concreto, el sentimiento de bienestar en el puesto de trabajo se vio favorecido por el aumento de la motivación intrínseca.

2.1.2 El neuroticismo

El neuroticismo también ha sido tratado en un amplio número de trabajos y existen evidencias biológicas que respaldan su existencia. En el inventario de personalidad para adultos de Eysenck (1954: 4) se dice de la dimensión de neuroticismo que:

Está ligada a la excitabilidad del sistema nervioso autónomo (...) Una alta tasa de neuroticismo (inestabilidad), indica una emotividad muy intensa y tono elevado. Son persona generalmente tensas, ansiosas, inseguras, tímidas, con fuertes bloqueos de conducta, respuestas inadecuadas e interrupciones emocionales desordenadas, tienden a presentar trastornos psicósomáticos de tipo ansioso o úlceras pépticas, etc.

En cuanto a los rasgos definitorios del factor, se encuentran los siguientes: ansiedad, hostilidad, depresión, ansiedad social, impulsividad y vulnerabilidad.

El estudio de este factor tiene gran relevancia a nivel clínico, puesto que se han encontrado correlaciones entre puntuaciones altas en neuroticismo y distintos trastornos de personalidad: esquizoide, esquizotípico, límite y evitativo (Cuponi, 2013), depresión mayor, ansiedad generalizada y distintas fobias (Hettema, 2006). También se han encontrado efectos de los rasgos neuróticos en el estado general de salud, siendo estos responsables de un mayor deterioro cognitivo y un peor pronóstico en enfermos crónicos de edad avanzada (Manning, 2017).

2.1.3 El psicoticismo

A diferencia de la extraversión y el neuroticismo, que están reconocidos en todos los estudios de personalidad, el psicoticismo no aparece siempre como factor. Apareció tardíamente en el modelo de factores de Eysenck y ha sido el menos estudiado; además, los correlatos biológicos encontrados son más débiles que los encontrados para los dos anteriores. En cuanto a sus características psicológicas y los rasgos que lo definen, se pueden mencionar las siguientes: “solitario, no le importa la gente, problemático y no encaja en ningún sitio, se muestra insensible, hostil y agresivo con los demás, le gustan las cosas inusuales y extrañas y no le preocupa el peligro. Disfruta ridiculizando y contrariando a los demás. Muestra elevada creatividad” (Vicente, 2018)

En otros modelos de personalidad reconocidos, como el modelo de los cinco factores (Costa y McRae, 1986) aparecen los rasgos apertura a la experiencia, cordialidad y responsabilidad que, en conjunto, cubren el polo opuesto del psicoticismo.

2.2. Relación entre lenguaje y características psicológicas: de los primeros estudios a los más recientes

La relación entre el lenguaje y distintos aspectos psicosociales lleva siendo estudiada desde que Freud empezó a escribir sobre los *lapsus linguae* y su vínculo con el subconsciente.

A partir de la década de los cincuenta del siglo pasado, la investigación se expandió hacia otros campos y fue claramente influenciada en sus inicios por dos corrientes distintas. Por un lado, hay que señalar el trabajo del físico Louis Gottschalk (Gottschalk, 1969, apud: Pennebaker y King, 1999: 1296), que enfocó su investigación en personas diagnosticadas clínicamente de alguna psicopatología. Basó los estudios en transcripciones de conversaciones orales. Hasta la fecha, la mayor parte de los experimentos han seguido este procedimiento y se han centrado en el discurso oral, en parte porque constituye una muestra de fácil acceso. Además, la oralidad ofrece variables paralingüísticas como la pronunciación, la entonación o el volumen que varían ostensiblemente entre individuos. Gottschalk estudió enunciados completos y buscó temas de conversación freudianos en los pacientes, por lo que siguió siendo un enfoque profundamente psicoanalítico.

Por otro lado, podemos señalar el método desarrollado por Stone, Dunphy, Smith y Ogilvy (1966), que se centró en el estudio de las palabras escogidas por el hablante para ser utilizadas en el discurso. Fueron los primeros en utilizar un programa informático para el análisis de los textos, originalmente llamado *General Inquirer*. También siguiendo la estela de Freud, buscaban palabras cuyo contenido semántico fuera congruente con el estado de ánimo supuesto en el paciente.

En un estudio con muestras orales, Scherer (1979) demostró que los extravertidos hablan en voz más alta y nasal y que los extravertidos americanos tienden a introducir menos pausas en el discurso, mientras que los extravertidos alemanes introducen más pausas que los introvertidos. De estos datos puede deducirse que los marcadores lingüísticos son dependientes de la cultura o el idioma, incluso entre sociedades occidentales.

Walter Weintraub, 1981 (apud: Pennebaker y King, 1999: 1296), ideó más tarde otro programa que analizaba el uso de las palabras funcionales, como preposiciones, pronombres, etc., anteriormente ignoradas. Sus hallazgos sobre el uso de pronombres personales en primera persona singular y la depresión resultaron ser muy avanzados, aunque, a pesar de eso, esta línea de investigación fue desatendida durante largo tiempo.

El principal problema de estas aproximaciones iniciales es que no supusieron un estudio teórico sobre la naturaleza del lenguaje, sino medios experimentales para demostrar las tesis sostenidas por sus autores. El primero en profundizar sobre la posible correlación entre el habla y las diferencias individuales fue Pennebaker (v., por ejemplo, Pennebaker y King, 1999) con la creación del LIWC (Linguistic Inquiry and Word Count). A partir de varios experimentos, estableció correlaciones estables y significativas entre las distintas categorías lingüísticas y algunas variables de personalidad. Mientras tanto, Allport (1996) establece la diferencia entre comportamiento adaptativo y comportamiento estilístico: lo que se hace y cómo se hace, respectivamente. En el caso del lenguaje, el comportamiento adaptativo sería el mensaje y el estilístico, el código.

Pennebaker y King (1999) proponen que las palabras que se seleccionan para elaborar los mensajes – el comportamiento estilístico – pueden ofrecer tanta información sobre el hablante como el mensaje en sí mismo (o más, incluso, ya que el tema del texto puede no ser autobiográfico, pero la información contenida en el código siempre estará presente).

Para aceptar esta tesis (la relación entre los rasgos de personalidad y los rasgos lingüísticos) hay que asumir antes dos premisas fundamentales: (i) que la personalidad es un constructo estable y (ii) que los rasgos lingüísticos asociados a ella se mantienen también estables a lo largo del tiempo y son independientes de los contextos lingüístico y extralingüístico. La primera premisa cuenta con la aceptación general de la comunidad científica (Bermúdez, 1986). La segunda es estudiada por primera vez por Pennebaker y King (1999), que demuestran en un experimento la consistencia del uso del lenguaje a lo largo de distintos temas y momentos temporales y su validez como indicador fiable de algunos aspectos de personalidad o características individuales.

Aceptando entonces la relación existente entre rasgos lingüísticos y personalidad, queda dirimir cuál es la aproximación más adecuada para su estudio.

2.2.1 Métodos de análisis

Salvo algunas excepciones en los últimos años, los análisis de los textos se han hecho, tradicionalmente, siguiendo un modelo de “vocabulario cerrado”. Este modelo se llama así porque funcionan en base a una lista de palabras organizadas en categorías que se define antes de entrar en contacto con el texto. Cuentan con un sistema de reconocimiento de palabras y varios diccionarios (psicológicos y gramaticales). Cada palabra encontrada en el texto analizado es buscada en el diccionario y se clasifica dentro de una o varias de las categorías (es un sistema jerárquico, “yo” entrará en la categoría de *pronombres*, en la de *pronombres personales* y en la de *pronombres personales de primera persona singular*). Toda la información extraída del texto es organizada en forma de porcentajes; no se trabaja con frecuencias absolutas, sino con frecuencias relativas.

El programa de análisis cerrado con mayor reconocimiento es el LIWC. Cuenta con un total de 64 categorías y hace una distinción fundamental: palabras de contenido y palabras de estilo. Las primeras, entre las que se encuentran adjetivos y sustantivos, son las que dotan a las oraciones de significado, y fueron las primeras en ser estudiadas cuando empezaron a buscarse relaciones entre estado psicológico y habla. Las segundas comprenden las llamadas *function words* en inglés (pronombres, preposiciones, artículos, etc.) con menor contenido semántico, pero que han demostrado ser más valiosas a la hora de revelar datos sobre la personalidad de los emisores en un estudio lingüístico. Su producción es independiente del contexto y casi inconsciente y, además, es difícil

modificar su uso voluntariamente, por lo que resultan también indicadores más fiables que las palabras de contenido, que sí dependen en cierta manera del control del emisor.

Diferenciándose del análisis cerrado, Schwartz et al. (2013) diseñaron recientemente un modelo de vocabulario abierto. Al analizar los textos a posteriori, este modelo pone solución a muchos de los problemas que presentaba el anterior planteamiento. Así, en lugar de elaborar una lista cerrada y asignar cada palabra a las categorías correspondientes, el programa busca relaciones en el texto. La frecuencia absoluta y relativa de las palabras son tenidas en cuenta, pero también se analizan las oraciones y los temas, la frecuencia con la que dos palabras aparecen juntas o la longitud de los enunciados. Además, al no emplear listas cerradas, permite la inclusión en el estudio de neologismos, errores tipográficos, palabras malsonantes, abreviaciones y emoticonos, que no eran contemplados en los análisis anteriores (por no aparecer en los diccionarios) y que pueden ser también valiosos para el estudio.

Aunque los sistemas de análisis cerrado permiten averiguar si un texto corresponde a uno u otro tipo de personalidad (o género, rango de edad, estado de salud mental, etc.), los sistemas de vocabulario abierto permiten hacer también atribuciones más generales, como, por ejemplo, los temas preferidos por un determinado grupo. Además, se ha demostrado (Schwartz et al. 2013) que el análisis abierto es superior al cerrado en cuanto a predicción de rasgos.

2.2.2 Resultados obtenidos

De los resultados de experimentos anteriores en lengua inglesa se desprende que el discurso de las personas extrovertidas tiende a ser más fluido e informal que el de las personas introvertidas (Gill y Oberlander, 2002) y más abstracto, mientras que el de los introvertidos es más concreto (Beukeboom, Tanis y Vermeulen, 2012). También, que es un lenguaje implícito y dependiente del contexto y se utilizan, por lo tanto, más pronombres, adverbios y verbos.

Dentro de los pronombres, los más utilizados son los de la primera persona del plural y los de segunda persona (Gonzales, Hancock y Pennebaker, 2010). Esto se debe a que ambas categorías incluyen a terceras personas, a diferencia de los pronombres de primera persona singular, característicos de los introvertidos (Liu, Perez y Nowson,

2016). Son menos utilizadas las palabras que expresan negación (*no, nunca*), tentativa y causalidad (*porque, puesto que, ya que*), artículos (*el, la, etc.*) y palabras de emoción negativas (*tristeza, depresión, melancolía*). Las palabras inclusivas (*y, con*) se encuentran más a menudo que las exclusivas (*aunque*). El recuento absoluto de palabras suele ser superior en extravertidos.

En cuanto al contenido de las palabras, hay mayor uso de palabras de emociones y sentimientos positivos (*alegría, amor, esperanza*), referencias a procesos sociales (*quedar, abrazar*) y comunicación (*llamar, hablar*), amistad (*amigo*) y otros seres humanos (*padre, madre, novia*). Se encuentran correlaciones negativas con palabras pertenecientes al campo semántico del trabajo o los estudios (*jefa, deberes, alumno*).

En base a todo lo anterior, se ha elaborado una serie de características que deberían cumplir los textos redactados por personas que puntúen alto en extraversión: el número de palabras total debería ser más elevado que el de personas con puntuaciones medias o bajas. Probablemente, este efecto se deba a la característica de locuacidad que caracteriza a los extrovertidos. La proporción de palabras gramaticales también debería ser superior, así como los pronombres personales de primera persona del plural (en relación con el gregarismo y la asertividad), pero la proporción de adverbios de negación, causalidad y tentativa (*no, porque, quizás*) debería ser menor. Asimismo, se espera que haya en estos textos un mayor número de palabras con referencias sociales (*hablar, fiesta, quedar*) y menos palabras negativas que en textos de personas introvertidas, así como un menor número de referencias al campo semántico del trabajo o los estudios.

A pesar de que los estudios relacionados con el rasgo de extroversión son los más abundantes y los que han dado resultados más claros, también encontramos trabajos que han estudiado la relación entre el lenguaje y los otros rasgos de personalidad antes comentados. Así, como se puede ver en Oberlander (2004), los rasgos característicos del neuroticismo dan lugar a correlaciones diferentes y se pueden encontrar algunas explicaciones psicológicas para estos fenómenos. Un rasgo típico de los neuróticos es la focalización en el “yo” y la consiguiente rumiación. Esta puede ser expresada mediante la preferencia por pronombres de primera persona singular (*yo, mi*). Su necesidad de apego puede traducirse en un mayor número de palabras inclusivas.

En cuanto al número de palabras por unidad de texto, existen datos que relacionan situaciones ansiógenas con una menor producción léxica. Dado que, como hemos visto,

los neuróticos tienden a la ansiedad, puede deducirse que las personas que puntúan alto en ansiedad producen un menor número de palabras. Por último, los neuróticos, que puntúan alto en emotividad, utilizan más palabras emocionales, aunque el recuento de palabras emocionales negativas es muy superior al de palabras emocionales positivas. Esta característica propicia también un lenguaje generoso en intensificadores, adverbios y adjetivos.

Oberlander (2004) también presenta un resumen de las correlaciones encontradas para el psicoticismo, relacionándolas con los rasgos de personalidad propios de este factor. Su alta creatividad explica el empleo de un léxico diverso, construcciones gramaticales poco comunes y numerosos adjetivos. El desapego característico de las personas psicóticas predice que utilizarán pocas palabras para referirse a otros, es decir, un bajo uso de pronombres. Uno de los rasgos más característicos, que es la dureza, se manifiesta en la escasa utilización de palabras de emoción que, de aparecer, serán negativas. No se han encontrado correlaciones para verbos o adverbios.

Por último, como se ha comentado en el epígrafe del psicoticismo, para su estudio es necesario contemplar tanto las correlaciones encontradas para este factor, como las encontradas para los factores apertura a la experiencia, cordialidad y responsabilidad, teniendo en cuenta que representan polos opuestos y que deben ajustarse los resultados.

Se ha encontrado que las personas con puntuaciones altas en apertura a la experiencia utilizan más artículos, palabras más largas, pocos pronombres de primera persona, verbos en tiempo presente y palabras de causalidad y que indican reflexión (*pensar, saber*). Las puntuaciones altas en cordialidad correlacionan con un alto uso de palabras de emoción positiva y bajo uso de palabras de emoción negativa y artículos. Las personas responsables escriben también con más palabras de emoción positiva, y pocas palabras de emoción negativa, pocas negaciones, palabras que expresen discrepancias (*debería, habría*) y palabras que expresen causación o exclusión.

Todas estas correlaciones deberían ser encontradas en sentido inverso en los estudios realizados teniendo las puntuaciones en psicoticismo.

2.3. Conclusiones

Hasta este momento, hemos ido viendo que, efectivamente, existe un vínculo entre la personalidad de un individuo y su discurso. Esta relación ha sido estudiada a través de distintos experimentos, que permiten tanto predecir los rasgos lingüísticos que presentará un texto a raíz de la personalidad de su autor como acertar el carácter del mismo mediante el análisis de su redacción. A este respecto, plantearemos dos cuestiones: la primera atiende a las aplicaciones prácticas que pueden desarrollarse a partir de estos resultados; la segunda trata las cuestiones éticas a las que todo análisis científico debe atenerse.

Los avances teóricos en el novedoso campo de estudio que nos ocupa invitan a pensar en vías para rentabilizar el conocimiento obtenido. Una de ellas podría ser internet, donde todo acto es informativo. Cualquier herramienta que permita descifrar la ingente cantidad de mensajes que se emiten en la red resulta útil en la era digital en que vivimos.

El análisis lingüístico en bruto ya se emplea para conocer los flujos de opinión o los intereses de la población, pero también el efecto que determinadas palabras pueden provocar en campañas políticas o publicitarias. Un estudio exhaustivo, como el que se ha presentado, podría favorecer que esas campañas sean dirigidas a grupos definidos en función de su personalidad, lo que permitiría tanto un ahorro de recursos como un aumento significativo de su efectividad. Sin embargo, no debemos olvidar los problemas de tipo ético que estos hallazgos pueden plantear. Esta clase de estudios permite la obtención de datos personales de los usuarios que estos, en principio, no desean compartir. Más allá del interés científico, ¿suponen este tipo de investigaciones algún tipo de beneficio para la sociedad? En el contexto de constantes denuncias contra corporaciones como Facebook a causa de la venta de información privilegiada, ¿dónde debemos poner los límites a la privacidad?

3. ¿Es posible replicar los resultados con muestras lingüísticas pequeñas? Un análisis experimental sobre el rasgo de extroversión.

3.1. Objetivos e hipótesis de partida

El objetivo de esta segunda parte del Trabajo de Fin de Grado es averiguar si se puede encontrar una correlación entre el factor extraversión/introversión y los rasgos lingüísticos en muestras pequeñas. Se ha elegido el factor extraversión por tres razones: (i) porque constituye el factor más estudiado y del cual se dispone de más bibliografía; (ii) porque las correlaciones encontradas para los extravertidos son más fuertes que las encontradas para los neuróticos o los psicóticos y (iii) porque los estudios demuestran que su interés supera lo clínico, pues es un rasgo relevante en lo cotidiano y en lo profesional.

El foco es el análisis del discurso. Este campo de estudio ha sido explotado, principalmente, por los estudios de mercado y las grandes compañías que estudian los flujos de opinión. Como se ha comentado anteriormente, para estos estudios se precisan grandes muestras de información, Big Data, con las que no todo el mundo puede operar. Por eso, el objetivo de esta segunda parte del TFG es muy concreto: descubrir si, a través de muestras pequeñas e individualizadas, se pueden obtener datos estadísticamente relevantes sobre la personalidad de los participantes.

Para ello, nos hemos basado en el experimento de Oberlander (2004), en el que los participantes completan primero un test de personalidad que permite clasificar a los individuos, y después se les pide que redacten dos *e-mails*, de los cuales se extraen los rasgos lingüísticos pertinentes. Metodológicamente, el trabajo que hemos desarrollado aquí es muy similar a este, aunque Oberlander (2004) analiza las tres dimensiones de personalidad, mientras que en este estudio se ha trabajado únicamente con la extraversión. El test utilizado para medir las puntuaciones en este rasgo es el Inventario de personalidad de Eysenck, forma B para adultos, el mismo que en el experimento original.

La diferencia más relevante se encuentra, además, en el idioma de los participantes. Algunos experimentos de los citados en el trabajo se han realizado con hablantes de italiano, español o alemán, pero tanto en el de Oberlander (2004) como en la mayor parte de los citados anteriormente, se ha trabajado con datos obtenidos de hablantes de inglés.

En este caso, hemos trabajado con hablantes nativos de español peninsular y todo el experimento se ha realizado en este idioma¹.

Dado que este experimento está basado en el de Oberlander (2004), vamos a realizar una breve revisión de sus hipótesis y de los resultados obtenidos y posteriormente presentaremos los de este estudio.

Las predicciones de Oberlander (2004) se concentran en torno a tres características propias de los extrovertidos: la fluidez verbal, la positividad y el habla implícita. Así pues, en cuanto a la fluidez verbal, espera encontrar frases largas, pocos signos de puntuación y un uso elevado de conjunciones. En relación con la positividad, predice un amplio uso de palabras de afecto positivo y un bajo uso de palabras negativas, tentativas y de causalidad. Por último, como signos del lenguaje implícito, presupone la aparición de numerosos pronombres personales y más verbos y adverbios que nombres y adjetivos. Para realizar el análisis de datos, agrupa a los participantes en tres grupos según su puntuación: extrovertidos altos, medios y bajos, y utiliza las puntuaciones medias de cada grupo para extraer las correlaciones. Las hipótesis relativas a la fluidez verbal y la positividad son verificadas. En el caso de las relacionadas con el habla implícita, los resultados son menos consistentes. Oberlander atribuye esta divergencia a las diferencias metodológicas respecto a otros experimentos.

A continuación, se expondrán las hipótesis de partida de este trabajo:

Hipótesis 1: se va a encontrar una correlación entre el rasgo de extroversión y el número de palabras de los textos construidos, de tal modo que ‘a mayor índice de extroversión, más palabras’.

Hipótesis 2: como reflejo de la fluidez verbal de los extrovertidos, utilizarán menos signos de puntuación y frases más largas hiladas mediante conjunciones, es decir ‘a mayor índice de extroversión, más largas serán las frases’ y ‘a mayor índice de extroversión, más conjunciones’.

Hipótesis 3: los extrovertidos se caracterizan por un tipo de habla más implícita que los introvertidos, de modo que ‘a mayor índice de extroversión, más uso de pronombres’ y ‘a mayor índice de extroversión, menor número de sustantivos’.

¹ Cuando se hable de los resultados, se verá que este hecho ha podido influir en los mismos.

Hipótesis 4: los extrovertidos son más gregarios que los introvertidos y tienden a participar en más actividades sociales, por lo tanto ‘a mayor índice de extroversión, más utilización de la primera persona plural’.

Hipótesis 5: el rasgo de positividad de los extrovertidos se reflejará en el lenguaje, de modo que ‘a mayor índice de extroversión, más adverbios positivos’ y ‘a mayor índice de extroversión, menos adverbios negativos’.

3.2. Participantes

Los participantes fueron 30 estudiantes de la Facultad de Filosofía y Letras de la Universidad de Zaragoza a los que se les ofreció participar voluntariamente en un experimento relacionado con sus estudios. Para todos ellos el español es su lengua materna. El 61,1% eran mujeres, el 33,3% hombres y el resto prefirió no especificar el género. La media de edad fue de 22,64 años.

3.3. Materiales

Se elaboró una encuesta (V. Anexo 1), con la herramienta Formularios de Google, que consta de 48 ítems pertenecientes al test de personalidad de Eysenck. 24 de las preguntas son sobre la dimensión extraversión-introversión y las 24 restantes, sobre la dimensión neuroticismo-estabilidad emocional. Las preguntas objeto de análisis fueron las de extroversión y las de neuroticismo se usaron de relleno y no se analizaron sus resultados. En todas las preguntas las posibles respuestas fueron sí y no. Las tareas de la segunda parte se extrajeron del experimento de Oberlander.

3.4. Procedimiento

Los voluntarios fueron convocados y se les solicitó que acudieran con un ordenador portátil, tableta o móvil desde los que poder acceder a la encuesta. Los correos electrónicos también se enviaron desde esos dispositivos. Todos los participantes fueron reunidos en una misma sala. Se les facilitó el enlace desde el que acceder a las preguntas y se les dieron 15 minutos para contestarlas.

Antes de las preguntas, en la encuesta aparecía un texto aclaratorio en el que se explicaba que no se trataba de un test de inteligencia ni de habilidad, que no había respuestas correctas o incorrectas, se pedían respuestas intuitivas y se garantizaba el tratamiento confidencial de los datos.

Para poder realizar la encuesta se solicitaba iniciar sesión con una cuenta de correo electrónico y de este modo fue posible relacionar las respuestas del test con los correos electrónicos de la segunda parte de la tarea. Se sugirió a los voluntarios que crearan una cuenta de Gmail específicamente para este experimento, en la que no apareciera su nombre, para garantizar el anonimato y así promover la sinceridad en sus respuestas del test.

Cuando todos los participantes terminaron de contestar las preguntas, se proyectaron tres diapositivas de Power Point con las instrucciones de la siguiente tarea, que consistió en escribir dos *e-mails*. En primer lugar, se garantizaba la confidencialidad y se sugería que, en caso de dar demasiados datos personales, los participantes podían alterar los hechos narrados o cambiar nombres de personas y/o lugares:

Esta tarea debería durar 20 minutos como máximo (10 minutos cada email). Toda la información será tratada de forma confidencial. Si durante la tarea te preocupa dar datos demasiado personales, puedes alterar los hechos o cambiar nombres de personas o lugares.

Muchas gracias por vuestra participación.

En la segunda diapositiva aparece el siguiente mensaje:

Imagina que hace mucho tiempo que no ves a un/a amigo/a y decides mandarle un email para ponerle al día. Escribe sobre lo que has hecho o te ha ocurrido durante la semana pasada, intentando contarle tanto como sea posible en 10 minutos. Escribe de manera natural, no pasa nada si hay errores gramaticales o faltas de ortografía. Una vez empieces una frase, acábala, y no revises o corrijas el contenido. El encabezado o la despedida no son importantes, céntrate en el cuerpo del mensaje. Debes enviar el email a ccuestafontela@gmail.com

El objetivo era que el mensaje se emitiera de la manera más natural posible.

Transcurridos los 10 minutos, se pidió que enviaran los correos electrónicos y se proyectó la última diapositiva, en la que se pedía que realizaran una tarea casi idéntica, pero en la que debían hablar de los planes que tenían para la semana siguiente.

Escribe un mensaje parecido al anterior, pero contándole a tu amigo/a tus planes para la semana que viene. Escribe tanto como sea posible en 10 minutos. Escribe de manera natural,

no pasa nada si hay errores gramaticales o faltas de ortografía. Una vez empieces una frase, acábala, y no revises o corrijas el contenido. El encabezado o la despedida no son importantes, céntrate en el cuerpo del mensaje. Debes enviar el email a ccuestafontela@gmail.com

Tras finalizar las dos tareas, se agradeció su participación a todos los que habían asistido a la sesión.

3.5. Resultados y análisis de datos

Aunque el análisis abierto ofrece resultados más completos, por cuestiones de tiempo y falta de medios, en este trabajo se siguió un método de análisis parecido al LIWC. Se contabilizaron las categorías gramaticales básicas que aparecen en los sistemas de análisis cerrado y se incluyeron también algunas categorías del sistema abierto, como la longitud de las frases.

El primer paso para realizar el análisis fue la preparación de los materiales, que consistió en ordenar a los participantes en función de su puntuación en el test y unir los dos *e-mails* enviados por cada participante en un mismo documento de Word.

La mayor parte del análisis se realizó manualmente. Utilizando la herramienta de búsqueda de Word se contó el número de palabras de cada categoría que aparecía en cada uno de los *e-mails*. Las categorías analizadas con este método fueron los pronombres (totales), los pronombres de primera persona singular y plural, segunda persona singular y tercera persona singular. También se contaron los adverbios de afirmación y negación, la aparición de dígitos, preposiciones, artículos y conjunciones.

Gracias a la página web www.linguakit.com, que ofrece un análisis básico de textos, fue posible añadir más categorías. Con este método se obtuvieron los siguientes datos: número de palabras y caracteres (lo cual permitió calcular también la longitud de las palabras y la longitud de las frases), sustantivos, verbos, adjetivos, adverbios y variabilidad léxica. Todos los datos fueron introducidos en forma de porcentaje –la frecuencia relativa en función del número total de palabras – en una tabla de Excel.

Utilizando el programa R-Commander se obtuvo el coeficiente de correlación de Pearson para cada rasgo lingüístico en relación con la puntuación obtenida en el test de personalidad. El objetivo era medir si la puntuación obtenida en el test podía predecir el

porcentaje de los distintos elementos lingüísticos medidos en los textos que habían producido los informantes.

Porcentaje de pronombres: $t = -0.37155$, $df = 18$, $p\text{-value} = 0.7146$
Porcentaje de pronombres 1ª persona singular: $t = -0.97407$, $df = 18$, $p\text{-value} = 0.3429$
Porcentaje de pronombres 1ª persona plural: $t = 1.8682$, $df = 18$, $p\text{-value} = 0.0781$
Porcentaje de pronombres 2ª persona: $t = 0.02387$, $df = 18$, $p\text{-value} = 0.9812$
Porcentaje de pronombres 3ª persona: $t = -1.0848$, $df = 18$, $p\text{-value} = 0.2923$
Porcentaje de adverbios de negación: $t = -0.87952$, $df = 18$, $p\text{-value} = 0.3907$
Porcentaje de adverbios de afirmación: $t = NA$, $df = 18$, $p\text{-value} = NA$
Porcentaje de preposiciones: $t = 1.3026$, $df = 18$, $p\text{-value} = 0.2091$
Porcentaje de números: $t = -1.3009$, $df = 18$, $p\text{-value} = 0.2097$
Porcentaje de artículos: $t = -0.40549$, $df = 18$, $p\text{-value} = 0.6899$

Tabla 1. Correlaciones de Pearson obtenidas para las variables contadas manualmente en el primer análisis. Resultados significativos $P < 0,005$.

Número de palabras: $t = -1.6649$, $df = 17$, $p\text{-value} = 0.1142$
Longitud media palabras: $t = 1.5185$, $df = 17$, $p\text{-value} = 0.1473$
Longitud media frases: $t = 0.8222$, $df = 17$, $p\text{-value} = 0.4223$
Porcentaje sustantivos: $t = 3.2352$, $df = 17$, $p\text{-value} = 0.004865$
Porcentaje verbos: $t = 1.148$, $df = 17$, $p\text{-value} = 0.2669$
Porcentaje adverbios: $t = -0.25105$, $df = 17$, $p\text{-value} = 0.8048$
Porcentaje adjetivos: $t = 0.63366$, $df = 17$, $p\text{-value} = 0.5347$
Variabilidad léxica: $t = 0.18659$, $df = 17$, $p\text{-value} = 0.8542$

Tabla 2. Correlaciones de Pearson obtenidas para las variables contadas con la página web www.linguakit.com en el primer análisis. Resultados significativos $P < 0,005$.

En primer lugar, el análisis se realizó incluyendo las puntuaciones de los 30 sujetos (Tabla 1, Tabla 2). Como puede comprobarse, la única correlación significativa que

encontramos fue en el porcentaje de sustantivos (Tabla 2). Encontramos, además, una tendencia ($p = 0,078$) en el porcentaje de pronombres de primera persona de plural.

De este modo, la correlación entre las palabras estudiadas y el rasgo Extraversión/Introversión es muy limitada y solo es significativa en el caso del porcentaje de sustantivos en el primer análisis. Sorprendentemente, esta correlación se presenta en sentido inverso al esperado.

Debido a las pobres correlaciones obtenidas, se realizó un segundo análisis, en el que se agrupó a los sujetos en categorías, del mismo modo que en el experimento de Oberlander (2004). Así, se organizaron tres grupos de sujetos, de modo que los sujetos con puntuaciones inferiores a 10 formaron parte del grupo 'Extroversión baja', los sujetos con puntuaciones entre 10 y 14 puntos formaron parte del grupo 'Extroversión media' y los sujetos con puntuaciones superiores a 14, 'Extroversión alta'. Una vez formados estos tres grupos, se realizó una correlación entre las distintas variables descritas anteriormente (tablas 1 y 2) y la pertenencia al grupo 1 (extroversión baja), 2 (extroversión media) y 3 (extroversión alta). Dado que en esta ocasión la correlación se establece entre un conjunto de variables continuas y una variable ordinal, en lugar de recurrir al Coeficiente de Pearson, se calculó el Coeficiente tau de Kendall.

En esta segunda fase del análisis se encontraron dos tendencias: **en el porcentaje de adverbios de negación: ($p = 0.075$)** y **en el porcentaje de números ($p = 0.077$)**. Se encontró también una **correlación significativa ($p = 0.01$) en el caso del porcentaje de sustantivos**. De todo ello tratará el siguiente apartado.

3.6. Discusión de resultados y conclusiones

En total, se ha trabajado con 19 categorías lingüísticas, de entre las cuales se han encontrado tres tendencias (en porcentaje de pronombres de primera persona plural, porcentaje de adverbios de negación y porcentaje de números) y una correlación significativa (en porcentaje de sustantivos) presente en las dos fases del análisis.

Existen varias circunstancias – el idioma, la falta de aleatoriedad de la muestra, las puntuaciones medias y la longitud del texto – que pueden explicar que no se hayan encontrado correlaciones significativas en este estudio.

En primer lugar, todos los estudios revisados que sirven de base de este trabajo (incluido el experimento que se toma como modelo) provienen de muestras de hablantes de inglés. Es probable que el modo en el que la personalidad del emisor se manifiesta en su expresión lingüística varíe interlingüísticamente. Si este fuera el caso, la escasez de estudios previos sobre lengua española ha podido dificultar el objetivo de este trabajo, que es corroborar que en muestras pequeñas se encuentran datos similares a muestras de Big Data. Efectivamente, creemos que es absolutamente necesario que se realicen estudios correlacionales con grandes cantidades de datos en español, para comprobar si existen diferencias interlingüísticas en este tipo de correlaciones.

En segundo lugar, la muestra experimental de este trabajo, además de ser pequeña no se ha formado de modo aleatorio, sino por conveniencia. Esto conlleva la imposibilidad de controlar múltiples variables – como las circunstancias personales o el estado de ánimo de los sujetos –, que pueden haber influido en la realización del test. Sería interesante replicar este estudio con una muestra mayor y con mayor control de variables (quizá podrían utilizarse cohortes, como control experimental).

En tercer lugar, la mayor parte de los participantes (todos excepto dos) han obtenido puntuaciones medias en el test de personalidad. La ausencia de extrovertidos e introvertidos extremos dificulta en gran medida el estudio, pues podría considerarse que las débiles correlaciones se deben a que todos los sujetos pertenecen al grupo intermedio. Sería, por ello, interesante, volver a realizar este trabajo con una muestra que presente una mayor variabilidad en el test de personalidad, con respecto al rasgo de extroversión. Quizá se podría realizar, de hecho, la muestra experimental utilizando como fase previa los resultados obtenidos en el test de personalidad.

A estos obstáculos hay que añadir los limitados recursos de los que se disponía a la hora de realizar el análisis de los textos. En la bibliografía consultada se mencionan instrumentos informáticos de análisis que ofrecen todo tipo de datos lingüísticos. En el caso de este trabajo, la mayor parte de los datos se han obtenido manualmente, con el coste de tiempo que ello implica. De haber dispuesto de una herramienta adecuada, se habría llevado a cabo un análisis más pormenorizado, con más participantes y más categorías analizadas.

Una vez presentadas las propuestas que justifican que no hayamos encontrado más resultados significativos, pasemos a comentar lo que sí hemos encontrado. En primer

lugar, se han hallado dos tendencias que corroboran las previsiones hechas a partir de la lectura de la bibliografía al uso: el mayor uso de pronombres de primera persona del plural (en el primer análisis) y el menor uso de adverbios de negación (en el segundo). La explicación de por qué aparecen como simples tendencias y no como resultados estadísticamente significativos sería la misma que para la falta de correlación que veíamos anteriormente (el tamaño y naturaleza de la muestra y el idioma fundamentalmente).

Estas tendencias encontradas en el análisis parecen resultar una evidencia a favor de alguna de las hipótesis presentadas en el apartado 3.1. En concreto, el mayor número de pronombres de primera persona del plural concuerda con lo esperado según la hipótesis 4, que reproducimos de nuevo aquí:

Hipótesis 4: los extrovertidos son más gregarios que los introvertidos y tienden a participar en más actividades sociales, por lo tanto ‘a mayor índice de extroversión, más utilización de la primera persona plural’.

El menor número de adverbios de negación por parte de los extrovertidos tendría que ver con su positividad. Por ello, podemos relacionarlo con la hipótesis 5:

Hipótesis 5: el rasgo de positividad de los extrovertidos se reflejará en el lenguaje, de modo que ‘a mayor índice de extroversión, más adverbios positivos’ y ‘a mayor índice de extroversión, menos adverbios negativos’

Y, por otra parte, hemos encontrado una tendencia y una correlación significativa que no se corresponden con lo esperado. De acuerdo con la teoría y anteriores experimentos (Furnham, 1990), el lenguaje de las personas extrovertidas tiende a ser implícito, mientras que el de las personas introvertidas es explícito. Esto se materializa en un mayor uso de sustantivos por parte de los introvertidos y mayor dependencia del contexto, los pronombres y la deixis en los extrovertidos. Los resultados de este estudio, por el contrario, indican una correlación positiva entre la puntuación en extroversión y el uso de sustantivos: a mayor índice de extroversión, más sustantivos y una tendencia a un menor uso de los números (pronombres o adjetivos deícticos) por parte de los extrovertidos. Se estaría incumpliendo, por tanto, la hipótesis 3:

Hipótesis 3: los extrovertidos se caracterizan por un tipo de habla más implícita que los introvertidos, de modo que ‘a mayor índice de extroversión, más uso de pronombres’ y ‘a mayor índice de extroversión, menor número de sustantivos’.

Esta contradicción entre lo esperado (hipótesis 3) y lo encontrado con respecto a la cantidad de números o sustantivos utilizados puede estar basado, como decíamos antes, en una diferencia interlingüística, de tal modo que los hablantes de español presentaran la correlación inversa a los de inglés. No obstante, es posible que la explicación sea más bien de naturaleza diafásica: una diferencia fundamental entre los estudios de los Big Data y la tarea analizada aquí es la naturaleza de los textos que sirven de base para la investigación. Efectivamente, los estudios de Big Data se basan, fundamentalmente, en mensajes de redes sociales como Twitter, cuya longitud es mucho menor que la de los textos aquí analizados. Tal vez, para el formato tweet, un extravertido quiere decir más cosas (no olvidemos su locuacidad) y por eso depende más de pronombres, que ahorran espacio (por su limitación de caracteres). En cambio, en un *e-mail*, dispone de más espacio y emplea más sustantivos. Sería interesante realizar más estudios al respecto, en español o en inglés, para corroborar esta hipótesis.

Por último, no hemos podido probar la mayor fluidez verbal de los extrovertidos, pues ni parecen presentar un mayor número de palabras (Hipótesis 1), ni frases más largas (Hipótesis 2).

Para concluir la parte experimental del trabajo, podemos decir que nuestros resultados no permiten asegurar que con el análisis lingüístico de pequeñas muestras (como son un par de *e-mails*) podamos extraer información relevante sobre la personalidad del autor del texto. Dados los resultados obtenidos, es posible que la ausencia de más correlaciones estadísticamente significativas se deba al diseño de esta investigación (en concreto a la muestra utilizada). De este modo, sería interesante tratar de repetir este estudio una vez mejorado el diseño del mismo. Ahora bien, también es posible que sea necesario cuestionar la viabilidad del estudio en términos absolutos: quizás, al fin y al cabo, los rasgos lingüísticos contenidos en un texto tipo *e-mail* sean insuficientes para determinar la personalidad de su autor.

4. Conclusiones

En este trabajo hemos estudiado, en dos partes, la relación entre los rasgos lingüísticos y los rasgos de personalidad. Los objetivos de la primera parte han sido demostrar la existencia de esa relación y realizar una breve recopilación de las

correlaciones significativas para cada factor de personalidad. Para ello, se ha realizado, en primer lugar, una breve introducción a la teoría de personalidad y un repaso de los rasgos fundamentales de los tres factores que más se han mencionado en el trabajo: la extraversión, el neuroticismo y el psicoticismo. En segundo lugar, se han recopilado los estudios más relevantes en este campo desde el siglo pasado y se han comentado los métodos de análisis empleados. A partir de los resultados obtenidos en todos ellos, se ha confeccionado una lista de rasgos lingüísticos que correlacionan positivamente con cada uno de los factores.

Así, se ha encontrado que la extroversión correlaciona positivamente con un número elevado de palabras, el uso de pronombres – en particular los de primera persona plural y segunda persona –, adverbios, verbos, palabras inclusivas y palabras emoción positiva, de procesos sociales y comunicación. Los adverbios de negación, tentativa y causalidad, los artículos, sustantivos, las palabras de emoción negativa y las del campo semántico académico o laboral correlacionan negativamente con las puntuaciones altas en extraversión. El total de palabras de función es superior al total de palabras de contenido. Por otra parte, las personas neuróticas producen un menor número de palabras en total y las puntuaciones altas en neuroticismo correlacionan con un mayor uso de pronombres de primera persona, palabras inclusivas, palabras emocionales – más de emoción negativa que de emoción positiva –, adverbios de intensidad, adjetivos y adverbios. Por último, los psicóticos se caracterizan por la creatividad en el uso del léxico y las estructuras gramaticales. Las personas con puntuaciones altas en este factor correlacionan positivamente con el uso de adjetivos y negativamente con el uso de pronombres y palabras de emoción. Para el uso de verbos y adverbios en personas psicóticas no se han encontrado correlaciones. De los estudios que trabajan con los factores responsabilidad, apertura a la experiencia y cordialidad, se deduce también una correlación positiva de las personas psicóticas con los pronombres de primera persona, verbos en presente, palabras que indiquen causalidad, discrepancia, causación, exclusión y reflexión, palabras de emoción negativa y adverbios de negación. La correlación sería negativa en el caso de los artículos, palabras largas y palabras de emoción positiva.

El objetivo de la segunda parte del trabajo consistía en averiguar, mediante un experimento, si los resultados que se han visto en la primera parte, procedentes de muestras de Big Data, eran replicables a pequeña escala en textos pequeños de tipo *e-mail* para el factor extraversión. Para hacerlo, nos hemos basado en el experimento llevado a

cabo por Oberlander en 2004. El experimento, llevado a cabo con la participación de alumnos de la Universidad de Zaragoza, consistía en la redacción de dos *e-mails* tras completar un test de personalidad. Los resultados obtenidos tras el análisis de los textos no resultan prometedores. Tan solo se han encontrado tres tendencias y una correlación significativa (y además esta y una de las tendencias se presentan en sentido contrario al esperado). No obstante, aunque los resultados del trabajo no suponen una confirmación de la hipótesis, dejan la puerta abierta a nuevas investigaciones que, superando las limitaciones de este estudio, comprueben de manera fehaciente la plausibilidad de este análisis con pequeños datos. Ahora bien, las aplicaciones, en el caso de conseguirse la correlación son tan interesantes como peligrosas. Como en tantos otros aspectos de la ciencia, será un conocimiento que puede permitir facilitar la vida de los ciudadanos. En concreto, puede ayudar, con poco esfuerzo, en tareas como la selección de personal de las empresas, la autoría de textos, la selección de parejas o grupos de afinidad, etc. No obstante, el tratamiento de esta información deberá estar gestionado, necesariamente, desde un control ético y legal de los datos.

5. Bibliografía

- Allport, G. W. (1996). *Pattern and Growth in Personality*, New York: Holt Rinehart & Winston.
- Bermúdez, J. (1986). Concepto de personalidad. *Psicología de la personalidad (I)* (pp. 17-38). Madrid: UNED.
- Bermúdez, J. (1986). Psicología de la personalidad: Área de estudio y teoría. *Psicología de la personalidad (I)* (pp. 39-71). Madrid: UNED.
- Beukeboom, C. J., Tanis, M., & Vermeulen, I. E. (2013). The language of extraversion: Extraverted people talk more abstractly, introverts are more concrete. *Journal of Language and Social Psychology*, 32(2), 191-201.
- Costa, P. T., & McCrae, R. R. (1980). Influence of extraversion and neuroticism on subjective well-being: happy and unhappy people. *Journal of personality and social psychology*, 38(4), 668.
- Eysenck, H. J., & Eysenck, M. W. (1987). *Personalidad y diferencias individuales*. Pirámide.
- Golbeck, J., Robles, C., Edmondson, M., & Turner, K. (2011, October). Predicting Personality from Twitter. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on* (pp. 149-156). IEEE.
- Goldberg, L. R. (1990). An alternative "description of personality": The big-five factor structure. *Journal of Personality and Social Psychology*, 59(6), 1216.
- Goldberg, L. R. (1992). The development of markers for the big-five factor structure. *Psychological Assessment*, 4(1), 26.
- Gonzales, A. L., Hancock, J. T., & Pennebaker, J. W. (2010). Language style matching as a predictor of social dynamics in small groups. *Communication Research*, 37(1), 3-19.
- Hogan, R., Curphy, G. J., & Hogan, J. (1994). What we know about leadership: Effectiveness and personality. *American psychologist*, 49(6), 493.

- Liu, F., Perez, J., & Nowson, S. (2016). A language-independent and compositional model for personality trait recognition from short texts. *ArXiv Preprint arXiv:1610.04345*.
- Manning, K. J., Chan, D. G., David, D., & Steffens, C. (2017). Neuroticism traits selectively impact long term illness course and cognitive decline in late-life depression. *The American Journal of Geriatric Psychiatry*, 25(3), 220-229.
- McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, 60(2), 175-215.
- Oberlander, J., & Gill, A. J. (2006). Language with character: A stratified corpus comparison of individual differences in e-mail communication. *Discourse Processes*, 42(3), 239-270.
- Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., et al. (2015). Automatic personality assessment through social media language. *Journal of Personality and Social Psychology*, 108(6), 934-952.
- Pelechano, V. (1996). El modelo de H.J. Eysenck. *Psicología de la personalidad. 1. teorías* (pp. 59-88). Barcelona: Ariel.
- Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, 77(6), 1296-1312.
- Pennebaker, J. W., & Stone, L. D. (2003). Words of wisdom: Language use over the life span. *Journal of personality and social psychology*, 85(2), 291.
- Sánchez Turet, M., & Cuadras, C. M. (1972). Adaptación española del cuestionario E.P.I. de Eysenck. *Universitat De Barcelona Anuario De Psicología*, 6(1), 31-60.
- Schmidt, V. e. a. Modelo psicobiológico de personalidad de Eysenck: Una historia proyectada hacia el futuro. *II* (2), 1018-1023.
- Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., et al. (2013). *Personality, gender, and age in the language of social media: The open-vocabulary approach*. *PloS One*, 8(9), e73791.
- Stogdill, R. M. (1948). Personal factors associated with leadership: A survey of the literature. *The Journal of psychology*, 25(1), 35-71.

- Sumner, C., Byers, A., Boochever, R., & Park, G. J. (2012, December). Predicting dark triad personality traits from twitter usage and a linguistic analysis of tweets. In *Machine learning and applications (icmla), 2012 11th international conference on* (Vol. 2, pp. 386-393). IEEE.
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, 29*(1), 24-54. doi:10.1177/0261927X09351676
- Yarkoni, T. (2010). Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of research in personality, 44*(3), 363-373.

Anexo I: Inventario de personalidad de Eysenck para adultos – Formato B

INVENTARIO DE PERSONALIDAD EYSENCK FORMATO B CUESTIONARIO

INSTRUCCIONES

Aquí tiene algunas preguntas sobre el modo como usted se comporta, siente y actúa. Después de cada pregunta, conteste en la hoja de respuestas con un “SI” o con una “NO” según sea su caso.

Trate de decir “SI” o “NO” representa su modo usual de actuar o sentir, entonces, ponga un aspa o cruz en el círculo debajo de la columna “SI” o “NO” de su hoja de respuestas. Trabaje rápidamente y no emplee mucho tiempo en cada pregunta; queremos su primera reacción, en un proceso de pensamiento prolongado.

El cuestionario total no debe de tomar más que unos pocos minutos. Asegúrese de no omitir alguna pregunta.

Ahora comience. Trabaje rápidamente y recuerde contestar todas las preguntas. No hay contestaciones “correctas” ni “incorrectas” y esto no es un test de inteligencia o habilidad, sino simplemente una medida de la forma como usted se comporta.

1. ¿le gusta abundancia de excitación y bullicio a su alrededor?
2. ¿tiene a menudo un sentimiento de intranquilidad, como si quisiera algo, pero sin saber que?
3. ¿tiene casi siempre una contestación lista a la mano cuando se le habla?
4. ¿se siente algunas veces feliz, algunas veces triste, sin una razón real?
5. ¿permanece usualmente retraído (a) en fiestas y reuniones?
6. cuando era niño(a) ¿hacía siempre inmediatamente lo que le decían, sin refunfuñar?
7. ¿se enfada o molesta a menudo?
8. cuando lo(a) meten a una pelea ¿prefiere sacar los trapitos al aire de una vez por todas, en vez de quedar callado(a) esperando que las cosas se calmen solas?

9. ¿es usted triste, melancólico (a)?
10. ¿le gusta mezclarse con la gente?
11. ¿a perdido a menudo el sueño por sus preocupaciones?
12. ¿se pone a veces malhumorado (a)?
13. ¿se catalogaría a si mismo(a) como despreocupado (a) o confiado a su buena suerte?
14. ¿se decide a menudo demasiado tarde?
15. ¿le gusta trabajar solo (a)?
16. ¿se ha sentido a menudo apático (a) y cansado(a) sin razón?
17. ¿es por lo contrario animado(a) y jovial?
18. ¿se ríe a menudo de chistes groseros?
19. ¿se siente a menudo hastiado(a), harto, fastidiado?
20. ¿se siente incomodo(a) con vestidos que no son del diario?
21. ¿se distrae (vaga su mente) a menudo cuando trata de prestar atención a algo?
22. ¿puede expresar en palabras fácilmente lo que piensa?
23. ¿se abstrae (se pierde en sus pensamientos) a menudo?
24. ¿esta completamente libre de prejuicios de cualquier tipo?
25. ¿le gusta las bromas?
26. ¿piensa a menudo en su pasado?
27. ¿le gusta mucho la buena comida?
28. cuándo se fastidia ¿necesita algún(a) amigo(a) para hablar sobre ello?
29. ¿le molesta vender cosas o pedir dinero a la gente para alguna buena causa?
30. ¿alardea (se jacta) un poco a veces?
31. ¿es usted muy susceptible (sensible) por algunas cosas?
32. ¿le gusta mas quedarse en casa, que ir a una fiesta aburrida?
33. ¿se pone a menudo tan inquieto(a) que no puede permanecer sentado(a) durante mucho rato en una silla?
34. ¿le gusta planear las cosas, con mucha anticipación?
35. ¿tiene a menudo mareos (vértigos)?
36. ¿contesta siempre una carta personal, tan pronto como puede, después de haberla leído?
37. ¿hace usted usualmente las cosas mejor resolviéndolas solo(a) que hablando a otra persona sobre ello?

38. ¿le falta frecuentemente aire, sin haber hecho un trabajo pesado?
39. ¿es usted generalmente una persona tolerante, que no se molesta si las cosas no están perfectas?
40. ¿sufre de los nervios?
41. ¿le gustaría mas planear cosas, que hacer cosas?
42. ¿deja algunas veces para mañana. Lo que debería hacer hoy día?
43. ¿se pone nervioso(a) en lugares tales como ascensores, trenes o túneles?
44. cuando hace nuevos amigos(as) ¿es usted usualmente quien inicia la relación o invita a que se produzca?
45. ¿sufre fuertes dolores de cabeza?
46. ¿siente generalmente que las cosas se arreglaran por si solas y que terminaran bien de algún modo?
47. ¿le cuesta trabajo coger el sueño al acostarse en las noches?
48. ¿ha dicho alguna vez mentiras en su vida?
49. ¿dice algunas veces lo primero que se le viene a la cabeza?
50. ¿se preocupa durante un tiempo demasiado largo, después de una experiencia embarazosa?
51. ¿se mantiene usualmente hénrico(a) o encerrado (a) en si mismo(a), excepto con amigos muy íntimos?
52. ¿se crea a menudo problemas, por hacer cosas sin pensar?
53. ¿le gusta contar chistes y referir historias graciosas a sus amigos?
54. ¿se le hace más fácil ganar que perder un juego?
55. ¿se siente a menudo demasiado consciente de si mismo(a) o poco natural cuando esta con sus superiores?
56. cuando todas las posibilidades están contra usted, ¿piensa aun usualmente que vale la pena intentar?
57. ¿siente “sensaciones” en el abdomen, antes de un hecho importante?

Anexo II: Datos del análisis lingüístico obtenidos en el recuento manual

Puntuación	Número total	Pronombres	Pron. 1ª sin plul	Pron. 1ª plul	Pron. 2ª	Pron. 3ª	Negación	Asentimiento	Preposicion	Números	Artículos	Conjuncione
6	254	22	11	2	9	0	7	0	17	1	11	17
	356	19	14	0	3	2	5	0	40	5	25	13
8	236	14	7	3	2	2	5	0	26	2	14	8
	239	7	5	2	0	0	4	1	26	0	19	0
9	219	5	3	2	0	0	0	1	41	13	25	10
	179	6	4	1	1	0	1	0	29	9	24	10
9	335	17	7	2	8	0	7	0	33	1	20	8
	394	22	15	1	6	0	5	2	36	1	30	15
10	169	7	3	0	4	0	4	2	17	0	11	9
	270	16	10	0	6	0	5	0	26	0	26	12
10	248	8	1	0	5	2	6	0	29	0	12	9
	277	16	8	3	5	0	2	0	34	1	24	10
10	225	18	10	1	5	2	1	0	28	0	16	7
	256	16	8	3	4	1	1	0	30	0	24	8
10	328	16	14	1	0	1	8	0	39	0	20	24
	302	12	9	0	0	3	11	1	24	0	17	19
10	260	19	10	1	2	6	5	0	35	0	25	7
	371	16	11	1	3	1	7	0	48	0	25	24
10	296	21	12	2	5	2	5	0	44	6	25	16
	229	11	7	0	2	2	4	0	25	1	20	10
11	327	18	11	3	2	2	5	2	37	0	23	15
	500	23	17	1	2	3	6	2	71	0	39	20
11	118	6	2	2	0	2	3	0	9	0	10	4
	144	7	1	1	3	2	1	0	18	0	4	8
11	106	3	1	2	0	0	0	0	12	0	12	10
	146	6	1	3	2	0	4	0	17	0	15	5
12	425	30	13	2	14	1	11	0	61	0	26	8
	553	31	20	5	4	2	6	1	80	0	42	20

Puntuación	Número	to	Pronombre	Pron. 1ª	sin	Pron. 1ª	pl	Pron. 2ª	Pron. 3ª	Negación	Asentimie	Preposicio	Números	Artículos	Conjuncioni
12	143	9	3	1	3	2	1	3	2	1	0	13	0	5	5
	180	4	0	1	2	1	1	2	1	2	0	28	0	11	6
12	256	17	10	0	6	1	0	6	1	6	1	31	1	12	9
	328	12	4	2	5	1	2	5	1	5	0	42	0	26	16
12	271	7	3	3	0	1	3	0	1	4	0	30	0	21	9
	334	7	4	0	2	1	0	2	1	9	0	41	0	26	9
12	264	15	4	3	8	0	3	8	0	7	1	34	0	24	9
	278	24	20	0	4	0	0	4	0	12	1	26	0	21	10
13	205	18	8	2	7	1	2	7	1	3	0	22	0	15	7
	201	6	2	0	2	2	0	2	2	1	0	25	0	12	7
14	325	13	9	1	3	0	1	3	0	5	0	29	1	23	18
	302	4	4	0	0	0	0	0	0	3	0	31	2	24	14
14	185	7	5	0	2	0	0	2	0	0	0	22	0	12	8
	290	10	3	3	4	0	3	4	0	4	0	38	0	18	11
15	324	13	9	0	3	1	0	3	1	5	0	32	0	22	18
	367	16	9	5	1	1	5	1	1	5	0	43	0	30	14
15	193	6	2	1	3	0	1	3	0	1	0	22	1	10	14
	165	5	5	0	0	0	0	0	0	1	0	20	0	16	8
15	206	7	6	0	0	1	0	0	1	2	1	21	0	14	11
	153	6	2	1	3		1	3		2	0	17		14	6
16	252	16	0	8	7	1	8	7	1	4	0	29	0	22	14
	291	13	2	1	9	1	1	9	1			40		28	12
17	210	20	9	3	7	1	3	7	1	4	0	20	1	18	6
	255	19	13	2	4	0	2	4	0	4	0	33	4	18	10
17	293	12	8	2	2		2	2		4	0	33	0	23	15
	280	11	8	0	2	1	0	2	1	2	1	37	0	27	13
17	90	2	1	1	0	0	1	0	0	1	0	13	0	2	3
	93	4	2	1	1	0	1	1	0	0	0	13	0	12	1
18	212	10	4	3	3	0	3	3	0	9	0	23	0	1	8
	286	23	13	3	7	0	3	7	0	5	0	32	0	7	10

Anexo III: Datos del análisis lingüístico obtenidos de la página web www.linguakit.com

test	total pala	longitud p	longitud f	sustantivo	verbos	adjetivos	adverbios	variacion
6	682	3,9	32,5	8,8	8,4	4	3,1	54,5
8	538	4,7	22,4	7,2	7,4	2,8	1,7	50,7
9	805	3,96	61,9	9,1	8,3	2,6	2,9	54,6
9	457	4,43	45,7	6,3	7,2	2,4	1,3	51,7
10	729	3,78	26	7	8,9	3,1	1,9	52,4
10	709	3,87	15,8	8,1	9,3	2,8	2,7	49,2
10	556	3,97	25,3	8,8	11	3,6	1,8	58,4
10	603	3,72	50,25	8,5	9,1	2,7	2,3	55,7
10	496	3,74	31	10,9	10,1	3	2,8	68,2
10	586	4,53	58,6	13,8	9,4	2,7	3,1	64,5
11	336	4	30,55	11,3	11,3	2,1	3	69,9
11	890	4,7	445	12,6	8	2,4	2,6	53,5
11	301	3,67	30,1	10,3	13,6	3,3	3,7	75,61
12	605	3,87	33,6	9,3	9,3	3	3	59,8
12	685	3,94	31,1	10,8	8,8	3,5	2,9	58,6
12	681	3,93	22,7	9,8	11,3	3,1	2,8	64,1
12	374	3,46	46,75	9,9	9,9	1,6	2,7	61,4
12	1066	3,82	20,9	10	8,4	2,6	2,8	55,2
13	438	4,28	23,8	9,6	10,7	4,3	3,2	62,5
14	533	3,9	66,6	9	8,8	2,4	3,4	53,9
14	722	4,88	12	11,5	9,4	2,8	3	60,1
15	201	4,14	28,7	12,4	14,4	2,5	2,5	72,7
15	428	3,6	35,7	9,8	7,9	3,7	1,4	60,1
15	772	4,67	38,6	10,4	7,9	3	2,3	52,7
15	421	3,77	16	10	11,1	2,9	3,6	67,8
16	573	4,1	114,6	11,2	8,7	2,8	2,3	61,6
17	208	4,7	29,7	10,1	12	2,9	1,4	66,3
17	645	4,9	38	10,4	8,5	2,9	4,5	57,2
17	486	4	69,4	10,3	8,4	3,9	3,3	57,8
18	554	4,51	30,8	9	9	2,7	1,8	25,1