

Analysis of content and personality: A model for profiling

Análisis de Contenido y Personalidad: Modelo para la elaboración de perfiles.

Luís Sobrino Conde
Investigador independiente, Madrid, España.

Email correspondencia: lsobrino@hotmail.es

Resumen

Es evidente que la forma en que las personas se comunican no es azarosa, sino que viene determinada por distintas variables. Para el análisis de estas relaciones, se emplea el Análisis de Contenido. En el campo de la criminología, los investigadores han venido utilizando estas técnicas para predecir características o conductas en los autores de un texto. Sin embargo, la investigación al respecto es muy limitada. Bajo esta necesidad, en el presente estudio se lleva a cabo un análisis de las relaciones entre la elección de distintos elementos lingüísticos a la hora de comunicarse, y variables personales como son el sexo, la edad y los rasgos de personalidad, con el objetivo de ampliar el conocimiento en este ámbito. Los resultados obtenidos a partir de distintos análisis, ponen de manifiesto diferencias en la utilización de distintas categorías gramaticales, según el sexo, la edad y los rasgos de personalidad del autor. Mediante estos hallazgos, se generó un modelo que permite inferir características de un sujeto en base a un documento suyo, para posteriormente ser aplicado a un conjunto de documentos amenazantes pertenecientes a casos del Cuerpo Nacional de Policía, en los cuales no se conocía la personalidad del autor. En definitiva, este estudio pone de manifiesto la utilidad y potencialidad de estas técnicas, y la importancia que podrían llegar a tener para los investigadores, como herramienta complementaria, a la hora de realizar perfiles.

Palabras Clave

Perfilado, Análisis, Contenido, Personalidad, Amenazas, MAXQDA

Abstract

It is obvious that the way in which people communicate is not random, but is determined by different variables. For the analysis of these relationships, the Analysis of Content is used. In the field of Criminology, crime investigators have been using them to predict characteristics or behaviors of a text authors. However, the research in this regard is very limited. Under this need, in the present study is done an analysis of the relationship takes place between the choice of different linguistic items at the time to communicate something, and personal variables such as sex, age and personality traits, with the aim of expanding knowledge in this sphere. The results obtained from different analyzes, reveal differences in the use of different grammatical categories, by sex, age and personality traits of the author. Through these findings, it has generated a model that allows us to infer features of an author by analyzing their documents, to be subsequently applied to the cases of threats, belonging to the Cuerpo Nacional de Policía, in which the personality of the author was not known. In short, this study highlights the utility and potentiality of these techniques, and the importance that they could have for crime investigators, as a complementary tool, at the time to make profiles.

Keywords

Profiling, Analysis, Content, Personality, Threats, MAXQDA

Con el agradecimiento al Dr. Juan Enrique Soto Castro: Jefe de la Sección de Análisis de Conducta del Cuerpo Nacional de Policía.

Fecha de envío: 18/10/2014

Fecha de aceptación: 06/04/2015

I. INTRODUCCIÓN

Hoy en día, parece claro que la personalidad es uno de los factores que influyen en la manera de comunicarnos. Concretamente, en la comunicación verbal, la manera en que se utilizan las palabras podría revelar aspectos importantes del mundo social y psicológico de las personas. De manera más o menos consciente, elegimos unas palabras u otras a la hora de expresarnos, dependiendo del momento, del lugar, e incluso de la persona con la que nos comunicamos. Esta selección de palabras puede proporcionarnos una gran cantidad de información sobre el contexto, el emisor o los emisores, y el receptor o los receptores.

Bardin (1986) define el Análisis de Contenido como “un conjunto de técnicas de análisis de comunicaciones tendente a obtener indicadores (cuantitativos o no) por procedimientos sistemáticos y objetivos de descripción del contenido de los mensajes, permitiendo la inferencia de conocimientos relativos a las condiciones de producción/recepción (variables inferidas) de estos mensajes” (p. 32).

Dönges (2010), recoge una afirmación del psicólogo estadounidense Pennebaker, según el cual, “el contenido de un texto no supone la clave para desentrañar la personalidad de su autor, sino que lo son las palabras aparentemente insignificantes (como pronombres, artículos o conjunciones), ya que su uso resulta menos consciente”.

Generalmente, se distinguen dos tipos de análisis de contenido: Análisis cualitativo y análisis cuantitativo. En el primero de ellos se lleva a cabo una descripción detallada y completa de un fenómeno lingüístico, o del comportamiento de una palabra o grupo de palabras. Los autores que siguen este enfoque, defienden que el lenguaje es, por definición, contextual, y por ello, frases o palabras deben ser consideradas dentro del contexto de los objetivos del hablante. Por otro lado, en el análisis cuantitativo se asignan índices de frecuencia a los fenómenos lingüísticos observados en el cuerpo del texto (Mendoza y Angel, 2000), esto es, estadísticas basadas en el recuento de unidades (Raigada, 2002). La mayoría de estudios realizados siguen este último enfoque, que implica básicamente el conteo de palabras, tales como unidades de reglas gramaticales (p.e. pronombres, adjetivos, preposiciones), o dimensiones lingüísticas derivadas psicológicamente (p.e. “palabras violentas”, “palabras

emocionales positivas”). Esta metodología se fundamenta en la hipótesis de que la elección de unas determinadas categorías de palabras, proporciona información psicológica, más allá del contexto semántico y del significado literal de la palabra (Bardin, 1986). El problema de este análisis de codificación simple, son las limitaciones debidas a la polisemia de ciertas palabras, las cuales se pueden minimizar, con un análisis exhaustivo de los textos y un tratamiento individual de los datos, pero muy difícilmente se podrán llegar a eliminar.

Las técnicas de análisis de contenido presentan multitud de posibilidades y aplicaciones en diversos campos. En el área que nos ocupa, la criminología, han sido utilizadas principalmente como herramienta complementaria a la hora de elaborar perfiles, ayudando a desentrañar características o intenciones del autor a partir del análisis de sus documentos. Sin embargo, su uso se basa en un análisis superficial de los textos por parte del investigador, poco sistemático, y con un carácter eminentemente cualitativo. Pocos son los estudios dedicados a desentrañar las relaciones entre la forma de hablar y la personalidad. Uno de ellos es el realizado por Pennebaker y King (1999), en el cual, utilizando como instrumento de medición de la personalidad el Five Factor Inventory (John, Donahue y Kentle, 1991), y como instrumento de análisis cuantitativo el software Linguistic Inquiry and Word Count (LIWC), las correlaciones más altas obtenidas entre variables de personalidad (Agradabilidad, Neuroticismo, Extraversión, Apertura a la experiencia y Responsabilidad) y variables lingüísticas, oscilaban entre 0,10 y 0,16, siendo consideradas por los autores como “modestas, pero fiables”.

Es necesario por ello desarrollar metodologías que permitan fortalecer su fiabilidad y amplíen su funcionalidad. Concretamente, estas técnicas podrían llegar a ser de gran utilidad en el análisis de documentos de autor desconocido, ya que, si lográramos establecer modelos de relaciones entre la forma de comunicarse y características de diverso índole en el autor, sería posible generar perfiles a partir del análisis lingüístico, que complementen los realizados a partir de otras técnicas de investigación.

Porque la forma de utilizar las palabras es un fenómeno relativamente sin estudiar, en la presente investigación se plantea un modelo de desarrollo de las técnicas de análisis de contenido. En primer lugar, se pretende poner de manifiesto las relaciones existentes entre la forma en que elegimos las palabras que utilizamos al comunicarnos, y los

rasgos de personalidad postulados en el modelo de Eysenck (1967) –Extroversión (E), Neuroticismo (N) y Psicoticismo (P)-, el sexo y la edad, para posteriormente, aplicar los resultados y conclusiones obtenidas, de tal manera que se logre generar un modelo que permita inferir estas características del autor en cualquier texto. En segundo lugar, se realiza una aplicación práctica del modelo obtenido, a una serie de casos del Cuerpo Nacional de Policía de España (en adelante CNP), que incluían algún escrito amenazante, y en los cuales no se conocían las características del autor.

II. MÉTODO

A. Muestra

Debe distinguirse entre la muestra utilizada para el análisis de relaciones y elaboración del modelo, de la empleada para la aplicación del mismo. En lo que respecta a la primera parte, la elección de la muestra se ha realizado atendiendo a los criterios de pertinencia, representatividad y homogeneidad. El primero de ellos se cumple, al ajustarse la información recogida a los objetivos del trabajo. El segundo se ve reflejado en que los sujetos participantes se distribuyen aleatoriamente y de manera proporcional según el género y grupo de edad. En el caso de la edad, lo hacen únicamente en dos grupos -menores de 30 y mayores de 30-, siguiendo así la categorización utilizada en el Revised Eysenck Personality Questionnaire (1997) (en adelante EPQ-R) a la hora de establecer los centiles para las puntuaciones directas obtenidas. Por último, la homogeneidad se manifiesta en el tipo de información recogida: Todos los sujetos completan el mismo test de personalidad, y siguen la indicación de 1 hoja mínima de extensión a la hora de redactar el texto, tratando además el mismo de un tema común (“Las vacaciones”). Para esta primera parte del estudio, se ha contado con una muestra de 27 sujetos, que voluntariamente accedieron a formar parte de la investigación. Se distribuyen según género y edad como se aprecia en la *Tabla 1*.

Tabla 1. Muestra de sujetos

| | Hombres | Mujeres | Total |
|------------------|---------|---------|-------|
| Menos de 30 años | 7 | 9 | 16 |
| 30 años o más | 4 | 7 | 11 |
| Total | 11 | 16 | 27 |

En cuanto a la segunda parte, esto es, la muestra en la que fue aplicado el modelo, se utilizaron 18 documentos de amenazas escritas, relativos a 6 casos del CNP en los cuales no se conocía la identidad del autor.

La totalidad de documentos procesados entre ambos apartados del trabajo, contienen 4238 palabras diferentes, de las cuales 4218 (99,5%) han sido asignadas a, al menos, una de las categorías de estudio propuestas.

B. Variables de estudio

En cuanto a la selección de las variables de estudio, los dos apartados mencionados anteriormente cuentan con una serie de variables comunes, pero existen otras que, por su propia naturaleza, no pueden ser analizadas en ambos.

Comunes son las variables lingüísticas que reflejan las distintas categorías gramaticales reconocidas por la Real Academia de la Lengua Española (en adelante RAE), así como los signos de puntuación, y seis categorías creadas expresamente para el estudio, que han sido denominadas como “Palabras violentas”, “Verbos violentos”, “Insultos”, “Palabras de emocionalidad positiva”, “Palabras de emocionalidad negativa” y “Risas”.

Todas ellas se analizan mediante el porcentaje de utilización de las mismas, obtenido al dividir el número de apariciones entre el total de palabras empleadas en cada texto. Algunas de las categorías utilizadas, cuentan a su vez con subcategorías (p.e. “verbos” incluye gerundios). Por ello, se han analizado tanto las frecuencias generales de cada categoría (en el ejemplo, porcentaje de “verbos”, que contiene también el de “gerundios”), como la frecuencia de cada subcategoría (en el ejemplo, porcentaje de “gerundios”).

Es necesario explicar que determinadas categorías gramaticales no pueden ser tenidas en cuenta en un análisis de este tipo, debido a que existen palabras que actúan de diferente forma en según qué frases, y el software informático únicamente detecta cada palabra como unidad simple, sin evaluar su función. Por ejemplo, “nuestro” puede actuar como adjetivo posesivo (“nuestro” coche), y

también como pronombre posesivo (este coche es “nuestro”). Debido a estas limitaciones, no se tienen en cuenta en el análisis, las categorías generales de “Pronombres” y “Adjetivos”, así como sus respectivas subcategorías de “posesivos” y “demostrativos”, pues ambas presentarían frecuencias de palabras que no están actuando como miembros de estas categorías. Para evitar la pérdida de información que proporcionarían estas variables lingüísticas, se han empleado dos categorías –que ya habían sido usadas en otros estudios (Pennebaker y Stone, 2003)-, denominadas “Posesivos” y “Demostrativos”. En la primera de ellas, se incluyen tanto los adjetivos como los pronombres posesivos, con subcategorías de primera, segunda y tercera persona. La segunda recoge tanto los adjetivos demostrativos como los pronombres demostrativos.

Dejando a un lado las categorías comunes en ambos casos, el análisis de contenido de documentos de autor conocido, cuenta con una serie de variables de estudio propias: Por una parte, los distintos rasgos psicológicos propuestos por Eysenck (1967) y medidos por el EPQ-R, es decir, Extroversión, Neuroticismo y Psicoticismo). Por otra parte, la edad y el sexo. En las variables psicológicas, se utilizan para los análisis tanto la puntuación directa como el centil. Respecto a la edad, se emplean como variables el número de años, y el subgrupo de edad al que pertenecen –coincidiendo con los propuestos por el EPQ-R- recodificados numéricamente como “1”: <30 años, y “2”: ≥ 30 años. En cuanto a la variable sexo, el subgrupo de hombres es recodificado numéricamente como “1”, y el de mujeres como “2”.

C. Procedimiento

En primer lugar, por un lado, se pidió a cada sujeto que participaba en el estudio, que realizara una redacción, con una extensión mínima de una cara en Microsoft Office (letra “Arial”, tamaño 11), que tratase sobre un tema seleccionado previamente; las vacaciones. Asignar una temática común (homogeneidad), favorece el análisis de diferencias en la elección de palabras, ya que si cada documento versara sobre un tema distinto, las diferencias halladas podrían deberse a esta disparidad. Además, el hecho de tratarse de un asunto autobiográfico, evita la posible incidencia de diferencias culturales. También es necesario establecer una extensión mínima en los escritos, pues textos muy cortos no permitirían la obtención de resultados concluyentes. Sin embargo, no se impone una extensión máxima, dejando a elección del sujeto el número

de palabras que puede alcanzar su redacción. Por otro lado, se facilitó a cada participante el test de personalidad EPQ-R, para que lo cumplimentara. La elección de este test y modelo de personalidad, se debe a que es el que generalmente utilizan los investigadores a la hora de realizar perfiles.

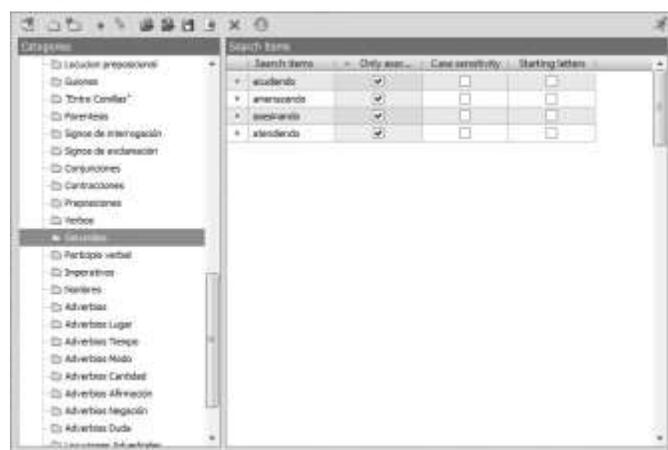
En segundo lugar, se realizó el análisis de documentos mediante el software MAXQDA versión 11. Éste permite obtener la frecuencia de aparición de cada palabra en un texto (*Imagen 1*), así como de categorías de palabras.

Imagen 1. Captura del MAXQDA v11. Frecuencia de palabras

| Word | Word length | Frequency | % |
|-----------|-------------|-----------|------|
| + año | 3 | 39 | 0,24 |
| + donde | 5 | 39 | 0,24 |
| + cuando | 6 | 37 | 0,23 |
| + siempre | 7 | 36 | 0,23 |
| + sin | 3 | 36 | 0,23 |
| + verano | 6 | 36 | 0,23 |
| + fue | 3 | 35 | 0,22 |
| + porque | 6 | 34 | 0,21 |
| + si | 2 | 34 | 0,21 |
| + también | 7 | 34 | 0,21 |
| + casa | 4 | 33 | 0,21 |
| + son | 3 | 33 | 0,21 |

Sin embargo, al no existir un diccionario en español para el software, que reconozca cada palabra como perteneciente a una clase gramatical, fue necesaria la elaboración de este “diccionario”, creando las categorías y asignando cada palabra a aquella a la que pertenece, de tal manera que pudiera hallarse la frecuencia de aparición de las diferentes categorías. (*Imagen 2*)

Imagen 2. Captura de MAXQDA v11. Creación del Diccionario.



Posteriormente, las frecuencias de aparición de cada categoría, en el correspondiente documento, se transformaron en porcentajes respecto al número de palabras del mismo.

En tercer lugar, se obtuvieron las correlaciones entre el porcentaje de aparición de cada categoría lingüística en un texto, y las variables personales de su autor (rasgos de personalidad, edad y sexo).

En base a estas relaciones, se elaboró el modelo de generación de perfiles a partir de documentos escritos. Los perfiles están formados por cinco variables personales: Extroversión alta/Introversión alta, Neuroticismo alto/Estabilidad emocional, Psicoticismo alto/Psicoticismo bajo, Edad por grupo –Menos de 30 años/30 años o más-, y Sexo –Varones/Mujeres-, de modo que el sujeto pueda tender hacia un polo u otro, en el caso de los rasgos de personalidad, y pertenecer a un grupo u otro, en el caso de la edad y el sexo.

Para obtener las variables personales, el modelo utiliza como variables predictoras aquellas categorías lingüísticas que han presentado correlaciones significativas con las variables del perfil en sujetos conocidos, de tal manera que cada relación supone un indicador de tendencia hacia un polo, o de pertenencia a un grupo (*ver Tablas 12, 13, 14, 15 y 16*). En los casos en que alguna variable del perfil no presentaba al menos 7 correlaciones significativas con categorías lingüísticas, y por tanto no se podían obtener 7 indicadores, se utilizaron aquellas categorías con las siguientes correlaciones más altas, hasta llegar al mínimo de 7. En la variable Neuroticismo, a pesar de existir 8 correlaciones significativas, se añadió un noveno indicador –con la correlación más alta de las no significativas- para evitar que, al ser par el número inicial de éstos, se produjera una hipotética igualdad -4 indicadores de neuroticismo y 4 de estabilidad emocional- que impidiera inferir hacia qué grupo tendía el sujeto. Se debe puntualizar que, para los indicadores de rasgos de personalidad, se han utilizado tanto las correlaciones de categorías gramaticales con la puntuación directa del rasgo, como con la de su centil.

El modelo se aplica de la siguiente manera: Para obtener la predicción acerca de la variable personal del sujeto desconocido, se atiende a las relaciones descritas en los indicadores, y se compara el porcentaje en que el sujeto desconocido utiliza la categoría gramatical de cada indicador, con el porcentaje medio de utilización de esa categoría en toda la muestra (Tabla 2). Cuando un indicador refleja una correlación positiva entre una categoría lingüística y una variable personal, y el sujeto desconocido utiliza esa categoría lingüística por encima de la media, se infiere que, respecto al indicador utilizado,

dicho sujeto tenderá a puntuaciones altas en ese rasgo –si se trata de variable de personalidad-, o a pertenecer al subgrupo recodificado como “2” –si el indicador es de sexo o grupo de edad-. Si, por el contrario, su porcentaje de utilización está por debajo de la media, se infiere que el sujeto tiende hacia puntuaciones altas en el otro polo del rasgo, o a pertenecer con mayor probabilidad al otro subgrupo. El proceso es inverso para aquellos indicadores que presentan correlaciones negativas entre la categoría lingüística y la variable personal, es decir, en estos casos, si el sujeto desconocido utiliza la categoría en mayor medida que la media, dicho indicador nos hace inferir que el sujeto presentará puntuaciones bajas en ese rasgo –en variables de personalidad- o pertenecerá al subgrupo recodificado como “1” –en las variables sexo y edad por grupo-. Cuando el sujeto utiliza por debajo de la media la categoría lingüística de un indicador de correlación negativa, se infiere que éste tenderá hacia puntuaciones altas en la variable – en el caso de ser un rasgo-, o a pertenecer al subgrupo “2” –si se trata de las variables edad por grupo o sexo-.

Tabla 2. Interpretación de los indicadores

| | Correlación positiva | Correlación negativa |
|------------------------------------|---|---|
| Utilización por encima de la media | Tendencia a puntuaciones altas en rasgo, o a pertenecer al subgrupo “2” | Tendencia a puntuaciones bajas en rasgo (altas en el otro polo), o a pertenecer al subgrupo “1” |
| Utilización por debajo de la media | Tendencia a puntuaciones bajas en rasgo (altas en el otro polo), o a pertenecer al subgrupo “1” | Tendencia a puntuaciones altas en rasgo, o a pertenecer al subgrupo “2” |

Posteriormente, el número de indicadores de tendencia o pertenencia hacia cada polo o subgrupo, es transformado en porcentaje, obteniendo así la probabilidad de que cada uno de los polos o de los subgrupos, estén presentes en el autor de cada documento anónimo –esto es, que el sujeto tienda a presentar puntuaciones altas en dicho polo, o con mayor probabilidad pertenezca a determinado subgrupo-.

Por ejemplo, para el Caso 1 (*Imagen 3*), de las 9 categorías utilizadas para la variable “Extroversión”, 2 (22,2%) indican tendencias a la “Extroversión alta”, y 7 (77,7%) a la “Introversión alta”; de las 9 relacionadas con “Neuroticismo”, las 9 reflejan tendencias al “Neuroticismo alto” (100%) y 0 a la “Estabilidad emocional” (0%); de las 7 relacionadas con Psicoticismo, 5 reflejarían “Psicoticismo alto” (71,4%), y 2 “Psicoticismo bajo” (28,5%); respecto a la edad por grupo, se observa que, de las 7 categorías

relacionadas con esta variable, 6 lo están con el grupo “Menos de 30” (85,7%), y 1 de ellas con el de “30 o más” (14,2%); por último, en la variable sexo, 5 de las 7 variables indican la pertenencia al grupo “Varones” (71,4%), y 2 al de “Mujeres” (28,5%).

Imagen 3. Ejemplo de elaboración del perfil en Caso 1

| | Extroversión | Neuroticismo | Psicoticismo | Edad por grupo | Sexo |
|--------|--------------|--------------|--------------|----------------|------|
| Caso 1 | 2/7 | 9/0 | 5/2 | 6/1 | 5/2 |

| | Extroversión | Neuroticismo | Psicoticismo | Edad por grupo | Sexo |
|--------|---------------------------|--------------|--------------|----------------|---------------|
| Caso 1 | Introversión alta (77,7%) | Alto (100%) | Alto (71,4%) | 30 (85,7%) | Varón (71,4%) |



III. RESULTADOS

Utilizando como referencia las correlaciones obtenidas por Pennebaker y King (1999), -oscilaban entre 0,10 y 0,16-, y dado que estas correlaciones pudieran ser algo bajas, en el presente estudio se considerará como relevantes, y se emplearán para la elaboración del modelo, aquellas correlaciones iguales o superiores a 0,25

A. Relaciones entre variables personales y categorías lingüísticas

A continuación se exponen de manera detallada, las correlaciones, superiores a 0,25, que presentan las variables de personalidad, edad y sexo, con las distintas categorías lingüísticas:

Tabla 3. Correlaciones superiores a 0,25 de la variable “Extroversión” con otras variables

| Categoría lingüística | Variable personal Extroversión |
|----------------------------|-----------------------------------|
| Número de palabras | 0,534** |
| Locuciones Adverbiales | -0,457* |
| Guiones | 0,376 |
| Signos de exclamación | 0,355 |
| Entre comillas | 0,340 |
| Locuciones preposicionales | -0,322 |
| Emocionalidad positiva | -0,297 |
| Puntos suspensivos | 0,279 |
| Adverbios Tiempo | 0,264 |

Tabla 4. Correlaciones superiores a 0,25 de la variable “Centil Extroversión” con otras variables.

| Categoría lingüística | Variable personal Centil Extroversión |
|------------------------|--|
| Locuciones Adverbiales | -0,504** |
| Entre comillas | 0,447* |
| Signos de exclamación | 0,429* |

| | |
|----------------------------|---------|
| Guiones | 0,420* |
| Número de palabras | 0,419* |
| Participios verbales | -0,409* |
| Puntos suspensivos | 0,358 |
| Emocionalidad negativa | 0,265 |
| Locuciones preposicionales | -0,296 |

Tabla 5. Correlaciones superiores a 0,25 de la variable “Neuroticismo” con otras variables.

| Categoría lingüística | Variable personal Neuroticismo |
|----------------------------------|-----------------------------------|
| Pronombres Personales 2ª Persona | 0,632** |
| Adverbios negación | 0,480* |
| Adverbios Duda | 0,412* |
| Imperativos | 0,383* |
| Puntos suspensivos | 0,326 |
| Verbos violentos | -0,321 |
| Adjetivos calificativos | -0,316 |
| Locuciones conjuntivas | 0,286 |
| Risas | -0,275 |
| Números | -0,263 |
| Posesivos 3ª Persona | 0,261 |
| Guiones | -0,252 |

Tabla 6. Correlaciones superiores a 0,25 de la variable “Centil Neuroticismo” con otras variables.

| Categoría lingüística | Variable personal Centil Neuroticismo |
|----------------------------------|--|
| Pronombres Personales 2ª Persona | 0,554** |
| Adverbios Duda | 0,435* |
| Adverbios Negación | 0,401* |
| Puntos suspensivos | 0,382* |
| Pronombres Personales 1ª Persona | -0,289 |
| Posesivos 2ª Persona | -0,286 |
| Entre comillas | 0,286 |
| Verbos violentos | -0,285 |
| Nombres | 0,282 |
| Números | -0,272 |

Tabla 7. Correlaciones superiores a 0,25 de la variable “Psicoticismo” con otras variables.

| Categoría lingüística | Variable personal Psicoticismo |
|----------------------------------|-----------------------------------|
| Posesivos 1ª Persona | -0,440* |
| Risas | 0,392* |
| Elementos compositivos | 0,378 |
| Artículos | -0,36 |
| Puntos | 0,334 |
| Participios verbales | -0,327 |
| Pronombres Personales 2ª Persona | 0,287 |
| Números | 0,276 |
| Preposiciones | 0,271 |

Tabla 8 Correlaciones superiores a 0,25 de la variable “Centil Psicoticismo” con otras variables.

| Categoría lingüística | Variable personal Centil Psicoticismo |
|----------------------------------|--|
| Posesivos 1ª Persona | -0,383* |
| Pronombres Personales 2ª Persona | 0,357 |
| Artículos | -0,337 |
| Puntos | 0,337 |
| Elementos compositivos | 0,324 |
| Signos de interrogación | 0,308 |
| Participio verbal | -0,307 |
| Adverbios cantidad | -0,288 |
| Interjecciones | 0,287 |
| Nombres | 0,282 |
| Gerundios | -0,267 |
| Posesivos | -0,266 |
| Risas | 0,265 |

Tabla 9. Correlaciones superiores a 0,25 de la variable “Sexo” con otras variables.

| Categoría lingüística | Variable personal Sexo |
|----------------------------------|---------------------------|
| Locuciones preposicionales | -0,593** |
| Palabras violentas | -0,567** |
| Adverbios comparativos | 0,481* |
| Comas | 0,387* |
| Expresiones | 0,343 |
| Verbos violentos | -0,341 |
| Posesivos Segunda Persona | 0,339 |
| Locuciones conjuntivas | 0,339 |
| Adverbios de lugar | 0,308 |
| Número de palabras | -0,307 |
| Signos de interrogación | 0,27 |
| Pronombres Personales 3ª Persona | -0,257 |

Tabla 10. Correlaciones superiores a 0,25 de la variable “Edad” con otras variables.

| Categoría lingüística | Variable personal Edad |
|----------------------------------|---------------------------|
| Verbos | -0,596** |
| Nombres | 0,543** |
| Adverbios | -0,538** |
| Adverbios de Modo | -0,492** |
| Adverbios de cantidad | -0,475* |
| Posesivos | -0,400* |
| Emocionalidad positiva | 0,397* |
| Gerundios | -0,332 |
| Imperativos | -0,323 |
| Artículos | 0,314 |
| Pronombres Personales 3ª Persona | -0,313 |
| Adverbios de tiempo | -0,312 |
| Insultos | 0,289 |
| Contracciones | 0,279 |
| Adjetivos calificativos | 0,267 |

| | |
|----------------------------------|--------|
| Entre comillas | 0,262 |
| Pronombres Personales 1ª Persona | -0,259 |

Tabla 11. Correlaciones superiores a 0,25 de la variable “Grupo de edad” con otras variables.

| Categoría lingüística | Variable personal Grupo de edad |
|----------------------------------|------------------------------------|
| Verbos | -0,540** |
| Posesivos | -0,534** |
| Adverbios Cantidad | -0,487* |
| Adverbios Modo | -0,437* |
| Emocionalidad positiva | 0,434* |
| Adverbios | -0,422* |
| Nombres | 0,389* |
| Gerundios | -0,377 |
| Contracciones | 0,373 |
| Entre comillas | 0,319 |
| Comas | 0,298 |
| Pronombres Personales 2ª Persona | 0,293 |
| Pronombres Personales 1ª Persona | -0,29 |
| Adverbios Duda | 0,287 |
| Imperativos | -0,251 |
| Adjetivos calificativos | 0,250 |

B. Modelo para la generación de perfiles

En este apartado se presentan los indicadores utilizados en el modelo para la obtención de cada variable del perfil, en base a la correlación obtenida (Tablas 12, 13, 14, 15 y 16).

Tabla 12. Indicadores utilizados para la variable “Extroversión/Introversión”

| | Extroversión/Introversión | | |
|----|---------------------------|------------------------|-------------|
| | Variable personal | Categoría lingüística | Correlación |
| I1 | Extroversión | Locuciones adverbiales | -0,457* |
| I2 | Centil Extroversión | Locuciones adverbiales | -0,504** |
| I3 | Centil Extroversión | Entre comillas | 0,447* |
| I4 | Centil Extroversión | Signos de exclamación | 0,429* |
| I5 | Centil Extroversión | Guiones | 0,420* |
| I6 | Centil Extroversión | Participios verbales | -0,409* |
| I7 | Extroversión | Guiones | 0,376 |

Tabla 13. Indicadores utilizados para la variable “Neuroticismo/Estabilidad emocional”

| | Neuroticismo/Estabilidad emocional | | |
|----|------------------------------------|----------------------------------|-------------|
| | Variable personal | Variable lingüística | Correlación |
| I1 | Neuroticismo | Pronombres personales 2ª persona | 0,632** |
| I2 | Centil Neuroticismo | Pronombres personales 2ª persona | 0,554** |
| I3 | Neuroticismo | Adverbios de negación | 0,480* |
| I4 | Centil Neuroticismo | Adverbios de duda | 0,435* |

| | | | |
|----|---------------------|-----------------------|--------|
| 15 | Neuroticismo | Adverbios de duda | 0,412* |
| 16 | Centil Neuroticismo | Adverbios de negación | 0,401* |
| 17 | Neuroticismo | Imperativos | 0,383* |
| 18 | Centil Neuroticismo | Puntos suspensivos | 0,382* |
| 19 | Neuroticismo | Puntos suspensivos | 0,326 |

Tabla 14. Indicadores utilizados para la variable “Psicoticismo Alto/Psicoticismo Bajo”

| Psicoticismo Alto/Psicoticismo Bajo | | | |
|-------------------------------------|---------------------|----------------------------------|-------------|
| | Variable personal | Variable lingüística | Correlación |
| I1 | Psicoticismo | Posesivos 1ª Persona | -0,440* |
| I2 | Centil Psicoticismo | Posesivos 1ª Persona | -0,383* |
| I3 | Psicoticismo | Risas | 0,392* |
| I4 | Psicoticismo | Elementos compositivos | 0,378 |
| I5 | Psicoticismo | Artículos | -0,36 |
| I6 | Centil Psicoticismo | Pronombres Personales 2ª persona | 0,357 |
| I7 | Centil Psicoticismo | Puntos | 0,337 |

Tabla 15. Indicadores utilizados para la variable “Sexo”

| Sexo | | | |
|------|-------------------|----------------------------|-------------|
| | Variable personal | Variable lingüística | Correlación |
| I1 | Sexo | Locuciones preposicionales | -0,593** |
| I2 | Sexo | Palabras violentas | -0,567** |
| I3 | Sexo | Adverbios comparativos | 0,481* |
| I4 | Sexo | Comas | 0,387* |
| I5 | Sexo | Expresiones | 0,343 |
| I6 | Sexo | Verbos violentos | -0,341 |
| I7 | Sexo | Posesivos 2ª persona | 0,339 |

Tabla 16. Indicadores utilizados para la variable “Edad por grupo”

| Edad por grupo | | | |
|----------------|-------------------|------------------------|-------------|
| | Variable personal | Variable lingüística | Correlación |
| I1 | Edad por grupo | Verbos | -0,540** |
| I2 | Edad por grupo | Posesivos | -0,534** |
| I3 | Edad por grupo | Adverbios Cantidad | -0,487* |
| I4 | Edad por grupo | Adverbios Modo | .0,437* |
| I5 | Edad por grupo | Emocionalidad positiva | 0,434* |
| I6 | Edad por grupo | Adverbios | -0,422* |
| I7 | Edad por grupo | Nombres | 0,389* |

Como ya se ha explicado ampliamente en el punto II. C. (Procedimiento), el modelo se aplica a cada caso concreto, obteniendo el número de indicadores para cada polo o grupo de la variable (Tabla 17), y posteriormente se transforma en porcentajes, obteniendo así un perfil probabilístico para cada uno de los casos (Tabla 18).

Tabla 17. Indicadores de variables en cada Caso.

| Caso | E | | N | | P | | Edad | | Sexo | |
|------|----|----|----|----|----|----|------|------|------|---|
| | EA | IA | NA | EM | PA | PB | <30 | ≥ 30 | V | M |
| 1 | 2 | 7 | 9 | 0 | 5 | 2 | 6 | 1 | 5 | 2 |
| 2 | 7 | 2 | 5 | 4 | 3 | 4 | 3 | 4 | 3 | 4 |
| 3 | 3 | 6 | 5 | 4 | 5 | 2 | 3 | 4 | 5 | 2 |
| 4 | 3 | 6 | 9 | 0 | 5 | 2 | 4 | 3 | 6 | 1 |
| 5 | 2 | 7 | 7 | 2 | 5 | 2 | 4 | 3 | 5 | 2 |
| 6 | 4 | 5 | 5 | 4 | 5 | 2 | 5 | 2 | 4 | 3 |

EA: Extroversión alta IA: Introversión alta NA: Neuroticismo alto
EM: Estabilidad emocional
PA: Psicoticismo alto PB: Psicoticismo bajo <30: Menos de 30 años
≥30: 30 años o más

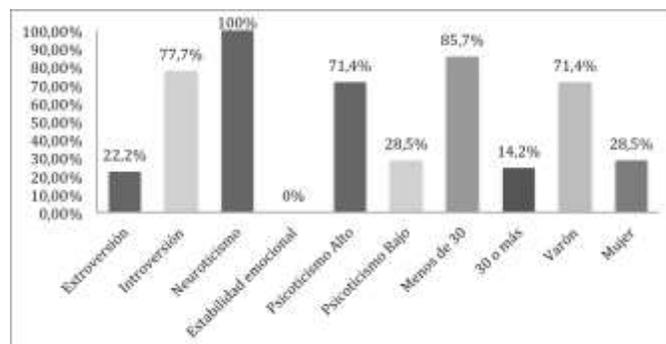
Tabla 18. Perfil probabilístico para cada Caso

| | Extroversión | Neuroticismo | Psicoticismo | Grupo Edad | Sexo |
|---|----------------------|----------------------|----------------------|-------------|-----------|
| 1 | IA (77,7%) | NA (100%) | PA (71,4%) | <30 (85,7%) | V (71,4%) |
| 2 | EA (77,7%) | N Medio-Alto (55,5%) | P Medio-Bajo (57,1%) | ≥30 (57,1%) | M (57,1%) |
| 3 | IA (66,6%) | N Medio-Alto (55,5%) | PA (71,4%) | ≥30 (57,1%) | V (71,4%) |
| 4 | IA (66,6%) | NA (100%) | PA (71,4%) | <30 (57,1%) | V (85,7%) |
| 5 | IA (77,7%) | NA (77,7%) | PA (71,4%) | <30 (57,1%) | V (71,4%) |
| 6 | I media-alta (55,5%) | N Medio-Alto (55,5%) | PA (71,4%) | <30 (71,4%) | V (57,1%) |

En base a estos datos obtenidos es posible extraer las siguientes conclusiones:

Para el Caso 1 (Tabla 19), es posible inferir que se trataría de un varón (con una probabilidad del 71,4%), con menos de 30 años (85,7%), con tendencias hacia la introversión (77,7%), alto neuroticismo (100%) y psicoticismo (71,4%).

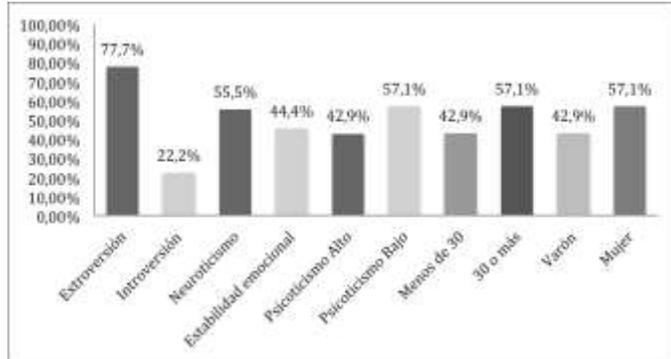
Tabla 19. Perfil Caso I



En cuanto al Caso 2 (Tabla 20), el autor sería, con

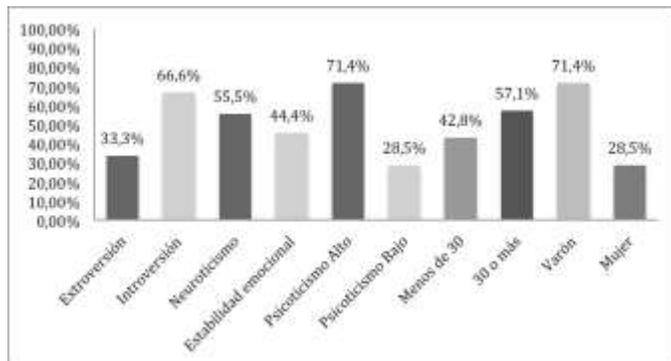
mayor probabilidad una mujer (57,1%), de 30 años o más (57,1%), tendente a la extroversión (77,7%), con puntuaciones medias-altas en neuroticismo (55,5%) y en psicoticismo bajo (57,1%).

Tabla 20. Perfil Caso 2



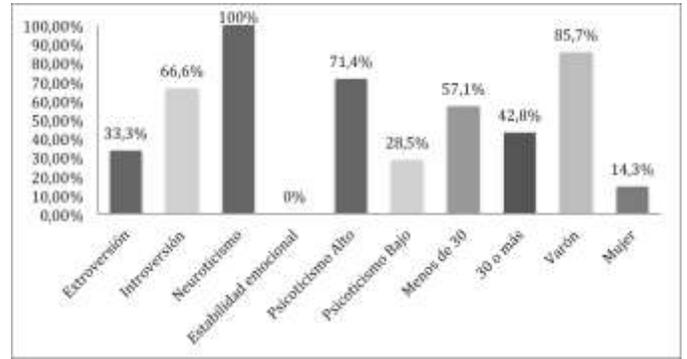
Detrás de los escritos del Caso 3 (Tabla 21) habría, con mayor probabilidad, un varón (71,4%), de más de 30 años (57,1%), con tendencia hacia la introversión (66,6%), neuroticismo medio-alto (55,5%) y psicoticismo alto (71,4%).

Tabla 21. Perfil Caso 3



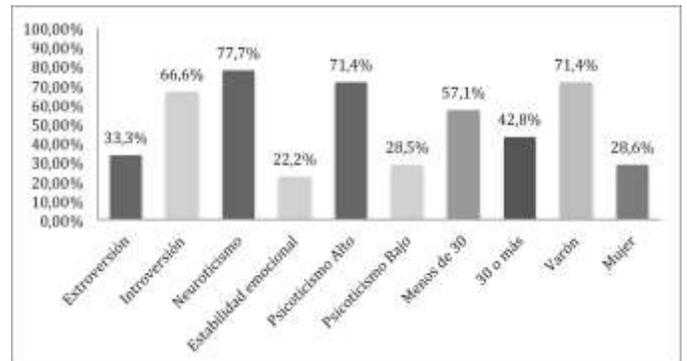
En el Caso 4 (Tabla 22), se postula la autoría de un Varón (85,7%), de menos de 30 años (57,1%), con inclinación hacia la introversión (66,6%), alto neuroticismo (100%) y alto psicoticismo (71,4%).

Tabla 22. Perfil Caso 4



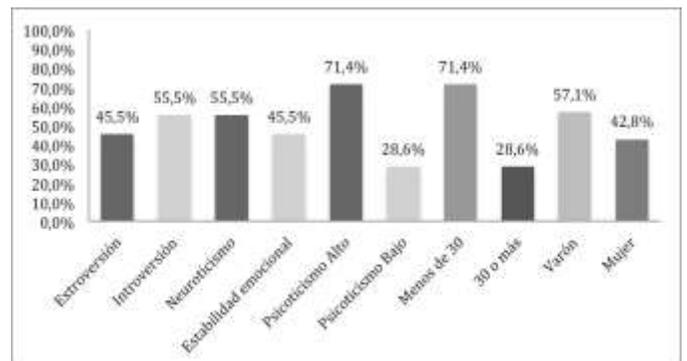
Para el Caso 5 (Tabla 23), los resultados permiten hipotetizar que se trataría de un varón (71,4%), con menos de 30 años (57,1%), tendente a la introversión (77,7%), neuroticismo (77,7%) y psicoticismo alto (71,4%).

Tabla 23. Perfil Caso 5



Por último, en el Caso 6 (Tabla 24) se puede inferir que, con mayor probabilidad, el autor es un Varón (57,1%), con puntuaciones medias-altas en introversión (55,5%) y neuroticismo (55,5%), así como puntuaciones altas en psicoticismo (71,4%).

Tabla 24. Perfil Caso 6



IV. DISCUSION

La presente investigación se desarrolla bajo el objetivo de mostrar las potencialidades del análisis cuantitativo de contenido, generando un modelo estandarizado que permita inferir distintas características de un sujeto (rasgos de personalidad, edad y sexo), a partir del análisis de sus documentos escritos. En lo que respecta a los rasgos de personalidad, se ha utilizado como base teórica el modelo de tres macrorasgos de Eysenck (1967) por los siguientes motivos: Por un lado, porque el número reducido de macrorasgos utilizados permite una mayor facilidad en la evaluación de las variables. Frente a las posturas de otros autores que postulan la existencia de numerosos rasgos de personalidad, como la de Catell (Catell y Kline, 1982), que propone la existencia de 16 factores o rasgos, el modelo de Eysenck plantea la posibilidad de describir la estructura de la personalidad con únicamente tres macrorasgos, cuantitativos y bipolares, entendidos como unidades amplias de organización de la conducta (Extroversión, Neuroticismo y Psicoticismo). En segundo lugar, el hecho de poseer una fundamentación biológica, nos aporta información adicional, gracias a las asociaciones establecidas entre los rasgos y las características neurofisiológicas subyacentes a los mismos.

Los resultados obtenidos, en los distintos análisis realizados en este trabajo, reflejan lo siguiente:

En primer lugar, la existencia de diferencias en las frecuencias de utilización de las distintas categorías gramaticales, según variables como la personalidad, la edad o el sexo. Tomando como punto de corte para el análisis las correlaciones superiores a 0,25, los resultados reflejan la existencia de 9 categorías relacionadas con la extroversión (desde una correlación de 0,264 con “adverbios de tiempo”, hasta 0,534* con “número de palabras”); 9 con el centil de extroversión (desde -0,296 en “locuciones preposicionales”, hasta -5,04** en “locuciones adverbiales”); 12 con neuroticismo (desde -0,252 en “guiones”, hasta 0,632** en “pronombres personales 2ª persona”); 10 con centil de neuroticismo (desde -0,272 en “números”, hasta 0,554** en “pronombres personales 2ª persona”); 9 con psicoticismo (desde 0,271 en “preposiciones”, hasta -0,440* en “posesivos 1ª persona”); 13 con centil de psicoticismo (desde 0,265 en “risas”, hasta -0,383* en “posesivos 1ª persona”); 12 con sexo (desde -0,257 en “pronombres personales 3ª persona”, hasta -0,593** en “locuciones preposicionales”); y 17 con la variable edad (desde -0,259 en “pronombres personales 1ª

persona, hasta -0,596** en “verbos”). Ante la ausencia de datos empíricos, que permitan establecer un punto de corte adecuado para otorgarles cierta importancia a las correlaciones obtenidas, en el presente estudio se ha tomado como referencia el criterio de Pennebaker y King (1999), quienes, en uno de los pocos estudios de estas características, realizados hasta la fecha, consideraron como fiables unas correlaciones que oscilaban entre 0,10 y 0,16. Sin embargo, teniendo en cuenta que éstas pueden ser algo bajas, y dada la cantidad de correlaciones superiores halladas en el presente estudio, se ha optado por fijar el punto de corte para el análisis en 0,25. Por otro lado, la comparación de medias realizada, entre los subgrupos de cada variable, proporciona resultados que confirman la existencia de diferencias en las frecuencias de utilización de cada categoría.

En segundo lugar, el gran número de altas correlaciones obtenidas entre las distintas variables de personalidad, sexo y edad, con las categorías lingüísticas utilizadas en este estudio, ha permitido la generación de un modelo de obtención de perfiles en documentos de autor desconocido, y su posterior aplicación a distintos casos reales pertenecientes al CNP.

Este modelo debe ser mejorado en futuras investigaciones, entre otros aspectos, prestando mayor atención en el análisis a variables sociodemográficas que pueden tener gran importancia en la comunicación verbal, como son el nivel de escolaridad o el tipo de zona (rural o urbana) en la que reside el sujeto. También debe tenerse en cuenta que respecto a la edad, los análisis se realizan distinguiendo únicamente entre dos grupos amplios (mayores y menores de 30 años), lo que supone una pérdida de información que puede ser relevante, por lo que esta variable deberá ser controlada en futuros trabajos, estableciendo grupos de edades con menor amplitud que nos permitan obtener más información, y más específica, sobre la misma.

Pese a la necesidad de mejorar estos aspectos, para seguir mejorando el modelo, el estudio cumple con el objetivo de mostrar las potencialidades de las técnicas de análisis de contenido, a la vez que sienta las bases para su aplicación a un campo tan importante como es la investigación criminal.

V. REFERENCIAS

- Bardin, L. (1996) Análisis de contenido (2). Akal.
- Cattell, R. B., Kline, P., Zamora, J. M. P., & Calabozo, J. E. (1982).

El análisis científico de la personalidad y la motivación. Ediciones Pirámide.

Dönges, J. (2010). Eres lo que hablas. *Mente y cerebro*, (43), 22-27. Psychological aspects of natural language use: Our words, our selves. *Annual review of psychology*, 54(1), 547-577.

Eysenck, H. J. (1967). *The biological basis of personality* (Vol. 689). Transaction publishers.

Eysenck, H. J., i Fabregat, G. O., & Eysenck, S. B. (1997). EPQ-R: cuestionario revisado de personalidad de Eysenck: versiones completa (EPQ-R) y abreviada (EPQ-RS): manual. España: TEA Ediciones.

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The big five inventory—versions 4a and 54*. Berkeley: University of California, Berkeley, Institute of Personality and Social Research.

Mendoza, M. A. G., & Angel, M. (2000). Análisis de contenido

cualitativo y cuantitativo: definición, clasificación y metodología. *Revista de Ciencias Humanas*. Fecha de consulta 12 de Febrero de 2014. [URL:http://www.utp.edu.co/~chumanas/revistas/revistas/rev20/gomez.htm](http://www.utp.edu.co/~chumanas/revistas/revistas/rev20/gomez.htm)

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: language use as an individual difference. *Journal of personality and social psychology*, 77(6), 1296.

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.

Raigada, J. L. P. (2002). Epistemología, metodología y técnicas del análisis de contenido. *Sociolinguistic Studies*, 3(1), 1-42.