D E L T A

Analysing deception in a psychopath's speech: a quantitative approach

Analisando a mentira no discurso de um psicopata: uma abordagem quantitativa

Ángela ALMELA (Centro Universitario de la Defensa, UPCT)
Gema ALCARAZ-MÁRMOL (Universidad de Castilla-La Mancha)
Pascual CANTOS (Universidad de Murcia)

ABSTRACT

Psychopathy involves a series of specific cognitive, social and emotional features which make the psychopath different from the general population; the two most significant characteristics are extreme selfishness and deep emotional deficit that is reflected in apathy. Notably, psychopaths are skilled communicators who that use language to lie. As there has been little examination of the speech associated specifically with psychopaths, especially in the Spanish language, the present study aims to contrast different veracious excerpts to others which are deceptive. The text analysis is framed within forensic computational linguistics, and complemented with some information related to the stylometric profile of the text. The investigation shows how the parameter mainly affected by the psychological condition of the psychopath subject is the distribution of grammatical persons; in addition, some further evidence includes the frequency of certainty adverbs and verbs related to cognitive processes.

Key-words: *Psychopathy; mental disorder; forensic linguistics; deception; computational analysis.*

Ángela Almela, Gema Alcaraz-Mármol, Pascual Cantos

RESUMO

Psicopatia envolve uma série de características cognitivas, sociais e emocionais específicas que diferenciam o psicopata do resto da população; as duas características mais notáveis são os déficits emocionais extremos e profundos refletidos na apatia do próprio egoísmo. Também vale destacar as habilidades de comunicação dos psicopatas, que usam a linguagem para mentir. Há uma falta de estudos sobre o discurso de psicopatas, especialmente em espanhol; portanto, o presente estudo objetiva contrastar declarações veridicas com outras declarações difamatórias verdadeiras. A análise textual feita enquadra-se no campo da linguística computacional forense, e é complementada com informações mais específicas para os procedimentos estilometria textual. A pesquisa mostra como o parâmetro de idioma predominante para a condição psicológica do psicopata tem a ver com a distribuição de pessoa gramatical; além disso, evidencia a frequência de uso de certos advérbios modais de segurança e verbos relacionados com os processos cognitivos.

Palavras-chave: *Psicopatía; transtorno mental; linguística forense; mentira; análise computacional.*

1. Introduction: Mental Diseases and Language

There has always been an interest in the study of mental diseases from different points of view. They have been normally studied from the clinical perspective. Nevertheless, some scholars have studied different psychological disorders by means of word analysis or linguistic patterns, most of them within the English speaking context. This type of studies are already found in the 50s such as the ones carried out by Lorenz and Cobb (1952; 1953) which were mostly related to psychoneurotic and maniac patients. Maniac people's speech was also of interest for Andreasen and Pfohl (1976). These studies explored different syntactic and grammatical patterns revealed by the linguistic behaviour of mentally disordered people. Lindenfeld (1973) observed changes in people's syntactic patterns according to their affective state. Schizophrenia has also been explored from a linguistic approach. To this respect, Chaika (1974) identified six linguistic aspects that were specific of schizophrenic speech. Thomas et al. (1987) focused on the

syntactic structures used by schizophrenic patients while speaking. They could find that those with negative symptoms used more simple syntactic constructions than those with positive symptoms, who, at the same time committed more syntactic and grammatical errors.

More recent examples use computational linguistics for their analysis. This is the case of Lott et al. (2002), who analyse the speech of three types of mental patients: schizophrenic, bipolar and depressive. Anderson et al. (2008) made a quantitative and qualitative analysis of the lexis used in narratives written by people who suffered from social phobia. The literary productions of some well-known writers have also been analysed given the fact that these writers suffered from some mental disease. Forgeard (2008) compared eminent writers diagnosed with unipolar disorder to those who were bipolar (Jane Austen, Charlotte Brönte, Henry James, Leon Tolstoy or Scott Fitzgerald). She found some differences between the two groups. Bipolar writers made more allusions to death, fewer references to other people rather than themselves and used fewer cognitive verbs in their speech. In a recent study Cantos (2014) focused his analysis on the controversial writer Poe. He observed how Poe's mental status was reflected in his linguistic behavior, affecting both lexis and syntax.

2. Psychopathy

The current study explores psychopathy. This mental disorder has been conceived in different ways by different experts. We are going to follow Woodworth and Porter (2002) in their definition of this mental disease as a "Personality disorder characterized by a profound affective deficit accompanied by a lack of respect for the rights of others and societal rules".

Psychopathy involves a series of specific neurobiological, social and emotional features that make the psychopath different from the general population. From a biological perspective, psychopaths' brain presents several structural and functional abnormalities such as grey matter reduction in frontal and temporal areas, as well as anomalies in the prefrontal cortex.

In addition, the psychopath's behavior is characterized by specific features such as pathological lying, poor behavioral controls, failure to accept responsibility for own actions, grandiose estimation of the self, shallow affect and lack of remorse or guilt. These five features can be considered to be the most relevant, although not the only ones.

As stated above, mental diseases such as Psychopathy can be explored from different viewpoints, among which Linguistics is found. However, limited research can be found in this regard. Scholars such as Cleckley (1976), Williamson (1993) and Brinkley et al. (1999) have focused on psychopath's discourse cohesion and coherence. In recent years, Hancock et al. have gone beyond by examining more specific cues such as lexis and morphology.

One of the main behavioural features of psychopaths is pathological lying. The procedures for the detection of deception have been divided into those based on physiological methods and those based on behavioural methods. Among the former, we can highlight the polygraph, brain activity analysis, thermal analysis or voice stress analysis. As for the latter, we can find nonverbal or verbal assessment tools. The most interesting for us are the second ones, that is, those which are based on communication, namely, statement validity assessment, reality monitoring and linguistic analysis, which is the one that is followed here.

3. Deception in Language

The linguistic analysis aims to distinguish fabricated messages from truthful ones; it operates independently of the message meaning, and it uses sophisticated statistical text analysis tools. One of these software applications is LIWC which stands for Linguistic Inquiry and Word Count. LIWC counts and classifies words into psychologically meaningful categories. 2200 words and word stems are grouped into 72 broad categories which are relevant to psychological processes. Some of these categories are used in other deception detection methods such as Reality Monitoring. LIWC lexicon has shown correlation with human ratings of a large number of written texts, which suggests that it is a valid tool for analysis. Indeed, we find comprehensive accounts

of LIWC dimensions such as Pennebaker et al. (2001) and the Spanish equivalent Ramírez-Esparza et al. (2007). Significantly enough, LIWC was first tested by Newman et al. (2003) on a corpus of university students' deceptive and truthful written and spoken language purposely produced. Furthermore, in this process of data collection the participants are not biased towards the concealment of the lies, which, according to Bull et al. (2006), is highly frequent among *professional* liars. The cost of the lies being detected would not be high in this case, opposite to what happens in high-stakes situations.

One of the key issues in psycholinguistics is the reflection of the emotional and cognitive frames of humans on the oral and written language they produce. Early approaches to psycholinguistic concerns involved almost exclusively qualitative philosophical analyses. More recent research in this field provides empirical evidence on the relation between language and the state of mind of subjects, or even their mental health (Rosenberg & Tucker, 1978). In this regard, further studies by Pennebaker and his team have dealt with the therapeutic effect of verbally expressing emotional experiences and memories. LIWC was developed precisely for providing an efficient method for studying these psycholinguistic concerns, and has been considerably improved since its first version (Francis & Pennebaker, 1993).

Within the first dimension, namely standard linguistic processes, most categories involve function words and grammatical information; thus, the selection of words is straightforward, as in the case of articles, which are made up of three words in English –*a, an* and *the*– and of nine words in Spanish –*el, la, los, las, uno, un, una, unos* and *unas.*

On the other hand, the second and fourth dimensions are more subjective, especially those denoting emotional processes within the second dimension. These categories indeed required human judges to make the lexical selection. For all subjective categories, an initial list of word candidates was compiled from dictionaries and thesauri for all subjective categories.

Similar to the first dimension, the third dimension, relativity, comprises a category concerning time, which is quite clear-cut: past, present, and future tense verbs. Within the same dimension, this is

also the case of the category space, in which spatial prepositions and adverbs have been included.

Finally, the fourth dimension involves word categories related to personal concerns intrinsic to the human condition. As explained below, this dimension has often been excluded in deception detection studies, on the basis that it is too content-dependent (Hancook et al., 2011; Newman et al., 2003).

4. Description of the Case Study

Turning now to our case study, our subject is a 27-year-old Spanish single woman, who, at the moment the data were collected, was studying a Master's degree in Translation and Interpreting. One morning, she alleged she had been raped the previous night, but she commented she could not remember almost anything. She suspected that she was drugged with Scopolamine, which is a substance that destroys free will and may cause amnesia. The case becomes highly well-known, mainly due to the use of this drug in Spain as a new way to commit this type of crime. The police asked her to write everything she can remember as an attempt to help the presumed victim to offer more details. At this point we have to comment that it is this written description what we use as the material for the analysis.[1]

After the allegation was formalized, the subject starts having problems with her attorney. He observes a strange behavior in his client. The subject attacks him verbally. She even threatens with accusing him of sexual harassment. That is why she is evaluated by an expert and she is diagnosed with psychopathic behavior. It is confirmed that the subject suffers a dissocial personality disorder that is characterized, among other things, as psychopathic behavior. In fact, it was demonstrated that the subject had lied and she had not been raped.

---

1. Due to the non-disclosure agreement signed with the expert who provided the data, the original text cannot be included as an appendix.

5. Methodology

5.1. Research Aim

Having described the facts that lead to this study, we delve into the aim of the study. The present work goes beyond the analysis of a psychopath's speech: it focuses on the act of lying, by contrasting different veracious excerpts to others which are deceptive. In other words, it explores deception in a psychopath's speech through quantitative analysis by means of psycholinguistic categories in a descriptive fashion. This means that the results are not to be statistically projected because of the small size of the sample.

5.2. Data

As stated above, the material for the analysis was the statement that was written by the subject, and that was required by the Catalan Police (called Mossos d'Esquadra). This type of procedure is not usual in European Civil Law, but it was admitted as an expert proof. For the analysis, the text has been divided into two files: one comprising the deceptive excerpts in the statement, and another file with the rest of the text used as the control sample.

5.3. Results

Thus, the two text fragments: "true text" *versus* "deceptive text" have been the basis of our analysis, by means of contrasting various aspects and dimensions. The basic extension parameters of both texts are given below (Table 1):

Table 1 – Extension parameters of true text and deceptive text

	Tokens	Types	Standardised Type-token Ratio (STTR)
True text	13,275	2,514	39.95
Deceptive text	5,585	1,058	39.20

Although both texts diverge in extension, their STTR are very similar, and therefore the comparison becomes very consistent as the percentage of new types for every n tokens is virtually identical. Similarly, another contrastive parameter: mean word length (in characters) is also stable in both texts (Table 2 and Figure 1):

Table 2 – More extension parameters of true text and deceptive text

	Mean word length (in characters)	**Mean sentence length (in tokens)**	**Use of numbers (total figure)**
True text	4.18	21.86	32
Deceptive text	4.16	19.23	2
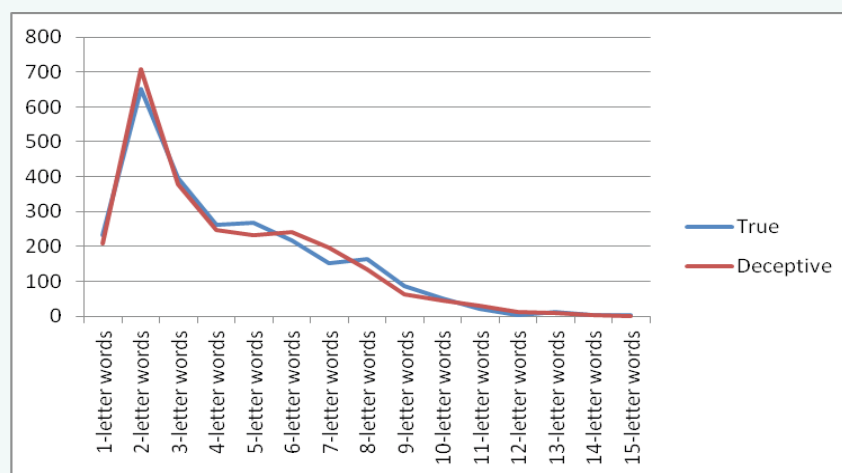
Figure 1 – Word selection regarding word length .

Other, more subtle parameters (mean sentence length and figure-use), reveal the first noticeable differences: true-text sentences are on average 2.5 words longer and a more prone to contain numbers: potential hints of using completely different sentence generation strategies when telling the truth or deceiving.

Focusing on the dimensions analysis (Table 3), we have discarded those with a relative frequency lower than 2 in the true text (Table 4).
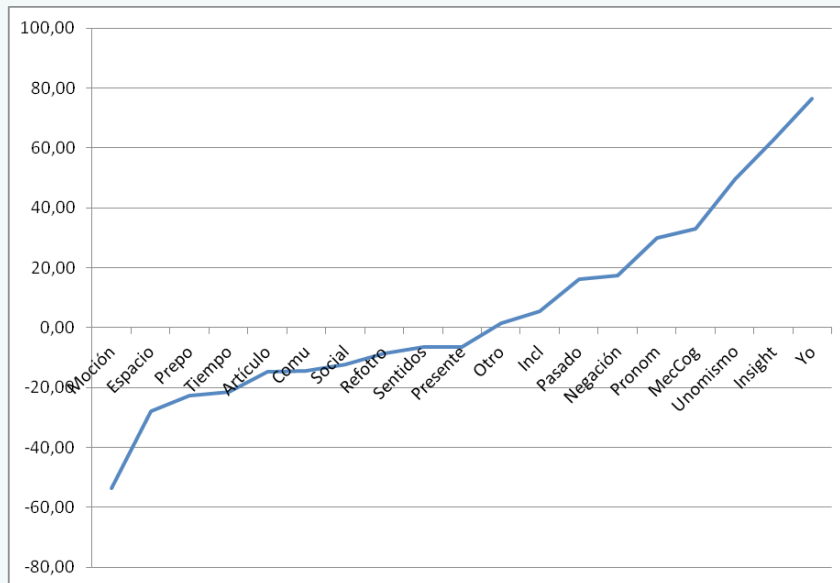
Table 3 – Dimensions of true text and deceptive text

	True	Deception			True	Deception
Pronom	10,80	14,02		Amigos	0,16	0,00
Yo	4,19	7,39		Familia	0,16	0,28
Nosotros	1,05	0,38		Humanos	0,32	0,19
Unomismo	5,20	7,77		Tiempo	3,75	2,94
Tú	0,60	0,57		Pasado	7,82	9,09
Otro	4,39	4,45		Presente	2,94	2,75
Negación	2,18	2,56		Futuro	0,08	0,00
Asentir	0,28	0,28		Espacio	2,50	1,80
Artículo	10,00	8,52		Arriba	0,12	0,28
Prepo	14,47	11,17		Abajo	0,04	0,09
Número	1,09	0,57		Incl	6,01	6,34
Afectiva	1,85	3,69		Excl	1,61	2,18
Emopos	1,01	2,08		Moción	3,06	1,42
Sentpos	0,28	0,66		Ocupa	1,45	0,95
Optimi	0,32	0,28		Escuela	0,44	0,19
Emoneg	0,60	1,33		Trabajo	0,24	0,19
Ansiedad	0,32	0,57		Logro	0,56	0,38
Enojo	0,08	0,09		Placer	1,09	0,38
Tristeza	0,20	0,28		Casa	1,29	1,04
MecCog	10,40	13,83		Deportes	0,00	0,00
Causa	1,29	1,61		TV	0,00	0,00
Insight	2,62	4,26		Música	0,00	0,09
Discrep	0,12	0,09		Dinero	0,44	0,19
Inhib	0,16	0,19		Metafo	0,00	0,09
Tentat	1,49	1,99		Relig	0,00	0,00
Certeza	0,56	1,04		Muerte	0,00	0,09
Sentidos	2,74	2,56		Físico	1,69	2,75
Ver	0,60	0,38		cuerpo	0,69	1,99
Oír	1,65	1,14		Sexual	0,00	0,19
Sentir	0,28	0,85		Comer	0,36	0,28
Social	9,39	8,24		Dormir	0,16	0,47
Comu	2,10	1,80		Asearse	0,24	0,00
Refotro	6,21	5,68		Maldecir	0,00	0,09

Table 4 and Figure 2 show the most prominent dimension differences between the true text and deceptive text; negative percentage difference evidence a more solid usage among true statements, whereas positive values more prominent habit in deception.

Table 4 – Dimensions of true text and deceptive text (>2,00)

	True	Deceptive	% Difference
Moción	3,06	1,42	-53,59
Espacio	2,50	1,80	-28,00
Prepo	14,47	11,17	-22,81
Tiempo	3,75	2,94	-21,60
Artículo	10,00	8,52	-14,80
Comu	2,10	1,80	-14,29
Social	9,39	8,24	-12,25
Refotro	6,21	5,68	-8,53
Sentidos	2,74	2,56	-6,57
Presente	2,94	2,75	-6,46
Otro	4,39	4,45	1,37
Incl	6,01	6,34	5,49
Pasado	7,82	9,09	16,24
Negación	2,18	2,56	17,43
Pronom	10,80	14,02	29,81
MecCog	10,40	13,83	32,98
Unomismo	5,20	7,77	49,42
Insight	2,62	4,26	62,60
Yo	4,19	7,39	76,37

Figure 2 – Dimensions of true text and deceptive text (>2,00).

6. Discussion and Final Remarks

It is worth noting that in the studies on statement veracity and validity, the psychological dimension is the most informative one. However, it has been observed that the data related to this dimension on some of the deceptive excerpts do not correspond to the prototypical parameters, probably because of the mental disorder of the subject. First, regarding cognitive processes, the linguistic elements categorized as 'insight' –e.g. think, consider– are most frequently found in truthful statements, and such is the case in the present analysis. However, there is a statistically significant difference between the values of certainty words –e.g. always, never–, which are considerably more frequent in untruthful excerpts, which happens to be the norm, since this kind of words are normally used as a strategy for concealing lies (Newman et al., 2003; Almela, 2012).

As far as affective processes are concerned, positive emotions have been associated to truthful statements in Spanish (Almela, 2012), which is similar to the results in the English language (Mihalcea & Strapparava, 2009; Newman et al., 2003). On the contrary, negative emotions have traditionally been associated to deception in English, although in Spanish some of them are positively correlated with truthfulness, namely anxiety and sadness. Interestingly enough, in the psychopath's speech both groups of emotions are more frequent when she tells lies, probably in an attempt to empathize with her readers. The overall percentage of affection words in untruthful excerpts doubles the category in the truthful ones.

Furthermore, words related to social processes are less frequent in untruthful excerpts as a global category. Since this category is typically associated to truthfulness (Newman et al., 2003; Zhou et al., 2004), it matches the normal parameter.

One of the defining cues to untruthfulness is the relative abundance of 2nd and 3rd person, whereas 1st person singular is used rather moderately, since the speaker prefers not to identify him/herself with the lies he/she is telling. However, this does not apply to subjects with a psychopathic personality; these are unable to feel any remorse or guilt, and have a marked egocentrism and a grandiose estimation

of self. Linguistically, this is reflected in a strong presence of the first person singular along the whole statement, with a considerable increase in untruthful excerpts (more than 3 points).

The values of average sentence length are also indicative. Unlike the results from studies with spoken language corpora, in the written medium untruthful statements are characterized by a shorter sentence length, probably because of the speaker's fear to fall into contradiction (Almela, 2012; Zhou et al., 2004). This means in practice that there is a tendency to simplify sentences and to reduce the use of subordination and coordination.

Last but not least, it is worth noting that there is a decrease in the amount of words related to number in deceptive excerpts. This fits into the usual pattern of deceptive communication, since the speaker prefers not to add specificity to their speech.

References

ALMELA, Á. 2012. *Featuring Deception in Written Language: A Contrastive Study of English and Spanish*. Doctoral Dissertation, Universidad de Murcia, Murcia, España.

ANDERSON, B.; GOLDIN, P.; KURITA, K. & GROSS, J. 2008. Self-Representation in Social Anxiety Disorder: Linguistic Analysis of Autobiographical Narratives. *Behaviour Research Theory* 46/10: 1119-1125.

ANDREASEN, N. & PFOHL, B. 1976. Linguistic Analysis of Speech in Affective Disorders. *Arch Gen Psychiatric* **33/11**: 1361-1367.

BRINKLEY, C. A.; NEWMAN, J.; HARPUR, T. & JOHNSON, M. 1999. Cohesion in Texts Produced by Psychopathic and Non-psychopathic Criminal Inmates. *Personality and Individual Differences* **26**: 873-885.

BULL, R.; COOKE, C.; BILBY, C. & GRANT, T. 2006. *Criminal Psychology: A Beginner's Guide.* Oxford: Oneworld Publications.

Cantos, P. 2014. Sketching a "Low-cost" Text-classification Technique for Text Topics in English. *Ibérica* **27**: 165-184.

Chaika, E. 1974. A Linguist Look at Schizophrenic Language. *Brain and Language* **1/3**: 257-276.

Cleckley, H. 1976. *The Mask of Sanity* (5th ed.). St. Louis, MO: Mosby.

Forgeard, M. 2008. Linguistic Styles of Eminent Writers Suffering From Unipolar and Bipolar Mood Disorder. *Creative Research Journal* **20/1**: 81-92.

Francis, M. E. & Pennebaker, J. 1993. *LIWC: Linguistic Inquiry and Word Count. Technical Report.* Dallas, TX: Southern Methodist University.

Hancock, J. T.; Woodworth, M. & Porter, S. 2011. Hungry Like the Wolf: A Word-pattern Analysis of the Language of Psychopaths. *Legal and Criminological Psychology* **18/1**: 1-13.

Hare, R. D. 2003. *The Hare Psychopathy Checklist. International Statistical Classification of Disseases and Related Health Problems* [en línea]. Available at: http://apps.who.int/classifications/icd10/browse/2010/en.

Jensen, M. L. 2007. *The Effects of an Expert System on Novice and Professional Decision Making with Application in Deception Detection.* Tesis Doctoral, University of Arizona, Tucson, AZ, EEUU.

Kanin, E. J. 1994. False Rape Allegations. *Archives of Sexual Behaviour* **23/1**: 81-93.

Lidenfeld, J. 1973. Affective States and the Syntactic Structure of Speech. *Semiótica* **4/8**: 368-376.

Lorenz, M. & Cobb, S. 1952. Language Behavior in Manic Patients. *Archives of Neurology and Psychiatry* **67**: 763-770.

Lott, P. R.; Guggenbühl, S.; Schneeberger, A.; Pulver, A. E. & Stassen, H. H. 2002. Linguistic Analysis of the Speech Output of Schizophrenic, Bipolar, and Depressive Patients. *Psychopathology* **35/4**: 220-227.

Masip, J.; Bethencourt, M.; Lucas, G.; Sánchez-San Segundo, M. & Herrero, C. 2012. Deception Detection from Written Accounts. *Scandinavian Journal of Psychology* **53**: 103-111.

Mihalcea, R. & Strapparava C. 2009. The Lie Detector: Explorations in the Automatic Recognition of Deceptive Language. Paper presented at the *47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural*

Language Processing of the Asian Federation of Natural Language Processing, Suntec, Singapur.

NEWMAN, M. L.; PENNEBAKER, J.; BERRY, D. & RICHARDS, J. 2003. Lying Words: Predicting Deception from Linguistic Styles. *Personality and Social Psychology Bulletin* **29**: 665-675.

OLIVERIA-SOUZA, R.; HARE, R.; BRAMATI, I.; GARRIDO, G.; IGNACIO, A.; TOVAR-MOLL F. & MOLL, J. 2008. Psychopathy as a Disorder of the Moral Brain: Fronto-temporo-limbic Grey Matter Reductions Demostrated by Voxel-based Morphometry. *NeuroImage* **40**: 1202-1213.

PARKER, A. D. & BROWN, J. 2000. Detection of Deception: Statement Validity Analysis as a Means of Determining Truthfulness or Falsity of Rape Allegations. *Legal and Criminological Psychology* **5**: 237-259.

PENNEBAKER, J. W.; FRANCIS, M. & BOOTH, R. 2001. *Linguistic Inquiry and Word Count.* Mahwah (NJ): Erlbaum Publishers.

RAINE, A.; ISHIKAWA, S.; ARCE, E.; LENCZ, T.; KNUTH, K.; BIHRLE, S. & COLLETII, P. 2004. Hippocampal Structural Asymmetry in Unsuccessful Psychopaths. *Biological Psychiatry* **60**: 1134-1142.

RAMÍREZ-ESPARZA, N.; PENNEBAKER J. & GARCÍA, F. 2007. La psicología del uso de las palabras: un programa de computadora que analiza textos en español. *Revista Mexicana de Psicología* **24**: 85-99.

ROSENBERG, S. D. & TUCKER, G. 1978. Verbal Behavior and Schizophrenia: The Semantic Dimension. *Archives of General Psychiatry* **36**: 1331-1337.

THOMAS, P.; KING, K. & FRASER, W. 1987. Positive and Negative Symptoms of Schizophrenia and Linguistic Performance. *Acta Psychiatrica Scandinava* **76/2:** 144-151.

TORRUBIA, R. & CUQUERELLA, Á. 2008. Psicopatía: una entidad clínica controvertida pero necesaria en psiquiatría forense. *Revista española de medicina legal: Órgano de la Asociación Nacional de Médicos Forenses* **34/1**: 25-35.

WILLIAMSON, S. E. 1993. Cohesion and Coherence in the Speech of Psychopathic Criminals. *Dissertation Abstracts International* **53**: 65-79.

WOODWORTH, M. & PORTER, S. 2002. In Cold Blood: Characteristics of Criminal Homicides as a Function of Psychopathy. *Journal of Abnormal Psychology* **111/3:** 436-445.

ZHOU, L.; BURGOON, J.; NUNAMAKER, J. & TWITCHELL, D. 2004. Automating Linguistics–based Cues for Detecting Deception in Text–based Asynchronous Computer–mediated Communications. *Group Decision and Negotiation* **13/2:** 81-106.