Seeing through deception: A computational approach to deceit detection in written communication

Ángela Almela
English Department
Universidad de Murcia
30071 Murcia (Spain)
angelalm@um.es

Rafael Valencia-García
Faculty of Computer Science
Universidad de Murcia
30071 Espinardo, Murcia (Spain)
valencia@um.es

Pascual Cantos
English Department
Universidad de Murcia
30071 Murcia (Spain)
pcantos@um.es

Abstract

The present paper addresses the question of the nature of deception language. Specifically, the main aim of this piece of research is the exploration of deceit in Spanish written communication. We have designed an automatic classifier based on Support Vector Machines (SVM) for the identification of deception in an *ad hoc* opinion corpus. In order to test the effectiveness of the LIWC2001 categories in Spanish, we have drawn a comparison with a Bag-of-Words (BoW) model. The results indicate that the classification of the texts is more successful by means of our initial set of variables than with the latter system. These findings are potentially applicable to areas such as forensic linguistics and opinion mining, where extensive research on languages other than English is needed.

1 Introduction

Deception has been studied from the perspective of several disciplines, namely psychology, linguistics, psychiatry, and philosophy (Granhag & Strömwall, 2004). The active role played by deception in the context of human communication stirs up researchers' interest. Indeed, DePaulo et al. (1996) report that people tell an average of one to two lies a day, either through spoken or written language. More recently, researchers in the field of opinion mining have become increasingly concerned with the detection of the truth condition of the opinions passed on the Internet (Ott et al., 2011). This issue is particularly challenging, since the researcher is provided with no information apart from the written language itself.

Within this framework, the present study attempts to explore deception cues in written language in Spanish, which is something of a novelty. The remainder of this paper is organized as follows: in Section 2, related work on the topic is summarized; in Section 3, we explain our methodology for analyzing data; in Section 4, the evaluation framework and experimental results are presented and discussed; Section 5 presents the results from a Bag-of-Words model as a basis for comparison; finally, in Section 6 some conclusions and directions for further research are advanced.

2 Related Work

There are verbal cues to deception which form part of existing verbal lie detection tools used by professional lie catchers and scholars (Vrij, 2010). Automated linguistic techniques have been used to examine the linguistic profiles of deceptive language in English. Most commonly, researchers have used the classes of words defined in the Linguistic Inquiry and Word Count or LIWC (Pennebaker et al., 2001), which is a text analysis program that counts words in psychologically meaningful categories. It includes about 2,200 words and word stems grouped into 72 categories relevant to psychological processes. It has been used to study issues like personality (Mairesse et al., 2007), psychological adjustment (Alpers et al., 2005), social judgments (Leshed et al., 2007), tutoring dynamics (Cade et al., 2010), and mental health (Rude et al., 2004). The validation of the lexicon contained in its dictionary has been performed by means of a comparison of human ratings of a large number of written texts to the rating obtained through their LIWC-based analyses.

LIWC was firstly used by Pennebaker's group for a number of studies on the language of deception, being the results published in Newman et al. (2003). For their purposes, they

collected a corpus with true and false statements through five different studies. In the first three tests, the participants expressed their true opinions on abortion, as well as the opposite of their point of view. The first study dealt with oral language, hence the videotaping of the opinions, whereas in the second and the third ones the participants were respectively asked to type and handwrite their views. In the fourth study, the subjects orally expressed true and false feelings about friends, and the fifth one involved a mock crime in which the participants had to deny any responsibility for a fictional theft. The texts were analyzed using the 29 variables of LIWC selected by the authors. Of the 72 categories considered by the program, they excluded the categories reflecting essay content, any linguistic variable used at low rates, and those unique to one form of communication (spoken vs. written language). The values for these 29 variables were standardized by converting the percentages to z scores so as to enable comparisons across studies with different subject matters and modes of communication. For predicting deception, a logistic regression was trained on four of the five subcorpora and tested on the fifth, which entails a fivefold cross-validation. The authors obtained a correct classification of liars and truth-tellers at a rate of 67% when the topic was constant and a rate of 61% overall. However, in two of the five studies, the performances were not better than chance. Finally, the variables that were significant predictors in at least two studies were used to evaluate simultaneously the five tests, namely self-reference terms, references to others, exclusive words, negative emotion elements and motion words. The reason for the poor performance in some of the studies may lie with the mixing of modes of communication, since, as stated by Picornell (2011), the verbal cues to deception in oral communication do not translate across into written deception and *vice versa*.

From this study, LIWC has been used in the forensic field mainly for the investigation of deception in spoken language. There are some early studies in this line which are concerned with the usefulness of this software application as compared to Reality Monitoring technique (RM). First, Bond and Lee (2005) applied LIWC to random samples from a corpus comprising lie and truth oral statements by sixty-four prisoners, only taking into consideration the variables selected by Newman et al. (2003) for the global evaluation. Overall, the results show that deceivers score significantly lower than truth-tellers as regards sensory details, but outstandingly higher for spatial aspects. The latter finding goes against previous research in RM theory; such is the case of Newman et al. (2003), where these categories did not produce significant results. Apart from this difference, both studies share common ground: despite considering RM theory, the authors did not perform manual RM coding on their data. Thus, they do not draw a direct comparison between the effectiveness of automatic RM coding through LIWC software and manual RM coding.

This gap in research was plugged by Vrij et al. (2007). Their hypothesis predicts that LIWC coding is less successful than manual RM coding in discriminating between deceivers and truth-tellers. In order to test this theory, they collected a corpus of oral interviews of 120 undergraduate students. Half the participants were given the role of deceivers, having to lie about a staged event, whereas the remainder had to tell the truth about the action. The analysis revealed that RM distinguished between truth-tellers and deceivers better than Criteria-Based Content Analysis. In addition, manual RM coding offered more verbal cues to deception than automatic coding of the RM criteria. There is a second experiment in this study assessing the effects of three police interview styles on the ability to detect deception, but the results will not be presented here because the subject lies outside the scope of this work.

More recently, Fornaciari & Poesio (2011) conducted a study on a corpus of transcriptions of oral court testimonies. This work presents two main novelties: first, the object of study is a sample of spontaneously produced language instead of statements uttered *ad hoc* or laboratory-controlled; moreover, it deals with a language other than English, namely Italian. The authors continue Newman et al.'s (2003) idea of a method for classifying texts according to their truth condition instead of simply studying the language in descriptive terms, their analysis unit being the utterance instead of the text. Their ultimate aim is a comparison between the efficiency of the content-related features of LIWC and surface-related features, including the frequency and use of function words or of certain n-grams of words or parts-of-speech. They used five kinds of vectors, taking the best features from their experiment, from Newman et al. (2003), and all LIWC categories. The latter

results in slightly better performance than the former, but they do not obtain a statistically significant difference.

LIWC has been also used for the investigation of deception in written language. Curiously enough, research in this line has been approached by computational linguists and not from the perspective of the forensic science. First, Mihalcea & Strapparava (2009) used LIWC for *post hoc* analysis, measuring several language dimensions on a corpus of 100 false and true opinions on three controversial topics – the design of the questionnaire is indeed similar to Newman et al.'s (2003). As a preliminary experiment, they used two ML classifiers: Naïve Bayes and Support Vector Machines, using word frequencies for the training of both algorithms, similar to a Bag-of-Words model. They achieved an average classification performance of 70%, which is significantly higher than the 50% baseline. On the basis of this information, they calculate a dominance score associated with a given word class inside the collection of deceptive texts as a measure of saliency. Then, they compute word coverage, which is the weight of the linguistic item in the corpora. Thus, they identify some distinctive characteristics of deceptive texts, but purely in descriptive terms.

In this strand of research, Ott et al. (2011) used the same two ML classifiers. For their training, apart from comparing lexically–based deception classifiers to a random guess baseline, the authors additionally evaluated and compared two other computational approaches: genre identification through the frequency distribution of part-of-speech (POS) tags, and a text categorization approach which allows them to model both content and context with n-gram features. Their ultimate aim is deceptive opinion spam, which is qualitatively different from deceptive language itself. Findings reveal that n-gram-based text categorization is the best detection approach; however, a combination of LIWC features and n-gram features perform marginally better.

These studies deal with written language as used in an asynchronous means of communication. In contrast, Hancock and his group explore deceptive language in synchronous computer-mediated communication (CMC), in which all participants are online at the same time (Bishop, 2009). Specifically, they use chat rooms. In their first study using LIWC, Hancock et al. (2004) explored differences between the sender's and the receiver's linguistic style across truthful and deceptive communication. For the analysis, they selected the variables deemed relevant to the hypotheses, namely word counts, pronouns, emotion words, sense terms, exclusive words, negations, and question frequency. Results showed that, overall, when participants told lies, they used more words, a larger amount of references to others, and more sense terms. Hancock et al. (2008) reported rather similar results from a comparable experiment. Apart from this, they introduced the element of motivation, and observed that motivated liars tended to avoid causal terms, while unmotivated liars increased their use of negations.

All these studies coincide in their exploration of a set of variables, but none of them take LIWC features as a whole for the automatic classification of both sublanguages on written statements. Furthermore, researchers usually take the language of deception as a whole, ignoring the particular features which may distinguish a speaker from the others, assuming that everybody lies similarly. Instead of comparing each individual sample of deceptive language to its corresponding control text, the whole set of statements labelled as "false" is contrasted with the set comprising "true" statements. This idiolectal comparison certainly permeates the practitioner lore within the forensic context, hence its interest for computational approaches to deception detection. It is worth noticing that the main disadvantage of a corpus of "authentic" language is precisely the difficulty to obtain a control sample of language in which the same speaker tells the truth for the sake of comparison.

3   Methodology

A framework based on a classifier using a Support Vector Machine (SVM) has been developed in order to detect deception in our opinion corpus. SVM have been applied successfully in many text classification tasks due to their main advantages: first, they are robust in high dimensional spaces; second, any feature is relevant; third, they are robust when there is a sparse set of samples; finally, most text categorization problems are linearly separable (Saleh et al., 2011).

We have used LIWC to obtain the values for the categories for the subsequent training of the abovementioned classifier. This software application provides an efficient method for

studying the emotional, cognitive, and structural components contained in language on a word by word basis (Pennebaker et al., 2001). The LIWC internal dictionary comprises 2,300 words and word stems classified in four broad dimensions: standard linguistic processes, psychological processes, relativity, and personal concerns. Each word or word stem defines one or more of the 72 default word categories. The selection of words attached to language categories in LIWC has been made after hundreds of studies on psychological behaviour (Tausczik & Pennebaker, 2010). Within the first dimension, linguistic processes, most categories involve function words and grammatical information. Thus, the selection of words is straightforward; such is the case of the category of articles, which is made up of nine words in Spanish: *el, la, los, las, un, uno, una, unos,* and *unas.* Similarly, the third dimension, relativity, comprises a category concerning time which is clear-cut: past, present and future tense verbs. Within the same dimension, that is also the case of the category space, in which spatial prepositions and adverbs have been included. On the other hand, the remaining two dimensions are more subjective, especially those denoting emotional processes within the second dimension. These categories indeed demanded human judges to make the lexical selection. For all subjective categories, an initial list of word candidates was compiled from dictionaries and thesauruses, being subsequently rated by groups of three judges working independently. Finally, the fourth dimension involves word categories related to personal concerns intrinsic to the human condition. As mentioned above, this dimension has been often excluded in deception detection studies, on the basis that it is too content-dependent (Hancock et al., 2004, 2008; Newman et al., 2003).

Table 1 provides an illustrative summary of the list of the dictionary categories −a comprehensive account is included in Pennebaker et al. (2001:17-21), and the equivalences in Spanish can be found in Ramírez-Esparza et al. (2007:37-39).

We implemented our experiments using the Weka library (Bouckaert et al., 2010). We applied a linear SVM with the default configuration set by the tool. In order to train the classifier, the corpus is divided into true and false samples. For their analysis, we have considered the attributes of each dimension of LIWC previously described.

I. Standard linguistic dimension	II. Psycholog. processes	III. Relativity	IV. Personal concerns
Total pronouns	Causation	Space	Job or work
% words captured by the dictionary	Affective or emotional processes	Inclusive	Physical states and functions
% words longer than six letters	Negative emotions	Exclusive	Religion
Word Count	Cognitive processes	Time	Money and financial issues
First-person singular	Positive emotions	Motion verbs	Leisure activity

Table 1: Summary of the variables used in LIWC2001

Several classifiers have been obtained by using the categories of each dimension. For each classifier a tenfold cross-validation has been done and all sets have an equal distribution between true and false statements.

4 Evaluation framework and results

To study the distinction between true and deceptive statements, a corpus with explicit labelling of the truth condition associated with each statement was required. For this purpose, the design of the questionnaire for the compilation of the corpus was similar to that used by Mihalcea and Strapparava (2009). Data were produced by 100 participants, all of them native speakers of Peninsular or European Spanish. We focused on three different topics: opinions on homosexual adoption, opinions on bullfighting, and feelings about one's best friend. A similar corpus was used in (Almela, 2011), where a pilot study on the discriminatory power of lexical choice was conducted. The corpus used included a further data set, comprising opinions on a good teacher. However, it was disregarded in the present paper, since the statements were shorter and false and true opinions were not so effectively differentiated.

As mentioned above, since it was not spontaneously produced language, it was deemed necessary to minimize the effect of the observer's paradox (Labov, 1972) by not explaining the ultimate aim of the research to the participants. Furthermore, they were told that they had to make sure that they were able to convince their partners on the topics that they were lying about, so as to have them highly motivated, like in Hancock et al. (2008).

For the first two topics (homosexual adoption and bullfighting), we provided instructions that asked the contributors to imagine that they were taking part in a debate, and had 10-15 minutes available to express their opinion about the topic. First, they were asked to prepare a brief speech expressing their true opinion on the topic. Next, they were asked to prepare a second brief speech expressing the opposite of their opinion, thus lying about their true beliefs about the topic. In both cases, the guidelines asked for at least 5 sentences and as many details as possible. For the other topic, the contributors were asked to think about their best friend, including facts and anecdotes considered relevant for their relationship. Thus, in this case, they were asked to tell the truth about how they felt. Next, they were asked to think about a person they could not stand, and describe it as if s/he were their best friend. In this second case, they had to lie about their feelings towards these people. As before, in both cases the instructions asked for at least 5 detailed sentences.

We collected 100 true and 100 false statements for each topic, with an average of 80 words per statement. We made a manual verification of the quality of the contributions. With three exceptions, all the other entries were found to be of good quality. Each sample was entered into a separate text file, and misspellings were corrected. Each of the 600 text files was analyzed using LIWC to create the samples for the classifier. It is worth noting that the version used was LIWC2001, since this is the one which has been fully validated for Spanish across several psycholinguistics studies (Ramírez-Esparza et al., 2007). The whole LIWC output was taken for the experiment, except for two categories classified as experimental dimensions (Pennebaker et al., 2001): nonfluencies (e.g. *er, hm, umm*) and fillers (e.g. *blah, Imean, youknow*), since they are exclusive to spoken language. The remaining experimental dimension, swear words, has been included for

our purposes in the first dimension, linguistic processes, since this is the case for the subsequent version of this software application.

The results from the ML experiment are shown in Table 2. In the first column, the number of LIWC dimensions used for each classifier is indicated. For example, 1_2_3_4 indicates that all the dimensions have been used in the experiment, and 1_2 indicates that only the categories of dimensions 1 and 2 have been used to train the classifier. The scores shown in the table stand for the F-measure, the weighted harmonic mean of precision and recall.

	Homos. adoption	Bullfight.	Best friend	Total
1	0.638	0.679	0.763	0.683
1_2	0.709	0.655	0.83	0.736
1_2_3	0.698	0.669	0.835	0.726
1_2_3_4	0.718	0.66	0.845	0.734
1_2_4	0.728	0.63	0.83	0.728
1_3	0.64	0.68	0.82	0.701
1_3_4	0.657	0.643	0.815	0.698
1_4	0.631	0.651	0.738	0.661
2	0.678	0.624	0.78	0.702
2_3	0.724	0.619	0.81	0.723
2_3_4	0.724	0.609	0.81	0.716
2_4	0.703	0.59	0.78	0.706
3	0.62	0.62	0.695	0.616
3_4	0.611	0.595	0.684	0.654
4	0.506	0.525	0.639	0.561

Table 2: Results from the experiment

Findings reveal that the dimension which performs overall best irrespective of topic is the second one, psychological processes (70.2%). This is in line with Newman et al.'s (2003) study, where belief-oriented vocabulary, such as *think*, is more frequently encountered in truthful statements, since the presence of real facts does not require truth-related words for emphasis. As regards dominant words in deceptive texts, previous research highlights words related to certainty, probably due to the speaker's need to explicitly use truth-related words as a means to conceal the lies (Bond & Lee, 2005; Mihalcea & Strapparava, 2009). Furthermore, according to Burgoon et al. (2003), other feature associated with deception is the high frequency of words

denoting negative emotions. All these categories are included in the second dimension, and their discriminant potential in deception detection is indeed confirmed in our classification experiment.

The first dimension shows a relatively high performance (68.3%). It is natural that it should be so, bearing in mind the considerable potential of function words, which constitutes a substantial part of standard linguistic dimensions. The prime importance of these grammatical elements has been widely explored, not only in computational linguistics, but also in psychology. As Chung and Pennebaker (2007:344) have it, these words "can provide powerful insight into the human psyche". Variations in their usage has been associated to sex, age, mental disorders such as depression, status, and deception.

On the contrary, and as could be expected from previous research (Newman et al., 2003; Fornaciari & Poesio, 2011), the fourth dimension is the least discriminant on its own. The reason may lie with the weak link of the topics involved in the questionnaire with the content of the personal concerns categories. However, there is not much difference with the third one, relativity –just 0.055 points in the total score.

As shown in Table 2, when the classifier is trained with certain combinations of dimensions, its performance improves noticeably. This finding is supported by Vrij's words: "a verbal cue uniquely related to deception, akin to Pinocchio's growing nose, does not exist. However, some verbal cues can be viewed as weak diagnostic indicators of deceit" (2010:103). In this way, it seems clear that a combination of lexical features is more effective than isolated categories. The grouping of the first two dimensions is remarkably successful (73.6%). Nevertheless, the addition of the other two dimensions to this blend is counterproductive, since it makes the score worse instead of improving it, probably due to their production of noise. No doubt that the factor loadings of the four dimensions play a considerable part in here. Overall, considering the total column, it seems as if the fourth LIWC dimension is the one cutting off the discrimination power.

Furthermore, it is worth noting that the results from the classification with these dimensions are strongly dependent on the topics of each subcorpus. The topics dealt with in our experiment show that the interaction of LIWC dimensions 1_2_4 (72.8%) and 2_3 (72.4%)

discriminates better true-false statements related to homosexuality adoption; similarly, the dimension selection of LIWC's 1_2_3 (83.5%) and 1_2_3_4 (84.5%) perform very positively regarding the topics related to the best friend. On the opposite scale, we get that true-false statements on bullfighting (1_3: 68%) are more difficult to tell apart by means of LIWC dimensions. A plausible explanation emerges here: when speakers refer to their best friend, they are likelier to be emotionally involved in the experiment; they are not just telling an opinion on a topic which is alien to them, but relating their personal experience with a dear friend and lying about a person they really dislike. This personal involvement is probably reflected on the linguistic expression of deception.

5 Comparison with a Bag-of-Words model

In this section we will present the results from a Bag-of-Words (BoW) representation to provide a basis for comparison with our methodology. In this model, a text is represented as an unordered collection of words, disregarding any linguistic factor such as grammar, semantics or syntax (Lewis, 1998). It has been successfully applied to a wide variety of NLP tasks such as document classification (Joachims, 1998), spam filtering (Provost, 1999), and opinion mining (Dave et al., 2003). However, its basis is not too sophisticated, hence the average scores obtained through this method in terms of precision and recall. Table 3 shows the F-measure scores obtained with this model.

Homosexual adoption	Bullfighting	Best friend	Total
0.654	0.622	0.715	0.648

Table 3: Results from the BoW model

Curiously enough, despite the simplicity of the method, in the first two topics the F-measure scores are better than the ones obtained from 6 LIWC dimension combinations (see Table 2). When it comes to the third topic, the number is reduced to three combinations. It is worth noting that, although the scores in this topic are good with this simple model (71.5%), a difference of 13 points is observed in the application of our methodology to this subcorpus.

By means of the comparison, it is confirmed that the third and the fourth dimensions, both on

their own and combined, perform worse than the BoW model, irrespective of the topic involved. However, as regards the total results, the only two scores which are worse than BoW's are derived from the application of these two dimensions on their own. Specifically, there is a difference of 8.8 points between the best total result from our experiment (73.6%), obtained by means of the combination of the two first dimensions, and the total result from BoW (64.8%). This means that, in general terms, the classification by means of our variables is more successful than with the BoW model.

6 Conclusions and further research

In the present paper we have showed the high performance of an automatic classifier for deception detection in Spanish written texts, using LIWC psycholinguistic categories for its training. Through an experiment conducted on three data sets, we have checked the discriminatory power of the variables as to their truth condition, being the two first dimensions, linguistic and psychological processes, the most relevant ones.

For future research in this line, we will undertake a contrastive study of the present results and the application of the same methodology to an English corpus, in order to identify possible structural and lexical differences between the linguistic expression of deceit in both languages.

Acknowledgements

References

Ángela Almela. 2011. Can lexical choice betray a liar? Paper presented at the *I Symposium on the Sociology of Words*, University of Murcia, Spain.

Georg W. Alpers, Andrew Winzelberg, Catherine Classen, Heidi Roberts, Parvati Dev, Cheryl Koopman, and Barr Taylor. 2005. Evaluation of computerized text analysis in an Internet breast cancer support group. *Computers in Human Behavior*, 21, 361-376.

Jonathan Bishop. 2009. Enhancing the understanding of genres of web-based communities: The role of the ecological cognition framework. *International Journal of Web-Based Communities*, 5(1), 4-17.

Gary D. Bond and Adrienne Y. Lee. 2005. Language of lies in prison: Linguistic classification of prisoners' truthful and deceptive natural language. *Applied Cognitive Psychology*, 19, 313-329.

Remco R. Bouckaert, Eibe Frank, Mark A. Hall, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2010. WEKA-experiences with a java open-source project. *Journal of Machine Learning Research*, 11:2533-2541.

Judee K. Burgoon, J. P. Blair, Tiantian Qin, and Jay F. Nunamaker. 2003. Detecting deception through linguistic analysis. *Intelligence and Security Informatics*, 2665, 91–101.

Whitney L. Cade, Blair A. Lehman, and Andrew Olney. 2010. An exploration of off topic conversation. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 669-672. Association for Computational Linguistics.

Cindy Chung and James W. Pennebaker. 2007. The psychological functions of function words. In K. Fiedler (Ed.), *Social Communication*, 343–359. New York: Psychology Press.

Malcolm Coulthard. 2004. Author identification, idiolect, and linguistic uniqueness. *Applied Linguistics*, 25(4):431-447.

Kushal Dave, Steve Lawrence, and David M. Pennock. 2003. Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international conference on World Wide Web (WWW '03)*. ACM, New York, NY, USA, 519-528.

Bella M. DePaulo, Deborah A. Kashy, Susan E. Kirkendol, Melissa M. Wyer, and Jennifer A. Epstein. 1996. Lying in everyday life. *Journal of Personality and Social Psychology*, 70: 979-995.

Tommaso Fornaciari and Massimo Poesio. 2011. Lexical vs. Surface Features in Deceptive Language Analysis. In Wyner, A. and Branting, K. *Proceedings of the ICAIL 2011 Workshop Applying Human Language Technology to the Law*.

Pär A. Granhag and Leif A. Strömwall. 2004. *The detection of deception in forensic contexts*. Cambridge, UK: Cambridge University Press.

Jeffrey T. Hancock, Lauren E. Curry, Saurabh Goorha, and Michael T. Woodworth. 2004. Lies in conversation: an examination of deception using automated linguistic analysis. *Annual Conference*

of the Cognitive Science Society. Taylor and Francis Group, Psychology Press, Mahwah, NJ.

Jeffrey T. Hancock, Lauren E. Curry, Saurabh Goorha, S. & Michael T. Woodworth. 2008. On lying and being lied to: A linguistic analysis of deception in computer-mediated communication. *Discourse Processes*, 45, 1-23.

Thorsten Joachims. 1998. Text categorization with support vector machines: learning with many relevant features. *ECML-98*, 137-142.

William Labov. 1972. *Sociolinguistic Patterns*. Oxford, UK: Blackwell.

Gilly Leshed, Jeffrey T. Hancock, Dan Cosley, Poppy L. McLeod, and Geri Gay. 2007. Feedback for guiding reflection on teamwork practices. In *Proceedings of the GROUP'07 conference on supporting group work*, 217-220. New York: Association for Computing Machinery Press.

David D. Lewis. 1998. Naive (Bayes) at Forty: The Independence Assumption in Information Retrieval. In *Proceedings of ECML-98, 10th European Conference on Machine Learning*, Springer Verlag, Heidelberg, Germany.

François Mairesse, Marilyn A. Walker, Matthias Mehl, and Roger K. Moore. 2007. Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research*, 30(1), 457-500.

Rada Mihalcea and Carlo Strapparava. 2009. The Lie Detector: Explorations in the Automatic Recognition of Deceptive Language. In *Proceedings of the Association for Computational Linguistics (ACL-IJCNLP 2009)*, Singapore, 309-312.

Matthew L. Newman, James W. Pennebaker, Diane S. Berry, and Jane M. Richards. 2003. Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, 29: 665-675.

Myle Ott, Yejin Choi, Claire Cardie, and Jeffrey T. Hancock. 2011. Finding deceptive opinion spam by any stretch of the imagination. In *Proceedings of ACL*, 309-319.

James W. Pennebaker, Martha E. Francis, and Roger J. Booth. 2001. *Linguistic Inquiry and Word Count.* Erlbaum Publishers, Mahwah, NJ.

James W. Pennebaker, Cindy K. Chung, Molly Ireland, Amy L. Gonzales, and Roger J. Booth, R. J. 2007. *The development and psychometric properties of LIWC2007*. LIWC.net, Austin, TX.

Isabel Picornell. 2011. The Rake's Progress: Mapping deception in written witness statements. Paper presented at the *International Association of Forensic Linguists Tenth Biennial Conference*, Aston University, Birmingham, United Kingdom.

Jefferson Provost. 1999. Naive-bayes vs. rule-learning in classifcation of email. *Technical Report AI-TR-99-284*, University of Texas at Austin, Artificial Intelligence Lab.

Nairán Ramírez-Esparza, James W. Pennebaker, and Florencia A. García. 2007. La psicología del uso de las palabras: Un programa de computadora que analiza textos en español [The psychology of word use: A computer program that analyzes texts in Spanish]. *Revista Mexicana de Psicología*, 24, 85-99.

Stephanie S. Rude, Eva-Maria Gortner, and James W. Pennebaker. 2004. Language use of depressed and depression-vulnerable college students. *Cognition and Emotion*, 18, 1121-1133.

Mohammed Rushdi-Saleh, Maria Teresa Martín-Valdivia, Arturo Montejo Ráez, and Luis Alfonso Ureña López. 2011. Experiments with SVM to classify opinions in different domains. *Expert Systems with Applications,* 38(12):14799-14804.

Yla R. Tausczik and James W. Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29, 24-54.

Aldert Vrij. 2010. *Detecting lies and deceit: Pitfalls and opportunities.* 2nd edition. John Wiley and Sons, Chischester, UK.

Aldert Vrij, Samantha Mann, Susanne Kristen, and Ronald P. Fisher. 2007. Cues to deception and ability to detect lies as a function of police interview styles. *Law and human behavior*, 31(5), 499-518.