Studying Dishonest Intentions in Brazilian Portuguese Texts

Francielle Alves Vargas and Thiago Alexandre Salgueiro Pardo

Interinstitutional Center for Computational Linguistics (NILC)
Institute of Mathematical and Computer Sciences, University of São Paulo
São Carlos, Brazil
francielleavargas@usp.br, taspardo@icmc.usp.br

Abstract. Previous work in the social sciences, psychology, and linguistics has shown that liars have some control over the content of their stories. Nevertheless, their underlying state of mind may "leak out" through the way that they tell them. To the best of our knowledge, no previous systematic effort exists to describe and model deception language for Brazilian Portuguese. To fill this important gap, we carry out a pioneering corpus study on false statements in Brazilian Portuguese texts. We methodically analyze linguistic features using a large deceptive corpus, which includes both fake and truthful news. The results show that Brazilian Portuguese deceptive and truthful present substantial lexical, syntactic, and semantic variations, as well as punctuation and emotion distinctions.

Keywords: deception detection, linguistic features, natural language processing

1 Introduction

According to the standard philosophical definition, lying is saying something that you believe to be false with the intent to deceive [12]. For deception detection, the FBI trains its agents in a technique named statement analysis, which attempts to detect deception based on parts of speech (i.e., linguistics style) rather than the facts of the case or the story as a whole [1]. This method is used in interrogations, where the suspects were interpolated to make a written statement. In [26], the authors report an example proposed by [1] of a man accused of killing his wife. In this statement, the accused consistently refers to "my wife and I" rather than "we", suggesting distance between the couple. Thus, for [26], linguistic style checking may be useful in the hands of a trained expert who knows what to look for and how to use language to reveal inconsistencies.

In this context, the deception spread through fake news and reviews is a relevant current problem. Due to their appealing nature, they spread rapidly [37]. Nevertheless, what makes fake content a complex problem to solve is the difficulty in identifying unreliable content. Fake news detection is defined as the prediction of the chances of a particular news article being intentionally deceptive

[33], and fake reviews or opinion spam are inappropriate or fraudulent reviews [28].

Psychologists and other social scientists are working arduously to understand what drives people to believe in fake news. Unfortunately, there is not yet a consensus on this issue. As claimed by [31], much of the debate among researchers falls into two opposing camps. One group claims that our ability to reason is hijacked by our partisan convictions. The other group claims that the problem is that we often fail to exercise our critical faculties: that is, we are mentally lazy.

The rationalization camp, which has gained considerable prominence in recent years, is built around a set of theories contending that, when it comes to politically charged issues, people use their intellectual abilities to persuade themselves to believe in what they want to be true, rather than attempting to discover the truth. In the context of social media, various pieces of evidence suggest that the main factor explaining the acceptance of fake news could be cognitive laziness, mainly where news items are often skimmed or merely glanced at.

[14] calls attention to a lack of non-laboratory studies related to deception. In [20], the authors comment that their study, examining the deceptive and truthful statements of a convicted murderer, was, at the time, the only known study of its type in a "high-stakes realistic setting". Moreover, as believed by [22], we do not know much about the embedded lies in texts or discourses. With the notable exception of a paper published by [16] and several studies proposed by [23] dealing with fictional discourse analysis in the American television show, there is a lack of empirical research.

Therefore, in this paper, we introduce a pioneering corpus study on false statements in Brazilian Portuguese texts. We methodically analyze linguistic features using a deceptive corpus called Fake.Br, which includes both fake and truthful news. Our main goal is to investigate predictive deception clues from texts. In particular, in this paper, we aim to provide linguistically motivated resources and computationally useful strategies for the development of automatic deception detection classifiers for the Portuguese language.

The remainder of this paper is organized as follows. In Section 2, we present the main related work. Section 3 describes an overview of our data. In Section 4, we show the entire empirical linguistic-based study. In Section 5, final remarks and future works are presented.

2 Related Work

[9] define deception as a deliberate attempt to mislead others. There are relatively few studies that have focused, specifically, on deceptive language recognition with speech or writing style, especially for Portuguese. Most of the available works have been used to aid in authorship attribution and plagiarism identification [7]. Recent studies have been valuable for detecting deception, especially in the Fake News classification.

[26] examined lying in written communication, finding that those deceptive utterances used more total words but fewer personal pronouns. The linguistic-based features have been employed for fake news detection. In [26], the authors listed a set of linguistic behaviors that predict deception, such as tones of words, kinds of prepositions, conjunctions, and pronouns. In addition, the deception linguistic style includes weak employment of singular and third-person pronouns, negative polarity, and frequent use of movement verbs. [9] also presents a long study on clues to deception. For [25], the basic assumption is that liars differ from truth-tellers in their verbal behavior, making it possible to classify the news by inspecting their verbal accounts. Accordingly, they present insights, decisions, and conclusions resulting from the deception research conference at the legal and criminologist psychology society. In [5], the authors proposed a set of features using several linguistic analysis levels. They employed lexical, syntax, semantic, and discourse linguistic features. At the lexical level, the authors explored the ba- of-words (BWO) approach using bi-grams. At the syntax level, a probability context-free grammar was implemented. For semantic analysis, the context information (such as profile content) has been incorporated. To model discourse features, the authors used the Rhetorical Structure Theory (RST) [21] analytical framework.

3 Corpus Overview

Providing a linguistic analysis of false statements in texts, the first challenge concentrates on data availability. The identification of corpora for each language is a relevant task. Most of the research has developed computational linguistic resources for English. In general, few resources are available for Portuguese. As we commented before, for the Brazilian Portuguese language, there is Fake.Br corpus [24], which includes fake and true news in Brazilian Portuguese. An overview of this corpus is shown in Tables 1, 2 and 3.

Table 1. Corpus Overview: Fake.Br [24].

Subjects	**Number of Texts**	**%**
Politics	4,180	58.0
TV & celebrities	1,544	21.4
Society & daily news	1,276	17.7
Science & technology	112	1.5
Economy	44	0.7
Religion	44	0.7

The Fake.Br corpus was composed in a semi-automatic way. The fake news was collected from sites that gather such content and the true ones were extracted from major news agencies in Brazil, such as G1, Folha de São Paulo, and Estadão portals. A crawler searched in the corresponding web pages of these agencies for

Table 2. Number of tokens.

News	Tokens	%
Fake	796.364	50.80
True	771.510	49.20

Table 3. Number of news texts.

News	Number of Texts	%
Fake	3,600	50.0
True	3,600	50.0

keywords of the fake news, which were nouns and verbs that occurred in the fake news titles and the most frequent words in the texts (ignoring stopwords). The authors have performed a final manual verification to guarantee that the fake and true news was subject-related.

4 Linguistic Features

Most of the false statements present linguistic features that are different than true statements. According to [6], most liars use their language strategically to avoid being caught. Despite the attempt to control what they are saying, language "leakage" occurs with certain verbal aspects that are hard to monitor, such as frequencies and patterns of pronouns, conjunctions, and negative emotion word usage [13].

In this section, we aim at understanding the relevant linguistic properties of fake and true statements in Brazilian news. We used Python 3.6.9 and the spaCy [1] library to automatically annotate [2] the corpus. We divided our analysis into two main groups: word-level and sentence-level analyses. In the first group, we analyzed the occurrence patterns of (i) sentiment and emotion words, (ii) part-of-speech tags, (iii) pronoun classification, (iv) named-entity recognition, and (v) punctuation behavior. In the second group, we evaluated the number of sentences in fake and true news and the average of words for each sentence. We also analyzed the occurrence of causal relations in syntactical dependency trees on fake and true statements. We present the results in what follows.

4.1 Word-Level Analysis

In the word-level analysis, our goal is to identify differences among word usage behavior and variations in fake and true news.

Sentiment and Emotion Words. According to [41], deception language involves negative emotions, which are expressed in language in terms of psychological distance from the deception object. The psychological distance and emotional experience reflect an attempt to control the negative mental representation. Therefore, we have identified the incidence of sentiment and emotion words in fake and true statements. We used the sentiment lexicon for Portuguese Sentilex-PT [34] and WordNetAffect.BR [30] to account for the sentiment words. Table

[1] `https://spacy.io/`
[2] `https://spacy.io/api/annotation`

4 shows the results. Note that the incidence of sentiment and emotion words in fake news overcame the ones in true news, except for surprise emotion.

Table 4. Word-level sentiment and emotion occurrence.

Sentences	True News	Fake News
Positive	103,376	115,260
Negative	102,54	115,431
Joy	4,941	5,657
Sadness	2,596	3,347
Fear	1,757	1,895
Disgust	1,561	1,667
Angry	2,865	3,232
Surprise	423	419
Total	642,636	665,489

In the fake news, we have observed a difference of 11,49 % and 12,57 % in positive and negative sentiment when compared to the true news; for joy, sadness, fear, disgust, angry and surprise emotions, the difference amounts to 14,49 %, 28,92 %, 7,85 %, 6,79 % 12,80 % and 0,95 % when compared to the true news. Accordingly, we evidence that in our corpus the fake statements presented more negative and positive sentiments and emotions than true statements, confirming what some relevant literature [10] [5] [26] [33] have found, i.e., that dishonest texts have more negative than positive sentiments and emotions.

Part-of-Speech. The growing body of research suggests that we may learn a great deal about people's underlying thoughts, emotions, and reasons by counting and categorizing the words they use to communicate. For [26], several aspects of linguistic styles, such as pronoun usage, preposition, and conjunctions that signal cognitive work, have been linked to many behavioral and emotional outcomes. To exemplify, in [39], the authors identified that poets who use a high frequency of self-reference but a lower frequency of other-reference in their poetry were more likely to commit suicide than those who showed the opposite pattern.

In this present study, we extracted the frequency of part-of-speech in our corpus to examine the grammatical manifestations of false behavior in text. The obtained results for part-of-speech occurrence are shown in Table 5. The results show an impressive increase in the number of interjections in fake news compared to true news. We must also point out that, for many authors, it is clear that interjections do not encode concepts as nouns, verbs or adjectives do. Interjections may and do refer to something related to the speaker or the external world, but their referential process is not the same as that of lexical items belonging to the grammatical categories mentioned, as the referents of interjections are difficult to pin down [29]. Similarly, the use of space characters has shown a relevant occurrence difference. We found 25,864 spaces in fake news and 3,977 spaces in true news. Furthermore, in true statements, the use of the NOUN category is 9,79 % larger than in fake statements. The verbal use is also 13,81 % more frequent in the true statements.

Table 5. Part-of-speech occurrence.

Label	Definition	True News	Fake News
NOUN	noun	140,107	127,609
VERB	verb	86,256	98,168
PROPN	proper noun	109,501	98,757
ADP	adposition	109,613	92,166
ADJ	adjective	33,433	32,535
DET	determiner	77,660	83,169
ADV	adverb	25,384	31,534
SPACE	**space**	**3,977**	**25,864**
PRON	pronoun	20,994	24,348
AUX	auxiliary	13,529	16,999
CCONJ	coordinating conjunction	17,263	16,352
NUM	numeral	16,951	12,596
SCONJ	subordinating conjunction	8,870	12,392
SYM	symbol	10,065	9,458
OTHER	other	2,684	3,113
INTJ	**interjection**	**66**	**220**
PART	particle	29	23

Pronouns. A wide range of studies related to deception shows that the use of the first-person singular is a subtle proclamation of one's ownership of a statement. In other words, liars tend to distance themselves from their stories and avoid taking responsibility for their behavior [15]. Therefore, deceptive communication should be characterized by fewer first-person singular pronouns (e.g., I, me, and my) [26]. In addition, when people are self-aware, they are more "honest" with themselves [4] [11] [36] and self-reference increases [8].

In accordance with deception literature, we investigate the pronoun behavior in our corpus. We identify the occurrence for first, second, and third persons of singular and plural pronouns. Table 6 exhibits the results. Surprisingly, the pronoun occurrence in fake news overcame the ones in true news, except in the 3rd person singular (tonic oblique). An unusual behavior, considering the literature on deception, may be noted on the 1st person singular (subject). In fake statements, there has been a jump in the occurrence of the "eu" pronoun (1,097) related to true statements (495). 3rd person singular (subject) and 3rd person singular (unstressed oblique) represent 34,16 % and 42,80 % respectively on the total occurrence of pronouns in the corpus for the fake news. Differently, for true news, the 3rd person singular (subject) and 3rd person singular (unstressed oblique) represent 36,88% and 47,74 % respectively. In other words, the 3rd person occurrence in true news overcame fake news considering the total occurrence of pronouns in the corpus.

Table 6. Pronoun occurrence.

Pronoun Classification	Example	True News	Fake News
1st person singular (subject)	**eu**	**495**	**1,097**
1st person singular (unstressed oblique)	me	233	447
1st person singular (tonic oblique)	mim	39	87
2nd person singular (subject)	**você, tu**	**390**	**683**
2nd person singular (unstressed oblique)	te	4	24
2nd person singular (tonic oblique)	ti, contigo	2	2
3rd person singular (subject)	ele, ela	3,344	4,006
3rd person singular (Unstressed oblique)	se, o, a, lhe	4,329	5,019
3rd person singular (tonic oblique)	si, consigo	44	41
1st person plural (subject)	nós	52	71
2nd person plural (subject)	vocês	7	26
3rd person plural (subject)	**eles,elas**	**128**	**222**

We must also point out that we found that the occurrence differences of 2nd person plural (unstressed oblique) and (tonic oblique) pronouns in fake and true news are statistically irrelevant.

Named-Entity Recognition. According to [22], most scholars in the field of deception research seem to accept standard truth-conditional. In addition, semantic assumptions on deception are rarely made explicit [40]. Moreover, the implicit content extraction is a hard task in natural language processing area, as [35] comments. Nevertheless, we propose a superficial semantic analysis based on named-entity recognition categories. Table 7 shows the results.

Table 7. Named-entity occurrence.

Named-Entity	Label	True News	Fake News
Person	(PER)	19,398	22,151
Localization	(LOC)	19,232	15,250
Organization	(ORG)	9,503	8,851
Miscellaneous	(MISC)	8,427	9,119

Based on the obtained results, the true statements present a larger number of localization occurrences (LOC) than fake statements. Otherwise, the true statements are overcome in a larger number of person occurrences (PER) when compared to the fake statements. Organization (ORG) has occurred more frequently in true statements, while miscellaneous (MISC) in fake statements.

Punctuation. [10] assumes that punctuation patterns could distinguish fake and true texts. Consequently, the punctuation behavior would be a "clue to deception". We evaluate the occurrence of each punctuation mark. The obtained data are shown in Table 8. In agreement with the literature, our results show a noticeable change among the punctuation setting in fake and true news. Note that in fake statements there has been an expressive use of interrogation, exclamation, endpoint, double quotes, two points, three consecutive points, square brackets, bar, and asterisks. For the true news, we observed the larger use of the comma, single trace, single quotes, and two consecutive points. We also point out that the "Error" label consists of annotation mistakes.

Table 8. Punctuation occurrence.

Punctuation		True News	Fake News
Comma	,	**43,244**	**31,610**
End point	.	**26,311**	**31,911**
Double quotes	" "	**5,004**	**17,200**
Parentheses	( )	12,170	11,027
Two points	:	**1,674**	**5,138**
Interrogation	?	**687**	**1,808**
Exclamation	!	**277**	**2,226**
Square brackets	[ ]	**237**	**4,638**
Three consecutive points	...	**71**	**2,746**
Two consecutive point	..	**1,619**	**127**
Error	n/a	1,389	1,320
Single trace	-	**910**	**13**
Long trace		80	46
Asterisk	*	**91**	**650**
Bar	/	**123**	**358**
Single quotes	' '	**443**	**3**
Four consecutive points	....	**0**	**54**
Circle		4	4
Double trace	–	12	0
Five consecutive points	.....	0	5

4.2 Sentence-level Analysis

According to the standard philosophical definition of lying, the intention to deceive is an important aspect of deception [2] [19] [38] [3]. For that reason, we carry out an initial understanding of dishonest intents in text, we analyze the sentence structure in true and fake statements. To achieve that, we evaluate the number of sentences and the average number of words, which is shown in Table 9.

Table 9. Sentence-level analysis.

Sentences	True News	Fake News
Total	43,066	50,355
Avg of words	15,45	13,24

Based on data displayed by Table 9, we may note that, in fake statements, there are 14,47% more sentences than in true statements. It is interesting to realize that, despite the greater number of sentences in fake news, the average number of words by sentence is smaller than in true news.

Moving forward, [6] suggest that the analysis of word usage is often not enough for deception prediction. Deeper language structure (syntax) must also be analyzed to predict instances of deception. A dependency tree, according to [18], is a syntactic structure corresponding to a given natural language sentence. This structure represents hierarchical relationships between words. Figure 1 shows a dependency tree example. Notice that the relations among the words are illustrated above the sentence with directed labeled arcs from heads to dependents. According to [17], we call this a typed dependency structure because the labels are drawn from a fixed inventory of grammatical relations. It also includes a *root* node that explicitly marks the root of the tree, i.e., the head of the entire structure. For the interested reader, the Universal Dependencies project [27] provides an inventory of dependency relations that are cross-linguistically applicable.

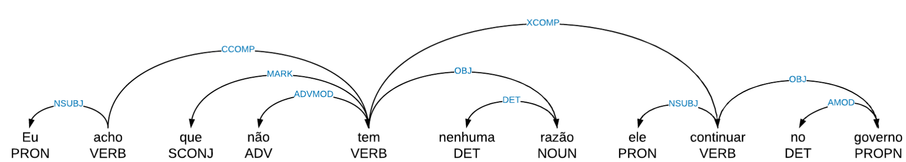

Fig. 1. Dependency tree example

As shown in Figure 4.2, the syntactical dependency structure for the following sentence extracted from our corpus: *Eu acho que não tem nenhuma razão ele continuar no governo.* ("I think there is no reason for him to remain in the government"). The NSUBJ relation identifies the subject; CCOMP identifies the complement of the main verb; ADVMOD identifies the adverb modifier; MARK is the word introducing a finite clause subordinate to another clause; OBJ identifies the direct object; DET identifies determinants; AMOD exhibits the adjectival modifier of a noun phrase (NP), and XCOMP consists of an open clausal complement for a verb [3].

Therefore, in order to investigate anomalies or divergences in syntactic structure of false and true statements, we also analyzed the dependency relation occurrences in our corpus. We present the results in the Table 10. Based on the obtained results (see Table 10), in an initial analysis, we found a relevant difference among the syntactic structures in fake and true news. For example, one may notice a significant difference in the occurrence of CASE, OBJ, OBL, NMOD, ROOT, DET, ADVCL, AUX, FLAT:NAME, CSUBJ and PARATAXIS structures. In the future, we intend to perform a deeper syntactical analysis of the dependency trees, looking for argument structure differences, for instance.

Table 10. Clausal dependency relations occurrence

Label	**Definition**	**True News**	**Fake News**
CASE	case marking	106,964	89,177
DET	determiner	71,070	78,013
AMOD	adjectival modifier	29,486	29,580
NMOD	nominal modifier	62,406	50,913
ROOT	root	43,055	50,035
FLAT:NAME	flat multiword expression (name)	45,955	40,025
NSUBJ	nominal subject	43,091	49,321
OBJ	object	39,787	45,063
OBL	oblique nominal	38,901	33,550
ADVMOD	adverbial modifier	22,741	28,834
CONJ	conjunct	21,316	20,907
APPOS	appositional modifier	21.146	20,587
CC	coordinating conjunction	18,603	17,586
MARK	marker	17,108	19,932
ACL	clausal modifier of noun (adjectival clause)	14,239	13,072
NUMMOD	numeric modifier	10,034	8,427
COP	copula	8,767	11,359
ADVCL	adverbial clause modifier	8,210	9,437
ACL:RELCL	relative clause modifier	8,177	7,720
CCCOMP	clausal complement	7,610	9,380
AUX	auxiliary	6,902	9,989
XCOMP	open clausal complement	6,208	7,411
AUX:PASS	auxiliary	5,931	6,485
NSUBJ:PASS	passive nominal subject	5,574	6,014
DEP	unspecified dependency	3,017	2,357
EXPL	expletive	2,139	2,874
NMOD:NPMOD	nominal modifier	2,055	2,093
OBL:AGENT	agent modifier	1,329	1,084
COMPOUND	compound	1,251	1,119
NMOD:TMOD	temporal modifier	1,193	368
FIXED	fixed multiword expression	1,135	1,259
PARATAXIS	parataxis	934	1,491
CSUBJ	clausal subject	769	1,053
IOBJ	indirect object	328	466
FLAT:FOREIGN	foreign words	13	15

[3] The typed dependency manual is available at `https://nlp.stanford.edu/software/dependencies_manual.pdf`

5 Final Remarks and Future Work

The current context of social media usage is unique, with diversity in format, and relatively new. However, lying and deceiving have been at play in other forms of human communication for ages [32]. In this paper, we presented a pioneering corpus study over a large deceptive corpus in Brazilian Portuguese. We automatically annotated a set of linguistic features to investigate actionable linguistics inputs and relevant differences among fake and true news. Based on the obtained results, we found that fake and true news present relevant differences in structural, lexical, syntactic, and semantic levels.

For future work, we intend to deepen our investigation of syntactical behavior and to explore discourse markers, and sophisticated machine learning techniques to provide deception detection classifiers for different tasks, such as fake news and reviews detection in several languages. For the interested reader, more information may be found at the OPINANDO project webpage (at `https://sites.google.com/icmc.usp.br/opinando/`).

References

1. Adams, S.H.: Statement analysis: What do suspects'words really reveal? FBI Law Enforcement Bulletin **1**(1), 12–20 (1996)
2. Augustine, S.: Lying. In: Deferrari, R.J. (ed.) Treatises on Various Subjects (The Fathers of the Church, vol. 16, p. 47–110. Catholic University of America Press (1952)
3. Bok, S.: Lying: moral choice in public and private life (1978)
4. Carver, C. S., S.M.F.: Attention and Self-Regulation: A Control-Theory Approach to Human Behavior (1981)
5. Conroy, N.J., Rubin, V.L., Chen, Y.: Automatic deception detection: Methods for finding fake news. In: Proceedings of the 78th ASIST Annual Meeting: Information Science with Impact: Research in and for the Community. American Society for Information Science, USA (2015)
6. Conroy, N.J., Rubin, V.L., Chen, Y.: Automatic deception detection: Methods for finding fake news. In: Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community. pp. 82:1–82:4. ASIST '15, American Society for Information Science, Silver Springs, MD, USA (2015)
7. Cristani, M., Roffo, G., Segalin, C., Bazzani, L., Vinciarelli, A., Murino, V.: Conversationally-inspired stylometric features for authorship attribution in instant messaging. In: Proceedings of the 20th ACM International Conference on Multimedia. pp. 1121–1124. ACM, New York, NY, USA (2012)
8. Davis, D., Brock, T.: Use of first person pronouns as a function of increased objective self-awareness. Journal of Experimental Social Psychology **11**, 381–388 (1975)
9. DePaulo, B., Lindsay, J.J., Malone, B., Muhlenbruck, L., Charlton, K., Cooper, H.: Cues to deception. Psychological bulletin **129**, 74–118 (2003)
10. DePaulo, B., Lindsay, J.J., Malone, B., Muhlenbruck, L., Charlton, K., Cooper, H.: Cues to deception. Psychological bulletin **129**, 74–118 (02 2003)
11. Duval, S., Wicklund, R.A.: A theory of objective self awareness. Academic Press, Oxford, England, 1 edn. (1972)

12. Fallis, D.: What is lying? Journal of Philosophy **106**(1), 29–56 (2009)
13. Feng, V.W., Hirst, G.: Detecting deceptive opinions with profile compatibility. In: Proceedings of the Sixth International Joint Conference on Natural Language Processing. pp. 338–346. Asian Federation of Natural Language Processing, Nagoya, Japan (Oct 2013)
14. Finney, K.L.: Detecting deception through rst: A case study of the casey anthony trial. Proceedings of the 31st annual NorthWest Linguistics Conference **1**(1), 12–23 (2015)
15. Friedman, H. S., T.J.S.: Language and deception. In: Robinson, G..W.P. (ed.) Handbook of language and social psychology, p. 257–270. John Wiley Sons, 1 edn. (1990)
16. Galasińki, D.: The language of deception: a discourse analytical study. SAGE Knowledge, SAGE Publications (2000)
17. Jurafsky, D., Martin, J.H.: Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition, 2nd Edition. Prentice Hall series in artificial intelligence, Prentice Hall, Pearson Education International (2009)
18. Jurafsky, D., Ranganath, R., McFarland, D.: Extracting social meaning: Identifying interactional style in spoken conversation. In: Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics. pp. 638–646. NAACL '09, Association for Computational Linguistics, Stroudsburg, PA, USA (2009)
19. Kupfer, J.: The moral presumption against lying. The Review of Metaphysics **36**(1), 103–126 (1982)
20. Mann, S., Vrij, A.: Telling and detecting lies in a high-stake situation: The case of a convicted murderer. Applied Cognitive Psychology **15**(2), 187–203 (2001)
21. Mann, W.C., Thompson, S.A.: Rhetorical Structure Theory: A Theory Of Text Organization (1987)
22. Meibauer, J.: The linguistics of lying. Annual Review of Linguistics **4**(1), 357–375 (2018)
23. Meibauer, J., Dynel, M.: Empirical approaches to lying and deception. International Review of Pragmatics **8**(3)
24. Monteiro, R., Santos, R., Pardo, T., Almeida, T., Ruiz, E., Vale, O.: Contributions to the Study of Fake News in Portuguese: New Corpus and Automatic Detection Results, pp. 324–334 (01 2018)
25. Nahari, G., Ashkenazi, T., Fisher, R., Granhag, P., Hershkowitz, I., Masip, J., Meijer, E., Nisin, Z., Sarid, N., Taylor, P., Verschuere, B., Vrij, A.: 'language of lies': Urgent issues and prospects in verbal lie detection research. Legal and Criminological Psychology **24**, 1–23 (01 2019)
26. Newman, M.L., Pennebaker, J.W., Berry, D.S., Richards, J.M.: Lying words: Predicting deception from linguistic styles. Personality and Social Psychology Bulletin **29**(5), 665–675 (2003)
27. Nivre, J., de Marneffe, M.C., Ginter, F., Goldberg, Y., Hajič, J., Manning, C.D., McDonald, R., Petrov, S., Pyysalo, S., Silveira, N., Tsarfaty, R., Zeman, D.: Universal dependencies v1: A multilingual treebank collection. In: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16). pp. 1659–1666. European Language Resources Association (ELRA), Portorož, Slovenia (2016)
28. Ott, M., Choi, Y., Cardie, C., Hancock, J.T.: Finding deceptive opinion spam by any stretch of the imagination. In: Proceedings of the 49th Annual Meeting

of the Association for Computational Linguistics: Human Language Technologies. pp. 309–319. Association for Computational Linguistics, Portland, Oregon, USA (Jun 2011)

29. Padilla Cruz, M.: Might interjections encode concepts? more questions than answers. Lodz Papers in Pragmatics **5**, 241–270 (12 2009)
30. Pasqualotti, P.R.: Reconhecimento de expressões de emoções na interação mediada por computador. Master's thesis, Porto Alegre, Brasil (2008)
31. Pennycook, G., Rand, D.: Why do people fall for fake news? New York Times Edition **1**, 1–12 (2019)
32. Rubin, V.L.: Deception detection and rumor debunking for social media. The SAGE Handbook of Social Media Research Methods (2017)
33. Rubin, V.L., Chen, Y., Conroy, N.J.: Deception detection for news: Three types of fakes. In: Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community. pp. 83:1–83:4. ASIST '15, American Society for Information Science, Silver Springs, MD, USA (2015)
34. Silva, M.J., Carvalho, P., Sarmento, L.: Building a sentiment lexicon for social judgement mining. In: Caseli, H., Villavicencio, A., Teixeira, A., Perdigão, F. (eds.) Computational Processing of the Portuguese Language. pp. 218–228. Springer Berlin Heidelberg, Berlin, Heidelberg (2012)
35. Vargas, F.A., Pardo, T.A.S.: Aspect clustering methods for sentiment analysis. In: 13th International Conference on the Computational Processing of Portuguese. pp. 365–374. Canela, RS, Brazil (2018)
36. Vorauer, J.D., Ross, M.E.: Self-awareness and feeling transparent: Failing to suppress one's self. Journal of Experimental Social Psychology volume = 35, number = 5, pp. 415–440 (1999)
37. Vosoughi, S., Roy, D., Aral, S.: The spread of true and false news online. Science **359**, 1146–1151 (03 2018)
38. williams, B.: Truth and Truthfulness: An Essay in Genealogy (2002)
39. Ye, S., Chua, T.S.: Learning object models from semistructured web documents **18**(3), 334–349 (2006)
40. Zimmermann, T.F.: Model-theoretic semantics. In: Maienborn, C., H.K.v.P.P. (ed.) Semantics: An International Handbook of Natural Language Meaning, p. 762–802. Berlin: de GruyterTo the best of our knowledge, 1 edn. (2011)
41. Zipitria, I., Sierra, B., Sopena-Garaikoetxea, I.: Emotion in deceptive language. In: Proceedings of the 39th Annual Meeting of the Cognitive Science Society, CogSci 2017, London, UK, 16-29 July 2017 (2017)