Linguistic Cues of Deception in a Multilingual April Fools' Day Context

Katerina Papantoniou[1,2], **Panagiotis Papadakos**[2], **Giorgos Flouris**[2], **Dimitris Plexousakis**[1,2]
1. Computer Science Department, University of Crete, Greece
2. Institute of Computer Science, FORTH, Greece
{papanton, papadako, fgeo, dp}@ics.forth.gr

Abstract

In this work we consider the collection of deceptive April Fools' Day (AFD) news articles as a useful addition in existing datasets for deception detection tasks. Such collections have an established ground truth and are relatively easy to construct across languages. As a result, we introduce a corpus that includes diachronic AFD and normal articles from Greek newspapers and news websites. On top of that, we build a rich linguistic feature set, and analyze and compare its deception cues with the only AFD collection currently available, which is in English. Following a current research thread, we also discuss the individualism/collectivism dimension in deception with respect to these two datasets. Lastly, we build classifiers by testing various monolingual and crosslingual settings. The results showcase that AFD datasets can be helpful in deception detection studies, and are in alignment with the observations of other deception detection works.

1 Introduction

April Fools' Day (for short AFD) is a long standing custom, mostly in Western societies. It is the only day of the year when practical jokes and deception are expected. This is the case for all social interactions, including journalism, which is generally considered to aim at the presentation of truth. Every year on this day, newspapers and news websites take part in an unofficial competition to invent the most believable, but untrue story. In this respect, AFD news articles fall into the deception spectrum, as they satisfy widely acceptable definitions of deception as in Masip et al. (2005).

The massive participation of news media in this custom establishes a rich corpus of deceptive articles from a diversity of sources. Although AFD articles may exploit common linguistic instruments with satire news, like exaggeration, humour, irony and paralogism, they are usually considered a distinct category. This is mainly due to the fact that they also employ other mechanisms which characterize deception in general, like sophisms, and changes in cognitive load and emotions (Hauch et al., 2015) to deceive their audience. AFD articles are often believable, and there exist cases where sophisticated AFD articles have been reproduced by major international news agencies worldwide[1].

This motivated us to extend our previous work on linguistic cues of deception and their relation to the cultural dimension of individualism and collectivism (Papantoniou et al., 2021), in the context of the AFD. That work examines if differences in the usage of linguistic cues of deception (e.g., pronouns) across cultures can be identified and attributed to the individualism/collectivism divide.

Specifically, the contributions of this work are:

- A new corpus that includes diachronic AFD and normal articles from Greek newspapers and news websites[2], adding one more AFD collection to the currently unique one in English (Dearden and Baron, 2019).
- A study and discussion of the linguistic cues of deception that prevail in the Greek and English collection, along with their similarities.
- A discussion on whether the consideration of the individualism/collectivism cultural di-

[1] https://www.nationalgeographic.com/history/article/150331-april-fools-day-hoax-prank-history-holiday

[2] The collection is available in: `https://gitlab.isl.ics.forth.gr/papanton/elaprilfoolcorpus`

mension in the context of AFD aligns with the results of our previous work.

- An examination of the performance of various classifiers in identifying AFD articles, including multilanguage setups.

2 Related work

The creation of reliable and realistic ground truth datasets for the deception detection task is a challenging task (Fitzpatrick and Bachenko, 2012). Crowdsourcing, in the form of online campaigns in which people express themselves in truthful and/or deceitful manner for a small payment are a well established way to collect deceptive data (Ott et al., 2011). Real-life situations such as trials (Soldner et al., 2019) or the use of data from board games have also been employed (Peskov et al., 2020). Also a popular approach is the reuse of content from sites that debunk articles like fake news and hoaxes (Wang, 2017; Kochkina et al., 2018). Lastly, satire news are another way to collect deceptive texts, but with some particularities due to humorous deception (Skalicky et al., 2020).

The only work that explores AFD articles is that of Dearden et al. (2019). They collected 519 AFD and 519 truthful stories and articles in English for a period of 14 years. A large set of features was exploited to identify deception cues in AFD stories. Structural complexity and level of detail were among the most valuable features while the exploitation of the same feature set to a fake news dataset resulted in similar observations.

To the best of our knowledge, the only deception related dataset for the Greek language is that of Karidi et al. (2019). This work proposed an automatic process for the creation of a fake news and hoaxes articles corpus, but unfortunately the created corpus over Greek websites is not available. If we also consider that the creation of a Greek dataset for deception through crowdsourcing is a cumbersome and expensive task, that is further hindered by the exceptionally limited number of native Greek crowd workers, it is easy to understand why there is a lack of datasets.

Regarding the individualism/collectivism cultural dimension, it constitutes a well-known division of cultures that concerns the degree in which members of a culture value more individual over group goals and vice versa. In individualism, ties between individuals are loose and individuals are expected to take care of only themselves and their immediate families, whereas in collectivism ties in society are stronger. In Papantoniou et al. (2021) there is an preliminary effort driven by prior work in psychology discipline (Taylor et al., 2017) to examine if deception cues are altered across cultures and if this can be attributed to this divide. Among the conclusions were that people from individualistic cultures employ more third and less first person pronouns to distance themselves from the deceit when they are deceptive, whereas in the collectivism group this trend is milder, signalling the effort of the deceiver to distance the group from the deceit. In addition, in individualistic cultures positive sentiment is employed in deceptive language, whereas in collectivists there is a restraint of expression of sentiment both in truthful and deceptive texts.

To this end, this work explores the deception-related characteristics of a new Greek corpus based on AFD articles from a variety of sources, and compares them with the English ones[3]. Further, since related studies (Triandis and Vassiliou, 1972; Hofstede, 1980; Koutsantoni, 2005) describe Greece as a culture with more collectivistic characteristics (by using country as proxy from culture), we also discuss differences in deception cues along this cultural dimension.

3 Corpus creation

The AFD articles have been hand gathered because a crawling based collection approach was not applicable in our case. Since the news web sites industry in Greece is not huge to establish an acceptable number of crawled AFD articles, we had to additionally collect articles from the press, including articles from the pre-WWW era. Specifically, we visited the local library that maintains a printed archive of newspapers and searched for disclosure articles in the issues after the 1st April, took photos of the AFD articles, and then used OCR and manual inspection to extract the text. In addition we contacted national and local news media providers to get access in their digitalized archives. The rest were gathered from the Web.

The articles were categorized thematically into the following five categories: society, culture, politics, world, and sports. If no category was pro-

[3] We also experimented with data from the limited number of satirical and hoaxes sources of the Greek Web. We do not discuss them here though, since the classifiers reported excellent accuracy showcasing the lack of diversity and the existence of domain specific information in the collected data.

vided by the original source, we manually annotated the articles. For each article we kept the title, the main body, the published date, the name, the type of the source (newspaper or news website), and (if available) the caption, the subtitle and the author. As preprocesing steps we applied spellcheck and normalization. The correction of spelling mistakes was necessary primarily for articles extracted through OCR tools, although spelling errors were identified in other articles too. Normalization was performed for homogeneity reasons in the texts retrieved from the 80's, since we observed language differences in some forms (e.g., in the suffix of genitive case), which are remains of an old form of Modern Greek[4].

For the truthful collection we used the same manual procedure and we tried to have a balanced dataset in terms of thematic categories. The truthful collection consists of articles that have been published in days relatively close to the 1st of April in order to have articles that do not differ significantly in respect to their topics, mentioned named entities, etc.

Since the AFD tradition is vivid in Greece, we were able to locate a lot of such articles from various newspapers and new websites for our corpus (112 different sources). Specifically, we managed to collect 254 truthful and 254 deceptive articles spanning over the period 1979 - 2021. In Tables 1 to 2 some statistics of the corpus are depicted.

Table 1: Overview of the dataset.

Measure	Truthful	Deceptive
Num. of articles	254	254
Avg. length	336	255
Min. length	57	33
Max. length	1347	1163

Table 2: Distribution of articles per topic.

Topic	Truthful	Deceptive
culture	20	24
politics	85	78
society	86	118
sports	22	29
world	41	5

4 Features analysis

For the analysis of AFD articles we adapt and build upon the feature set used in Papantoniou et al. (2021), but for the Greek language. The resulting feature set consists of 64 features for the Greek language and 75 for the English, due to the smaller availability of linguistic resources for Greek (e.g., in sentiment lexicons). For the analysis we performed the non-parametric Mann–Whitney U test (two-tailed) with a 99% confidence interval (CI) and $\alpha = 0.01$. Table 3 depicts the results of this analysis for elAFD and enAFD datasets[5].

In both datasets, positive sentiment is related to the deceptive articles, while negative sentiment with the truthful articles. The only exception concerns the enAFD dataset, where for the NRC lexicon the opposite holds (NRC is one of the six sentiment lexicons used for features in English). In addition, negative emotions like anger, fear and sadness are related to truthful news articles in both datasets. The use of positive emotive language during deception may be a strategy for deceivers to maintain social harmony as noticed also by other studies (Newman et al., 2003; Pérez-Rosas et al., 2018). The difference in the use of emotional language between truthful and deceptive news is more intense in the enAFD dataset, where five out of the eight emotions in the NRC lexicon are found statistical significant. This is in alignment with the results in Papantoniou et al. (2021) for individualistic and collectivistic cultures.

Further, deceptive texts seem to be related with an increased use of adverbs in both datasets. This can be related to the less concreteness of deceptive texts as discussed in Kleinberg et al. (2019) and it is in line with many theories of deception like the Reality Monitoring (Johnson et al., 1998), Criteria based Content Analysis (Undeutsch, 1989) and Verifiability Approach (Nahari et al., 2014). This also explains the prevalence of the number of named entities, spatial related words, conjunctions and WDAL imagery score in truthful texts in the enAFD dataset and the use of more motion verbs in deceptive texts in the elAFD dataset. According to cognitive load theory (Sweller, 2011) in deceptive texts the language is less specific and consists of simpler constructs. The same holds for modality, another common feature among the datasets, that is considered a signal of subjectivity that pro-

[4] https://en.wikipedia.org/wiki/Katharevousa

[5] All the features are described in https://gitlab.isl.ics.forth.gr/papanton/elaprilfoolcorpus

vides a degree of uncertainty. In addition, hedges in enAFD dataset, also express some feeling of doubt or hesitancy.

Lexical diversity as expressed by the token-type ratio (TTR), that is the ratio of unique words to the total number of tokens, is related to the deceptive texts. This seems to contradict all the above, but could be attributed to the fact that deceptive texts are shorter. Although this is more evident in the case of the enAFD dataset, it also holds for elAFD dataset (see Table 1).

Boosters, which are words that express confidence (e.g., certainly) are quite discriminative for deceptive texts for the enAFD dataset. Moreover we observe the connection of the future tense with deception and of the past with truth. The above were also marked in Papantoniou et al. (2021) in different domain from the news articles domain.

Finally, first personal pronouns have been found to be rather discriminative of deceptive texts in various deception detection and cultural studies, including Papantoniou et al. (2021). However, in this study pronouns are statistical important only for the enAFD dataset. This probably reflects idiosyncrasies of the news domain, since articles mainly present objectively facts and not opinions, and as a result the use of first personal pronouns is avoided. This holds for the elAFD dataset that includes AFD articles from the news sites and the press, and not for the enAFD dataset that consists of various types of AFD articles and stories collected from the web through crowdsourcing[6].

Table 3: The statistical significant features (p<0.1) with at least a small effect size (r>0.1) for the elAFD and enAFD datasets. The features are in ascending p value order. We also report the effect size. Features with moderate effect size (r>0.3) are bold, while common features between the datasets are underlined. pp denotes personal pronouns.

Deceptive	Truthful
elAFD	
adverbs (0.31)	punctuation (-0.17)
adj. & adv. (0.27)	nrc sadness(-0.17)
TTR (0.27)	plosives (-0.16)
pos. sentiment (0.21)	nrc anger (-0.15)
modal verbs (0.17)	nrc fear (-0.14)
motion verbs (0.117)	vowels (-0.14)
	consonants (-0.14)
enAFD	
boosters (0.39)	NE num. (-0.27)
modal verbs (0.35)	spatial num. (-0.26)
TTR(0.31)	conjuctions (-0.24)
future (0.27)	nrc fear (-0.23)
adverbs (0.2)	past (-0.23)
1st pers. pp (0.2)	nrc sadness (-0.23)
mpqa pos. (0.2)	nrc anger (-0.21)
nrc neg.* (-0.2)	nrc trust (-0.21)
2nd pers. pp (0.19)	avg. word len. (-0.17)
1st pers. pp pl. (0.18)	collectivism (-0.16)
sentiwordnet pos. (0.17)	nrc pos.* (-0.16)
demonstrative (0.17)	wdal imagery (-0.15)
hedges (0.17)	mpqa neg. -0.14)
adj & adv (0.16)	nasals (-0.14)
present (0.15)	fbs neg. (-0.14)
vader sentiment (0.14)	consonants (-0.13)
verb num. (0.14)	anew arousal (-0.13)
pers. pron. (0.12)	prepositions (-0.12)
total pronouns (0.11)	fricatives (-0.11)
	3rd per. pp sg. (-0.11)
	avg. preverb len. (-0.11)
	nrc disgust (-0.1)

5 Classification

We evaluated the predictive performance of different feature sets and approaches for AFD datasets, including logistic regression experiments[7] and fine-tuned monolingual BERT models for each language[8] (Devlin et al., 2019; Koutsikakis et al., 2020). We also performed cross lingual experiments by exploiting the multilingual BERT model (mBERT) to examine if there are similarities among AFD datasets captured by the BERT.

A stratified split to the datasets was used to create training, testing, and validation subsets with a 70-20-10 ratio. For the cross lingual experiment we trained and validated a model over the 80% and 20% of a language specific dataset respectively, and then tested the performance of the model over the other dataset. We report the results on test sets, while validation subsets were used for fine-tuning the hyper-parameters of the algorithms. For the logistic regression the tuned through brute force parameters were: a) Weka algorithm ($SimpLog|Log$: *simple logistic* (Landwehr et al., 2005) or *logistic* (Le Cessie and Van Houwelingen, 1992)) b) all n-grams of size in $[a, b]$, with $a \geq b$ and $a, b \in [1, 3]$ ((a, b)), c) stemming ($stem$), d) attribute selection ($attrsel$) (applicable only to Log algorithm since it is the de-

[6] https://aprilfoolsdayontheweb.com/2004.html

[7] We employ the Weka API (Hall et al., 2009)

[8] We used tensorflow 2.2.0, keras 2.3.1, and the bert-for-tf2 0.14.4 implementation of google-research/bert, over an AMD Radeon VII card and the ROCm 3.7 platform.

fault for $SimpLog$), e) stopwords removal ($stop$) and, f) lowercase conversion ($lowercase$). For the BERT experiments, the hyperparameters were tuned by random sampling 60 combinations of values, keeping the combination that gave the minimum validation loss. Early stopping with patience 4 was used and the max epochs number was set to 20. The tuned hyperparameters were: learning rate, batch size, dropout rate, max token length, and randomness seeds.

In all cases, we report Recall (R), Precision (P), F-measure (F), Accuracy (A) and AUC (A'). Since the datasets are balanced the majority baseline is 50%. The input for the models consists of the concatenation of the title, the subtitle, the body of the articles and the caption text. Since titles are important for deception detection (Horne and Adali, 2017) and BERT processes texts of up to 512 wordpieces, we placed the title first.

5.1 Logistic regression experiments

The examined features sets were: a) the features presented in section 4 (ling), b) n-grams features i.e., phoneme-gram (ph-gram), character-gram (char-gram), word-gram (w-gram), POS-gram (pos-gram), and syntactic-gram (sn-gram) (the latter for the enAFD only), and c) the linguistic+ model that represents the best model that combines the linguistic features with any of the n-gram features. The results are presented in Tables 4 and 5. With * we mark the setups with a statistically significant difference to the best setup regarding accuracy, based on a two proposition z-test (1-tailed) with a 99% CI. We observe that the combination of *lingustic* features with uni/bi/tri-grams for the elAFD dataset and the unigrams for the enAFD are the best setups. For the enAFD dataset, the second best model is the combination of *linguistic* features with trigrams. $SimpLog$ seems to perform better, while stemming, lowercase conversion and stopwords removal are generally beneficiary.

Table 4: Logistic regression results for el AFD .

Best setup	**R**	**P**	**F**	**A'**	**A**
ling.$_{SimpLog}$	62	76	68	82	71
ph-gram$_{(1,2),attrsel,Log}$*	70	67	68	77	68
char-gram$_{(3,3),SimpLog}$*	72	68	70	76	69
w-gram$_{(1,2),SimpLog}$	68	73	71	80	72
pos-gram$_{(2,3),SimpLog}$*	72	65	68	75	67
ling.+$_{word,(1,3),stop,lowercase,SimpLog}$	**74**	**79**	**76**	**85**	**77**

Table 5: Logistic regression results for en AFD .

Best setup	**R**	**P**	**F**	**A'**	**A**
ling.$_{Log}$*	66	80	72	**87**	75
ph-gram$_{(1,1),SimpLog}$	80	77	78	84	78
char-gram$_{(1,3),attrsel,Log}$*	76	72	74	80	73
w-gram$_{(1,1),stem,SimpLog}$	**79**	**81**	**80**	**87**	**80**
pos-gram$_{(3,3),SimpLog}$*	71	69	70	76	69
sn-gram$_{(2,2),SimpLog}$*	80	68	73	77	71
ling.+$_{Word,(3,3),stop,lowercase,SimpLog}$	74	80	77	**87**	78

Table 6: BERT models evaluation results.

	R	**P**	**F**	**A'**	**A**
el$_{bert}$	76	83	79	90	**80**
el$_{bert+ling}$	72	80	76	86	77
el$_{mbert}$	70	73	71	81	72
el$_{mbert+ling}$	50	81	62	83	69
en$_{bert}$	88	85	87	94	**86**
en$_{bert+ling}$	74	89	81	91	83
en$_{mbert}$	50	95	66	91	73
en$_{mbert+ling}$	50	86	63	86	71
en→el $_{mbert}$	38	76	50	72	63
el→en $_{mbert}$	24	87	37	71	60

5.2 BERT experiments

In these experiments, we fine-tuned BERT by adding a task-specific linear classification layer on top, using the sigmoid activation function. We also combined BERT with linguistics features by concatenating the embedding of the [CLS] token with the linguistic features, and pass the resulting vector to the task-specific classifier (with a slightly modified architecture). The results of the experiments are presented in Table 6. Although it outperformed logistic regression experiments in both datasets, the differences are not statistical significant. In addition, the combination with linguistic features is not beneficial. Multilingual BERT models perform worse, especially for Greek. In the cross lingual experiments the classifiers performance is limited to about 60% accuracy in both experiments, showcasing that the BERT layers are not able to capture language agnostic information from our datasets.

6 Conclusion and Future work

We introduced a new dataset with AFD news articles in Greek and analyzed and compared its deception cues with another English one. The results showcased the use of emotional language, especially of positive sentiment, for deceptive articles which is even more prevalent in the individualistic English dataset. Further, deceptive articles use less concrete language, as manifested by the increased use of adverbs, hedges, and boosters and less usage of named entities, spatial related words and conjunctions compared to the truthful ones. The future and past tenses were correlated with deceptive and truthful articles respectively. All the above, mainly align with previous work (Papantoniou et al., 2021), except from some differences in the usage of pronouns for the Greek dataset, which is attributed to the idiosyncrasies of the news domain. The accuracy of the deployed classifiers offered adequate performance, with no statistically significant differences between the best logistic regression and the BERT models.

In the future we aim at creating even more crosslingual datasets for deception detection tasks through crowdsourcing and by employing the Chattack platform (Smyrnakis et al., 2021).

Acknowledgement

This work has received funding from the Hellenic Foundation for Research and Innovation (HFRI) and the General Secretariat for Research and Technology (GSRT), under grant agreement No 4195.

References

[Dearden and Baron2019] Edward Dearden and Alistair Baron. 2019. Fool's Errand: Looking at April Fools Hoaxes as Disinformation through the Lens of Deception and Humour. April. 20th International Conference on Computational Linguistics and Intelligent Text Processing, CICLing 2019 ; Conference date: 07-04-2019 Through 13-04-2019.

[Devlin et al.2019] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–86, Minneapolis, Minnesota, June. Association for Computational Linguistics.

[Fitzpatrick and Bachenko2012] Eileen Fitzpatrick and Joan Bachenko. 2012. Building a Data Collection for Deception Research. In *Proceedings of the Workshop on Computational Approaches to Deception Detection*, pages 31–8, Avignon, France, April. Association for Computational Linguistics.

[Hall et al.2009] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The WEKA data mining software: an update. *SIGKDD Explorations*, 11(1):10–18.

[Hauch et al.2015] Valerie Hauch, Iris Blandón-Gitlin, Jaume Masip, and Siegfried L. Sporer. 2015. Are Computers Effective Lie Detectors? A Meta-Analysis of Linguistic Cues to Deception. *Personality and Social Psychology Review*, 19(4):307–342. PMID: 25387767.

[Hofstede1980] Geert Hofstede. 1980. *Culture's consequences: International differences in work-related values*. Sage Publications.

[Horne and Adali2017] Benjamin D. Horne and Sibel Adali. 2017. This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News. *ArXiv*, abs/1703.09398.

[Johnson et al.1998] Marcia K. Johnson, Julie G. Bush, and Karen J. Mitchell. 1998. Interpersonal Reality Monitoring: Judging the Sources of Other People's Memories. *Social Cognition*, 16(2):199–224.

[Kleinberg et al.2019] Bennett Kleinberg, Isabelle van der Vegt, Arnoud Arntz, and Bruno Verschuere. 2019. Detecting deceptive communication through linguistic concreteness, Mar.

[Kochkina et al.2018] Elena Kochkina, Maria Liakata, and Arkaitz Zubiaga. 2018. PHEME dataset for Rumour Detection and Veracity Classification.

[Koutsantoni2005] Dimitra Koutsantoni. 2005. Greek Cultural Characteristics and Academic Writing. *Journal of Modern Greek Studies*, 23:97–138, 05.

[Koutsikakis et al.2020] John Koutsikakis, Ilias Chalkidis, Prodromos Malakasiotis, and Ion Androutsopoulos. 2020. GREEK-BERT: The Greeks Visiting Sesame Street. In *11th Hellenic Conference on Artificial Intelligence*, SETN 2020, page 110–117, New York, NY, USA. Association for Computing Machinery.

[Landwehr et al.2005] Niels Landwehr, Mark Hall, and Eibe Frank. 2005. Logistic Model Trees. *Machine Learning*, 59(1):161–205, May.

[Le Cessie and Van Houwelingen1992] S. Le Cessie and J.C. Van Houwelingen. 1992. Ridge Estimators in Logistic Regression. *Applied Statistics*, 41(1):191–201.

[Masip et al.2005] Jaume Masip, Siegfried L. Sporer, Eugenio Garrido, and Carmen Herrero. 2005. The detection of deception with the reality monitoring approach: a review of the empirical evidence. *Psychology, Crime & Law*, 11(1):99–122.

[Nahari et al.2014] Galit Nahari, Aldert Vrij, and Ronald P. Fisher. 2014. The Verifiability Approach: Countermeasures Facilitate its Ability to Discriminate Between Truths and Lies. *Applied Cognitive Psychology*, 28(1):122–128.

[Newman et al.2003] 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(5):665–75. PMID: 15272998.

[Ott et al.2011] Myle Ott, Yejin Choi, Claire Cardie, and Jeffrey T. Hancock. 2011. 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 - Volume 1*, HLT '11, pages 309–19, Stroudsburg, PA, USA. Association for Computational Linguistics.

[Papantoniou et al.2021] Katerina Papantoniou, Panagiotis Papadakos, Theodore Patkos, Giorgos Flouris, Ion Androutsopoulos, and Dimitris Plexousakis. 2021. Deception detection in text and its relation to the cultural dimension of individualism/collectivism. *Natural Language Engineering*. Also appeared as an arXiv preprint arXiv:2105.12530.

[Pérez-Rosas et al.2018] Verónica Pérez-Rosas, Bennett Kleinberg, Alexandra Lefevre, and Rada Mihalcea. 2018. Automatic Detection of Fake News. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3391–3401, Santa Fe, New Mexico, USA, August. Association for Computational Linguistics.

[Peskov et al.2020] Denis Peskov, Benny Cheng, Ahmed Elgohary, Joe Barrow, Cristian Danescu-Niculescu-Mizil, and Jordan Boyd-Graber. 2020. It Takes Two to Lie: One to Lie, and One to Listen. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3811–3854, Online, July. Association for Computational Linguistics.

[Pla Karidi et al.2019] Danae Pla Karidi, Harry Nakos, and Yannis Stavrakas. 2019. Automatic Ground Truth Dataset Creation for Fake News Detection in Social Media. In Hujun Yin, David Camacho, Peter Tino, Antonio J. Tallón-Ballesteros, Ronaldo Menezes, and Richard Allmendinger, editors, *Intelligent Data Engineering and Automated Learning – IDEAL 2019*, pages 424–436, Cham. Springer International Publishing.

[Skalicky et al.2020] Stephen Skalicky, Nicholas Duran, and Scott A Crossley. 2020. Please, Please, Just Tell Me: The Linguistic Features of Humorous Deception. *Dialogue & Discourse*, 11(2):128–149, December.

[Smyrnakis et al.2021] Emmanouil Smyrnakis, Katerina Papantoniou, Panagiotis Papadakos, and Yannis Tzitzikas. 2021. Chattack: A Gamified Crowdsourcing Platform for Tagging Deceptive & Abusive Behaviour. In *European Conference on Information Retrieval*, pages 549–553. Springer.

[Soldner et al.2019] Felix Soldner, Verónica Pérez-Rosas, and Rada Mihalcea. 2019. Box of Lies: Multimodal Deception Detection in Dialogues. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1768–1777, Minneapolis, Minnesota, June. Association for Computational Linguistics.

[Sweller2011] John Sweller. 2011. Chapter Two - Cognitive Load Theory. volume 55 of *Psychology of Learning and Motivation*, pages 37–76. Academic Press.

[Taylor et al.2017] Paul J. Taylor, Samuel Larner, Stacey M. Conchie, and Tarek Menacere. 2017. Culture moderates changes in linguistic self-presentation and detail provision when deceiving others. *Royal Society Open Science*, 4(6):170128, June.

[Triandis and Vassiliou1972] Harry C. Triandis and Vasso Vassiliou. 1972. Interpersonal influence and employee selection in two cultures. *Journal of Applied Psychology*, 56:140–145.

[Undeutsch1989] Udo Undeutsch, 1989. *The Development of Statement Reality Analysis*, pages 101–19. Springer Netherlands, Dordrecht.

[Wang2017] William Yang Wang. 2017. "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection. In Regina Barzilay and Min-Yen Kan, editors, *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 2: Short Papers*, pages 422–426. Association for Computational Linguistics.