Automatic Detection of Cross-language Verbal Deception

Pasquale Capuozzo[1] **(pasquale.capuozzo@phd.unipd.it)**
Ivano Lauriola[2,3] **(ivano.lauriola@phd.unipd.it)**
Carlo Strapparava[3] **(strappa@fbk.eu)**
Fabio Aiolli[2] **(aiolli@math.unipd.it)**
Giuseppe Sartori[1] **(giuseppe.sartori@unipd.it)**

[1] University of Padova, Department of General Psychology, Padova, Via Venezia 8, 35131, IT
[2] University of Padova, Department of Mathematics, Padova, Via Trieste 63, 35121, IT
[3] Fondazione Bruno Kessler, Trento, Via Sommarive 18, 38123, IT

Abstract

The assessment of how a deceptive message is produced in different languages has received little attention, with the majority of studies focused on the English language. Moreover, there is no agreement about the stability of linguistic clues of deceit across different languages. In this paper, we address this issue by analysing both theory-driven linguistic markers of deception (cognitive load hypothesis) and standard text categorisation features. After compiling a multilingual corpus of both honest and deceitful first-person opinions regarding five different topics, we assessed the cross-language applicability of four different features sets in within-topic, cross-topic and cross-language binary classification experiments. Results showed promising classification performances in all the three experiments with few exceptions. Interestingly, linguistic markers of deceit linked to the cognitive load hypothesis exhibited the same trend in the two languages under investigation and the cross-language evaluation highlighted their usefulness in spotting deceit between different languages.

Keywords: deception; multilingual; cognitive load; computational linguistics; machine learning

Introduction

Lying is a ubiquitous phenomenon across societies (Serota, Levine, & Boster, 2010), and the spread of internet usage increased web-based cross-cultural interactions, providing new opportunities for deceiving. Moreover, with the globalisation advancement, the interactions between investigators and potential deceivers coming from different cultural backgrounds increased in the last decade (Giebels & Taylor, 2009). It is to be noted that the interactions above often occur between people speaking different languages, making harder the task of spotting possible deceitful intentions (Da Silva & Leach, 2013).

Up to now, far too little attention has been paid to how a deceptive message is yielded across different languages. Indeed, the majority of studies analysed deceit-related verbal characteristics within one language, where English is the most commonly studied language. Thus, what is known about linguistic markers of deceit is restricted to the English language and their applicability to other languages is still to be verified. Accordingly, some authors highlighted the need for shifting the focus also on languages other than English (Spence, Villar, & Arciuli, 2012; Pérez-Rosas & Mihalcea, 2014), assessing deception-related linguistic features in different ethnicities speaking in their native language (Potapova & Lykova, 2016). However, the cross-linguistic applicability of markers of deceit tested in previous studies remains an unsolved research subject and, to date, there is no agreement about the stability of linguistic clues of deception across different languages. Indeed, although some studies highlighted possible stability across languages of verbal clues of deceit (Matsumoto, Hwang, & Sandoval, 2015b; Matsumoto & Hwang, 2015; Matsumoto, Hwang, & Sandoval, 2015a), others led to the opposite conclusion (Leal et al., 2018; Rungruangthum & Todd, 2017). Therefore, the assessment of how a deceptive message is yielded in different languages seem to deserve further attention, considering the possibility of relying on new investigation methodologies.

Interestingly, the last decades have seen a growing interest from the scientific community in the automatic detection of deceit through text and speech analysis (Nunamaker et al., 2012; Fitzpatrick, Bachenko, & Fornaciari, 2015; Hauch, Blandón-Gitlin, Masip, & Sporer, 2015) and an increasing amount of studies is addressing verbal deception through computational linguistics and automated classification methods (Fusilier, Montes-y Gómez, Rosso, & Cabrera, 2015; Krishnamurthy, Majumder, Poria, & Cambria, 2018; Kleinberg, Mozes, Arntz, & Verschuere, 2018). Nonetheless, also here most of the mentioned studies focused on the English language while only a few works evaluated the automatic detection of deceit between different languages (Pérez-Rosas & Mihalcea, 2014; Levitan, Maredia, & Hirschberg, 2018). Moreover, to the best of our knowledge, no studies compared the predictive value across languages of theory-driven linguistic markers of deceit.

In this paper, we try to fill this gap by implementing a text representation, including linguistic clues of deceit related to a specific theory: the cognitive load hypothesis. Among several theories proposed for studying verbal deception, the cognitive load approach seems one of the most promising, leading to higher accuracy rates compared to standard methods (Vrij, Fisher, & Blank, 2017). The cognitive load approach assumes that lying is most of the times more mentally taxing than telling the truth (Zuckerman, DePaulo, & Rosenthal, 1981) and the higher cognitive effort accompanied to the act of lying should produce measurable linguistic clues of deceit.

Recently, one of the most comprehensive meta-analysis on the effectiveness of computers as lie detectors (Hauch et al., 2015) provided an extensive assessment of linguistic clues of deception linked to specific theories, including the cognitive load hypothesis. Based on that meta-analysis, we extracted the cognitive load related linguistic markers of deceit for the assessment of their cross-lingual applicability.

In this study, we aim to assess the effectiveness of a text representation composed of both theory-driven linguistic clues of deceit and standard linguistic features for text categorisation to spot deceitful narratives among different languages. After collecting a multilingual corpus of both deceiving and truthful first-person opinions about five various topics, we assessed the joint and individual performance of the features set considered in within-topic, cross-topic and cross-language binary classification experiments.

Method

To gather a multilingual corpus of both deceitful and truthful narratives, we considered two different samples from the US and Italy. We asked participants to provide in their language first-person opinions regarding five various topics. In this section, we will describe the recruiting methodology, participants' demographics and data collection procedure.

Participants

For collecting the sample from the US, we employed Amazon Mechanical Turk (AMT), setting a location restriction. The task could be performed only by Turkers with an approval rating equal to or higher than 80%. The time allotted for each task was 20 minutes, and only a single submission per participant was allowed. Each contribution was rewarded with 0.25$. In total 727 Turkers completed the task, but 227 contributions were rejected before analysis because participants haven't followed the instructions (i.e. unintelligible, unreasonably short or contained a description of the phenomenon instead of a first-person opinion). Therefore, five-hundred participants from the US (315 female; age 37.7±13.2) were included in the final analysis.

Regarding the Italian sample, since only 2% of AMT workers are located in Italy (Difallah, Filatova, & Ipeirotis, 2018), we decided to employ the Google form service for the data collection. Italian participants were recruited on a volunteer basis by spreading the Google form on social media and by e-mail. Furthermore, they have not received any monetary incentive for the participation, no limitation in time was allotted for completing the task, and only a single submission per participant was allowed. In total, 425 volunteers completed the task, but 90 contributions were rejected before analysis because participants haven't followed the instructions (by applying the same criteria employed for the US sample). Hence, three-hundred thirty-five participants from Italy (242 female; age 27.2±10.0) were included in the final analysis.

Data collection procedure

As mentioned before, for compiling a multilingual corpus of both deceptive and truthful narratives, we focused on first-person opinions about five different topics: Abortion (Abo), Cannabis legalisation (CL), Euthanasia (Eut), Gay marriage (GM) and Policy on migrants (PoM). The rationale behind topics selection relies on the assumption that the majority of people are likely to have a polarised position towards these highly debated topics. Consequently, they can easily express or deny it.

All the participants were asked to type in their native language both truthful and deceptive first-person opinions about the topics mentioned above in a free text response modality. The applied paradigm is based on an experimental ground-truth, meaning that each participant provided a truthful or a deceptive opinion according to specific instructions. The honest first-person opinions were generated by asking participants to provide in at least 4-5 lines their actual attitude towards a given topic. Contrarily, for gathering deceptive statements, participants were instructed to describe in at least 4-5 lines a fake opinion, not corresponding to their own opinion with the primary purpose to convince a hypothetical reader that the deceptive opinion represents their real point of view about the topic.

To maintain the overall proportion between truthful and deceptive narratives as balanced as possible, four Human Intelligence Tasks (HITs) were created. The HITs were balanced for ground-truth in a way that, overall, half of the first-person opinions gathered would have been deceptive and the other half truthful for each topic (Table 1).

Topic	HIT1	HIT2	HIT3	HIT4
Abo	D	T	D	T
CL	T	T	D	D
Eut	T	D	T	D
GM	T	D	T	D
PoM	D	T	D	T

Table 1: HITs' structure description. T = Truthful opinion; D = Deceptive opinion; Abo = Abortion; CL = Cannabis Legalization; Eut = Euthanasia; GM = Gay Marriage; PoM = Policy on migrants.

For instance, regarding the HIT1, participants typed their genuine attitude towards Gay marriage, Euthanasia and Cannabis legalisation. At the same time, for Policy on migrants and Abortion, they provided a deceptive opinion according to the instructions. The final result of the data collection procedure is shown in Table 2. The dataset will be made publicly available upon request for research purposes.

Features

The present study aims to assess the effectiveness of both standard text categorisation features and theory-driven linguistic markers of deceit in detecting deception across different languages. For doing so, we considered both the individ-

Topic	IT		EN	
	T	D	T	D
Abo	160	175	250	250
CL	158	177	250	250
Eut	175	160	250	250
GM	175	160	250	250
PoL	160	175	250	250
All	828	847	1250	1250

Table 2: The number of truthful (T) and deceptive (D) narratives per topic, gathered from the Italian and the US sample.

ual and merged contribution of various features sets, where each of them is independent of the others and expresses different aspects of the phenomenon under investigation. This section describes in detail the feature sets examined.

Bag-Of-Words (BOW)

The Bag-Of-Words is a popular orderless representation of documents, which are described as the (multi) set of words that compose it. This representation is commonly used in document classification tasks, where the content is useful information. The simplest extension of the BOW representation is the *n-grams*, where features consist of the frequency of contiguous sequences of *n* words. However BOW, or *n*-grams in general, is a domain-specific document representation. Hence, it could not be suitable to catch relations between different topics, or especially between different languages. In this work, the *n*-grams (uni- and bi-grams) representation has been computed by the Scikit-learn package (Pedregosa et al., 2011). Stop-words have been removed.

Part-of-Speech (POS)

The Part-Of-Speech tagging is the process of assigning a category (e.g. adjective, verb) for each word of a document, relying on both, syntax and context. The sequence of associated categories provides useful information regarding the structure of the document and sentences instead of the content. The POS tagging procedure has been computed by means of the TreeTagger[1] system, which exploits the Penn Treebank tagset (36 different tags). Uni- and bi-grams POS tags have been extracted from the tagged documents.

Linguistic Features (LF)

Based on the meta-analysis mentioned previously (Hauch et al., 2015), we extracted the verbal clues which demonstrated to be effective in spotting the higher cognitive effort accompanied by the act of deceiving.

The resulting linguistic cues are word quantity, type-token ratio, average sentence length, and exclusive words. The first three linguistic features were calculated as described in (Hauch et al., 2015), while the exclusive words were extracted from the Oxford Thesaurus of English (Waite, 2009), translated in Italian for the cross-language comparison and then computed as specified by (Hauch et al., 2015). According to the meta-analysis results, all the cognitive load-based linguistic features should have higher average values in truthful statements than in deceptive ones. The LF computation on the collected data is shown in Table 3.

Function Words (FW)

Function words (e.g., pronouns, prepositions, articles, conjunctions, and auxiliary verbs) express grammatical relationships between content words within a sentence. They signal the structural relationships that content words have to one another and are considered one of the most critical stylometric feature (Kestemont, 2014). Indeed, how people use function words reveals their linguistic style (Chung & Pennebaker, 2007) and can account what they are thinking and feeling (Newman, Pennebaker, Berry, & Richards, 2003). In the present study, a set of 318 English function words has been extracted from the scikit-learn package (i.e. the stopwords list). Then, the representation has been computed by considering the occurrences of each function word as an individual feature. For the cross-language comparison, the function words list has been translated into Italian by two experts.

Experimental Assessment

An extensive empirical assessment has been carried out to evaluate the capability of the above features sets in the automatic classification of truthful and deceptive narratives. Three different types of binary classification experiments and evaluations have been conducted, which are within-topic, cross-topic and cross-language. In this section, after portraying the general characteristics of the employed classifier, a detailed description of the three classification tasks is provided.

A SVM has been used as binary classifier. The hyperparameters have been selected using a 5-fold cross-validation procedure. These hyper-parameters are the regularization value *C* and the kernel function. The *C* value has been selected from the set $\{10^i, i : -2, \dots, 4\}$. The kernel function is the Homogeneous Polynomial (HP), with form $k(\boldsymbol{x},\boldsymbol{z}) = \langle \boldsymbol{x},\boldsymbol{z}\rangle^d, d \in \{1,\dots,5\}$. Other kernel functions have been used in a preliminary experimentation phase, such as the popular RBF kernel, without concrete improvements. However, each base representation and feature set is able to express different information, emphasizing a specific aspect of the main problem, i.e. the content, the cognitive load, or the linguistic structure. These representations are virtually orthogonal, and their combination could improve the performance of the classifier. To this end, feature sets have been considered both individually and combined via Multiple Kernel Learning (MKL; (Gönen & Alpaydın, 2011)). In short, the MKL is a popular framework used to learn the kernel as a principled combination of several weak feature sets. In this work, the EasyMKL (Aiolli & Donini, 2015) algorithm has been used. The algorithm learns the convex combination of base kernels which maximizes the margin between classes, i.e.: $k_{\boldsymbol{\mu}}(\boldsymbol{x},\boldsymbol{z}) = \sum_r \mu_r k_r(\boldsymbol{x},\boldsymbol{z})$, with constraints $\|\boldsymbol{\mu}\|_1 = 1$ and

[1] http://www.cis.uni-muenchen.de/∼schmid/tools/TreeTagger/

Linguistic features	IT		EN	
	T	D	T	D
Word quantity	$46.24_{\pm 32.73}$	$39.99_{\pm 23.41}$	$62.55_{\pm 30.15}$	$48.82_{\pm 25.92}$
Exclusive words	$7.47_{\pm 7.98}$	$6.49_{\pm 6.90}$	$0.48_{\pm 0.80}$	$0.40_{\pm 0.68}$
Average sentence length	$25.83_{\pm 12.40}$	$23.30_{\pm 10.57}$	$18.09_{\pm 6.52}$	$15.10_{\pm 6.51}$
Type-token ratio	$0.87_{\pm 0.08}$	$0.88_{\pm 0.08}$	$0.76_{\pm 0.09}$	$0.80_{\pm 0.09}$

Table 3: Linguistic features computed on the proposed datasets for truthful and deceptive opinions.

$\mu_r \geq 0$, where k_r is the r-th kernel, and $\boldsymbol{\mu}$ is the weights vector which parametrizes the combination. For each feature set, 5 HP kernels with degrees $1 \ldots 5$ and the identity kernel have been used as base kernels. In the MKL setting, base kernels have been normalized to prevent scaling issues. The F_1 measure has been used to evaluate the effectiveness of the models.

Within-topic evaluation

In the within-topic evaluation, a binary classifier has been trained for each topic individually. This experiment allows understanding the complexity of deception classification problem and the potential of the dataset. Furthermore, results achieved on this evaluation task provide a useful baseline to quantify and to compare the cross-topic and the cross-language effectiveness. A stratified nested 5-fold cross-validation has been applied to each topic of a given language. Specifically, the outer cross validation has been used to isolate the test set, whereas the inner cross-validation (i.e. applied to remaining training data) has been used to train the model and to select the hyper-parameters. The results computed on the outer cross-validation have been averaged and reported in Table 4.

The standard deviation is also reported. Note that, despite the global number of collected opinions, the classifier can rely on much less training examples, i.e. 268 and 400 opinions (80% of the available data) for the Italian and English datasets respectively.

Cross-topic evaluation

Classical content-based representations, as is the case of BOW and n-grams, are not suitable to analyze cross-topic patterns, and they may suffer for a significant decrease of efficacy. Indeed, different topics are virtually described by different terms. Terms related to different domains can be very different, making the comparison between examples difficult. On the other hand, structural representations, such as the linguistic features, POS tagging, or the occurrences of function words, are more general and content independent, and they can be applied to catch cross-domain dependencies and patterns.

In the cross-topic evaluation, four topics have been selected for training the machine learning model, whereas the fifth has been used as a test set for each language separately. The hyper-parameters have been selected with a 5-fold cross-validation, as is the previous case. This procedure has been applied for each task in rotation. Results of the comparison are shown in Table 5.

Cross-language evaluation

Structure-based representations can be able to reduce the aforementioned cross-topic issue, by catching high-level context-independent structural features. However, different languages may have different grammar structures, for which the transfer of such features could not be able to catch deception-related patterns. Considering the above limitation, only the cognitive load related Linguistic Features have been employed in this evaluation. Indeed, these linguistic measures are supposed to catch the cognitive effort linked to the act of deceiving, and they should be theoretically language/topic independent.

The cross-language evaluation has been conducted in two different phases. In the first phase, a topic and their associated Italian opinions have been used as training set. The test set consists of English opinions of the same topic. The validation has been conducted with a 5-fold cross-validation as the case of the previous experiments. Then, the same procedure has been used by learning from English opinions, and testing on Italian opinions. Furthermore, the ROC-AUC metric has been included to better evaluate the effectiveness of the machine learning models. This measure can be helpful in the cross-language experiment, where there is a reasonable drop in performance, which makes difficult to understand if the algorithm is able to learn something from data. Table 6 contains the results of the evaluation.

Results

All feature sets, considered separately, have proven their effectiveness in the within-topic classification, exceeding the 50% baseline performance that could have been obtained by random guessing (Table 4). The highest F_1 has been achieved using the BOW representation, that is 80.90 ±2.07 (IT) and 74.40±3.62 (EN) when using all the data, and 79.80±6.40 (IT) and 72.58±3.90 (EN) on average when focusing on single topics. The principled combination via MKL improves the results significantly, with a score of 82.06±0.6 (IT) and 77.02±0.70 (EN) when considering all the features sets while, excluding BOW, we reached a score of 79.35±2.11 and 71±1.97 for the IT and EN languages respectively.

In the cross-topic assessment depicted in Table 6, all the features considered separately suffered a drop in classification performances. In particular, the BOW representation maintains similar results compared to the within-topic evalua-

Dataset	Topic	BOW	LF	POS	FW	MKL	MKL w/o BOW
IT	Abo	$82.28_{\pm 5.17}$	$63.21_{\pm 6.28}$	$59.81_{\pm 4.77}$	$71.67_{\pm 4.39}$	$82.88_{\pm 7.11}$	$70.97_{\pm 5.13}$
	CL	$75.74_{\pm 6.06}$	$60.75_{\pm 4.09}$	$58.16_{\pm 9.79}$	$59.85_{\pm 5.10}$	$71.47_{\pm 4.21}$	$63.78_{\pm 6.98}$
	Eut	$84.05_{\pm 6.49}$	$68.63_{\pm 0.00}$	$69.78_{\pm 4.42}$	$78.73_{\pm 3.16}$	$80.93_{\pm 3.28}$	$77.09_{\pm 1.76}$
	GM	$87.42_{\pm 2.33}$	$68.63_{\pm 0.00}$	$74.41_{\pm 4.91}$	$78.13_{\pm 4.62}$	$88.17_{\pm 2.43}$	$77.83_{\pm 3.13}$
	PoM	$69.50_{\pm 8.72}$	$63.23_{\pm 4.61}$	$60.40_{\pm 1.99}$	$76.57_{\pm 4.01}$	$74.75_{\pm 3.87}$	$72.64_{\pm 3.59}$
	All	$80.90_{\pm 2.07}$	$60.44_{\pm 1.83}$	$66.31_{\pm 1.43}$	$65.63_{\pm 0.48}$	$82.06_{\pm 2.01}$	$79.35_{\pm 2.11}$
EN	Abo	$67.27_{\pm 2.87}$	$65.90_{\pm 2.61}$	$64.55_{\pm 5.29}$	$65.61_{\pm 4.74}$	$75.59_{\pm 4.00}$	$65.83_{\pm 3.01}$
	CL	$74.96_{\pm 2.66}$	$67.24_{\pm 1.18}$	$71.57_{\pm 4.44}$	$65.74_{\pm 3.88}$	$78.12_{\pm 2.99}$	$71.34_{\pm 5.04}$
	Eut	$74.59_{\pm 2.33}$	$66.26_{\pm 3.23}$	$68.87_{\pm 3.52}$	$64.08_{\pm 6.82}$	$74.57_{\pm 3.36}$	$68.59_{\pm 3.03}$
	GM	$77.45_{\pm 1.98}$	$68.19_{\pm 1.99}$	$70.61_{\pm 4.96}$	$72.49_{\pm 5.53}$	$81.12_{\pm 2.09}$	$73.59_{\pm 3.51}$
	PoM	$68.74_{\pm 3.07}$	$67.86_{\pm 2.65}$	$64.76_{\pm 3.32}$	$67.57_{\pm 5.27}$	$73.01_{\pm 3.11}$	$63.73_{\pm 1.91}$
	All	$74.40_{\pm 3.62}$	$67.33_{\pm 0.46}$	$68.80_{\pm 2.03}$	$69.10_{\pm 1.89}$	$77.02_{\pm 0.70}$	$71.00_{\pm 1.97}$

Table 4: F_1 (%) scores achieved for the within-topic evaluation.

Dataset	Topic	BOW	LF	POS	FW	MKL	MKL w/o BOW
IT	Abo	73.27	55.06	58.47	55.06	74.67	56.94
	CL	69.07	54.64	60.71	53.29	70.20	60.31
	Eut	80.80	55.12	69.23	63.07	81.74	68.57
	GM	75.27	57.07	63.37	65.66	74.49	61.19
	PoM	64.40	58.40	60.05	56.36	63.05	62.82
EN	Abo	56.78	66.37	67.61	67.16	72.61	69.92
	CL	65.58	64.24	62.02	61.74	73.10	65.95
	Eut	68.14	66.95	71.22	69.12	77.34	73.90
	GM	66.97	68.33	65.72	62.53	69.47	68.04
	PoM	50.40	65.14	63.56	54.41	61.06	68.09

Table 5: F_1 (%) scores achieved for the cross-topic evaluation.

	IT→EN		EN→IT	
Topic	F_1	AUC	F_1	AUC
Abo	58.47	67.24	58.42	63.55
CL	49.76	62.00	64.36	61.07
Eut	66.76	59.88	67.61	51.55
GM	66.67	56.60	50.27	55.69
PoM	38.76	61.18	64.77	60.86
All	64.52	58.48	66.10	65.99

Table 6: Cross-language evaluation using Linguistic Features. Training→Test.

tion for the Italian language while significantly decreasing for the English one. The LF feature set showed the opposite trend compared to the BOW performance, leading to higher performances on the English language compared to Italian. On the other hand, both the POS and FW representations maintain similar performances in both languages. Again, the MKL combination improves the results w.r.t. individual representations with a slight decrease in performances when excluding BOW.

As stated before, only the linguistic markers of deceit linked to the cognitive load hypothesis have been analysed in the cross-language evaluation. Results show F_1 scores of 64.52 and 66.10 when testing on the English and Italian complete datasets respectively (Table 6). The ROC-AUC shows that the algorithm can effectively learn patterns from data, exceeding the random guessing performance.

Discussion and conclusions

In the present study, the aim was to investigate the automatic detection of deception across different languages, a research question that received little attention and led to contrasting results so far. After compiling a multilingual (English and Italian) corpus of both truthful and deceitful first-person opinions regarding five different topics, we assessed the individual and collective performance of four features sets in detecting deceit, comparing results obtained on the two languages. The employed features sets included both theory-driven linguistic markers of deceit (LF) and standard linguistic measures for text categorization (BOW, POS and FW). The experimental set up included within-topic, cross-topic and cross-language binary classification experiments.

The first result to be highlighted refers to the computation on both the languages of the cognitive load related linguistic features: three out of four linguistic markers of deceit showed to be in line with the expected trend according to previous studies (Hauch et al., 2015). Indeed, except for the type-token ratio, deceitful narratives appear to be characterised by fewer words, fewer exclusive words and lesser average sentence length compared to the truthful ones in both the Italian and English language. This finding provides some support to the effectiveness across languages of the cognitive load hypothesis, suggesting that the mental effort in deceiving produces useful verbal cues in spotting deception regardless of the examined language.

About the three automatic classification tasks, in the within-topic and cross-topic experiments, all the feature sets demonstrated their usefulness in the correct classification of deceitful and truthful narratives. Indeed, both their individual and collective performance exceeded the 50% random guessing performance, where the MKL combination achieved the best results. However, while in the within-topic setting all the features considered separately led to comparable performances between languages, the BOW and LF text representations performed better in one language compared to the other in cross-topic experiments. Finally, the cross-language examination showed that it's possible to build an automatic classifier for detecting deceit regardless of the language under investigation. Indeed, the linguistic markers of deception related to the cognitive load hypothesis showed to be able to transfer the deception-related patterns from one language to the other.

Concluding, It's worth to notice that our results highlighted some differences in classification performance between the two languages. A possible explanation could be that the text representation derived from the cognitive load approach does not consider the cross-language variability in the occurrences of that specific class of words. In future studies, we aim to strengthen our models by considering this variability and assessing its impact on classification performances.

References

Aiolli, F., & Donini, M. (2015). Easymkl: a scalable multiple kernel learning algorithm. *Neurocomputing*, *169*, 215–224.

Chung, C., & Pennebaker, J. W. (2007). The psychological functions of function words. *Social communication*, *1*, 343–359.

Da Silva, C. S., & Leach, A.-M. (2013). Detecting deception in second-language speakers. *Legal and criminological psychology*, *18*(1), 115–127.

Difallah, D., Filatova, E., & Ipeirotis, P. (2018). Demographics and dynamics of mechanical turk workers. In *Proceedings of the eleventh acm international conference on web search and data mining* (pp. 135–143).

Fitzpatrick, E., Bachenko, J., & Fornaciari, T. (2015). Automatic detection of verbal deception. *Synthesis Lectures on Human Language Technologies*, *8*(3), 1–119.

Fusilier, D. H., Montes-y Gómez, M., Rosso, P., & Cabrera, R. G. (2015). Detecting positive and negative deceptive opinions using pu-learning. *Information processing & management*, *51*(4), 433–443.

Giebels, E., & Taylor, P. J. (2009). Interaction patterns in crisis negotiations: Persuasive arguments and cultural differences. *Journal of Applied Psychology*, *94*(1), 5.

Gönen, M., & Alpaydın, E. (2011). Multiple kernel learning algorithms. *Journal of machine learning research*, *12*(Jul), 2211–2268.

Hauch, V., Blandón-Gitlin, I., Masip, J., & Sporer, S. L. (2015). Are computers effective lie detectors? a meta-analysis of linguistic cues to deception. *Personality and Social Psychology Review*, *19*(4), 307–342.

Kestemont, M. (2014). Function words in authorship attribution. from black magic to theory? In *Proceedings of the 3rd workshop on computational linguistics for literature (clfl)* (pp. 59–66).

Kleinberg, B., Mozes, M., Arntz, A., & Verschuere, B. (2018). Using named entities for computer-automated verbal deception detection. *Journal of forensic sciences*, *63*(3), 714–723.

Krishnamurthy, G., Majumder, N., Poria, S., & Cambria, E. (2018). A deep learning approach for multimodal deception detection. *arXiv preprint arXiv:1803.00344*.

Leal, S., Vrij, A., Vernham, Z., Dalton, G., Jupe, L., Harvey, A., & Nahari, G. (2018). Cross-cultural verbal deception. *Legal and Criminological Psychology*, *23*(2), 192–213.

Levitan, S. I., Maredia, A., & Hirschberg, J. (2018). Linguistic cues to deception and perceived deception in interview dialogues. In *Proceedings of the 2018 conference of the north american chapter of the association for computational linguistics* (pp. 1941–1950).

Matsumoto, D., & Hwang, H. C. (2015). Differences in word usage by truth tellers and liars in written statements and an investigative interview after a mock crime. *Journal of Investigative Psychology and Offender Profiling*, *12*(2), 199–216.

Matsumoto, D., Hwang, H. C., & Sandoval, V. A. (2015a). Cross-language applicability of linguistic features associated with veracity and deception. *Journal of Police and Criminal Psychology*, *30*(4), 229–241.

Matsumoto, D., Hwang, H. C., & Sandoval, V. A. (2015b). Ethnic similarities and differences in linguistic indicators of veracity and lying in a moderately high stakes scenario. *Journal of Police and Criminal Psychology*, *30*(1), 15–26.

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and social psychology bulletin*, *29*(5), 665–675.

Nunamaker, J. F., Burgoon, J. K., Twyman, N. W., Proudfoot, J. G., Schuetzler, R., & Giboney, J. S. (2012). Establishing a foundation for automated human credibility screening. In *IEEE international conference on intelligence and security informatics* (pp. 202–211).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pérez-Rosas, V., & Mihalcea, R. (2014). Cross-cultural deception detection. In *Proceedings of the 52nd annual meet-*

ing of the association for computational linguistics (volume 2: Short papers) (Vol. 2, pp. 440–445).

Potapova, R., & Lykova, O. (2016). Verbal representation of lies in russian and anglo-american cultures. *Procedia-Social and Behavioral Sciences*, *236*, 114–118.

Rungruangthum, M., & Todd, R. W. (2017). Differences in language used by deceivers and truth-tellers in thai online chat. *Journal of the Southeast Asian Linguistics Society (JSEALS)*, *10*(2).

Serota, K. B., Levine, T. R., & Boster, F. J. (2010). The prevalence of lying in america: Three studies of self-reported lies. *Human Communication Research*, *36*(1).

Spence, K., Villar, G., & Arciuli, J. (2012). Markers of deception in italian speech. *Frontiers in psychology*, *3*, 453.

Vrij, A., Fisher, R. P., & Blank, H. (2017). A cognitive approach to lie detection: A meta-analysis. *Legal and Criminological Psychology*, *22*(1), 1–21.

Waite, M. (2009). *Oxford thesaurus of english*. Oxford University Press.

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In *Advances in experimental social psychology* (Vol. 14, pp. 1–59). Elsevier.