Deception Detection Within and Across Cultures

Veronica Perez-Rosas and Cristian Bologa and Mihai Burzo and Rada Mihalcea

Abstract In this paper, we address the task of cross-cultural deception detection. Using crowdsourcing, we collect four deception datasets, two in English (one originating from United States and one from India), one from Romanian speakers, and one in Spanish obtained from speakers from Mexico, covering three predetermined topics. We also collect two additional datasets, one for English from United States and one for Romanian, where the topic is not pre-specified. We run comparative experiments to evaluate the accuracies of deception classifiers built for each culture, and also to analyze classification differences within and across cultures. Our results show that we can leverage cross-cultural information, either through translation or equivalent semantic categories, and build deception classifiers with a performance ranging between 60-70%.

1 Introduction

The identification of deceptive behavior is a task that has gained increasing interest from researchers in computational linguistics. This is mainly motivated by the rapid growth of deception in written sources, and in particular in Web content, including product reviews, online dating profiles, and social networks posts [10].

Veronica Perez-Rosas
University of North Texas e-mail: veronicaperezrosas@my.unt.edu

Cristian Bologa
Universitate Babes-Bolyai e-mail: cristian.bologa@econ.ubbcluj.ro

Mihai Burzo
University of Michigan-Flint e-mail: mburzo@umich.edu

Rada Mihalcea
University of Michigan e-mail: mihalcea@umich.edu

To date, most of the work presented on deception detection has focused on the identification of deceit clues within a specific language, where English is the most commonly studied language. However, a large portion of the written communication (e.g., e-mail, chats, forums, blogs, social networks) occurs not only between speakers of English, but also between speakers from other cultural backgrounds, which poses important questions regarding the applicability of existing deception tools. Issues such as language, beliefs, and moral values may influence the way people deceive, and therefore may have implications on the construction of tools for deception detection.

In this paper, we explore within- and across-culture deception detection for four different cultures, namely United States, India, Romania, and Mexico. Through several experiments, we compare the performance of classifiers that are built separately for each culture, and classifiers that are applied across cultures, by using unigrams and word categories that can act as a cross-lingual bridge. Our results show that we can achieve accuracies in the range of 60-70%, and that we can leverage resources available in one language to build deception tools for another language.

1.1 Related work

Research to date on automatic deceit detection has explored a wide range of applications such as the identification of spam in e-mail communication, the detection of deceitful opinions in review websites, and the identification of deceptive behavior in computer-mediated communication including chats, blogs, forums and online dating sites [11, 16, 10, 15, 19].

Techniques used for deception detection frequently include word-based stylometric analysis. Linguistic clues such as n-grams, count of used words and sentences, word diversity, and self-references are also commonly used to identify deception markers. An important resource that has been used to represent semantic information for the deception task is the Linguistic Inquiry and Word Count (LIWC) dictionary [12]. LIWC provides words grouped into semantic categories relevant to psychological processes, which have been used successfully to perform linguistic profiling of true tellers and liars [20, 9, 14]. In addition to this, features derived from syntactic Context Free Grammar parse trees, and part of speech have also been found to aid the deceit detection [3, 17].

While most of the studies have focused on English, there is a growing interest in studying deception for other languages. For instance, Fornaciari and Poesio [5] identified deception in Italian by analyzing court cases. The authors explored several strategies for identifying deceptive clues, such as utterance length, LIWC features, lemmas and part of speech patterns. Almela et al. [1] studied the deception detection in Spanish text by using SVM classifiers and linguistic categories, obtained from the Spanish version of the LIWC dictionary. A study on Chinese deception is presented in [18], where the authors built a deceptive dataset using Internet news and performed machine learning experiments using a bag-of-words representation to train a classifier able to discriminate between deceptive and truthful cases.

It is also worth mentioning the work conducted to analyze cross-cultural differences. Lewis and George [7] presented a study of deception in social networks sites and face-to-face communication, where authors compare deceptive behavior of Korean and American participants, with a subsequent study also considering the differences between Spanish and American participants [6].

At difference from us, both studies analyze cultural differences using a statistical approach, where data was collected by interviewing participants and principal component analysis was applied to identify cultural aspects related with deception such as liars topic's choice, and gender differences. In this study we rely on machine learning techniques to build deception classifiers from written statements provided by true tellers and deceivers.

In general, related research findings suggest a strong relation between deception and cultural aspects, which are worth exploring with automatic methods.

2 Datasets

We collect four datasets for four different cultures: United States (English-US), India (English-India), Romania, and Mexico (Spanish-Mexico). Following [8], we collect short deceptive and truthful essays for three topics: opinions on Abortion, opinions on Death Penalty, and feelings about a Best Friend.

To collect both truthful and deceptive statements for the Abortion and Death Penalty topics we first instructed the participants to think they were participating in a debate, where they were asked to provide their truthful opinion about the topic. Secondly, we asked them to imagine a debate where they had to provide an opposite view from what they truly believed, thus generating false statements about the topic being discussed. In both cases, we asked them to provide plausible details and to be as convincing as possible. For the Best Friend topic, we collected the deceptive and truthful essays by first asking participants to provide a description of their best friend, and second asking them to describe someone they disliked as though he/she were their best friend.

In order to collect the English-US and English-India datasets, we used Amazon Mechanical Turk with a location restriction, so that all the contributors are from the country of interest (US and India). We collected 100 deceptive and 100 truthful statements for each of the three topics. To avoid spam, each contribution was manually verified by one of the authors of this paper.

For Spanish-Mexico, while we initially attempted to collect data also using Mechanical Turk, we were not able to receive enough contributions. We therefore created a separate web interface to collect data, and recruited participants through contacts of the paper's authors. The overall process was significantly more time consuming than for the other two cultures, and resulted in fewer contributions as shown in Table 1.

For the Romanian dataset we also used a separate web interface and participants were recruited through contacts of one of the paper's authors. Since participants were allowed to end their participation at any time, the final process resulted in a different number of contributions per each topic as shown in Table 1.

Table 1 Dataset distributions for four deception datasets

Topic	EnglishUS		EnglishIN		Romanian		Spanish	
	D	T	D	T	D	T	D	T
Abortion	100	100	100	100	139	139	39	39
Best Friend	100	100	100	100	151	151	42	42
Death Penalty	100	100	100	100	145	145	94	94

For all four cultures, the participants first provided their truthful responses, followed by the deceptive ones. Also, all contributors provided their responses for different topics in the same topic order: Abortion, Best Friend, and Death Penalty.

Table 2 shows sample statements from each dataset. Also, word count distributions for the four datasets are shown in Table 3. Interestingly, for all four cultures, the average number of words for the deceptive statements is significantly smaller than for the truthful statements, which may be explained by the added difficulty of the deceptive process, and is in line with previous observations about the cues of deception [2].

3 Experiments

Through our experiments, we seek answers to the following questions. First, what is the performance for deception classifiers built for different cultures? Second, can we use information drawn from one culture to build a deception classifier for another culture? Finally, what are the psycholinguistic classes most strongly associated with deception/truth, and are there commonalities or differences among languages?

In all our experiments, we formulate the deception detection task in a machine-learning framework, where we use an SVM classifier to discriminate between deceptive and truthful statements. [1]

[1] We use the SVM classifier implemented in the Weka toolkit, with its default settings.

Table 2 Sample stataments from four deception datasets

EnglishUS		
Topic	Deceptive	Truthful
Abortion	Abortion should not be an acceptable practice, ever. Precluding the life of an unborn child is dominating and nullifying their inalienable right to live ...	Abortion should be a legal option for pregnant mothers. Of course, it needs to be very early in the pregnancy and the mother must give significant ...
BestFriend	”John” Is a great person. John always puts himself before others. John never says derogatory remarks to people.	My best friend, we will call him ”Bob” is a truly exceptional person. I can talk to Bob about anything and everything.
DeathPenalty	Life is sacred. Who are we to end a life? People, even criminals, deserve to live. They deserve a second chance.	Sometimes, there are those who commit crimes so heinous that there is only one appropriate punishment.
English India		
Topic	Deceptive	Truthful
Abortion	I think abortion is needed. It should be done, if the life of the mother is in risk. It should also be done in other necessary circumstances. Abortion should ...	In my opinion, abortion is very cruel. It is another form of murder. We have no right to end the life of an innocent child. So, abortion should be banned.
BestFriend	He is one of the best people I have met in my life. He has never troubled be in any way. At work, he never competes with me. I ”hope” we remain friends ...	He is my best friend in my life. He helped me in all my downs in my life as guiding and gives suggestions. He can understand me as anyone can and
DeathPenalty	I disagree the act death penalty. No one has the rights to take the life of a human except God. Instead of death penalty...	Yes, of course I support death penalty. Only fear from death would prevent these crimes. In this modern era crime...
Spanish Mexico (Translated)		
Topic	Deceptive	Truthful
Abortion	Abortion is a legal thing.it needs to be appreciated in all the way. People should be encouraged to do an abortion.	Abortion is very cruel thing for all humans in the earth. Abortion is a big sin before God.
BestFriend	My best friend is very nice. I love spending time with her. We have always get along very well and we like each ...	My best friend always listen to me. We have a lot of things in common. We always find time to talk to each other.
DeathPenalty	Death penalty should be applied in all countries without mercy. Criminals should pay for what they have done	I think we should not decide about the life of another human being. The only one who can make such decision is ...
Romanian (Translated)		
Topic	Deceptive	Truthful
Abortion	I do not agree with abortion under any circumstances (or in exceptional cases, any request) because it is not moral ...	Abortion can help women to avoid giving birth a child that could affect their life’s. If a woman decides she does ...
BestFriend	This person give me a sense of confidence, always coming up with new ideas that I like. Always supports ...	My best friend knows me very well. He knows when I’m upset and something goes wrong. We got along ...
DeathPenalty	The death penalty is very brutal and should not take place in a civilized world. Although they are murderers...	I think the death penalty is the correct one because criminals do not think about the lives of others when they...

Table 3 Word count distribution between deceptive (D) and truthful (T) statements and average number of words per statement for four deception datasets

Topic	EnglishUS		EnglishIN		Romanian		Spanish	
	D	T	D	T	D	T	D	T
Abortion	52	72	64	76	68	91	76	106
Best Friend	51	64	67	75	65	89	60	87
Death Penalty	56	68	74	85	70	92	63	97
Average	53	68	69	78	68	90	66	97

3.1 What is the performance for deception classifiers built for different cultures?

We represent the deceptive and truthful statements using two different sets of features. First we use unigrams obtained from the statements corresponding to each topic and each culture. To select the unigrams, we use a threshold of 10, where all the unigrams with a frequency less than 10 are dropped. We choose this threshold due their best performance in the reported experiments. Also, since previous research suggested that stopwords can contain linguistic clues for deception, no stopword removal is performed.

Experiments are performed using a ten-fold cross validation evaluation on each dataset. Using the same unigram features, we also perform cross-topic classification, so that we can better understand the topic dependence. For this, we train the SVM classifier on training data consisting of a merge of two topics (e.g., Abortion + Best Friend) and test on the third topic (e.g., Death Penalty). The results for both within- and cross-topic are shown in the last two columns of Table 4.

Second, we use the LIWC lexicon to extract features corresponding to several word classes. LIWC was developed as a resource for psycholinguistic analysis [12]. The 2001 version of LIWC includes about 2,200 words and word stems grouped into about 70 classes relevant to psychological processes (e.g., emotion, cognition), which in turn are grouped into four broad categories[2] namely: linguistic processes, psychological processes, relativity, and personal concerns. We also used a Spanish version of the LIWC lexicon [13] as well as a Romanian version [4]. A feature is generated for each of the 70 word classes by counting the total frequency of the words belonging to that class. The resulting features are then grouped into four different sets containing the LIWC classes subset corresponding to each of the four broad categories. We perform separate evaluations using each of the feature sets derived from broad LIWC categories, as well as using all the categories together. The accuracy classification results obtained with the SVM classifier are shown in Table 4.

[2] http://www.liwc.net/descriptiontable1.php

Table 4 Within-culture classification, using LIWC word classes and unigrams. For LIWC, results are shown for within-topic experiments, with ten-fold cross validation. For unigrams, both within-topic (ten-fold cross validation on the same topic) and cross-topic (training on two topics and testing on the third topic) results are reported.

Topic	LIWC					Unigrams	
	Linguistic	Psychological	Relativity	Personal	All	Within-topic	Cross-topic
English-US							
Abortion	72.50%	68.75%	44.37%	67.50%	73.03%	63.75%	80.36%
Best Friend	75.98%	68.62%	58.33%	54.41%	73.03%	74.50%	60.78%
Death Penalty	60.36%	54.50%	49.54%	50.45%	58.10%	58.10%	77.23%
Average	69.61%	63.96%	50.75%	57.45%	69.05%	65.45%	72.79%
English-India							
Abortion	56.00%	48.50%	46.50%	48.50%	56.00%	46.00%	50.00%
Best Friend	68.18%	68.62%	54.55%	53.18%	71.36%	60.45%	57.23%
Death Penalty	56.00%	52.84%	57.50%	53.50%	63.50%	57.50%	54.00%
Average	60.06%	59.19%	52.84%	51.72%	63.62%	54.65%	53.74%
Spanish-Mexico							
Abortion	73.17%	67.07%	48.78%	51.22%	62.20%	52.46%	57.69%
Best Friend	72.04%	74.19%	67.20%	54.30%	75.27%	66.66%	50.53%
Death Penalty	73.17%	67.07%	48.78%	51.22%	62.20%	54.87%	63.41%
Average	72.79%	69.45%	54.92%	52.25%	67.89%	57.99%	57.21%
Romanian							
Abortion	61.87%	64.02%	64.02%	62.58%	63.30%	65.10%	58.99%
Best Friend	70.19%	68.21%	68.21%	68.54%	67.54%	68.80%	54.30%
Death Penalty	64.13%	66.55%	66.55%	64.48%	65.51%	63.79%	57.27%
Average	65.39%	66.26%	66.26%	65.20%	65.45%	65.89%	56.85%

Table 5 Cross-cultural experiments using LIWC categories and unigrams

Topic	Linguistic	Psychological	Relativity	Personal	All LIWC	Unigrams
Training: English-US Test: English-India						
Abortion	58.00%	51.00%	48.50%	51.50%	52.25%	57.89%
Best Friend	66.36%	47.27%	48.64%	50.45%	59.54%	51.00%
Death Penalty	54.50%	50.50%	50.00%	48.50%	53.5%	59.00%
Average	59.62%	49.59%	49.05%	50.15%	55.10%	55.96%
Training: English-US Test: Spanish-Mexico						
Abortion	70.51%	46.15%	50.00%	52.56%	53.85%	61.53%
Best Friend	69.35%	52.69%	51.08%	46.77%	67.74%	65.03%
Death Penalty	54.88%	54.88%	53.66%	50.00%	62.19%	59.75%
Average	64.92%	51.24%	51.58%	49.78%	61.26%	62.10%
Training: English-US Test: Romanian						
Abortion	61.15%	55.04%	56.47%	48.2%	57.19%	56.47%
Best Friend	64.56%	50.66%	63.90%	51.55%	52.98%	66.22%
Death Penalty	61.72%	48.96%	64.13%	47.93%	58.27%	60.34%

Overall, the results show that it is possible to discriminate between deceptive and truthful cases using machine learning classifiers, with a performance superior to a random baseline which for all datasets is 50% given an even class distribution. Considering the unigram results, among the four cultures, the deception discrimination works best for the English-US dataset, and this is also the dataset that benefits most from the larger amount of training data brought by the cross-topic experiments. In general, the cross-topic evaluations suggest that there is no high topic dependence in this task, and that using deception data from different topics can lead to results that are comparable to the within-topic data. An exception to this trend is the Romanian dataset, where the cross-topic experiments lead to significantly lower results than the within-topic evaluations, which may be partly explained by the high lexicalization of Romanian. Interestingly, among the three topics considered, the Best Friend topic has consistently the highest within-topic performance, which may be explained by the more personal nature of the topic, which can lead to clues that are useful for the detection of deception (e.g., references to the self or personal relationships).

Regarding the LIWC classifiers, the results show that the use of the LIWC classes can lead to performance that is generally better than the one obtained with the unigram classifiers. The explicit categorization of words into psycholinguistic classes seems to be particularly useful for the languages where the words by themselves did not lead to very good classification accuracies. Among the four broad LIWC categories, the linguistic category appears to lead to the best performance as compared to the other categories. It is notable that in Spanish, the linguistic category by itself provides results that are better than when all the LIWC classes are used, which may be due to the fact that Spanish has more explicit lexicalization for clues that may be relevant to deception (e.g., verb tenses, formality).

Concerning the specific accuracy for the deception class, we analyzed detailed accuracies per class, obtained by the best classifier from Table 4, which is the one built using only the Linguistic category from LIWC. Table 6 shows the precision, recall, and F-measure metrics obtained for the deceptive and truthful classes obtained by the classifier for each culture. From this table we can observe that for Spanish as well as for both English cultures, the identification of deceptive instances is slightly easier than the identification of truthful statements. For Romanian instead, the truthful instances are more accurately predicted than the deceptive ones. We further analyzed differences in word usage among true tellers and liars in each culture in Section 3.3.

3.2 *Can we use information drawn from one culture to build a deception classifier in another culture?*

In the next set of experiments, we explore the detection of deception using training data originating from a different culture. As with the within-culture experiments, we use unigrams and LIWC features. For consistency across the experiments, given

Table 6 Classification accuracy per class for Linguistic category classifier

Topic	Precision	Recall	F-measure	Class
		English US		
Abortion	0.73	0.71	0.72	Deceptive
	0.72	0.73	0.72	Truthful
BestFriend	0.74	0.79	0.76	Deceptive
	0.77	0.72	0.75	Truthful
Death Penalty	0.60	0.58	0.59	Deceptive
	0.60	0.62	0.61	Truthful
		English India		
Abortion	0.55	0.59	0.57	Deceptive
	0.56	0.53	0.54	Truthful
BestFriend	0.68	0.68	0.68	Deceptive
	0.68	0.68	0.68	Truthful
Death Penalty	0.55	0.58	0.56	deceptive
	0.56	0.54	0.55	Truthful
		Spanish		
Abortion	0.73	0.73	0.73	Deceptive
	0.73	0.73	0.73	Truthful
BestFriend	0.69	0.77	0.73	Deceptive
	0.75	0.67	0.70	Truthful
Death Penalty	0.73	0.73	0.73	Deceptive
	0.73	0.73	0.73	Truthful
		Romanian		
Abortion	0.66	0.55	0.60	Deceptive
	0.61	0.71	0.66	Truthful
BestFriend	0.66	0.61	0.63	Deceptive
	0.64	0.68	0.66	Truthful
Death Penalty	0.65	0.70	0.67	Deceptive
	0.67	0.62	0.65	Truthful

that the size of the Spanish and the Romanian datasets is different compared to the two English datasets, we always train on the English-US dataset.

To enable the unigram based experiments, we translate the two English datasets into either Spanish or Romanian by using the Bing API for automatic translation.[3] As before, we extract and keep only the unigrams with frequency greater or equal to 10. The results obtained in these cross-cultural experiments are shown in the last column of Table 5.

In a second set of experiments, we use the LIWC word classes as a bridge between languages. First, each deceptive or truthful statement is represented using features based on the LIWC word classes grouped into four broad categories: linguistic process, physiological process, relativity, and personal concerns. Next, since the same word classes are used in all three LIWC lexicons, this LIWC-based representation is independent of language, and therefore can be used to perform cross-cultural experiments. Table 5 shows the results obtained with each of the four broad LIWC categories, as well as with all the LIWC word classes.

[3] http://www.bing.com/dev/en-us/dev-center

Table 7 Top ranked LIWC classes for each culture, along with sample words

Class	Score	Sample words	Class	Score	Sample words
English-US					
Deceptive			Truthful		
Metaph	1.77	Die, died, hell, sin, lord	Friends	0.46	Buddies, friend
Other	1.46	He, her, herself, him	We	0.55	Our, ourselves, us, we,
You	1.41	Thou, you	Self	0.55	myself, our, ourselves, us
Humans	1.22	Baby, human, person	Optimism	0.65	accept, hope, top, best
Othref	1.18	He, her, herself, him	I	0.66	I, me, my, myself,
Negemo	1.18	Afraid, agony, awful, bad	Insight	0.68	Accept, believe, understand
English-India					
Deceptive			Truthful		
Negate	1.49	Cannot, neither, no, none	Friends	0.46	Buddies, companion, friend, pal
Physical	1.46	Heart, ill, love, loved,	We	0.55	Our, ourselves, us, we
Future	1.42	Be, may, might, will	Self	0.55	I, me, mine, my, myself
Negemo	1.37	Afraid, agony, alone, bad,	Optimism	0.65	Accept, accepts, best, bold,
Other	1.17	He, she, himself, herself	I	0.66	I, me, mine, my
Humans	1.08	Adult, baby, children, human	Past	0.78	Happened, helped, liked, listened
Spanish-Mexico					
Deceptive			Truthful		
Certain	1.47	Fiel(loyal), jamás (never)	School	0.32	Consejo(advice), estudiar(study)
Humans	1.28	Bebé(baby), persona(person)	Past	0.32	Compartimos(share), vivimos(lived)
You	1.26	Eres(are),estas(be), su(his/her)	Friends	0.37	Amigo/amiga(friend), amistad(friendship)
Negate	1.25	Jamás(never), tampoco(neither)	We	0.58	Estamos(are),somos(be), tenemos(have)
Other	1.22	Es(is), esta(are), otro(other)	Self	0.65	Conmigo(me), tengo(have), soy(am)
Othref	1.11	Eres(are),tiene(have), tuvo(had)	Optimism	0.66	Aceptar(accept), alegre(cheerfully)
Romanian					
Deceptive			Truthful		
Money	2.31	Bani(money), pret(price)	We	0.65	Ne(us,ourselves), noi(we), noastra(our)
Posfeel	1.95	Fericita(happy), zambetul(smile)	Religion	0.72	Cer(heaven), dumnezeu (god), suflet(soul)
Other	1.42	Ei/ele(they), insusi(oneself)	Family	0.73	Tata(dad),mamica(mother), familie(family)
Pronoun	1.34	Ei/le(they), ii(him), va(yourself)	Time	0.77	Oricand(always), momentul(time)
Optimism	1.29	Increderea(confidence), usoara(easy)	Past	0.80	Intalnit(met), ajutat(helped), traiasca(live)
Anx	1.23	Frica(fear), emotionala(emotional)	Friends	0.79	Prietenie(friendship), prieten(friend)

Note that we also attempted to combine unigrams and LIWC features. However, in most cases, no improvements were noticed with respect to the use of unigrams or LIWC features alone.

These cross-cultural evaluations lead to several findings. First, we can use data from a culture to build deception classifiers for another culture, with performance figures better than the random baseline, but weaker than the results obtained with within-culture data. An important finding is that LIWC can be effectively used as a bridge for cross-cultural classification, with results that are comparable to the use of unigrams, which suggests that such specialized lexicons can be used for cross-cultural or cross-lingual classification. Moreover, using only the linguistic category from LIWC brings additional improvements, with absolute improvements of 2-4% over the use of unigrams. This is an encouraging result, as it implies that a semantic bridge such as LIWC can be effectively used to classify deception data in other languages, instead of using the more costly and time consuming unigram method based on translations.

3.3 *What are the psycholinguistic classes most strongly associated with deception/truth?*

The final question we address is concerned with the LIWC classes that are dominant in deceptive and truthful text for different cultures. We use the method presented in [8], which consists of a metric that measures the saliency of LIWC classes in deceptive versus truthful data. Following their strategy, we first create a corpus of deceptive and truthful text using a mix of all the topics in each culture. We then calculate the dominance for each LIWC class, and rank the classes in reversed order of their dominance score. Table 7 shows the most salient classes for each culture, along with sample words.

This analysis shows some interesting patterns. There are several classes that are shared among the cultures. For instance, the deceivers in all cultures make use of negation, negative emotions, and references to others. Second, true tellers use more optimism and friendship words, as well as references to themselves. An interesting finding is the use of the Religion and Family classes by Romanian true-tellers, which seems to be very related to cultural background, as religion is an important cultural component. In contrast with the other cultures, Romanian speakers use more positive feeling (Posfeel) and Optimism related words when expressing deceptive statements.

These results are in line with previous research, which showed that LIWC word classes exhibit similar trends when distinguishing between deceptive and non-deceptive text [9]. Moreover, there are also word classes that only appear in some of the cultures; for example, time classes (Past, Future) appear in English-India and Spanish-Mexico, but not in English-US, which in turn contains other classes such as Insight and Metaph.

4 Deception detection using short sentences

One limitation of the experiments presented in the previous section is that they all rely on domain-specific datasets, which may bias the deception detection. To address this potential concern, as a final experiment, we explore the detection of deception in a less-constrained environment, where the topic of the deceptive statements is not set apriori.

We collect and experiment with two datasets consisting of short open-domain truths and lies, contributed by speakers of English-US and Romanian.

For English, we set up a Mechanical Turk task where we asked workers to provide seven lies and seven truths, each consisting of one sentence, on topics of their choice. For Romanian, we designed a web interface to collect data, and recruited participants through contacts of the paper's authors. Romanian speakers were asked to provide five truths and five lies, again on topics of their choice. In both cases, the participants were asked to provide plausible lies and avoid non-commonsensical statements such as "A dog can fly." In addition to the one-sentence truths and lies, we also collect demographic data for the contributors, such as gender, age, and education level. The class distribution for these datasets is shown in Table 8.

Table 8 Class distribution for the Romanian and English-US open-domain deception datasets

Language	Contributors	Male	Female	Truths	Lies	Total
English	512	214	298	3584	3584	7168
Romanian	136	35	101	680	680	1360

Similar to the domain-specific experiments, for these open-domain datasets we run within- and across culture experiments. Table 9 shows the results of the deception classification experiments run separately on the English and Romanian datasets, whereas Table 10 shows the results obtained in the cross-cultural experiments.

Table 9 Within-culture classification, using LIWC word classes and unigrams. Results are obtained using ten-fold cross validation.

Language	Linguistic	Psychological	Relativity	Personal	All LIWC	Unigrams
English	52.01%	52.92%	51.92%	50.33%	56.86%	58.33%
Romanian	56.76%	50.22%	52.35%	50.66%	55.29%	57.86%

Table 10 Cross-cultural experiments using LIWC categories and unigrams

Training: English-US Test: Romanian					
Linguistic	Psychological	Relativity	Personal	All LIWC	Unigrams
56.25%	51.69%	51.69%	50.07%	56.91%	59.70%

Not surprisingly, the accuracy of the deception detection method on the open-domain data is below the accuracy obtained on the domain-specific datasets. In addition to the domain-specific/no-domain difference, this drop in accuracy can also be attributed to the fact that the open-domain data consists of short sentences rather than full paragraphs, which could also further explain why using the LIWC derived features does not lead to noticeable improvements over the use of unigrams.

A similar trend is observed in the cross-culture experiments reported in Table 10, where unigrams outperform the use of LIWC classes. It is important to note however, that the use of linguistic classes is still preferable over the use of unigrams, with a rather small accuracy drop of only 2.79% over the use of costly and more time consuming translations.

To further analyze the nature of the lying process in the open-domain datasets, we obtained the psycholinguistic classes most strongly associated with deception

Table 11 Top ranked LIWC classes for English and Romanian, along with sample words

Class	Score	Sample words	Class	Score	Sample words
English-US					
Deceptive			Truthful		
Certain,	1.93	Completely, all, never, always	Sleep	0.87	Bed, tires, sleeps, wake, dream, asleep
Negate	1.79	Can't, cannot, not, without, nothing	Incl	0.86	Here, include, into, together, also, too
Anger	1.64	Fight, destruction, poisonous, lied	Posemo	0.84	Richest, enjoyed, fun, better, trust, honest
Down	1.42	Under, off, bottom, lowest, down	Relig	0.65	Church, minister, religion, faith, religious
Motion	1.41	Fly, take, traveled, ran, walk	Posfeel	0.73	Agrees, enjoy(ed), care, love(ed),happy
Money	1.37	Richest, buy, sell, dollars, bank	Music	0.73	Listening, songs, music, sing, song, radio
Friends	1.3	Friend, neighbor,(boy/girl)friend	See	0.74	Vision, see, look(ing), watch, eyes, shows
Otheref	1.35	They, yourself, you, we, someone	Family	0.82	Wife, sister, dad, father, parents, family
Other	1.25	They, he, them, she, himself, him	Tv	0.79	Film, channel, movie, tv, show, television
Romanian					
Deceptive			Truthful		
Negate	2.24	deloc,niciodata,nimic,fara,nu Not at all, nothing, without, not	Motion	0.62	Intregul,alergat,iei,fugit,intr,vizita Entire, running, take, ran, in, visit
Eating	1.91	gateste, mancarea, slabire, mancare Cook, food, weakening, food	Cause	0.66	Cum, judecati, reactii, scopul, deoarece Why, judgments, reactions, order, because
Past	1.85	Zbura, fost, invatat, facut, mintit, luat Flee, former, learned, made, lying, taken	We	0.72	Ne, noi, noastra, noua, noastre, nostru Us, we, our, us, our, our
Money	1.80	Cumparat, bogata, monede, bani Bought,rich,coins,money	Posemo	0.72	Fericita, bun, bucuria, fericirea, frumoasa Blessed, good, joy, happiness, beautiful
Anger	1.70	Nebunie, rau, mintit, urasc Madness, evil, lying, hate	Friends	0.74	Colega, fosta, prietena, iubita, prietenii Colleague, former, friend, girlfriend, friends
Senses	1.69	Apuc, simtit, mancat, simti, mananca Grab, felt, ate, feel, eat	Achieve	0.75	Pierd, prima, inainte, succesul, munca Lose, first, before, success, work
Physical	1.63	Trezesc, cap, degete, gata, picioare Walking, head, fingers, ready, feet	Tentav	0.76	Putea, orice, ori, doar, mult, multi, Can, any, and/or, only, much, many
Certain	1.58	Incredere, intotdeauna, niciodata Confidence, always, never	Home	0.76	Apartamentul, casa, familia, traieste, acasa Apartment, home, family, lives, at home
Body	1.51	Picioare, nascut, degete, limba Feet, born, fingers, language	Posfeel	0.78	Fericita, dragi, romantica, place, zambesti Blessed, dear, romantic, like, smile

and truth sentences. The results are presented in Table 11. Interestingly, the analysis confirm our findings for the domain-specific experiments, where shared lying patterns among cultures include the use of negation, negative emotions, and references to others. Furthermore, true-tellers related patterns are also shared among cultures, where the most salient classes are family, positive emotions, and positive feeling.

At the same time, we can observe interesting differences among cultures, for instance the use of the words associated with the classes We and Achieve by the Romanian speakers as indicative of truthful responses. Moreover, unlike the American deceivers, Romanian deceivers use Eating, Senses and Body classes more frequently.

5 Conclusions

In this paper, we addressed the task of deception detection within- and across-cultures. Using four datasets from four different cultures each covering three different topics, as well as two additional datasets from two cultures on free topics, we conducted several experiments to evaluate the accuracy of deception detection when learning from data from the same culture or from a different culture. In our evaluations, we compared the use of unigrams versus the use of psycholinguistic word classes.

The main findings from these experiments are: 1) We can build deception classifiers for different cultures with accuracies ranging between 60-70%, with better performance obtained when using psycholinguistic word classes as compared to simple unigrams; 2) The deception classifiers are not sensitive to different topics, with cross-topic classification experiments leading to results comparable to the within-topic experiments; 3) We can use data originating from one culture to train deception detection classifiers for another culture; the use of psycholinguistic classes as a bridge across languages can be as effective or even more effective than the use of translated unigrams, with the added benefit of making the classification process less costly and less time consuming; 4) Similar findings, although with somehow lower classification results, can be obtained for open-domain short sentence texts in both within- and across-cultures experiments, which confirm the portability of the classification method presented in this paper.

The datasets introduced in this paper are publicly available from http://lit.eecs.umich.edu.

Acknowledgments

This material is based in part upon work supported by National Science Foundation awards #1344257 and #1355633 and by DARPA-BAA-12-47 DEFT grant #12475008. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or the Defense Advanced Research Projects Agency.

References

1. Almela, A., Valencia-García, R., Cantos, P.: Seeing through deception: A computational approach to deceit detection in written communication. In: Proceedings of the Workshop on Computational Approaches to Deception Detection, pp. 15–22. Association for Computational Linguistics, Avignon, France (2012). URL http://www.aclweb.org/anthology/W12-0403
2. DePaulo, B., Lindsay, J., Malone, B., Muhlenbruck, L., Charlton, K., Cooper, H.: Cues to deception. Psychological Bulletin **129**(1) (2003)
3. Feng, S., Banerjee, R., Choi, Y.: Syntactic stylometry for deception detection. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers - Volume 2, ACL '12, pp. 171–175. Association for Computational Linguistics, Stroudsburg, PA, USA (2012). URL http://dl.acm.org/citation.cfm?id=2390665.2390708
4. Fofiu, A.: The romanian version of the liwc2001 dictionary and its application for text analysis with yoshikoder. Studia Universitatis Babes-Bolyai-Sociologia (2), 139–151 (2012)
5. Fornaciari, T., Poesio, M.: Automatic deception detection in italian court cases. Artificial Intelligence and Law **21**(3), 303–340 (2013). DOI 10.1007/s10506-013-9140-4. URL http://dx.doi.org/10.1007/s10506-013-9140-4

6. Lewis, C., George J., G.G.: A cross-cultural comparison of computer-mediated deceptive communication. In: Proceedings of Pacific Asia Conference on Information Systems (2009)
7. Lewis, C., George, J.: Cross-cultural deception in social networking sites and face-to-face communication. Comput. Hum. Behav. **24**(6), 2945–2964 (2008). DOI 10.1016/j.chb.2008.05.002. URL http://dx.doi.org/10.1016/j.chb.2008.05.002
8. Mihalcea, R., Strapparava, C.: The lie detector: Explorations in the automatic recognition of deceptive language. In: Proceedings of the Association for Computational Linguistics (ACL 2009). Singapore (2009)
9. Newman, M., Pennebaker, J., Berry, D., Richards, J.: Lying words: Predicting deception from linguistic styles. Personality and Social Psychology Bulletin **29** (2003)
10. Ott, M., Choi, Y., Cardie, C., Hancock, J.: 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, pp. 309–319. Association for Computational Linguistics, Stroudsburg, PA, USA (2011). URL http://dl.acm.org/citation.cfm?id=2002472.2002512
11. Peng, H., Xiaoling, C., Na, C., Chandramouli, R., Subbalakshmi, P.: Adaptive context modeling for deception detection in emails. In: Proceedings of the 7th international conference on Machine learning and data mining in pattern recognition, MLDM'11, pp. 458–468. Springer-Verlag, Berlin, Heidelberg (2011). URL http://dl.acm.org/citation.cfm?id=2033831.2033870
12. Pennebaker, J., Francis, M.: Linguistic inquiry and word count: LIWC (1999). Erlbaum Publishers
13. Ramírez-Esparza, N., Pennebaker, J.W., García, F.A., Suriá Martínez, R., et al.: La psicología del uso de las palabras: Un programa de computadora que analiza textos en español [the psychology of word use: A computer program that analyzes texts in spanish] pp. 85–99 (2007)
14. Rubin, V.: On deception and deception detection: Content analysis of computer-mediated stated beliefs. Proceedings of the American Society for Information Science and Technology **47**(1), 1–10 (2010). DOI 10.1002/meet.14504701124. URL http://dx.doi.org/10.1002/meet.14504701124
15. Toma, C., Hancock, J.: Reading between the lines: linguistic cues to deception in online dating profiles. In: Proceedings of the 2010 ACM conference on Computer supported cooperative work, CSCW '10, pp. 5–8. ACM, New York, NY, USA (2010). DOI 10.1145/1718918.1718921. URL http://doi.acm.org/10.1145/1718918.1718921
16. Toma, C., Hancock, J., Ellison, N.: Separating fact from fiction: An examination of deceptive self-presentation in online dating profiles. Personality and Social Psychology Bulletin **34**(8), 1023–1036 (2008). DOI 10.1177/0146167208318067. URL http://psp.sagepub.com/content/34/8/1023.abstract
17. Xu, Q., Zhao, H.: Using deep linguistic features for finding deceptive opinion spam. In: Proceedings of COLING 2012: Posters, pp. 1341–1350. The COLING 2012 Organizing Committee, Mumbai, India (2012). URL http://www.aclweb.org/anthology/C12-2131
18. Zhang, H., Wei, S., Tan, H., Zheng, J.: Deception detection based on svm for chinese text in cmc. In: Information Technology: New Generations, 2009. ITNG '09. Sixth International Conference on, pp. 481–486 (2009). DOI 10.1109/ITNG.2009.66
19. Zhou, L., Shi Y.and Zhang, D.: A statistical language modeling approach to online deception detection. IEEE Trans. on Knowl. and Data Eng. **20**(8), 1077–1081 (2008). DOI 10.1109/TKDE.2007.190624. URL http://dx.doi.org/10.1109/TKDE.2007.190624
20. Zhou, L., Twitchell, D., Qin, T., Burgoon, J., Nunamaker, J.: An exploratory study into deception detection in text-based computer-mediated communication. In: Proceedings of the 36th Annual Hawaii International Conference on System Sciences (HICSS'03) - Track1 - Volume 1, HICSS '03, pp. 44.2–. IEEE Computer Society, Washington, DC, USA (2003). URL http://dl.acm.org/citation.cfm?id=820748.821356