International Journal of Knowledge and Language Processing
Volume 9, Number 1, **2018**

www.ijklp.org
KLP International ©2018 ISSN 2191-2734
pp.1—16

Exploring Distinction between Deception and Truth in Chinese: An Analysis Based on Linguistic Features

Yingsu Zhou[1] and Qi Su[2*]

[1]Peking University
No.5, Yiheyuan Road, Haidian District
Beijing, China
1574722549@qq.com

[2]Peking University
No.5, Yiheyuan Road, Haidian District
Beijing, China
sukia@pku.edu.cn

Received May 2018; revised May 2018

ABSTRACT. *Deception is a pervasive psycholinguistic phenomenon in human society. Because of its potential damage and great harm, the identification of deception is very important and has long been an interesting and valuable research subject. In this paper, we build a Chinese deception corpus. Based on this corpus, we examine the effectiveness of several linguistic features from previous researches in differentiating deception from truth. We also employ the linguistic features from the appraisal theory and test if there is any statistical difference between lies and truths in the high-stake corpus in terms of the appraisal resources used. The results suggest that linguistic devices like particles, lacking in English, are proved to be of significantly different distributions in truths and lies in this Chinese deception corpus. In terms of appraisal resources used in lies and truths in high-stake corpus, the results suggest that liars tend to evaluate others' capacity, quote others' words to speak for themselves, express negative attitude and increase the evaluation graduation by using appraisal resources of sharpening. This result also suggests that features of appraisal resources can be a useful kind of verbal cues for future deception detection researches.*

Keywords: deception detection, verbal cues, appraisal theory, deception corpus

*Corresponding author

1. **Introduction.** Deception is a pervasive psycholinguistic phenomenon in human society, which includes a wide range of hyponyms: lies, fabrications, disinformation, etc. In the past few decades, many definitions have been given to describe this complicated concept. According to [1], deception is "an act that is intended to foster in another person a belief or understanding which the deceiver considers false". While [2] takes deception as "a message knowingly transmitted by a sender to foster a false belief or conclusion by the receiver" and approaches it in the context of communication through their interpersonal deception theory (IDT).

Deception can be found in almost every corner of social life, from lies on solemn occasions like legal trials and criminal statements to compliments against your will in small talks. Although some deceptive words can help to ease interpersonal relationship and would hardly cause substantial harm, there are cases where deception might lead to serious consequences, especially when it is related to public security. Therefore, deception detection is a practical need and of great value.

Given the fact that, partly due to a "truth-bias" – the tendency to identify a message as truthful instead of deceptive, humans incline to be credulous in deception detection as human judges (even those with special training) only perform slightly better than chance or even below chance [3], researchers strive to find useful deceptive cues with the hope that these cues or features may help to distinguish deception from truth-telling.

Psychologists mainly focus on the visual cues (e.g. facial expressions, blinks and body gestures) or biometric cues etc. [4-6]. While deception detection researches based on human languages (i.e. verbal cues) carried out by social linguists and computational linguists have just sprung up in only a few decades [7-8]. The previous researches have found some linguistic features that may help to identify truth telling and lying, such as first-person singular pronoun, negative emotion words, exclusive words and linguistic hedges [9-10]. Detecting deception using natural language processing techniques has turned out to be a beneficial supplement of human judgments, or even more competent to this task. Effective methods of tagging deception corpus based on observable cues would have lots of applications in real life scenario.

However, regardless of its promising future, there remain serious problems to be tackled in the field of deception detection based on verbal cues at the present stage.

First and foremost, researchers have to face the lack of suitable corpus. Unlike other natural language resources, corpora of verified deceptive and truthful statements are quite rare, because "lying impose a cognitive and emotional load on individuals which is not easy to reproduce artificially" [11] and may also involve ethical problems. Therefore, the high-stake corpora collected from the real-life situations are rare and almost inaccessible to researchers outside the team that builds the corpus.

Second, even if we leave that point out of account, so far, deception detection based on verbal cues and the related research results are, to a large extent, limited to English language. Would the linguistic features be useful when applied to deception detection in Chinese? Is there any difference between the linguistics cues that are proved to be useful in Chinese deception detection and those in Indo-European languages?

The third problem is verbal cues based on surface information and lexical information have been fully explored in the past fifteen years (since [9]) and have hit a bottleneck. There is evidence that surface linguistic features are insufficient for the task of distinguish lies based on known objects from truths, which are much harder than distinguishing fabrication about unknown objects from truthful statements. Linguistic features from deeper levels are needed for the following deception detection studies based on verbal cues.

In this paper, we build a deception corpus in Chinese. Based on this corpus, we examine the effectiveness of features from previous researches. What's more, we employ the linguistic features from the appraisal theory and test if there is any statistical difference between lies and truths in the corpus, with the hope that this might offer insights for future deception detection studies.

The reminder of this paper is as follows. We first review the previous work in deception detection, which includes some of the representative studies and datasets. Section 3 includes more details of the dataset created for this paper and our research method. In section 4, we introduce our two experiments, analyze and discuss the results obtained. Section 5 is the conclusion part where major findings are summarized and limitations are discussed.

2. **Literature Review.** When people try to deceive, there must be something different in their statements, actions, or even just in their mind, which thus should be traceable. Freud observes this in [12], "He that has eyes to see and ears to hear may convince himself that no mortal can keep a secret. If his lips are silent, he chatters with his finger-tips; betrayal oozes out of him at every pore." This idea is also known as Undeutsch Hypothesis which is first formalized by Udo Undeutsch, a German professor of psychology. Undeutsch asserts that "the memory of a real-life self-experienced event differs in content and quality from a fabricated or imagined event" and that "the cognitive elaboration of an untruthful narrative differs from the elaboration of a truthful one, therefore this difference should be traceable in the features of the narrative itself" [13-14].

For studies based on verbal cues, corpus is dispensable. One way of data collecting is through laboratory mode where participants, usually school students, are invited to lie. Several representative corpora are collected in this way, such as the Columbia-SRI-Colorado (CSC) Deception Corpus and the "Japanese Deception Corpus (JDC)" [7, 9, 15-16]. Apart from inviting college students as participants, another practical way of collecting truthful and deceptive statements is to collect data through the Amazon Mechanical Turk (AMT) service [8, 17-18]. Deceit in this situation is often sanctioned, even with spiritual encouragement or material incentive. Therefore, the participants have relatively less mental pressure, and they need not to worry about the result of being exposed. Yet that is exactly where the flaw of this kind of corpus lies – it "lack(s) the element of deception under stress" [19] and "the mock deceiver has nothing to fear if detected" [20]. This way of "lying" is more like a simulation. To what extent the findings from these studies could generalize to the real-life scenario remains a question.

Given the inherent limitations of deception produced by the laboratory mode, obviously, deceit that occurs in the real-life scenario is a much better choice for the study of it. One resource of this kind of deception data is the narratives from law enforcement and intelligence gathering, such as court, enquiry and testimony [11, 20-21]. The person who issues deceptive statements needs to take responsibility. Once the narrator is caught to be lying, there is a consequence, and that is exactly what is missed out from the mock crimes and "artificial" lies. Therefore, this kind of corpus is also called high-stake corpus, which is relatively rare, due to the limitation of the objective conditions.

In [9], Newman et al. put forward that there are at least three dimensions which should be associated with lies: fewer self-references, more negative emotion words, and fewer markers of cognitive complexity. The linguistic features that could reflect these dimensions are as follows: fewer first-person singular pronouns (e.g. I, me, my), more negative emotion words (e.g. hate, worthless, sad), fewer exclusive words (e.g. except, but, without), and more motion verbs (e.g. walk, move, go). And their hypotheses have been supported by their experimental results: deceptive communications do show the four linguistic features mentioned above, and one unexpected feature - fewer third-person pronouns, which they think might be influenced by the subject - abortion attitude.

As "the first and best known attempt to develop a computational method for deception detection relying entirely on verbal cues", their experiment has been replicated by many researches later and its findings have also been tested on different corpora. The results obtained in subsequent studies on English are generally consistent with [9], but there are also some interesting exceptions.

The researchers build a high-stake corpus in Italian in [11] and tests the findings of [9] on their corpus. This research advances the study of [9] and makes up for its limitations in two aspects. First, it tests the cross-linguistic validity of claims of [9] on Italian, a Romance language. Second, as a high-stake corpus, DeCour avoids the limitation of lacking "external motivation to lie successfully", and the results based on this high-stake corpus could help to evaluate the effectiveness of features found from laboratory-built corpus. In fact, this study does confirm the results of [9], as false utterances show high values in the dimensions of negative emotions, exclusive words, and cognitive/perceptual process. But one exception is the presence of first-person pronouns – in this Italian high-stake corpus, greater use of first-person pronouns is found.

Another inconsistent finding worth mentioning is from [11], which also builds their analysis on a high-stake corpus. Their corpus composes of transcripts of 911 homicide calls in America. The caller-side transcripts are labeled as truthful or deceptive according to the subsequent adjudication of the cases, and the results suggest that negative emotion words are found more in truthful calls rather than in deceptive ones. It is not difficult to understand - as reports of crimes happening in real-life, truthful callers would spontaneously show more negative emotion, such as anxiety, horror, and frustration. On the contrary, the affect of deceptive callers is rather flat. This is something the lies collected from laboratory settings would not tell us. And it reminds us the significance of real context, which needs to be taken into consideration in genuine deception detection.

Although fruitful results have been obtained in these studies, there are problems which cannot be answered by solely using these linguistic features, which are basically some surface features of the texts (e.g. word count, average sentence length) and statistics based on lexicon category. Researchers may get the results that words from a certain category have a significantly higher frequency, but whether it is because of some structural factors of deception or truth remains a mystery. Therefore, to better explain the distinctions, theories that can offer insights about the deep linguistic structures are needed, such as the appraisal theory.

Appraisal theory is developed by [22], as an extension of the interpersonal system of Systemic Functional Linguistic at the level of discourse semantics. According to [22], appraisal is "one of the three major discourse semantic resources construing interpersonal meaning (alongside involvement and negotiation)". Appraisal system itself includes three interacting domains – "**attitude**", "**engagement**" and "**graduation**". Together these three domains provide a collection of semantic resources for construing emotional reactions (**affect**), assessing behaviors (**judgement**), and construing the value of things (**appreciation**) in human experience (**attitude**), with sourcing attitudes (**engagement**) and grading phenomena (**graduation**). And an overview of the structure of appraisal resources is as follows (see FIGURE 1, quoted from [22]).

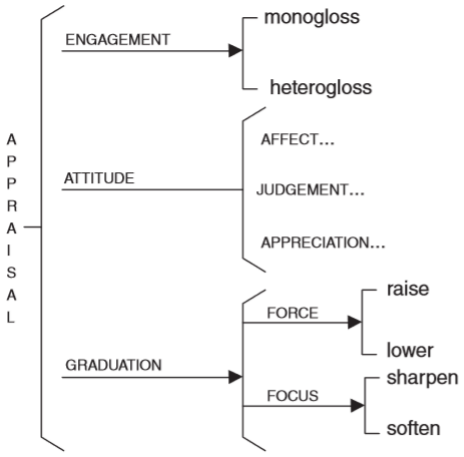

FIGURE 1. AN OVERVIEW OF APPRAISAL RESOURCES

Attitude is the core of appraisal resources. While **engagement** system refers to the ways in which linguistic resources (such as projection, modality) position the appraiser. Through engagement system, the appraiser can moderate the extent of commitment to the proposition expressed (e.g. by quoting, reporting, denying etc.). As for **graduation**, it is about the linguistic resources that help to adjust the degree of an appraisal/evaluation – "force" for this kind of gradable resources, and that help to adjust the strength of boundaries between categories – "focus" for this kind of un-gradable resources.

As a good and practical discourse analysis tool, the appraisal theory has been widely applied into the analyses of literary works, articles of popular science, as well as autographical texts [23-25]. It is also suitable for the analysis of deception happening in

conversations, as the essence of lying in conversations is about evaluating others' behaviors, expressing trust or suspicion and assessing the value of others' words, which would also inevitably include adopting engagement resources (e.g. projection, quotation, denying, etc.) and graduation system (e.g. intensifiers, repetition). The interpersonal meanings embedded in the language usage that the appraisal theory offers may provide some valuable cues and help distinguish deception from truth.

3. **Data Set and Research Design.** This deception corpus comprises the transcripts of the Werewolf game, in which players are divided into the camps of good and evil people and deceptive communications are ineluctable in the process of playing.

3.1. **The Werewolf Game Corpus.** The Werewolf game, also known as Mafia[1], is a party game created by Dmitry Davidoff in 1986. Its basic model is a conflict between an informed minority - the werewolves, and an uninformed majority - the villagers (cf. Wikipedia). In this game, the werewolves strive to hide their identity whereas the villagers need to find out the hidden werewolves. There are a good few of programs about playing werewolves games on the Internet nowadays in China, such as *Lying Man* and *God Lie*. As can be told from the name of these programs, the essence of this game is to deceive. Among them, *The Temptation of Dinner Party* is the first talk show with Werewolves game. So far it has two seasons, which have been viewed for more than 1 billion times on the Internet. In these two seasons, werewolves won more, which means that their lies had not been detected by other players in most cases.

For this corpus, we only chose the rounds of game when the good camp won to transcribe. Our hypothesis was that the deceptive statements made by the werewolves were successfully identified in these rounds as the good camp won, thus there might be some traceable linguistic features.

To obtain the transcripts of the Werewolf game played in *The Temptation of Dinner Party*, we first downloaded the videos of this show from the Internet. Then we used "Youtube to mp3"[2], transforming the videos into audio files, after which we clipped the audio file and reserved the part of people playing the Werewolf game only. These clips were first machine-transcribed into texts using the service of IFLYTEK[3], and then corrected manually.

The Werewolf game is a group conversational scenario where players are not just making a monologue but are involved in a continuous conversation. In this paper, we use the term "utterance" to refer to one uninterrupted statement made by a player. Utterances are annotated according to the following scheme:

[1] https://en.wikipedia.org/wiki/Mafia_(party_game)
[2] http://convert2mp3.net/en/
[3] https://www.iflyrec.com/

Truth The utterance is held as true if it is coherent with the fact. In the case of Werewolf game, it means that the player's statement is consistent with his/her identity and deeds.

Deception The utterance is held as deception if it is in contrast with the fact. In the case of Werewolf game, it means that the player claims to be another identity he/she is not or to have done something he/she did not do, or will do something he/she is not qualified to do. Sometimes only some clauses in one utterance are deceptive. In this case, the whole utterance will still be labeled as deception, although only the deceptive clauses would be analyzed in the following experiment.

Besides, for such a conversational game, there are lots of utterances whose truthfulness or untruthfulness is uncertain. They lack propositional value from the perspective of logic, thus they cannot be labeled as truth or deception, such as gossips, interjections, phatic communications. Also, one important part of uncertain utterances is when players only make speculations about others' identities or make comments on others' behaviors, without mentioning his/her own identity. In this case, we cannot judge these words of appraisal as true or deceptive.

To verify the agreement in the judgments of the truthfulness or untruthfulness of the utterances, half of this corpus were annotated by two annotators, and two labels were tagged: truth and deception, as the rest would automatically be classified as uncertain. Here we use Cohen's kappa coefficient (κ) to measure the agreement between raters [26], and we obtained a κ value of κ =. 73. For the inconsistent judgments, the annotators discussed later and decided the label.

The size of this high-stake corpus is 70,540 tokens composed of 2645 utterances (punctuations included) within which 294 utterances are annotated as deception and 203 as truth.

TABLE 1. STATISTICS OF HIGH-STAKE CORPUS

Label	Utterances	Tokens
Deception	294	16743
Truth	203	12402

3.2. **Analysis Tools.** For the research purpose of this paper, we used two analysis tools: TextMind (the Chinese version of LIWC) and UAM Corpus Tool.

Linguistic Inquiry and Word Count (LIWC) is a computerized tool, which could analyze written or spoken documents on a word-by-word basis. Researches have proved that the use of words in the daily language, indexed by the frequency of occurrence, can effectively reflect important psychological processes and reveal thoughts, feelings, personality and motivations [18], which makes it probably suitable for deception analysis. [9] uses LIWC to capture the linguistic profiles of deceptive and truthful texts, and our first experiment will adopt the linguistic features used in their study and test whether statistical significance exists between deception and truth of our corpus through the Chinese version of LIWC-TextMind.

Of the 102 dimensions available in TextMind, only 25 are used in this study, such as First-person singular, Positive emotions, Negative emotions, and Exclusive verbs (See TABLE 2). The rest of the dimensions are excluded for extremely low frequencies, reflecting content with no representativeness (e.g. words related to sex, leisure, death) and sparse distribution.

TABLE 2. TEXTMIND DIMENSIONS USED IN THIS STUDY

Output labels	Meanings of some Dimensions	Examples
WordCount		
WordPerSentence		
RateFourCharWord	percentage of words longer than 4 characters	
Pronoun		我(I), 他们(them)
I	1st person singular	我的(my),我(I)
We	1st person plural	我们(we)
You	2nd person	你们(you)
SheHe	3rd person singular	它(it), 她(she)
Ppron	personal pronoun	
Conj	Conjunctions	但是(but), 且(and)
QuanUnit	Quantity Unit	
Affect	Affective process	失望(disappoint)
PosEmo	Positive emotions	高兴(happy)
NegEmo	Negative emotions	讨厌(hate)
CogMech	Cognitive process	知道(know)
Discrep	Discrepancy	应该(should)
Tentative		或许(maybe)
Certainty		从不(never)
Percept	Perceptual process	听(hear), 感觉(feel)
Relative	Relativity	相比(compare to)
Time		天(day), 小时(hour)
Exclusive		除了(except)
Work		专业(major)
Assent	Assents	是(yes), 好(OK)
Interjunction	Particle indicating mood	

The UAM Corpus Tool (UAM) is designed for annotation of text corpora. It is widely used in corpus linguistics, as it provides a platform for annotation of multiple texts using the same annotation scheme at multiple levels (e.g. sentence, clause, whole document). Users can choose built-in schemes or design their own schemes, and the appraisal paradigm we plan to use in this paper is already included in the built-in scheme. Besides, it also offers file information and comparative statistics across subsets, which perfectly fits our need of finding discriminating linguistic features between deceptive texts and truthful ones.

4. **Experiments and Analysis.**

4.1 **The experiment based on TextMind dimensions.** According to the statistics, among the 25 TextMind dimensions employed, none of the three dimensions concerning surface information of the utterances, i.e. WordCount, WordPerSentence and RateFourCharWord, show any significant difference between deception and truth. There is statistical separation in 9 out of the 25 dimensions at the level of .05, within which 4 dimensions even show significant difference at the level of .01 (see TABLE 3).

TABLE 3. LIST OF DIMENSIONS AND DISTRIBUTIONS OF *P* VALUES LESS THAN 0.05

p value	Dimensions	Distributions
.01 < *p* < .05	Interjunction	D > T
	Assent	T > D
	CogMech	D > T
	Discrep	D > T
	Tentative	D > T
p < .01	I	T > D
	You	D > T
	Relative	D > T
	Work	T > D

Significant TextMind Dimensions in Deception. The frequencies of words in the following six dimensions are significantly higher in deceptive statements: You, Relative, CogMech, Discrep, Tentative, Interjunction. These results are generally consistent with the previous findings [11, 27-28].

The dimension of You refers to second-person pronouns. In this corpus, the average frequency of second-person pronouns in deceptive statements is twice as much as that in truthful statements. It becomes clear when we see this distinction together with the result of the dimension of I – the frequency of first-person singular pronoun used in truthful statements is significantly higher than that in deceptive statements at the level of .01. These two results suggest that in truthful statements, the speakers use self-reference words to proclaim their "ownership" of this statement [9], whereas in deceptive statements, the liars would try to dissociate themselves from their deceptive claims by avoiding using first-person singular pronoun and direct their accusations at others [8, 27]. Here are two examples from the corpus, one deceptive, the other truthful:

a. 我验了**你**是狼人，**你**说**你**这是，**你自己**是平民，所以**你**在说谎。这一把会把一号胡可投出去。

(I've checked **you** are a werewolf, **you** say **you** are, **you yourself** are a villager, so **you** are lying. This round (we) will vote No.1 Hu Ke out.)

b. **我**真的本来舍不得用的，但**我**看他是你**我**才用的，是别人**我**就不救了。**我**本来留给**自己**的。
(**I** was truly loath to use (the antidote) at first, but **I** used it (only because) **I** saw it was you, if it's others **I** would not save (them). **I** suppose to leave (the antidote) for **myself**.)

The dimension of Relative reflects comparison. Under the context of this corpus, it means that deceptive statements apply more linguistic resources to make comparisons between players. This might also result from their reluctance to directly talking about themselves, thus to draw others' attention away from them, they would talk more about others by making comparison. This dimension has not been used in previous studies, because the words included in dimension of Relative in the English version of LIWC are different from those in the Chinese version. Therefore, this linguistic feature is unique to Chinese deception corpus.

The higher frequencies of words reflecting cognitive processes and discrepancy in deception are consistent with the findings of [11]. CogMech refers to cognitive processes, including words concerning cognitive activity, e.g. *know*, *cause*, *because*. The liars would unconsciously use more words from the dimension of CogMech. One possible explanation is that although others do not know their identity and their intention to deceive, the liars themselves are bothered by the heavy psychological burden and could not resist providing proof for their sincerity by revealing their cognitive process to others. This tendency of liars can be summarized as an "image-and relationship-protecting behavior" [27], which refers to the "verbal and nonverbal behaviors used to make oneself appear sincere and trustworthy and to sustain the self-presentation one has created" [29]. The dimension of Discrep, i.e. discrepancy, refers to words concerning the difference between things that should be the same, e.g. *should*, *could*, *would*. Here are some examples from the corpus:

a. 你们不都是高手么，**怎么**玩成这个样子呢？
(Aren't you masters of this game, **how come** ending like this?)
b. 对，就是觉得，**应该**是，**应该**是跟我一样……但又有点不一样，因为我比你高一级。
(Yes, (I) just feel that, (your identity) **should** be, **should** be the same with me…. but a little bit different, as I am higher than you.)

The dimension of Tentative reflects the uncertainty and hesitation of a proposition, e.g. *maybe*, *perhaps*, *guess*. Its higher frequency in deception provides support for previous theoretical hypothesis that the language of liars tends to be ambiguous and uncertain [16, 27].

The dimension of Interjunction represents the particles indicating mood in Chinese, such as *吗*(*ma*),*呢*(*ni*),*呀*(*ya*). This word dimension is lacking in English. Adding these final particles to clauses could help to express the emotion swings, moderate the tone of speech or make the speaker sound more lively and affable. The significantly higher frequency of these words in deceptive statements might reflect the need of liars to construe a false image and pretend to be harmless, which also belongs to the "image-and relationship-protecting

behavior" [29] of deceivers.

In short, in this high-stake corpus, deceptive statements include more second-person pronouns and words in the dimension of Relative, through which the speakers aim to put others on spotlight and dissociate themselves from their deceptive claims. Besides, for the purpose of sustaining the false image they have created, liars would employ linguistic devices like particles indicating mood and words in the dimension of Tentative to moderate their speech tone, with the hope that these linguistic devices might soften their act of framing others.

Significant TextMind Dimensions in Truth. Except for the dimension of the first-person singular pronoun mentioned above, the frequencies of words in the following two dimensions are significantly higher in truthful statements: Work and Assent.

In this high-stake corpus, words about work are used in truthful statements twice as much as them in deceptive statements, which shows a significant difference at the level of .01. This might be attributed to the content of this corpus: in Werewolf game, the statements made by the players are mainly about their own identities, deeds, speculations about others' identities and judgments of others' deeds. A typical pattern of truthful statements is that the speaker reveals his/her own identity and tell others what he/she did or knew from the other night frankly. Whereas liars might lie about their identity and focus on the "abnormality" of others, without saying a word about what they did. About the reason behind, we conjecture that either this is partly due to the specific topic of this corpus; or it is not limited to a certain corpus, but a commonality – when people are telling lies, they avoid making statements about their own deeds, but incline to make judgments about others' behaviors.

Assent refers to words concerning agreement and passivity. The frequency of words expressing assent used in truthful statements is significantly higher than it in deceptive statements at the level of .05. In the dictionary of LIWC, the dimension of Assent belongs to the spoken categories and refers to expressions like *yes*, *ok*, *absolutely*. In Chinese, this paralinguistic dimension includes words like *对*(*dui*, right/yes), *是的*(*shi-de*, yes). Apart from using words of assent at the beginning of a statement to show that the speaker accepts others' statements, there is one interesting usage of assent words in truthful statements:

> a. **对**，我可以带走一个人。**当然**像我是有仇必报的，**对**，谁投我，我肯定带走你们俩中间的一个人。
>
> **(Yes**, I can take one person away. **Of course** (people) like me would give an eye for an eye, **yes**, (for those) who vote me, I will definitely take one of you two away.)

In this truthful utterance, the speaker unconsciously expresses her identification with what she is saying by inserting words of assent in. It suggests that the truth-tellers identify with what they are saying, whereas the liars may not.

To sum up, in truthful statements of this high-stake corpus, the narrators use significantly

more first-person singular pronouns, as they incline to proclaim their "ownership" of this statement; they would talk more about what they have done through words about work, and use more words in the dimension of Assent, which shows their subconscious identification with what they are saying.

4.2. **The experiment based on features of appraisal theory**. In this experiment, we annotated a sample of truths and lies in our high-stake corpus, using the scheme of the appraisal theory in UAM, in which appraisal resources are finely classified. Apart from the essential system of attitude, the appraisal theory also includes engagement system, which shows the extent of the appraiser's commitment to the proposition expressed, and graduation system to adjust the degree of an appraisal or evaluation.

UAM can output comparison results between truths and lies in terms of different features, and it has three levels of significance: weak significance (at the level of .1), medium significance (at the level of .05) and high significance (at the level of .02). In this paper, we only discuss the features showing medium and high significance (as all the features of high significance also show significant difference at the level of .01). Among the seven significant features (see TABLE 4), two of them are from the system of attitude (marked by the capital letter A), three from the system of engagement (marked by the capital letter E) and two from the system of graduation (marked by the capital letter G). From the perspective of distribution, five significant features appear more in deceptive statements, whereas the other two features salient in truthful statements are both from the system of engagement.

TABLE 4. LIST OF FEATURES FROM THE APPRAISAL THEORY OF P VALUES LESS THAN 0.05

p value	Features	Distributions
$.01 < p < .05$	A-capacity	D > T
	E-disclaim	T > D
	E-attribute	D > T
$p < .01$	A-negative attitude	D > T
	E-contract	T > D
	G-focus	D > T
	G-sharpen	D > T

Significant Appraisal Resources in Deception. The frequencies of five appraisal features are significantly higher in deceptive statements: capacity, attribute, negative attitude, focus, and sharpen.

(1) **Capacity** is under the system of **attitude**, and it belongs to **judgement**, which deals with attitudes towards behaviors. Words in the region of capacity are mainly

assessments of people's competence and ability, such as *skilled, brilliant, stupid*. Deceptive statements include significantly more words about capacity which is judgement of others' behaviors. This result confirms our conjecture in the first experiment based on the same corpus: in deceptive statements, words from the dimension of Work are significantly fewer than those in truthful statements, which might be the result of liars avoiding talking about their own behaviors but focusing on judging others. And through the annotation of the appraisal resources, it becomes clear that liars do use more words to judge others' behaviors. Here are some examples:

a. 因为那威老师又**太厉害**了，所以我……
(Because Na Wei is **too powerful**, so I …)
b. 我真的是好人，而且我属于那种**完全没有逻辑性，记性也不好**，就是**完全不记得大家都说了什么**。
（I am truly a good guy, and I am the kind (of people) with **absolutely no logicality**, **poor memory**, as (I) totally **cannot remember what everyone has said**.)

(2) **Attribute** belongs to the system of **engagement**. It is an act of dialogistic expansion through introducing the externalized proposition, by which the appraiser "dissociate the proposition from the text's internal authorial voice by attributing it so (to) some external source" [22]. Examples are *according to X, in X's view, X said*. And the significantly higher use of this resource accords with the need of liars. In the first experiment, similar needs of liars are marked by fewer uses of self-reference words, but that is not enough. With the engagement system of the appraisal theory, it is clearer that liars would avoid taking responsibility of what they say by introducing external sources.

(3) **Negative attitude** is generally believed to be a marker of deception. In the scheme of appraisal theory in UAM, negative attitude is evaluated through the whole system of attitude: of affect, of judgement and of appreciation. And the result of our experiment is consistent with the finding of [9].

(4) & (5) **Sharpen** belongs to **focus**, under the system of **gradation**, in which the term being graduated is a non-attitudinal term (e.g. *husband*, *villager*, *music*) and there is "a strong tendency for the cline of prototypicality to be invested with attitudinality" [22]. Sharpen refers to a positive attitudinal assessment. More words of sharpening a concept are used in deceptive statements, which might be the result of conscious stress – liars consciously stress their fake identities or stories. Here is an example:

a. 我**就**是一个平民。
(I am **purely** a Villager[4].)

[4] Notice that "Villager" refers to an identity in Werewolf game, and a Villager is a member of the good camp.

To sum up, the deceivers would employ more appraisal resources to judge others' behaviors, introduce external source to transfer the responsibility. Their evaluations incline to be negative. Besides, they would use expressions of the value of sharpening to stress their fake claims.

Significant Appraisal Resources in Truths. The two features salient in truthful statements – disclaim and contract, belong to the system of engagement. Contrary to deceptive statements (of more dialogic expansion), truthful statements employ more resources of dialogic contraction, especially the resources of disclaim. Dialogic contraction, i.e. contract, is the act of contracting the dialogic space, by excluding certain alternatives or views from subsequent communications. And disclaim is one way of dialogic contraction, which is the act of rejecting an alternative or replace it with another one, e.g. *never*, *not*, *but*. Truthful statements include significantly more appraisal resources expressing disclaiming, which might be attributed to the truthteller's certainty about their statements. Here is an example:

a. **不**。因为我觉得他，**但是**他真的**不是**预言家，因为我确实**是**一个村民，这我很肯定，他**不是**预言家。那他为什么要跳预言家呢？
(**No**. Because I think he, **but** he is truly **not** Seer, because I am indeed a Villager, this I am sure, he is **no** Seer. Then why did he pretend to be Seer?)

5. **Conclusions.** In this paper, we build a Chinese deception corpus of transcripts of the Werewolf game. Based on this corpus, linguistic features chosen from the dimensions of LIWC/TextMind are tested first, then features from the appraisal theory are annotated and compared based on this high-stake corpus.

In this high-stake corpus, deception is characterized by higher use of second-person pronouns, particles indicating mood, words of comparison, words about cognitive process, words showing difference, and words expressing tentativeness, whereas truthful statements contain more first-person pronouns and words of work and assent.In terms of the appraisal resources employed, our data show that liars tend to evaluate others' capacity, introduce others' voice to speak for themselves, express negative attitude and increase the evaluation graduation by using appraisal resources of sharpening, whereas truthful statements would use more appraisal resources of disclaim to contract the conversation space. The results of our experiment suggest that deception differs from truth in terms of the appraisal resources used, which can be taken as a useful verbal cue and might offer insights for future deception detection studies.

Acknowledgment. This research project is supported by Beijing Social Science Foundation (No.14WYC041).

REFERENCES

[1] Zuckerman, M., Depaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception 1. *Advances in Experimental Social Psychology, 14*, 1-59.

[2] Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. *Communication theory*, 6(3), 203-242.

[3] Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. Personality & Social Psychology Review an Official Journal of the Society for Personality & Social Psychology Inc, 10(3), 214.

[4] Ekman, P. (2001). Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage.

[5] Ekman, P., Friesen, W. V., & O'Sullivan, M. (1988). Smiles when lying. Journal of Personality & Social Psychology, 54(3), 414-20.

[6] Leal, S., & Vrij, A. (2008). Blinking during and after lying. Journal of Nonverbal Behavior, 32(4), 187-194.

[7] Enos, F., Shriberg, E., Graciarena, M., Hirschberg, J., & Stolcke, A. (2007). Detecting deception using critical segments. INTERSPEECH 2007, Conference of the International Speech Communication Association, Antwerp, Belgium, August (pp.2281-2284). DBLP.

[8] Mihalcea & Strapparava. (2009). The lie detector: explorations in the automatic recognition of deceptive language. Unt Scholarly Works, 309-312.

[9] Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: predicting deception from linguistic styles. Pers Soc Psychol Bull, 29(5), 665-675.

[10] Bachenko, J., Schonwetter, M., & Schonwetter, M. (2008). Verification and implementation of language-based deception indicators in civil and criminal narratives. International Conference on Computational Linguistics (Vol.28, pp.41-48). Association for Computational Linguistics.

[11] Fornaciari, T., & Poesio, M. (2013). Automatic deception detection in italian court cases. Artificial Intelligence & Law, 21(3), 303-340.

[12] Freud, Sigmund. “Fragment of an Analysis of a Case of Hysteria (1905),” Collected Papers, Vol. 3; Basic Books, 1959.

[13] UNDEUTSCH, & UDO. (1967). FORENSISCHE PSYCHOLOGIE : Handwörterbuch der Kriminologie, Band 1: Aberglaube - Kriminalbiologie. Aberglaube - Kriminalbiologie.

[14] Undeutsch, U. (1989). The Development of Statement Reality Analysis. Credibility Assessment. Springer Netherlands.

[15] Hirschberg, J., Benus, S., Brenier, J. M., Enos, F., Friedman, S., & Gilman, S., et al. (2005). Distinguishing Deceptive from Non-Deceptive Speech. INTERSPEECH 2005 - Eurospeech, European Conference on Speech Communication and Technology, Lisbon, Portugal, September (pp.1833-1836). DBLP.

[16] Chen, F., & Jokinen, K. (2010). Speech technology. Springer Science+ Business Media, LLC.

[17] Ott, M., Choi, Y., Cardie, C., & Hancock, J. T. (2011, June). 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 (pp. 309-319). Association for Computational Linguistics.

[18] Salvetti, F. (2012). Detecting deception in text: a corpus-driven approach. Dissertations & Theses - Gradworks.

[19] Adams, S. H. (2002). Communication under stress: indicators of veracity and deception in written narratives. Virginia Tech.

[20] Fuller, C. M., Biros, D. P., & Delen, D. (2011). An investigation of data and text mining methods for real world deception detection. Expert Systems with Applications, 38(7), 8392-8398.

[21] Burns, M. B., & Moffitt, K. C. (2014). Automated deception detection of 911 call transcripts. Security Informatics, 3(1), 1-9.

[22] Martin, J. R., & White, P. R. R. (2005). The Language of Evaluation: Appraisal in English.

[23] Bian & Gao (2006). EFL students' cultural stereotype change in their narrative. Foreign Languages in China, 3(1), 35-40.

[24] Zhao, L. (2015). The analysis on speech in the joy luck club from appraisal perspective. Overseas English.

[25] Zhang, J. Y., & Wang, X. (2011). Interpersonal meaning analysis of popular science english. Journal of North China Electric Power University.

[26] Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. Fam Med, 37(5), 360-363.

[27] Zhou, L., Burgoon, J. K., Nunamaker, J. F., & Twitchell, D. (2004). Automating linguistics-based cues for detecting deception in text-based asynchronous computer-mediated communications. Group Decision & Negotiation, 13(1), 81-106.

[28] Hancock, J. T., Curry, L., Goorha, S., & Woodworth, M. (2005). Automated Linguistic Analysis of Deceptive and Truthful Synchronous Computer-Mediated Communication. Proceedings of the, Hawaii International Conference on System Sciences (Vol.01, pp.22.3). IEEE Computer Society.

[29] Buller, D. B., & Burgoon, J. K. (1994). Deception: strategic and nonstrategic communication. Strategic Interpersonal Communication.