Tanulmány

Ahmad Adha

Linguistic Based Cues in Detecting Deception in Indonesian Language Use

Abstract

Language is used to communicate differently in various cultures, but is universally used to exchange rational information. Languages are also used to communicate interpersonal information; the information being communicated is both truthful and deceptive. Previous research suggests that there are several linguistic cues of deception when someone is lying. The present research tries to replicate and apply these cues to speakers of the Indonesian language. Thus, the aim of this research is to find out some linguistic cues of deception in Indonesian language use. The method used in this research was an interview of ten participants. The participants were asked to provide two stories based on their personal experiences; truthful and deceptive stories. Data obtained were then analyzed by calculating the mean and standard deviation to get value discriminating the two stories. The result indicates that some cues were significantly different in two stories. The cues are word quantity, verb quantity, sentence quantity, third person pronoun, group reference, and self-reference. The result shows that linguistic based cues of category of quantity, uncertainty and non-immediacy are the most obvious discrimination in truthful and deceptive information.
Keywords: cues, deception, lie, Indonesian

1   Introduction

Though languages are used to communicate differently in various cultures, there seems to be a universal aspect about language use. Languages are universally used to exchange rational information and to communicate interpersonal information, emotion, and feelings, such as guilt, fear, anger or happiness, including good memories. Walker and his colleagues (2003) claim that people have better abilities in recalling good memories, compared to bad ones since the latter will fade faster (see Jupe et al. 2018). McGrath (2014) also states that almost 60% of unpleasant experiences will be forgotten whereas only 42% of pleasant memories will be forgotten. We might remember and be able to describe the pleasant days, people we met or places we visited after a holiday, but forget about the flight delays or problems in immigration check. Thus, communicating pleasant stories to others will not be a difficult task for someone if it is done truthfully. However, on some occasions, people do not entirely use language to share truthful information. People may be deceptive when giving information. Dynel (2018) defines deception as an intentional verbal or nonverbal communicative act of attempting to cause the targeted listener to have a false belief. As one form of deception, she adds, lying is a

form of verbal communicative activity involving the speaker intending to deceive the listener. Coleman & Kay (1981) has a similar argument; lying is commonly defined as a statement that is used by the speakers to intentionally mislead the hearers. Furthermore, lies are statements that the speaker believes to be false and that are intended to mislead the listener (Bok 1978). Davis (1961) argues that when the speaker lies, he or she will show a strong psychological response because he or she wants to make the statement sound truthful and reliable even though he or she knows that it is false. These struggling emotional phenomena that the speaker has can be revealed through the examination of his or her verbal and nonverbal behavior. In fact, Vrij et al. (2000) even suggest three ways to detect lies, they are as follows; (1) observing how the speakers behave through movements, smile or gaze; (2) listening to what the speakers say through speech context analysis; and (3) measuring physiological response through polygraph test or control question test. Miller and Burgoon (1996) observe that it is important to note that some of the nonverbal cues also express things other than deception, since biological changes such as blood pressure and heart rate do not necessarily mean that the person is lying or telling the truth. Thus, analyzing people's verbal behavior is the most observable and reliable tool to detect whether they are telling untruthful statements.

The way people think is reflected through the language use. So, when people talk about personal topics, there will be some differences if the topics are truthfully communicated or not. This suggests that creating a false or imaginative story needs some efforts and results in different patterns of language use. Accordingly, it is possible to distinguish truth and lying through verbal cues. Fraser (1991) argues that this difference is the result of a feeling of stress, indicated in a decline in capacity for cognitive integration, in precision, in organization, and in ranking things.

There has been a lot of research conducted on deception. Most research before the 1980s focused on non-verbal deceptive cues, such as body language and the facial expression of a liar. Since then, researchers started to analyze paralinguistic cues of deception, for example liar's pauses and speech errors. In the last two decades, research on deception has included the examination of morphology, syntax, semantics, and discourse. However, there have been some issues that the previous research has not touched upon. For example, in his research, Dulaney (1982) examined his subjects in an interview setting. By doing this, he gave no room for the respondents to create a wholeness of a story since they only needed to answer questions of the interviewer. So, there was a lack of individual invention and motivation on behalf of the respondents in his research. Hancock et al. (2008) faced the similar inadequacy since the motivation to lie was not natural and the discourse examined was not a full text. Stellar and Koehnken (1989) focused on children but the method they applied only worked if there were non-verbal clues integrated in detecting deception. However, in Stellar and Koehnken's examination, it was not possible to separate the discourse created by the respondents to their non-verbal clues. Moreover, children have problems in differentiating real and imaginative events (Johnson & Raye 1981). There are some disadvantages to having children as respondents. Previous research was mostly using native speakers of English (such as Newman et al. 2003) and Hebrew (Dilmon 2009). Cross-cultural research examining verbal cues to deception is scarce. Sharon and colleagues (2018) are among the pioneers in this field. Respondents were speakers of English, Chinese and Arabic. However, the transcripts of Chinese and Arabic discourses were then translated which may result in loss of information. Aforementioned paragraphs indicate that there is a need for research which includes varying research populations, a strong motivation on behalf of the respondents, i.e. they have pseudo-

intention to manipulate a story, complete deceptive and truthful discourses to examine, a thorough examination of different linguistic areas, and of the linguistic tools significantly distinguishing truthful and untruthful stories. Based on the assumption that there is a difference between the truthful and untruthful forms of language use, the present paper aims to detect lying through the linguistic cues provided by the speakers of Indonesian language. In order to fulfill this aim, an experiment has been conducted in which subjects were asked to provide truthful and untruthful stories of their past pleasant experiences. These stories were then coded and analyzed according to a taxonomy created relying on the theories of deception to find differences on the basis of their linguistic cues.

The organization of the paper is as follows. After this introductory Section 1, in Section 2, I will briefly describe theories of deception I rely on in this paper. In Section 3, I will provide the research question and the hypothesis for the experiment. In Section 4, the methodology of the experiment will be discussed; I characterize the participants, the instrument, and the procedure of the experiment. In Section 5, I will provide the result of the experiment showing linguistic differences of truthful and untruthful stories. In addition, statistical measurement through means and standard deviations of the linguistic cues will also be provided. After that, I will discuss the linguistic differences between truthful and untruthful stories and make a comparison to findings of previous research in Section 6. Finally, in Section 7, I will summarize the conclusions.

2 Theory on deception

Lying is an anomaly of the conventional rule of communication. Mey (1993) argues that to communicate is the objective of the speakers, thus they strive to be understood correctly and aim not to make a misleading impression. He adds that communication is a matter of cooperation. Grice (1975) has dealt with the philosophical aspect of conversation by suggesting the Cooperative Principle with subdivided categories or maxims of quantity, quality, relation, and manner. According to Grice, the speakers need to be attentive to these maxims in order to take part in a logical and comprehensible conversation. However, there may be occurrences in which one of the maxims is not observed which interferes with the logic and comprehensibility of the conversation. The non-fulfillment of one of the maxims of the Cooperative Principle can occur through flouting or violation of the maxims. Dynel (2018) argues that violation of the first sub-maxim of quality (i.e. do not say what you believe to be false) leads to lying. When one is lying, it is not about the information he or she is giving being objectively false, but it is about the speaker not believing the truthfulness of the information.

There are points in common among speakers of different languages regarding deception. As for honesty, it is also universal when the truth is sufficient for achieving a goal. All cultures want their members to value honest communication and avoid deception; however humans also have abilities to enable deceit at communication. Regardless of their native languages, liars will always try to avoid being caught in a lie. Galanski (2000) claims that deceptive communication is reinforced by communication that is not deceptive. Thus, liars will try to conceal their message within a discourse of truth and in non-deceptive context. That is why it is believable to say that there are indeed differences in lying and truthful discourse. There have been several analytic tools for examining textual information that have

been proposed and accepted in deception research or even deception detection practice. Some of them that will be referred to in this paper are Information Manipulation Theory (IMT) (McCornack 1992), Interpersonal Deception Theory (IDT) (Buller & Burgoon 1996), Reality Monitoring (RM) (Johnson & Raye 1981), Verbal Immediacy (VI) (Mehrabian and Wiener 1966) and Self-Presentational Perspective (DePaulo et al. 2003).

Information Manipulation Theory offers a multidimensional approach to deceptive messages. According to McCornack (1992), deception will possibly arise from violation of one or more of Grice's four maxims. In deceptive communication, covert violations of quantity can result in 'lies of omission' in which the speaker does not give enough information or even withholds some. Thus, the amount of information given to the listener is being altered. Covert violations of quality involve falsification of information. These violations might be considered stereotypical deceptive messages in which information is deliberately distorted or fabricated. Deception by evading others' statements involves covert violations of relevance. By committing evasion, the speaker manipulates the relevance of information. For example, a liar fails to directly answer a question. The deceptive act through the covert violation of the maxim of manner is resulting in equivocation. Here, information is conveyed in an ambiguous fashion or lacks clarity. Thus, covert violation of one or more of Grice's conversational maxims is assumed to result in messages that are functionally deceptive.

Interpersonal Deception Theory was developed to explain and predict deception and its detection in interpersonal contexts. This theory was proposed by Buller and Burgoon (1996) who suppose that a liar will utilize several strategies to manipulate information so that it appears credible and the receiver of information will be unable to detect deception. A meta-analysis made by DePaulo and her colleagues (2003) has given clarification in strategies and specific verbal indicators of lying. There are six categories where liars manipulate in their information transmission. The strategies are as follows:

(a) Quality manipulation. When liars hide either complete or partial truthfulness, they are inclined to use more adjectives and adverbs to adequate the meaning of their statements.

(b) Quantity manipulation. Liars want to create incompleteness of their information by using less words or sentence, or they may also reduce the specificity of the information such as not giving appropriate information regarding the time or place of the event. Liars might also repeat the same words, reducing their vocabulary and diversity of their messages.

(c) Clarity manipulation. Liars also tend to create vagueness and uncertainty of their messages by using modal verbs such as 'may' or 'might', and using generalizing terms such as 'somebody', 'one time', etc.

(d) Relevance manipulation. Instead of direct responses, liars provide indirect responses that are diverged from the expected responses.

(e) Disassociation manipulation. Liars are trying to distance themselves from the messages and their contents. They do not want to be held responsible if they are once caught to be lying. Thus, they are using non-immediate language forms such as lack of pronouns or removing the author of the action through passive voice.

(f) Image and relationship protecting behavior. Liars apply verbal and nonverbal behaviors to make them look sincere and trustworthy, and also to sustain their self-

presentation that they have created. These behaviors can be done by avoidance of discrediting information and evading negative affect in their message.

Reality Monitoring suggests that a truthful memory will be likely to contain perceptual, contextual and affective information in comparison to an event that has been made up that is likely to contain cognitive operations (Johnson & Raye 1981). Since deception is sometimes based on imagined rather than self-experienced events, this theory has been applied in the context of deception detection. The perceptual information mentioned before is any kind of information regarding visual details, sounds, smells, tastes and physical situations of the experienced story. Contextual pieces of information are spatial details, details regarding the location where the story took place, and details about how objects and people were situated in relation to each other, e.g. *John sits beside me*, and temporal details or information about the time order of events and details about the duration of events. Furthermore, affective information is about how someone felt during a real event. In contrast, a story from imagination is taken from an internal source, thus it contains cognitive information as stated before such as thought and reasoning, for example, 'I remember I wore a blue coat that night because it was very cold.' In addition, Lindsay and Johnson (1987) (see in Zhou et al. 2004) state that Reality Monitoring was found to be more useful for analyzing adults' statements because they have been able to differentiate between fact and fantasy as opposed to children.

Verbal Immediacy was initially proposed as a means of inferring people's attitude or affect (Mehrabian & Wiener 1966 in Zhou et al. 2004). The general construct of immediacy and non-immediacy refers to verbal and nonverbal behaviors creating a psychological sense of closeness or distance. Verbal non-immediacy can be indicated through lexical choices, syntax and phraseology of separation, non-identity, directness, or change in the intensity of interaction between the speaker and the listener. Furthermore, Mehrabian and Wiener (1966) (see in Zhou et al. 2004) state that this theory can be classified in three major categories: spatio-temporal, denotative specificity, and agent-action-object categories, which will then be broken down into subcategories. Donohue (see Zhou et al. 2004) states that VI has been applied to conversation analysis and concludes that the speaker avoiding being immediate can be indicated by some non-immediacy sub-categories. The examples of the sub-categories are spatial and temporal terms, passive voice, presence of modifiers, and other expressions such as volitional words, politeness, and automatic phrasing.

Self-Presentational Perspective theory was born from a meta-analysis conducted by DePaulo et al. in 2003. According to this theory, there are several cues of deception indicating whether one is lying. These cues are organized into five categories representing nonverbal, verbal and paraverbal communication. The first category suggests that liars are less forthcoming. They will provide shorter and less detailed responses and may appear reticent. The second category predicts that liars will tell less compelling stories which will include more discrepancies, be less engaging, more passive, uncertain, and non-immediate. The third category states that liars will be less positive and pleasant. The fourth category shows that liars are more tense. And the fifth category suggests that liars will include fewer ordinary imperfections and unusual contents within their messages.

Even though all these theories were created separately and none of them were developed specifically for text analysis and deception detection, the theories provide the theoretical and evidentiary foundation for the cues included in the current paper. Furthermore, some of these theories share many common aspects, and in fact, they are overlapping. For example, passive voice will be found more in deceptive messages as suggested by Interpersonal Deception

Theory, Verbal Immediacy, and Self-Presentational Perspectives. The taxonomy in Table 1 shows the linguistic categories used to determine truthful and deceptive messages applied in the research presented here. This is a modified version of a taxonomy proposed by Zhou et al. (2004) and Fuller et al. (2013; 2015). Zhou and his colleagues (2004) included a category of informality into their analysis by measuring typographical errors occurring in written texts they examined. Fuller (2008) used official reports taken from people of interest as his data. Thus, he added a category of severity which is described as the consequence of being involved in the incident described by the speaker. The table also has more categories compared to previous research by Adha (2018) as the current research is a rethinking and continuation of the previous one.

Category	Theoretical Foundation
Quantity	IMT, IDT, Self-Presentational Perspective
Specificity	IMT, IDT, RM, VI
Uncertainty	IMT, IDT, VI, Self-Presentational Perspective
Complexity	IMT, IDT
Diversity	IMT, IDT
Expressivity	IDT
Non-immediacy	IMT, IDT, VI, Self-Presentational Perspective

Table 1. Text-Based Deception Categories

The first category is quantity, dealing with the length of a message given by a deceiver. If time is available and efforts at persuasion may be beneficial, the deceiver is likely to create a longer message (Zhou et al. 2004). This is due to one wanting to appear reliable and thus using more words, verbs and sentences in the communication. In the context of the research presented here, the subjects were asked to provide stories without interruption and to provide believable stories thus quantity might increase. The next category is specificity, reflecting on the type and amount of details within a message. It is expected that the subjects would mention less spatio-temporal and perceptual information when they told an untruthful story. The third category is uncertainty. It relates to attempts to avoid giving relevant answers to the context and circumstances. This suggests that the subjects would have ambiguous and general statements in their untruthful stories. Some linguistic cues indicating uncertainty are for example the use of modal verbs, uncertainty words, and third person pronouns. The subjects of the current experiment were presumed to have these linguistic cues when they presented untruthful stories. The next category is complexity which deals with the simplicity of a message (Zhou et al. 2004). Zhou and colleagues (2004) argue that a deceiver is assumed to use simpler message since his or her energy is consumed to manage the situation (see Fuller et al. 2015). Complexity is also about the comprehensibility of the message. Following Zhou et al.'s experiment, to measure the complexity of text, the average number of sentence length was counted. These numbers were expected to be lower in stories containing deception by the subjects of the present experiment. The fifth category is diversity, concerning lexical, content word diversity and redundancy. Lexical diversity is obtained from the number of different words or terms created by the speakers divided by the total number of words or terms that

they produce. As for the content word diversity, it is attained by dividing the number of different content words or terms and the total number of all content words. And, redundancy is from the number of function words per the number of sentences. According to Fuller et al. (2015), a deceiver is thought to use less diverse language and more repetition of words and phrases. It is presumed that a higher diversity level means the subjects were telling truthful stories. The sixth category as suggested by IDT is expressivity. It refers to how vivid and expressive language is used to create emotion in others. Fuller et al. (2015) explains that the number of adjectives and adverbs will show the expressivity of the language. Thus, in order to obtain expressivity, the number of adjectives and adverbs will be divided with the number of nouns and verbs created by the speakers. The speakers were expected to produce a higher amount of expressivity in their deceptive stories. Finally, the category of non-immediacy relates to language use creating psychological distance and to avoidance of taking responsibility of or claiming ownership of a message. It is one of the strategies for depersonalizing a message as described in some theories of deception. To distance oneself from the message, the communicator will use a more passive voice, objectification, generalizing terms and first plural pronouns. These cues were also presumed to be used by the subjects of the current research.

3 Research questions and hypotheses

The aim of this research is to examine the linguistic cues of deception provided by the speakers of the Indonesian language. The concept of deception may be universal, however through this current research, I want to find out if the speakers of Indonesian language produce similar linguistic cues of deception as the other language speakers. To assess this, the subjects will tell three different stories in which one of them is based on untruthful events. A comparison of these stories will be made and it is expected that the untruthful stories will contain more words and sentences, less spatio-temporal and perceptual information, use of modal verbs, uncertainty words, and third person pronouns, use simpler message indicated by lower average number of sentence length, use less diverse language and more repetition of words and phrases, use more adjectives and adverbs that increase the expressivity of the message, and use more passive voices, objectification, generalizing terms and first plural pronouns.

4 Method

4.1 Participants

In the present study, there were 10 participants (5 men and 5 women) between the ages of 24 and 37 (average age: 28.9), all of them have finished at least their bachelor study and 8 participants are now currently enrolling at either master or PhD study in a university in Hungary, while the others are housewives/househusbands. The experimental population was heterogeneous, coming from different places in Indonesia; 6 respondents are from Java, 2 are from Sulawesi, one is from Sumatra, and one is from Kalimantan. 9 of them speak at least one regional language, and also English. But, all of them are native speakers of Indonesian. All participants volunteered to participate in this study.

4.2 Instruments

Table 2 is obtained from the construction of categories from Table 1. This table was constructed from findings of previous research in which written or oral texts of other languages were analyzed. Previous research such as Zhou et al (2004), conducted their research by analyzing written texts of students, meanwhile Fuller et al (2015) analyzed data taken from a real legal case.

The linguistic cues to determine truthful and untruthful stories will be categorized according to the following table. The plus (+) and the negative (-) mean that these cues are predicted to be higher or lower respectively in untruthful stories.

Category	Cue	Predicted Change in Deceptive Messages
Quantity	Word Quantity	+
	Verb Quantity	+
	Sentence Quantity	+
Specificity	Spatial Close Information	-
	Spatial Far Information	+
	Temporal Immediate Information	-
	Temporal Non-immediate Information	+
	Perceptual Information	-
Uncertainty	Modal Verb	+
	Third Person Pronoun	+
	Uncertainty	+
Complexity	Average Number of Clause	-
	Average Sentence Length	-
Diversity	Content Word Diversity	-
	Function Word Diversity or Redundancy	-
	Lexical Diversity	-
Expressivity	Emotiveness	+
Non-Immediacy	Group Reference	+
	Self-Reference	-
	Objectification	+
	Passive Voice	+
	Generalizing term	+

Table 2. Linguistic Cues for Examination

4.3 Procedure

During the data collection, each of the subjects was contacted individually for a meeting. The meeting was conducted in a location where it was adequate for the subjects to tell the stories and for me to record their stories. As for the subjects who were not able to have a face-to-face

meeting, they were contacted via mobile applications, such as Skype or WhatsApp. In the beginning of the meeting, all subjects were asked to tell a preliminary story about a joyful activity this week. This story served as an icebreaker to form a relationship between the interviewer and the subjects, and also as true story elicitation before they were made aware of the subject of the study in which they were required to tell stories of deception.

The subjects were told that they would be assisting in research on deception and would be asked to provide two happiest stories from their past in which one of them was truthful and the other was fabricated. They were told that another group of subjects would listen to their stories and identify which of the stories were true. The subjects did not know in this phase of the research that it was their stories that would be studied. Thus, their speech was more natural and they were more motivated to deceive and attempt to mislead other people. After telling the stories, all subjects were asked to tell which of them were lying.

At the end of the meeting, the subjects were given more details about the particular purposes of the study and were asked if they would still allow their recordings to be used for the study. All of the subjects gave their permission.

The recorded stories were then transcribed. The transcription was done by typing all words and sentences produced by the subjects. Each of the transcribed stories was examined separately. A comparison was then made of the three stories told by each subject: the joyful activity this week, the happiest story from the past, and the fabricated story. The examination was conducted according to the linguistic cues on Table 2. Each cue was tallied and listed. For example, for the category of quantity, the number of words, verbs, modifiers, and sentences were tallied from all subjects. The average and standard deviation were also counted and then presented in the table of results.

5   Results

During the interview, the respondents were asked to provide stories without time restriction. As the icebreaker story was unplanned in the interview, all respondents spent less time in the stories compared to the other two types. This affects the entirety of the data from the icebreaker story in which the numbers in the following tables are lesser from the others. The numbers in the tables presented below are the mean and the standard deviation, where the latter is put in brackets.

5.1   Results of the category of quantity

Cue	Predicted Change in Deceptive Messages	Icebreaker Story	Truthful Story	Deceptive Story
Word Quantity	+	54.9 (28.98)	142.9 (82.93)	134.8 (75.82)
Verb Quantity	+	6.9 (3.84)	20.3 (15.8)	16.3 (13.56)
Sentence Quantity	+	5.4 (2.27)	13 (8.68)	12.9 (6.94)

The table above reveals that there are more words and verbs produced by the subjects in their deceptive stories. However, the mean of the sentences is almost the same in their truthful stories. Although this is not as it was expected, these results might be different if time limitation in the speakers' storytelling was applied.

5.2 Results of the category of specificity

Cue	Predicted Change in Deceptive Messages	Icebreaker Story	Truthful Story	Deceptive Story
Spatial Close Information	-	2 (1.63)	3.8 (2.44)	3.3 (2.71)
Spatial Far Information	+	0.6 (1.26)	1.3 (1.49)	0.9 (0.88)
Temporal Immediate Information	-	1.6 (1.42)	1.4 (2.37)	1.3 (2.16)
Temporal Non-immediate Information	+	0.4 (0.7)	0.7 (0.82)	0.9 (1.29)
Perceptual Information	-	0.6 (1.26)	0.9 (1.6)	0.3 (0.48)

As we can see in the table of specificity, findings on spatial far information are not as it was expected. This is because the subjects in the current experiment used many general locations in their truthful stories. The other cues such as spatial far information and temporal non-immediate information are in line with the prediction although the margin is not really high.

5.3 Results of the category of expressivity

Cue	Predicted Change in Deceptive Messages	Icebreaker Story	Truthful Story	Deceptive Story
Emotiveness	+	0.47 (0.34)	0.34 (0.17)	0.45 (0.16)

To obtain the results of the above cue, the number of adjectives and adverbs created by the subjects were divided by their number of nouns and verbs. As expected, the subjects produced more adjectives and adverbs in their deceptive stories as shown in the table above.

5.4   Results of the category of uncertainty

Cue	Predicted Change in Deceptive Messages	Icebreaker Story	Truthful Story	Deceptive Story
Modal Verb	+	0.5 (0.53)	0.7 (0.95)	2.1 (1.91)
Third Person Pronoun	+	0.7 (1.89)	0.3 (0.95)	1.7 (1.95)
Uncertainty	+	0.5 (0.85)	0.9 (1.29)	1.1 (0.74)

Results in the category of uncertainty are more significant compared to the results in the previous categories. The mean of modal verbs, third person pronouns and uncertainty words was found to be higher in the subjects' deceptive stories.

5.5   Results of the category of complexity

Cue	Predicted Change in Deceptive Messages	Icebreaker Story	Truthful Story	Deceptive Story
Average Sentence Length	-	10.39 (2.9)	11.99 (3.89)	10.53 (2.08)

To obtain the results for average sentence length, the total number of words is divided by the total number of sentences created by the subjects. The mean in this category proves that sentences in truthful stories are more complex than in deceptive stories.

5.6   Results of the category of diversity

Cue	Predicted Change in Deceptive Messages	Icebreaker Story	Truthful Story	Deceptive Story
Content Word Diversity	-	0.81 (0.11)	0.73 (0.12)	0.71 (0.06)
Function Word Diversity	-	4.16 (1.37)	5.01 (1.68)	4.81 (1.27)
Lexical Diversity	-	0.74 (0.09)	0.62 (0.12)	0.59 (0.1)

The content word diversity is acquired by dividing the number of different content words or terms and the total number of all content words. Meanwhile, function word diversity comes from the number of function words divided by the number of sentences. As for the lexical diversity, the number is obtained from the number of different words or terms created by the speakers divided by the total number of words or terms that they produced.

From the result table of diversity, the content word diversity and lexical diversity in general are higher in icebreaking and truthful stories. However, the function word diversity is slightly higher in truthful stories compared to deceptive ones.

5.7 Results of the category of non-immediacy

Cue	Predicted Change in Deceptive Messages	Icebreaker Story	Truthful Story	Deceptive Story
Group Reference	+	0.7 (1.34)	3.9 (4.68)	2.9 (3.14)
Self-Reference	-	1.8 (1.93)	4.5 (3.03)	7.5 (6.84)
Objectification	+	0.2 (0.42)	0 (0)	0.2 (0.42)
Passive Voice	+	0.6 (0.97)	1.5 (1.08)	1.8 (1.32)
Generalizing term	+	0.3 (0.48)	0.2 (0.63)	0.9 (1.2)

Deceptive story is expected to contain more group reference and less self-reference since a speaker wants to disassociate herself to the story. However, the results above show that there were less group reference and more self-reference in the respondents' deceptive story. As for the other cues, the numbers of objectification, passive voice and generalizing terms were higher when the respondents told imaginative stories.

6   Discussion

The research presented in this paper has attempted to uncover the possibilities of using cues to distinguish truthful stories from deceptive ones in Indonesian language use. Findings in this research support almost all cues predicted in Table 2. The findings in this paper are mostly not contradictory to previous ones. The only differences are in the categories of quantity and non-immediacy.

Previous research (De Paulo et al. 2003; Vrij, 2000; Hancock et al. 2008; Zhou et al. 2004) suggests that liars will produce more words and sentences because they want their story to become more believable. However, the subjects in this research produced fewer words and fewer sentences which made the overall quantity lower when they told untruthful stories. A deceptive speaker communicating less and rather briefly signifies one of the strategies of concealment as suggested by Dilmon (2009). Concealment is the removal of information essential in showing that what is being told is incompatible with the facts and reality. In this research, the subjects providing fewer sentences in their untruthful stories might be caused by the fact that they were more willing to tell the joyful stories since sharing positive experiences will heighten its impact on positive affect (Lambert et al. 2012). Though more words were produced in truthful stories, more adjectives and adverbs were created in deceptive ones. This is indicated by the higher mean of emotiveness cue in the later stories. Dilmon (2009) categorizes this as one of the techniques of persuasion where the deceiving speakers will exaggerate through the use of emotive words in order to impart credibility in their words.

Perceptual information in truthful stories has the highest mean in this paper. The information is produced in the form of sensory verbs, such as *lihat* 'to see' and *merasa* 'to feel'. Although these verbs appeared both in truthful and deceptive data, quoting was more common when the respondents told the truthful story. Quotation was made by respondent number 7.

(1) | *Jadi,* | *kayak* | *papasan* | *sebentar,* | *pas* | *trus,* | *pas* | *kita* | *sadar,* |
     |---------|---------|-----------|-------------|-------|--------|-------|--------|----------|
     | so | like | encounter | brief | when | then | when | we | realize |

'Itu si Anggun sudah datang!' Pas kita cek lagi,
that ART Anggun have come when we check again
ternyata si Anggun-nya sudah hilang, kayak gitu.
turn out ART Anggun-POSS have disappear like like that.

'So, it was like a brief encounter, and then when we realized, 'That's Anggun coming!' But when we checked it again, it turned out Anggun had gone, like that.'

Respondent number 8 was even able to produce a quotation of people that she had had a conversation with in her past experience.

(2) *'Dimana kamar mandi-nya. Kita mau mandi.' Trus teman saya yang*
where bathroom-POSS we want bath then friend I REL.PRON
orang Kanada bilang, 'Gak ada kamar mandi di sini. Cuman ada toilet aja.
person Canada say no exist bathroom in here only exist toilet just
'Trus, kalau kita mau mandi gimana?' Kebetulan ada danau
then if we want bath how by chance exist lake
'Itu berenang aja di danau!' Tapi hati-hati, suhu, suhu
that swim just in lake but beware temperature temperature
air-nya mungkin sekitar dua derajat selsius.
water-POSS may around two degree celcius.

'Where is the bathroom? We want to take a bath. Then my Canadian friend said, 'There is no bathroom here. There is only a toilet.' 'Then how do we bathe?' There happened to be a lake by chance there. 'Just swim in the lake. But, beware of the temperature; the water temperature may be around two degrees Celsius.'

Self-reference was predicted less but ended up being produced more in deceptive stories. Although on average, deceptive stories in this study contain more self-references, it appears to be a repetition in Indonesian texts. Unlike English, Indonesian has the same form for subjective, objective and possessive first person singular pronoun, that is *saya* 'I, me, my'. Example will be given below by respondent number 5.

(3) *Dan saya menjadi salah satu delegasi tersebut. Hal ini*
and I become one of delegation mentioned instance this
menyenangkan bagi saya karena sebelum~sebelum-nya saya mendapatkan
pleasure for I because before-POSS I get
penolakan, berupa email penolakan karena saya tidak pernah

rejection		in the form	email	rejection		because	I	no	have ever
berhasil	*mendapatkan*	*kesempatan*	*ke*	*luar*	*negeri*	*untuk*	*mengikuti*		*hal~ha*
success	get		chance	to	abroad	to	attend		instances
apa	*pun.*	*Dan*	*ini*	*merupakan*	*pengalaman*	*bagi*	*saya*	*yang*	
what	even	and	this	constitute	experience	for	I	REL.PRON	
cukup	*membawa*	*saya*		*hingga*	*sekarang,*	*di mana*	*saya*	*saat*	*ini sedang*
quite	bring	I		until	now	in where	I	time	this PROG
menempuh		*master degree.*							
take		master degree							

'And I became one of the delegations. This makes me happy because I have been rejected before by an email because I have never succeeded to get a chance to attend any event abroad. And for me, this experience quite affects me until now where I am studying my master degree.'

Group references were found to be higher in truthful story, which was predicted to be lower. This similar finding was also obtained by Fuller et al (2015). They argue that cue of group reference is highly context dependent. This can be the reason why the respondents in this study used many group references since their joyful experiences include friends and relatives so mentioning them could not be avoided.

While the majority of cues in Table 2 had different average margins between the truthful and deceptive stories, some of those cues resulted with such small margins that they do not make these cues effective enough to detect deception in Indonesian texts at least for this study. However, it may be that by using a different context and text in future research, these cues will become relevant. Thus, they should not be omitted at this early step at investigating deception in Indonesian texts. Furthermore, according to the previous results, cues that are relatively strong and applicable in detecting deception in Indonesian are the number of words and sentences, modal verbs, and pronouns or references. From these results, regardless of their different languages, it can be argued that the linguistic cues of deception are nearly identical.

The identification of cues of deception is the first stage in automation of deception detection in Indonesian. It is essential to identify those cues that can easily be implemented into the current technology. Since research and development of a technology in natural language processing continues to develop, it is arguably possible to say that all of the cues in this paper could be applicable using technologies in order to detect deception. A system that is created for automation deception detection can be based on machine learning techniques that drive weights for the various cues presented here. This can be done by taking advantage of machine learning algorithms. The machine will later be able to identify cues within a text before being able to justify whether the text is based on truthful or deceptive stories. Nonetheless, any machine, method or system for automation deception detection should be able to adapt to different contexts. For that reason, it is necessary to conduct similar research with respondents telling stories of different topics.

7   Conclusion and suggestion

The current research is based on the assumption that the speakers of Indonesian will have similar linguistic cues of deception compared to speakers of other languages. The results show that most of the categories presented above are aligned to previous research which was conducted mostly with English speakers. The hypothesis in the beginning of the research presented here is mostly supported, that when lying, the message of Indonesian speakers is more expressive, more uncertain, less specific, less complex, less diverse, and less immediate. Some hypotheses are not supported such as in the category of quantity. This is because there was no time limit given to the subjects and they were more willing to tell their real experience. The spatial far information in the category of specificity is not supported since the subjects used more general locations in their truthful story. The use of group and self-reference produced the opposite of what was expected due to context dependence of the stories that the subjects provided. Despite the results, it can be concluded that there are several categories of linguistic cues that can be great discriminators to differentiate truthful and deceptive stories in Indonesian language use. Those categories are quantity, uncertainty, and non-immediacy. This paper is among the very first research detecting linguistic based cues on deception speech in Indonesian language. However, I feel that overall description in this research is very broad and general. Hence, the future researchers who are interested in deception study can focus on specific categories in the cues. This way, the results will be more specific and focused. Finally, if this type of research has been done numerous times and in greater description, a machine for automatic deception detection could be further developed.

References

Adha, A. (2018): Linguistic based cues in detecting deception in Indonesian. *Prosiding Kongres Internasional Masyarakat Linguistik Indonesia (KIMLI) 2018*, 6-10.

Bok, S. (1978): *Lying: Moral Choice in Public and Private Life*. New York: Random House. https://doi.org/10.2307/1288137.

Buller, D.B. & Burgoon, J.K. (1996): Interpersonal deception theory. *Communication Theory* 6.1, 203-242. https://doi.org/10.1111/j.1468-2885.1996.tb00127.x.

Coleman, L. & Kay, P. (1981): Prototype semantics: the English word lie. *Language* 57.1, 26-44. https://doi.org/10.1353/lan.1981.0002.

Davis, R.C. (1961): Psychological responses as a means of evaluating information. In: Biderman, A.D. & Herbert, Z. (eds.): *The manipulation of Human Behavior*. New York: John Wiley & Sons Inc.

DePaulo, B.M., Malone, B.E., Lindsay, J.J., Muhlenbruck, L., Charlton, K. & Cooper, H. (2003): Cues to deception. *Psychological Bulletin* 129.1, 74-118. https://doi.org/10.1037/0033-2909.129.1.74.

Dilmon, R. (2009): Between thinking and speaking-Linguistic tools for detecting a fabrication. *Journal of Pragmatics* 41: 1152-1170. https://doi.org/10.1016/j.pragma.2008.09.032.

Donohue, W.A. (1991): *Verbal Intensity: Communication, Marital Dispute, and Divorce Mediation*. Hillsdale: Lawrence Erlbaum Associates.

Dulaney Jr., E.F. (1982): Changes in language behavior as a function of veracity. *Human Communication Research* 9, 75-82. https://doi.org/10.1111/j.1468-2958.1982.tb00684.x.

Dynel, M. (2018): *Irony, Deception and Humour: Seeking the Truth about Overt and Covert Untruthfulness*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9781501507922.

Fraser, B. (1991): Questions of witness credibility. *Working Papers Series Program on Negotiation*, 3-91.

Galasinski, D. (2000): *The Language of Deception*. Thousand Oaks: Sage.

Fuller, C.M. (2008): *High-stakes and Real World Deception*: *An Examination of the Process of Deception and Deception Detection Using Linguistic-based Cues*. Doctoral dissertation in Oklahoma State University.

Fuller, C.M., Biros, D.P., Burgoon J. & Nunamaker, J. (2013): An examination and validation of linguistic constructs for studying high-stakes deception. *Group Decis Negot* 22.10, 117-134. https://doi.org/10.1007/s10726-012-9300-z.

Fuller, C.M., Biros, D.P., Twitchell, D.P. & Wilson, R.L. (2015): Real-world deception and the impact of severity. *The Journal of Computer Information System* 55.2, 59-67. https://doi.org/10.1007/s10726-012-9300-z.

Grice, P. (1975): Logic and conversation. In: Cole, P. & Morgan, J.L. (eds.): *Syntax and Semantics*. New York: Academic Press.

Hancock, J.T., Curry, L.E., Goorha, S. & Woodworth, M. (2008): On lying and being lied to: A linguistic analysis of deception in computer-mediated communication. *Discourse Processes* 45.1, 1-23. https://doi.org/10.1080/01638530701739181,

Johnson, M.K. & Raye, C.L. (1981): Reality monitoring. *Psychology Review* 88.1, 67-85. https://doi.org/10.1037/0033-295X.88.1.67.

Jupe, L.M., Vrij, A., Leal, S. & Nahari, G. (2018): Are you for real? Exploring language use and unexpected process questions within the detection of identity deception. *Applied Cognitive Psychology* 32.5, 622-634. https://doi.org/10.1002/acp.3446.

Lambert, N.M., Gwinn, A.M., Baumeister, R.F., Strachman, A., Washburn, I.J., Gable, S.L. & Fincham, F.D. (2012): A boost of positive affect: The perks of sharing positive experiences. *Journal of Social and Personal Relationship* 30.1, 24-43. https://doi.org/10.1177/0265407512449400.

Lindsay, D.S. & Johnson, M.K. (1987): Reality monitoring and suggestibility: Children's ability to discriminate among memories from different sources. In: Zhou L., Burgoon, J.K., Nunamaker, J.F. & Twitchell, D. (eds.): *Group Decision and Negotiation*. https://doi.org/10.1007/978-1-4684-6338-5_6.

Mehrabian, A. & Wiener, M. (1966): Nonimmediacy between communicator and object of communication in a verbal message: Application to the inference of attitudes. In: Ceci, S.J., Toglia, J. & Ross, D.F. (eds.): *Children's Eyewitness Memory.* New York: Springer-Verlag. https://doi.org/10.1037/h0023813.

Mey, J. (1993): *Pragmatics: An Introduction*. Malden: Blackwell Publishers.

McCornack, S.A. (1992): Information manipulation theory. *Communication Monographs* 59, 1-16. (See in Yeung, L.N.T, Levine, T.R. & Nishiyama, K., 1999) https://doi.org/10.1080/03637759209376245.

McGrath, P. (2014): Why Good Memories Are Less Likely to Fade. Available from: http://www.bbc.com/news/health-27193607.

Miller, G.R. & Burgoon, J.K. (1996): Factors affecting assessments of witness credibility. In: Kerr, N.L., Bray, R.M. (eds.): *The Psychology of the Courtroom*. New York: New York Academic Press.

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. https://doi.org/10.1177/0146167203029005010.

Sharon, L., Vrij, A., Vernham, Z., Dalton, G., Jupe, L., Harvey, A. & Nahari, G. (2018): Cross-cultural verbal deception. *International Journal of Psychology* 23.2. https://doi.org/10.1111/lcrp.12131.

Stellar, M. & Koehnken, G. (1989): Criteria-based statement analysis. In: Raskin, D. (ed.): *Psychological Methods in Criminal Investigation*. New York: Springer.

Vrij A., Edward, K., Roberts K.P. & Bull, R. (2000): Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior* 24.4, 239-263. https://doi.org/10.1023/A:1006610329284.

Walker, W.R., Skowronski, J.J. & Thompson, C.P. (2003): Life is pleasant-and memory helps to keep it that way. *Review of General Psychology* 7.2, 203-210. (See in Jupe, L.M., Vrij, A., Leal, S. & Nahari, G., 2018). https://doi.org/10.1037/1089-2680.7.2.203.

Yeung, L.N.T, Levine, T.R. & Nishiyama, K. (1999): Information manipulation theory and perceptions of deception in Hong Kong. *Communication Reports* 12.1, 1-11. https://doi.org/10.1080/08934219909367703.

Zhou, L., Burgoon, J.K., Nunamaker, J.F. & Twitchell, D. (2004): Automating linguistic-based cues for detecting deception in text-based asynchronous computer-mediated communication. *Group Decision and Negotiation* 13.1, 81-106. https://doi.org/10.1023/B:GRUP.0000011944.62889.6f.

Ahmad Adha
University of Szeged
Theoretical Linguistics
H-6722 Szeged
Alföldi u. 2.
ahmad.adha@hung.u-szeged.hu