

**THESIS**

**MORPHOSYNTACTIC FEATURES OF DECEPTION ON ENGLISH TEXT  
WRITTEN BY EFL UNIVERSITY STUDENTS**

Written and Submitted by

**NURFAIZAH SAMSUR  
F022191025**



**ENGLISH LANGUAGE STUDIES  
FACULTY OF CULTURAL SCIENCES  
POST GRADUATE PROGRAM  
HASANUDDIN UNIVERSITY**

**MAKASSAR  
2022**

**MORPHOSYNTACTIC FEATURES OF DECEPTION ON ENGLISH  
TEXT WRITTEN BY EFL UNIVERSITY STUDENTS**

**Thesis**

**As a partial fulfillment to achieve Master Degree**

**Program**

**English Language Studies**

**Written and Submitted by**

**NURFAIZAH SAMSUR**

**to**

**POSTGRADUATE PROGRAM  
HASANUDDIN UNIVERSITY  
MAKASSAR  
2022**

THESIS

MORPHOSYNTACTIC FEATURES OF DECEPTION ON ENGLISH TEXT  
WRITTEN BY EFL UNIVERSITY STUDENTS

Written and Submitted by

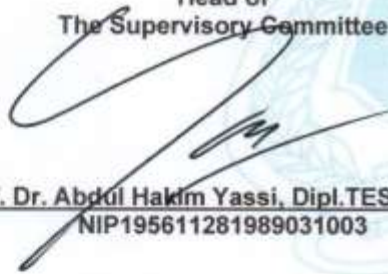
**NURFAIZAH SAMSUR**

**Register Number: F022191025**

has been defended in front of the thesis examination committee which was formed in order to complete the study of the Master Program in English Language Studies Faculty of Cultural Sciences Hasanuddin University on August, 15<sup>th</sup> 2022 and is declared to have met the graduation requirements.

Approved by:

Head of  
The Supervisory Committee



Prof. Dr. Abdul Hakim Yassi, Dipl.TESL., M.A.  
NIP195611281989031003

Member of  
The Supervisory Committee



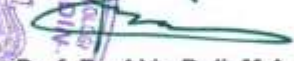
Dr. Harlinah Sahib, M.Hum.  
NIP196211281987032001

The Head of English Language Studies  
Program



Dr. Harlinah Sahib, M.Hum.  
NIP196211281987032001

The Dean of Faculty  
of Cultural Sciences



Prof. Dr. Akin Duli, M.A.  
NIP196407161991031010

## A STATEMENT OF THESIS AUTHENTICITY

The undersigned:

Name : NURFAIZAH SAMSUR

Register Number : F022191025

Study Program : English Language Studies

states truthfully that this thesis was the result of my own work. If it is proven that some part of this thesis is the work of others, I am willing to accept any sanctions for my dishonesty.

Makassar, August 1<sup>st</sup> 2022

The Researcher



Nurfaizah Samsur

## ACKNOWLEDGMENT



Alhamdulillah, Praise to Allah *Subhanahu Wataala*, the Most Gracious and High Merciful, for the health, strength, and guidance to the writer's presence in this world especially to accomplish her thesis as one of requirements to obtain her Magister degree in English Language Studies of Post Graduate Program in Hasanuddin University. *Shalawat* and *salam* also go to Prophet *Muhammad Sall Allahu 'alaihi wa sallam*, the greatest of humankind.

In this occasion, the writers' grateful thanks are due to the following people for their help, care, and advice, so that the writer can finish her study. Those people are as follows:

1. The writer would like to deliver her deepest gratitude to her beloved parents, **H. Samsur, S.Pd., M.Pd.** and **Dra. Hajrah, S.Pd., M.Pd.** for all the prayers, motivation, love, and care. Thank you for always being there to support every decision the writer takes. She is on her way to make you proud.
2. The writer expresses her deep appreciation to **Prof. Dr. Abdul Hakim Yassi, Dipl. TESL. MA** and **Dr. Harlinah Sahib, M.Hum** as the writer's first and second consultant for their time to help and guide the writer in writing her thesis.

3. The writer also expresses her deep gratitude to **Prof. Dr. Hamzah A. Machmoed, M.A, Dr. Abidin Pammu, M.A. and Dr. Sukmawaty, M.Hum** as the writer's examiners for their valuable comments, suggestions, corrections during examinations as well as their help during the process of this thesis compilation.
4. The writer would like to deliver her deep gratitude to **Dr. Harlinah Sahib, M.Hum** as the Head of English Language Studies of Hasanuddin University for all of motivation and care. Thank you for always being there to support the writer in completing her study.
5. All lecturers of **English Language Studies, Post Graduate Program, Hasanuddin University** for improving the writers' knowledge during her study. Not forgotten also, all the staffs that already help the writer during administration process.
4. The writer also dedicated her special thanks to all friends in **English Language Studies** especially for **Linguistics 2019/1** for all experiences, laugh and spirit.

As the conclusion, the writer hopes this thesis will give a good contribution for the students of English Language Studies and also for the future study about similar topic. The writer realizes that this thesis is still far from being perfect, therefore any suggestion or criticisms will be very much appreciated for the improvement of this writing. Thank you.

## ABSTRACT

**NURFAIZAH SAMSUR (F022191025):** *Morphosyntactic Features of Deception on English Text Written By EFL University Students* (Supervised by **Abdul Hakim Yassi** and **Harlinah Sahib**)

This research aims to identify: (1) the morphosyntactic features of Deception on English text written by EFL University students, and (2) the significant differences of morphosyntactic features of deception on English text written by native English and EFL University students.

The method used in this research was a descriptive-quantitative method. The data were taken from 10 participants who wrote deceptive and truthful stories. The words from the stories were calculated using corpus or word-count tools to obtain the total number of word sand the percentage of each category.

The results show that the Morphosyntactic features of deception on English text written by EFL University students consist of more quantity, less complexity, more uncertainty, more diversity, less specificity, and less emotiveness. Moreover, the differences of the cues shown by native English speakers and EFL university students lay on the category of quantity, diversity, non-immediacy, and the use of modality and third-person pronoun.

**Keywords:** **Morphosyntactic features of Deception; Deceptive Text; Truthful Text; EFL University students.**

## ABSTRAK

**NURFAIZAH SAMSUR (F022191025):** *Ciri-Ciri Teks Kebohongan Berbasis Morfosintaksis pada Teks Bahasa Inggris yang ditulis oleh Mahasiswa EFL (English as Foreign Language)* (Dibimbing oleh **Abdul Hakim Yassi dan Harlinah Sahib**)

Penelitian ini bertujuan untuk mengidentifikasi: (1) Ciri-Ciri Teks Kebohongan Berbasis morfosintaksis pada Teks Bahasa Inggris yang ditulis oleh Mahasiswa EFL (English as Foreign Language), dan (2) perbedaan ciri-ciri teks kebohongan berbasis morfosintaksis pada penutur bahasa Inggris asli dan mahasiswa EFL.

Metode yang digunakan pada penelitian ini adalah deskriptif-kuantitatif. Data dikumpulkan dari 10 responden yang menuliskan cerita bohong dan cerita jujur. Setiap kata dari cerita-cerita tersebut dijadikan korpus yang kemudian dikalkulasi untuk mendapatkan jumlah dan presentasi setiap indikator dari kategori-kategori yang mencirikan teks kebohongan.

Dari penelitian ini diperoleh bahwa ciri-ciri teks kebohongan yang ditulis oleh mahasiswa EFL adalah lebih banyak kuantitas, sedikit kompleksitas, lebih banyak ketidakpastian lebih banyak keragaman, sedikit kekhususan, dan sedikit emosi. Selanjutnya, perbedaan ciri-ciri kebohongan berbasis lingistik antara penutur bahasa inggris asli dan mahasiswa EFL terletak pada kategori kuantitas, keragaman, non-kedekatan, dan penggunaan kata kerja modalitas dan kata ganti orang ketiga.

Kata kunci; **Ciri-ciri Kebohongan Berbasis Morfosintaksis; Teks Kebohongan; Teks Jujur; Mahasiswa EFL.**



## TABLE OF CONTENTS

APPROVAL SHEET .....	<b>Error! Bookmark not defined.</b>
A STATEMENT OF THESIS AUTHENTICITY .....	<b>Error! Bookmark not defined.</b>
ACKNOWLEDGMENT .....	iii
ABSTRACT .....	vii
ABSTRAK.....	viii
TABLE OF CONTENTS.....	ix
LIST OF FIGURES .....	xi
LIST OF TABLES .....	xii
CHAPTER I INTRODUCTION .....	1
A. BACKGROUND OF RESEARCH.....	1
B. RESEARCH QUESTIONS .....	4
C. OBJECTIVES OF THE STUDY.....	4
D. SCOPE OF THE STUDY .....	4
E. SIGNIFICANCES OF THE STUDY .....	5
CHAPTER II REVIEW OF RELATED LITERATURE .....	7
A. PREVIOUS STUDIES .....	7
B. THEORETICAL BACKGROUND .....	11
1. DECEPTION.....	11
2. VERBAL AND NONVERBAL CUES OF DECEPTION.....	12
3. LINGUISTIC-BASED CUES OF DECEPTION IN ENGLISH .....	12
4. LINGUISTIC-BASED CUES OF DECEPTION IN INDONESIAN LANGUAGE USE .....	21
5. MORPHOSYNTACTIC FEATURES.....	22
6. WORD-COUNT TOOLS.....	23
C. CONCEPTUAL FRAMEWORK .....	28
D. OPERATIONAL DEFINITION .....	29

CHAPTER III RESEARCH METHODOLOGY .....	30
A. TYPE OF RESEARCH.....	30
B. RESEARCH TIME AND LOCATION.....	30
C. POPULATION AND SAMPLE .....	31
D. SOURCE OF THE DATA.....	31
E. METHOD OF DATA COLLECTION .....	31
F. METHOD OF DATA ANALYSIS.....	32
CHAPTER IV FINDINGS AND DISCUSSION.....	36
A. FINDINGS.....	36
A. RESULTS ON THE CATEGORY OF QUANTITY .....	36
B. COMPLEXITY .....	38
C. UNCERTAINTY .....	39
D. NON-IMMEDIACY .....	39
E. DIVERSITY.....	43
F. SPECIFICITY .....	43
G. EMOTIVENESS.....	44
B. DISCUSSION.....	45
CHAPTER V CONCLUSION AND SUGGESTIONS .....	51
A. CONCLUSION.....	51
B. SUGGESTIONS.....	52
BIBLIOGRAPHY .....	54
APPENDICES .....	57
A. DECEPTIVE AND TRUTHFUL TEXTS .....	57
B. DATA ACCUMULATION.....	64

## LIST OF FIGURES

Figure 1 Google Form to Submit Deceptive and Truthful Texts.....	30
Figure 2 Word-count Tools .....	33
Figure 3 TEXTINSPECTOR.....	34
Figure 4 LancsBox.....	34
Figure 5 Linguistic Inquiry Word Count (LIWC) .....	35
Figure 6 The Results on the Category of Uncertainty.....	37
Figure 7 The Distribution of Parts of Speech.....	37
Figure 8 Complexity.....	38
Figure 9 Modal Verbs .....	39
Figure 10 The Use of Passive Voice .....	40
Figure 11 The Distribution of Pronoun .....	41
Figure 12 The Accumulation of Pronoun Distribution .....	41
Figure 13 The Use of Self-reference, Group-reference, and Other- reference .....	42
Figure 14 Lexical Diversity.....	43
Figure 15 Specificity .....	44
Figure 16 Emotiveness .....	45

## LIST OF TABLES

Table 1 Deception category proposed by Zhou (2014) .....	15
Table 2 Tests Variation and its Cues to Deception.....	20
Table 3 Deception Platforms and their Cues to Deception .....	21
Table 4 Linguistic Cues of Deception in Indonesian Language Use .....	22

# **CHAPTER I**

## **INTRODUCTION**

### **A. BACKGROUND OF RESEARCH**

Deception is defined as a message or information transmission with an intention to create a false conclusion (Zhou: 2003). People deceive to influence others, make a good impression, convince others, maintain a relationship, manipulate others by disclosing their personal information and presenting the appealing ones, and calm down their partner when they quarrel (Ennis et al. 2008).

Messages can be deceived in many different ways. One of the deception cases that happened recently is an investigation conducted by CNBC which revealed that Mina Chang, the deputy assistant secretary in the State Department's Bureau of Conflict and Stability Operations, wrote her résumé containing misleading claims about her professional background to acquire a sophisticated job at the Trump Administration (DeLuce et al 2019). This supports a study showing that 63% of applicants lie on their CV and only 31% of them were caught (Gentle: 2020). Likewise, Forsyth and Anglim (2020: 1) stated that there have been various studies showing that job applicants lie to get the job they desire and that deception does not only occur on written curriculum vitae but also on a range of platforms such as social media profiles, personality tests, and employment interviews. This proves that some people lie to achieve

their life goals and are not afraid to get caught since deception cannot be easily detected at a glance, most particularly the deceptive text.

Detecting deception, however, has been widely investigated either in verbal cues of linguistic or non-verbal cues in psychology (Lowerse et al. 2010). Each verbal and non-verbal deception detection has its cues. The non-verbal behaviors deal with gaze, parts of the body movement, latency period, speech rate, frequency of smiles and laughs, and the frequency of stretching the head and wrist (Vrij et al. 2000: 246). On the other hand, verbal cues contain self-reference, emotion words, cognitive complexity (Newman et al. 2003: 666), unigram, psycholinguistic features, syntactic complexity (Perez-Rosas et al. 2015:2336), and a text analysis software applying NP model and RM Framework (Forsyth and Anglim 2019:5).

Linguistic features referring to as morphosyntactic features as verbal cues have been proven to be able to detect deceptive messages. However, there are still limited studies regarding detecting deception written by non-native speakers. Leach et al. (2019) state that people communicate in their non-native languages in many contexts such as border crossing or intelligent interviews. As a result, there are many cases related to deception on non-native speakers such as international purchase, border controls, international email communication, and international job application (Volz: 2019:256). Thus, the linguistic cues in detecting deceptive messages proposed by Zhou et al. (2004:94) which

contain morphosyntactic properties need to be further examined to find out the differences between native and non-native cues of deceptive messages. These linguistic-based cues consist of quantity, complexity, uncertainty, non-immediacy, expressivity, diversity, informality, specificity, and affect.

Besides, Adha (2020:25) stated that there are two significant differences in the cues between truthful and deceptive information given by the speakers of the Indonesian language. The first difference lies in the number of words, verbs, and sentences. Generally, people produce more words, verbs, and sentences when they deceive than tell the truth, but the result of the research is contradicted since deceivers produce less quantity than truth-tellers. Secondly, he found that Indonesian speakers' deceptive stories employ more group reference and less self-reference indicating that there is also a distinction in the category of non-immediacy.

Furthermore, the result of the observation conducted by Adha regarding the deceptive texts written by Indonesians in the English language shows that some linguistic cues specifically the morphosyntactic ones on deception are contradicted to the result of the previous research. Zhou (2003) claims that deceivers employ uncertain language when they tell lies. They tend to use more modal verbs, more modifiers, less self-reference, and more group reference. Meanwhile, English deceptive texts written by EFL University students contain fewer modal verbs, more self-reference, and less group-reference. This indicates that the

morphosyntactic features of deception cannot be generalized and needs further investigation most notably when dealing with non-native English speakers.

## **B. RESEARCH QUESTIONS**

The researcher formulates research questions to get the main purpose of this research as follows:

1. What are the morphosyntactic features of deception on English text written by EFL University Student?
2. How do the morphosyntactic features reveal the significant differences of deception by native speakers and EFL University students?

## **C. OBJECTIVES OF THE STUDY**

Based on the research questions stated above, the objectives of this research are as follows:

1. To identify the morphosyntactic features of deception on English text written by EFL University Student.
2. To examine the significant differences of morphosyntactic features of Deception by native speakers and EFL University students.

## **D. SCOPE OF THE STUDY**

This study deals with the deceptive texts written by EFL University students, specifically those who are currently studying at English



Language Studies (ELS) of Cultural Sciences Faculty of Hasanuddin University. To answer the first research question, the texts will be analyzed through their morphological and syntactical items and the frequency of their occurrences. The items comprise the number of words, phrases, and sentences, the length of words, clauses, and sentences, the parts of speech, the use of modals and modifiers as a sign of uncertainty, the misspelled words, the specificity, and the positive and negative emotion words. Besides, the answer to the second research question will be obtained from the comparison of Zhou's finding on the cues of deception and the cues found from the first research question. It should be noted that this study compares the cues of deception on the messages written by English native and non-native speakers and does not compare the text written by native and non-native speakers. Hence, the writer will not provide any data from native speakers.

## **E. SIGNIFICANCES OF THE STUDY**

From this study, it is expected that the results of the research contribute both theoretically and practically.

Practically, this research intends to be helpful for law enforcement most particularly in the field of forensic linguistics. Also, it is intended for the human resource personnel to be able to discover the fraudulent international applicants and for the immigration officers to be able to detect the potential criminal subject at borders.

As for the theoretical significance, this research will contribute to the reference of linguistic-based features of deceptive messages written by non-native English speakers. Moreover, the findings will hopefully be used to construct the Automating Linguistic-based features of deception detectors most notably when dealing with non-native speakers.

## **CHAPTER II**

### **REVIEW OF RELATED LITERATURE**

In this section, the writer presents some theories which relate to each other in supporting the analysis. This chapter consists of previous studies, theoretical background, and conceptual framework.

#### **A. PREVIOUS STUDIES**

The writer reviews some of the researches that are related to linguistic-based cues in detecting deception. For completing this research, the writer finds out some references which might be helping this research. Based on this part, the writer would like to present various journal articles and some theses that are relative and helpful to this analysis. The previous studies are as follows:

Zhou et al. 2004 suggested that the linguistic-based cues in detecting deception include the high rate of quantity, expressivity, positive affect, informality, uncertainty, and non-immediacy, and less complexity, diversity, and specificity for both senders and receivers of the text. The participants of the TA-CMC in the experiment were given roles randomly either as a sender or a receiver and either constructing deceptive text or telling truth after logging onto a designated web-based messaging system. The result showed that: (1) a systemic analysis of linguistic information can be utilized to detect deception, (2) while some linguistic features are effective

to be used as deception detector indicators, some others oppose the prediction of the prior research, and (3) the authors have added some other linguistic components which will be helpful to detect deceptive messages.

Evans and Michael (2013) further compared the ability to detect lies and truths in native and non-native English speakers and investigated the differences in the cues displayed via the *Psychologically Based Credibility Assessment Tool*. The research was conducted by assigning judges to analyze the stimulus videos containing truthful and deceptive statements. The result shows that different population characteristics and backgrounds influence the ability of deception detection.

Although they did not focus on the cues to deception, Mbaziira and Jones (2017) attempted to construct hybrid models for detecting cybercrime in the text. They used both computational linguistics and psycholinguistic features to analyze the native and non-native English speaking cybercriminal networks including fraud, scam, favorable fake reviews, and negative fake reviews. They found that computational linguistics and psycholinguistic features are reliable to detect cyber deception both in native and non-native English speakers.

On the other hand, Grondahl and Asokan (2019) studied the semantic domain of deceptive text based on writing styles so-called *stylometry*. The authors believe that detecting deception requires linguistic

and extra-linguistic analysis. They analyzed whether the deceptive messages can reveal the hidden author profile by investigating the writing styles. However, the result shows that stylistic markers of deception tend to be content-based and fail to reveal the liar identity.

Recently, Volz et al. (2019) claimed to be the first to examine the truthful and deceptive messages written by native and non-native English speakers. One hundred native English speakers judge the written messages and overall, they proved less likely to correctly classify either the non-native speakers' messages are truthful or deceptive. It is contrary to judging the native speakers' messages which they were more likely to be evaluated correctly. Unfortunately, this research is restricted to the classification either the judge can recognize the truthful or deceptive text. It does not reveal more on the cues the employ to judge the text.

Following Valz, Leach et al. (2019) examined whether observers' or judges' beliefs about deception were affected by the speakers' language proficiency. They employed verbal and nonverbal cues of deception modified from Akehurst's study of individual's belief and DePaulo's cues to deception. The result shows that the observers or judges believed that native and non-native speakers shared the same features of deception cues both verbal and non-verbal cues. Also, they claimed that non-native speakers would likely face several challenges due to the difficulties of understanding the interrogator's questions.

Besides, Adha (2020) attempted to replicate and apply the cues of deception proposed by the previous research to speakers of the Indonesian Language. He aimed to find out some linguistic cues of deception in Indonesian language use by asking participants to provide two stories (both deceptive and truthful stories) based on their personal experience. He found that some cues were significantly different in both stories such as word quantity, verb quantity, sentence quantity, third-person pronoun, group reference, and self-reference.

From the previous research listed above, it can be concluded that most of the research regarding deception in non-native speakers' messages focus on the detection through their psychological or non-verbal features of deception. The recent articles also tend to examine and analyze the judges' belief in non-native speakers' deceptive messages. There has not been any study which particularly focuses on the linguistic-based features of deception in non-native speakers. Thus, in this study, the cues of deceptive texts written by non-native speakers will be examined to identify either deception has similar features across languages.

## **B. THEORETICAL BACKGROUND**

### **1. DECEPTION**

A deception is an act of linguistic manipulation for certain goals. According to Zhou et al. (2004:140), deception in human communication happens when informants attempt to create a false impression in receivers. Likewise, Louwrese et al. (2010) stated that people lie to help them manipulate the impressions of other people such as thanking the waiters for guiding you to your table even though you are actually upset after waiting for 20 minutes but you only want to appreciate that they did their job. Besides, people lie to achieve social interaction goals, influence others, make a good impression, or reassure and support others (Ennis et al.: 2008: 106). Furthermore, Brundell (2011:16) defines deception in a more detailed description. He stated that

..... deception is a deliberate communicative act in which the intention is for one person (the sender) to bring about a false belief or conclusion in another person (the receiver) who does not know that they are being deceived.

Deception is regarded as fiction in which these two terms share similar definition. Fiction according to Deutch as elaborated in Standford Encyclopedia of Philosophy is the product of creative, imaginative activity made up by authors. Accordingly, Reeves-Everson (2016:135) distinguished deception and fiction where the former involves an intention to deceive or manipulate while the latter focuses on the presenting objects that are not

present in real life to be put in an artwork. Therefore, the difference lies on the intention of the sender to the receivers.

## **2. VERBAL AND NONVERBAL CUES OF DECEPTION**

According to Vrij et al. (2000:239), there are three ways to detect deception:

1. by observing how the suspect of liars behave (the particular body movement, the pitch of voice, speech rate, pause or repeat words)
2. by listening to what they say (the content of the speech)
3. by measuring their psychological responses.

## **3. LINGUISTIC-BASED CUES OF DECEPTION IN ENGLISH**

Newman et al. (2003) claimed that one way to differentiate between truthful and deceptive messages is by investigating the language people use. They also suggested that deceivers can be identified by their words, not by what they say but by how they say it. This, in turn, leads to the concept of linguistic-based cues of deception which will be elaborated as follows;

- a. Liars used first-person pronouns at a lower rate than truth-teller
- b. Liars used negative emotion words at a higher rate than truth-teller



- c. Liars used fewer “exclusive” word than truth-tellers
- d. Liars used more motion verbs than truth-tellers
- e. Liars used third-person pronouns at a lower rate than truth-tellers.

Additionally, Zhou et al. (2004:145) provide a more detailed measurement of deception detection. Those are on the following table;

Category	Component dependent variables
Quantity	Verbs: words that are the grammatical center of a predicate and express ac, occurrence, or model of being.
	Modifiers: adjectives and adverbs that describe words or makes the meaning of the words more specific.
Complexity	Average sentence length: $\frac{\text{Total number of word}}{\text{Total number of sentence}}$
	Average word length: $\frac{\text{Total number of characters}}{\text{Total number of words}}$
	Plausibility: $\frac{\text{Total number of punctuation marks}}{\text{Total number of sentences}}$
Uncertainty	Modal verbs: auxiliary verbs that are used with a

	verb of predication and express a modal modification.
	Passive voice: a form of the verb used when the subject is being acted upon rather than doing something.
Non-immediacy	Individual references: singular first and second personal pronoun.
	Group references: first personal plural pronoun.
Expressivity	Emotiveness: $\frac{\text{Total number of adjectives and adverbs}}{\text{Total number of nouns and verbs}}$
Diversity	Content diversity: $\frac{\text{Total number of different content words}}{\text{Total number of content words}}$ Where content words primarily express lexical meaning
	Redundancy: $\frac{\text{Total number of function words}}{\text{Total number of sentences}}$
Informality	Typographical error ratio: $\frac{\text{Total number of misspelled words}}{\text{Total number of words}}$
Specificity	Spatio-temporal information: information about locations of people or objects, or information about

	when the even happened or explicitly describes a sequence of events.
	Perceptual information: indicates sensorial experiences, such as sounds, smells, physical sensations, and visual details.
Affect	Affect: conscious subjective aspect of emotion apart from bodily changes
	Pleasantness: positive or negative feelings associated with the emotional state.
	Activation: the dynamics of emotional state.
	Imagery: words that provide a clear mental picture.

*Table 1 Deception category proposed by Zhou (2014)*

There are 9 (nine) categories proposed by Zhou as indicators to measure manipulative language. The first deals with the category of quantity derived from Interpersonal Deception Theory. The theory suggests that deceivers have an attempt to hold their words back resulting in the minimum words produced while telling lies. There is also possibility that their sentences are incomplete.

The second category is the complexity of the text. Liars tend to make up their stories which are lack of real and rich memory so they mostly get difficulties repeating the same words (Zhou et al., 2004). The more complex the text is, the more reliable the sentence is. Thus Zhou

(2004:98) and (Adda et al., 2019) proposed that deceptive language is less complex than the truthful one.

Uncertainty is one of deception cues which is derived from Interpersonal Deception Theory proposed by Buller and Burgoon in 1996 (Zhou et al., 2004). In verbal and non-verbal cues of deception proposed by DePaulo et al, it is stated that non-verbal cues of deception indicated uncertainty is when the speaker seems get difficulties answering the questions given by the interviewer and seems uncertain, insecure, or not very dominant, assertive, or emphatic; speakers seem while the verbal cue is indicated by the expressions of uncertainty such as “I’m not sure but” or “at least I believe it was like that” (DePaulo et al., 2003). Moreover, present research includes the use of modality and third person pronoun as Adda stated that the cues indicating uncertainty are the use of modal verbs, uncertainty words, and third person pronouns which he found as one of the most obvious categories to distinguish between deceptive and truthful texts. (Adda et al., 2019) also proposed that liars express greater uncertainty by using more modifiers and modal verbs in their text and found that deceptive text contain fewer modifier and modality.

The fourth category proposed by Zhou is non-immediacy. This category deals with the use of passive voice, self-reference, and group reference. The use of passive voice is assumed as an act to reduce a speaker’s ownership of a statement or to distance the speaker from the action being described. Another indicator to define non-immediacy is the

use of self-reference. Previous research found that liars dominantly used group reference instead of self-reference. However, recent findings regarding such indicator are variative. Adha (2020:25) found that liars used more self-reference and less group-reference when they were interviewed. On the other hand, Addawood et al. (2019:21) found that deceivers used less self-reference and group reference to indicate avoidance of having themselves on the story they told.

According to Zhou (2004) there are two extensions of the concept of reduced completeness when people lie. They include the reduced lexical (vocabulary) and content diversity. Lexical diversity is considered as one of Interpersonal deception theory (IDT) strategies and tactics proposed by Buller and Burgoon and have been summarized by Zhou. They believe that liars are less diverse when lying since they lack memories or they can not rely on the memory and this apparently affect the range of vocabulary used in describing events. Though Adha (2020) included this category as a cue to deception, it is not found on the research conducted by Newman et al. and Addawwod et al. This category refers to the range of different lexical used on truthful and deceptive texts.

It is expected that deceptive texts are less specific compared to the truthful ones. Specificity consists of spatio-temporal and perceptual information. Zhou et al. (2004:91) hypothesized that deceptive senders employ higher specify of language in their messages than truthful senders. However, they failed to present significant results and concluded that they

lack of dictionary of spatio-temporal and perceptual terms. On the other hand, Adha (2020) managed to find that deceptive stories have more spatial close information, more spatial far information, more temporal immediate information, less temporal non-immediate information, and less perceptual information although the margin between the truthful and the deceptive stories are not really high.

Liars, according to Newman et al (2003:672) used fewer “exclusive” words than truth-tellers. They employ fewer exclusive words indicated by the lower use of cognitive complexity. Speakers who use words such as “but, except, and without” are considered having rich memory as they are able to make distinction between what is happening in the story and what is not. Liars do not seem to have such typical cognitive load. Newman et al also proposed that liars tend to use negative emotion rather than positive ones. There is an indication that they feel guilty either because of telling lies or because the topic they lie about. The indicators on this category will be fully observed using LIWC as the corpus tool.

To date, Grondahl and Asokan (2019:8) conducted a literature review to summarize the linguistic cues of deception both experimentally elicited and non-elicited deception. The former means the researchers obtain data from experimental research or elicited deceptive text while the latter comprises the deception from real-world corpora such as the Enron e-mail corpus conference call transcripts, online dating profiles, and fraudulent scientific papers. The summary is displayed below;

<b>Tests variation</b>	<b>Deception Cues</b>
A theft-based game; a variant of the Desert Survival Problem	Quantity, reduced immediacy, expressiveness, informality, affect, reduced complexity, reduced verbs, modifiers, word length, punctuation, modal verbs, individual reference, group reference, emotiveness, content diversity, redundancy, perceptual information, spatio-temporal information, errors, affect, imagery, pleasantness, positive activation, positive imagery, negative activation diversity reduced specificity.
Two variants of the Desert Survival Problem	Verbs, modifiers, word length, punctuation, modal verbs, individual reference, group reference, emotiveness, content diversity, redundancy, perceptual information, spatiotemporal information, errors, affect, imagery, pleasantness, positive activation, positive imagery, negative activation.
Reported views about abortion, friendship, and a mock crime scenario.	Reduced first-person pronouns, reduced third-person pronouns, reduced exclusive words, negative emotion words, motion words.
Conversations	Quantity, questions, reduced first-person

between two participants	singular pronouns, other-directed pronouns, sense terms.
A questioner-responder game	Causation words, insight words, certainty words, first-person singular pronouns, present-tense verbs, and tenacity verbs.
Reported views about abortion, capital punishment, and friendship	Reduced self-related words and certainty-related words.

*Table 2 Tests Variation and its Cues to Deception*

Furthermore, they also categorize the cues of deception based on the deception platform. Below is the detail;

<b>Obtained data</b>	<b>Deception cues</b>
Enron e-mails	abstractness, negations, first-person pronouns
Conference call transcripts	general group references, reduced non-extreme positive emotion terms, reduced third-person plural pronouns
Online dating profiles	reduced first-person singular pronouns, negations, reduced word count, reduced negative emotion words
Fraudulent scientific papers	Words related to scientific methodology, amplifying terms, certainty-related words,



	emotional words, reduced terms, and reduced adjectives.
Enron e-mails	Modal verbs, base, and present tense verbs, second-person pronouns, function words.

*Table 3 Deception Platforms and their Cues to Deception*

#### **4. LINGUISTIC-BASED CUES OF DECEPTION IN INDONESIAN LANGUAGE USE**

Adha (2020:21) constructed the linguistic cues to differentiate truthful and deceptive stories which had been adapted from previous research. After examining the constructed linguistic cues, he then applied them to the Indonesian language context which led to a slightly significant difference. The cues of deception in the Indonesian language will be listed in the following table;

<b>Category</b>	<b>Linguistic cues</b>
Quantity	Less word quantity
	Less verb quantity
	More sentence quantity
Specificity	Less spatial close information
	Less spatial far information
	Less temporal immediate information
	Temporal non-immediate

	information
	Less perceptual information
Uncertainty	More modal verb
	More third-person pronoun
	More uncertainty
Complexity	Less average sentence length
Diversity	Less content word diversity
	Less function word diversity
	Less lexical diversity
Expressivity	More emotiveness
Non-immediacy	Less group reference
	More self-reference
	More objectification
	More passive voice
	More generalizing term

*Table 4 Linguistic Cues of Deception in Indonesian Language Use*

## **5. MORPHOSYNTACTIC FEATURES**

In systemic linguistics, the grammar or the linguistic system of a language comprises three levels; the semantic level, the lexicogrammatical level, and the phonological level (Morley:2000). The semantic level presents semantic and pragmatic matters of a language, lexicogrammatical level lies on the morphological and syntactical features

which later assigned as morphosyntactic features, and phonological level has phonology and phonetic description.

In detecting deception, Zhou et al. (2004) provided the Linguistic-based cues which focuses on the lexicogramatical or morphosyntactic level since the cues consist of words, phrases, clauses, sentences, parts of speech, modals, modifiers, and passive voices.

## **6. WORD-COUNT TOOLS**

### **a. LANCSBOX**

LancsBox is a new-generation software package commonly used to analyze language data and corpora developed at Lancaster University. It is used by linguists, language teachers, historians, sociologists, educators and anyone interested in the language field of science. There are features that can help the researchers or users to explore and analyze the data. The data is available for download on <http://corpora.lancs.ac.uk/lancsbox/> and it is free for all users.

The details of the feature and function taken from the official website of Lancsbox are as follow:

#### **1. KWIC**

The KWIC tool generates a list of all instances of a search term in a corpus in the form of a concordance. It can be used, for example, to:

- Find the frequency of a word or phrase in a corpus.

- Find frequencies of different word classes such as nouns, verbs, adjectives.
- Find complex linguistic structures such as the passives, split infinitives etc. using 'smart searches'
- Sort, filter and randomize concordance lines.

## **2. Whelk**

The Whelk tool provides information about how the search term is distributed across corpus files. It can be used, for example, to:

- Find absolute and relative frequencies of the search term in corpus files.
- Filter the results according to different criteria.
- Sort files according to absolute and relative frequencies of the search term.

## **3. Words**

The Words tool allows in-depth analysis of frequencies of types, lemmas and POS categories as well as comparison of corpora using the keywords technique. It can be used, for example, to:

- Compute frequency and dispersion measures for types, lemmas and POS tags.
- Visualize frequency and dispersion in corpora.
- Compare corpora using the keyword technique.

- Visualize keywords.

#### **4. GraphColl**

The GraphColl tool identifies collocations and displays them in a table and as a collocation graph or network. It can be used, for example, to

- Find the collocates of a word or phrase.
- Find colligations (co-occurrence of grammatical categories).
- Visualize collocations and colligations.
- Identify shared collocates of words or phrases.
- Summarize discourse in terms of its 'aboutness'.

#### **5. Text**

The Text tool enables an in-depth insight into the context in which a word or phrase is used. It can be used, for example, to

- View a search term in full context.
- Preview a text.
- Preview a corpus as a run-on text.
- Check different levels of annotation of a text/corpus.

#### **6. Ngrams**

The *Ngrams* tool allows in-depth analysis of frequencies of *ngram* types, lemmas and POS categories as well as comparison of corpora using the key *ngram* technique. It can be used, for example, to:

- Compute frequency and dispersion measures for *ngram* types, lemmas and POS tags.
- Visualize frequency and dispersion in corpora.
- Compare corpora using the key *ngram* technique.
- Visualize key *ngrams*.

### **7. Wizard**

The Wizard tool combines the power of all tools in #LancsBox, searches corpora and produces research reports for print (docx) and web (html). It can be used, for example, to:

- Carry out simple or complex research.
- Produce a draft report.
- Download all relevant data.

#### **b. LINGUISTIC INQUIRY WORD COUNT (LIWC)**

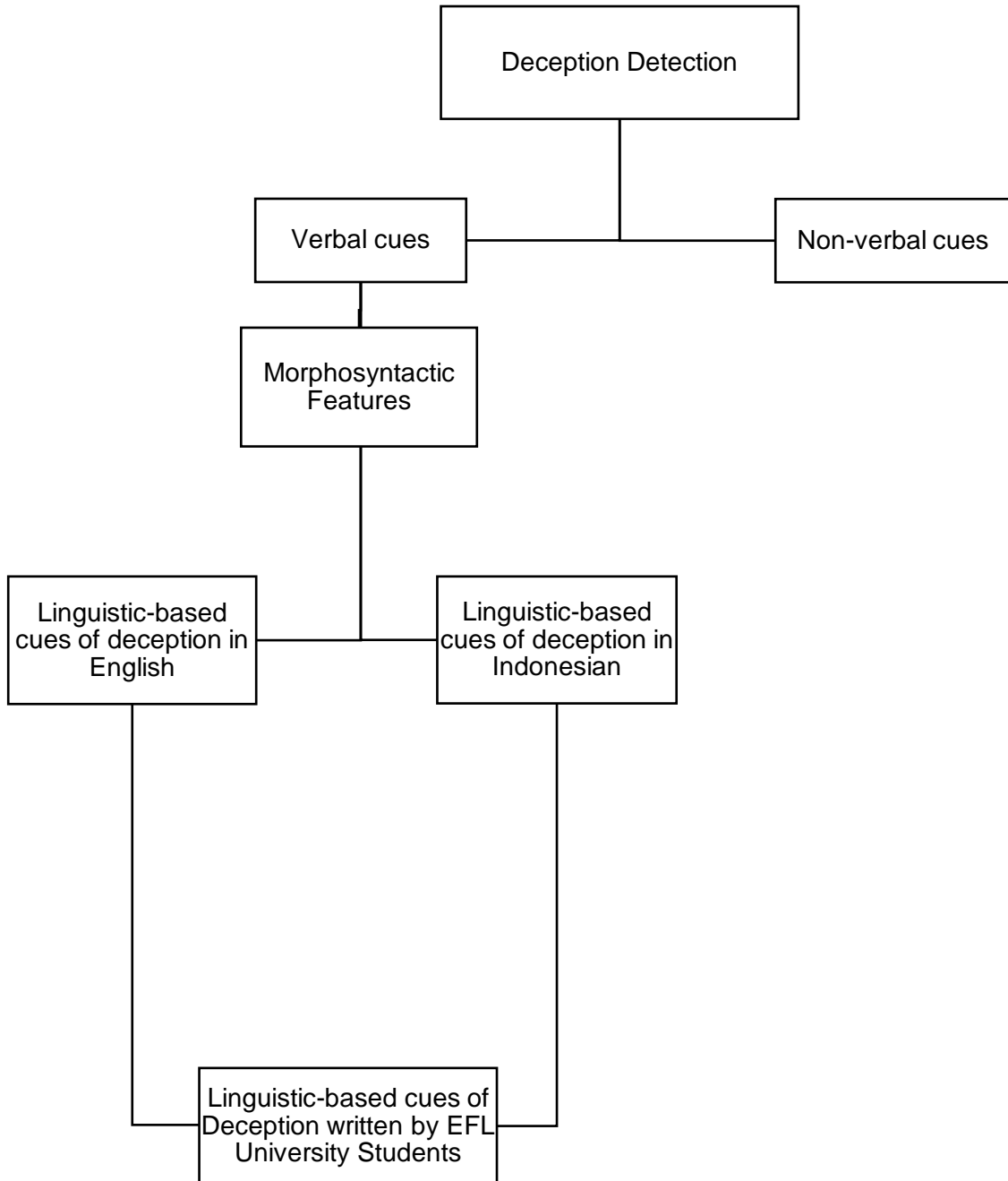
Another word-count tool used in this research is LIWC (Linguistic Inquiry Word Count). The LIWC program deals with the main text analysis module along with a group of built-in dictionaries. This text analysis program was created in the Java programming language and is available on PC and Mac computers. LIWC works with written or transcribed verbal texts which are stored in a digital, computer-readable form (such as text files). LIWC then compares each word in the text

against a user-defined dictionary. The dictionary identifies which words are associated with which psychologically-relevant categories.

According to the website, the software works by calculating the percentage of total words that match each of the dictionary categories. For example, if LIWC analyzed a single speech consisting 2,000 words and compared them to the built-in LIWC2015 dictionary, it might find that there were 150 pronouns and 84 positive emotion words used on the transcribed texts. In the end, it would convert these numbers to percentages, 7.5% pronouns and 4.2% positive emotion words.

This program has been widely used academically or commercially. Particularly, LIWC is a popular tool for text analysis (Forsyth and Anglim:2019). This can assist in several ways including operationalizing the assessment of language indicating deception. Hence many researchers on deception rely on this software.

### C. CONCEPTUAL FRAMEWORK





## **D. OPERATIONAL DEFINITION**

The definition of the terms used in this research is necessary to avoid any misconception. Below is the list of the terms;

1. Deception:

Deception is an act of manipulating others to believe as true or valid what is false or invalid. The word deception is synonymous to lying as they are used interchangeably in this research.

2. Deceptive stories:

Deceptive stories are fake stories written by the authors with an intention to make the readers believe that the stories are true and based on real experience.

3. Morphosyntactic features

Morphosyntactic features are the linguistic-based cues of deception comprising words, phrases, clauses, sentences, parts of speech, modals, modifiers, and passive voices.

4. EFL university students

English as Foreign Language University students are learners who are currently studying English in a university.

5. ELS students

ELS students are those who are currently studying English at English Language Studies in Faculty of Cultural and Sciences in Hasanuddin University