

Detecting Veracity in Selected Speeches of Egyptian Presidents (1956-2015) and American Presidents (1981-2015): A Psycholinguistic Corpus-based Study

Marina S. Badawy¹, Khaled A. ElGhamry², Radwa M. Kotait³

English Language Department, Faculty of Al-Asun, Ain Shams University

Abassey, Cairo

¹marinasameh1990@gmail.com

²elghamryk@gmail.com

³radwa_kotait@alsun.asu.edu.eg

Abstract: *Language and psychology have much in common in the sense that each of these two disciplines can be used to study the features of the other one. Words that people use every day carry a lot of explicit linguistic features which signal some implicit psychological traits, personal characteristics, social relations and cognitive processes. Since more attention has been recently paid to such an area of research, this study investigates linguistic and psychological features used as indicators of veracity or deception. The study examines selected speeches of the last five Egyptian and American presidents using a computerized content analysis tool called LIWC (Linguistic Inquiry and Word Count). Drawing on previous researches that examine the psychometrics of the language, this study follows an eclectic approach adopting Newman et al.'s model combined with other cues to deception concluded by other studies. The analysis detects the frequency of using pronouns, negation, exclusive words, details, conjunctions and big words in the speeches in order to reveal instances of deception.*

Keywords: *Psychometrics; Deception; Egyptian presidential speeches; American presidential speeches*

1 INTRODUCTION

Language that people use everyday is not only a medium of communication, but also a reflection and embodiment of people's feelings, opinions and attitudes. In other words, words that people use are representative of their thinking styles, emotional statuses, social relationships and personality traits. Investigating the psychology of language has been the interest of many psychologists, linguists and, most recently, psycholinguists who delve into how language has psychological and social dimensions. Sigmund Freud (1914) tackles this issue when he discusses the slips of the tongue which may reveal a lot about human intentions and hidden thoughts [1]. After Freud and throughout history, other researchers have developed different ways in order to examine what people say from a psychological perspective.

During the second half of the 20th century, as interest in such an area of study has rapidly grown, more advanced computer-based programs have been introduced for the same purpose. Quantitative approaches to text analysis have become highly popular. One of the most recently developed programs is Linguistic Inquiry and Word Count (LIWC) introduced by Pennebaker, Francis and Booth in 2001. It analyzes transcripts of everyday conversations, articles and literature, among others. It aims at categorizing the analyzed words into different user-defined dimensions of language, including standard language categories as well as psychological processes.

Given the effective results that such analyses provide, this text mining method can be used in different fields. Politics is one of the fields in which psycholinguistic analysis of language is widely applicable. The language of politicians, in general, and presidents, in particular, is constantly monitored and analyzed because it has aspects that are not common in other forms of language. In addition, it is, most of the time, linked to power and influence. Consequently, it merits analysis since it reveals a lot about the personalities of these politicians and their leadership skills.

The present study conducts a quantitative analysis of selected speeches of the last five Egyptian presidents (1956-2015); namely, Gamal Abdel Nasser, Muhammed Anwar El-Sadat, Muhammed Hosni Mubarak, Muhammed Morsi and Abdel Fattah El-Sisi, as well as the last five American presidents (1981-2015); namely, Ronald Reagan, George H. W. Bush,

Bill Clinton, George W. Bush and Barack Obama. Using LIWC as the tool of analysis, the study focuses on identifying cues of veracity and deception in the selected speeches providing comparisons when applicable. Thus, the significance of this study lies in the fact that it focuses on detecting aspects of language from a psycholinguistic perspective in the Arabic corpora; namely, Egyptian presidential speeches. This study addresses two major research questions. First, it investigates how far language features such as pronouns, negations, exclusive words and conjunctions contribute to depicting a picture of a president's authenticity. Second, it aims at highlighting significant differences between Egyptian and American presidents in their use of linguistic features, and in their authenticity.

Although computer-based psychoanalysis of language has been recently developed, psycholinguistic analysis has its roots back in early psychological theories. Basically, the father of psychoanalysis, Freud, is the first to pave the way. In one of his works (1914), he argues that slips of the tongue and parapraxes in general are expressions of people's motives and emotions [1]. Further contributions to psychoanalysis of language have been made by Gottschalk and Gleser (1969) who have developed a content analysis method which detects Freudian themes and the volume of mental or emotional states in verbal communication [2]. These early assumptions in psychology in relation to language have paved the way to the recent revolutionary approaches to content analysis. All of them are based on the hypothesis that all the types of the language that people produce are genuine manifestations of their inner selves. Since interest in psychoanalysis of language has recently grown, the development of computer-based tools facilitates the process of analysis and yields more authentic results.

One of the most recent content analysis tools is LIWC which has been developed in the framework of studying how health improvements can be predicted from the features of writing about unpleasant life experiences. The results yielded by LIWC are interpreted in terms of a number of findings made by studies in the psychometrics of language. These findings form the framework of the present study. In this sense, the study follows an eclectic approach based on a variety of sources which provide insights into how different linguistic categories can be possibly explained in terms of psychological and social dimensions to indicate authenticity or lack thereof.

Studying deception is an interdisciplinary quest. It has its significance and can be approached from different perspectives; social, psychological, political and linguistic. Verbal deception and veracity in this study are explained from a psycholinguistic point of view utilizing LIWC as a natural language processing tool. Ekman defines deception as 'a deliberate choice to mislead a target without any notification of the intent to do so' [3]. It has also been confirmed by Buller&Burgoon that the deceptive message intends 'to foster a false belief or conclusion by the receiver' [4]. Consequently, the authenticity or veracity of the speaker entails willingness to reveal the truth.

Several researchers participate in this area of research by developing different but supplementary models, for detecting veracity and deception. In Adams's model, she tests the efficiency of analyzing selected linguistic attributes in predicting veracity or deception. Her study depends on the narratives of suspects and criminals whose veracity or deception has already been determined [5]. Similarly, Bachenko et al. construct their model based on analyzing similar high stakes situations, such as; criminal testimonies and police transcripts [6]. Concerning Newman et al.'s model of detecting deception, it is developed in a study of the opinions of 200 undergraduate participants on abortion in three different ways: videotaping, typing and writing. Newman et al.'s model, which depends on LIWC in the analysis, is constructed to test whether these linguistic cues are genuine indicators of deception since the deception or veracity of the participants had already been determined [7]. Some of the findings of the three models on linguistic cues of deception and veracity are similar, while others are different.

A. *Use of Pronouns*

Adams as well as Newman et al. agrees on the low rates of using first person pronouns in deceptive communication [5-8]. Knapp et al. explain the reason behind using first-person singular pronouns at a low rate by stating that liars tend to "dissociate" themselves from what they describe [7]. Another reason behind this observation is that these liars lack personal experience [4]. Using first-person singular pronouns is also meant to omit the truth by referring to what did not occur [5]. Similar to almost all the previous findings, Pennebaker states that a sincere message is loaded with self-references as if claiming ownership [9]. Hancock et al. confirm the low rates of self-oriented pronouns, but notice more use of other-oriented pronouns in deceptive communication [10]; contrary to Newman et al.'s findings. This is consistent with previous literature on this matter. Knapp and Pennebaker state that using other-oriented pronouns aims at distancing the speaker and shifting responsibility [8]-[9]. The results are different in Newman et al.'s study due to the nature of the analyzed subject; attitudes towards abortion. One of the limitations of Newman et al.'s model is that it is only applicable to the English language because in Spanish, for instance, a sentence may lack a pronoun. Hence, the use of pronouns is not a reliable marker of deception or veracity [8].

B. Cognitive Complexity

Cognitive complexity is not related to deception as stated by Newman et al. Few exclusive words such as *except* and *but*, and some motion verbs such as *walk* and *move* mark inauthentic messages. Truth-tellers tend to make distinctions between what is true and what is false using more exclusive words causing more cognitive complexity [7]. Pennebaker confirms that truth-tellers are cognitively complex due to the high usage of exclusive words, big words, conjunctions and insight words [9]. Moreover, liars resort to using more motion verbs to help them make up their stories because they are easier and simpler than state verbs [7]. In addition, using these verbs keeps the fabricated story flowing and alive [9].

C. Negation

The three models conclude that high rates of negative expressions mark the inveracity of any communication. Negative expressions include explicit negations such as *no* and *not*, and negative emotions such *afraid* and *worried*. Ekman and Knapp suggest that the employment of negative emotions is explained in terms of sense of guilt or anxiety because of lying [3]-[8]. On the use of negation, Adams & Jarvis state that using negation as an answer to a yes-or-no question is not an indication of truthfulness or deception. Rather, in truthful statements, speakers need not use negations at high rates because they relate what they have already experienced, not what they did not do. On the contrary, liars may unconsciously use high rates of negations because they try to hide something they know, or because they are making up a false story about something they have not experienced before [11].

D. Other Linguistic Cues

Adams's model also associated deception to the use of equivocation, such as *maybe* and *probably*, as it reveals uncertainty. In her model, veracity is linked to the use of sensory details such as sounds and smells [5]. This is contradicted in the study conducted by Hancock et al. stating that sense-based words are used more by deceivers so that they can appear credible though it is difficult for a liar to mention such details [10]. It is more plausible that sensory words are mostly utilized by truth-tellers since it may need an efficient liar to flawlessly use them to prove authenticity. With regard to Bachenko et al., their model comprises 12 linguistic markers including linguistic hedges such as *I guess* and *sort of* which give a sense of vagueness and uncertainty like hedges. They also include unexplained lapses of time such as *later that day* and verb tense changes among other markers [6].

Furthermore, Pennebaker proposes that deceptive communications lack details of time, space, numbers, prepositions and motion. There is also minimal use of bigger words (more than 6 letters) which are not easily accessible to liars. Additionally, discrepancy verbs are highly used in deceptive communication because the speakers do not relate things that are true, but rather imagine things while describing them [9].

As shown, some linguistic cues of deception and veracity are agreed upon by a number of researchers. Upon testing them, they have yielded valid results with high rates of accuracy. This study utilizes Newman et al.'s model, in addition to some other indications of deception in analyzing the speeches in question. They mainly include pronouns, negation, exclusive words, details, conjunctions and big words. These linguistic cues are specifically selected for a number of reasons. First, they are applicable to the Arabic language as well as to the English language. Second, some of the Arabic texts included in the analysis are not delivered in standard Arabic so this may pose a problem in detecting some other cues because these colloquial words are not recognized by LIWC dictionary. Third, a number of the other linguistic features are not incorporated in the Arabic version of LIWC such as negative emotion, motion verbs and sensory words, which makes it insufficient to solely use Newman et al.'s model in analyzing the Arabic data.

2 RELATED WORKS

The findings of the studies discussed in the previous section, along with others that preceded them, have formed the theoretical framework of many of the studies that followed. Definitely, the three models discussed are considered a significant part of the literature of deception detection. Hancock et al. conduct a study on 66 participants randomly paired into 33 same-sex pairs. Their communication is computer-mediated and each one in a pair is asked to send text messages to the other via a software called *Net meeting* to discuss 4 topics; namely, a significant person in one's life, a mistake made recently, unpleasant job a person had before and responsibility. Each sender is required to deceive the receiver on two of the aforementioned topics without informing the receiver of the topics. The study produces 242 transcripts analyzed using eight LIWC linguistic indices. The study is based on 7 hypotheses expecting liars' verbal behavior.

The results reveal that deceivers use more words, more questions, less first-person pronouns and more other-reference pronouns as well as few causation words. No increase in using negative emotions or exclusive terms is detected. Some of

these findings may contradict other results of previous studies, but this can be explained in two ways. First, this study focuses on conversation context, unlike previous studies which depend more on monologues. Second, it highlights the linguistic correlation between the liar and the receiver because not only does it focus on the liars, but also on the reactions of the receiver. Whether liars are motivated, or not, is taken into consideration when accounting for the results. Actually, these characterizations of the study are what makes it distinguished among other studies conducted in this field [10].

Influenced by Hancock et al., Duran et al. conducted a similar study using the same transcripts of Hancock et al.'s study. In this experiment, another language-processing tool called Coh-Metrix is used to detect linguistic features related to deception from cognitive and social perspectives. The tool is similar to LIWC, but includes 700 linguistic indices and algorithm capable of evaluating collocations of words. The researchers chose a number of linguistic cues in Coh-Metrix that are somehow similar to those in LIWC to allow comparisons of the results of the two studies. Despite the fact that LIWC and Coh-Metrix share a number of similar results, Coh-Metrix may have an advantage over LIWC in some areas of deception detection such as accessibility, redundancy and complexity. Moreover, they do not produce the same results in some similar indices due to differences in their operationalization. On the contrary, LIWC's coded psychological dimensions, such as sensory information, have an advantage over Coh-Metrix since they account for how deceivers try to appear convincing [12].

Detecting (in)veracity in political texts using computerized text analysis tools has also been a growing interest. Slatcher et al. analyze 271 transcripts of interviews, debates and press conferences of the two presidential candidates John Kerry and George W. Bush, and the two vice-presidential candidates Dick Cheney and John Edwards. The study uses LIWC as a tool of analysis to examine differences among the four subjects across six linguistic categories; among which honesty and cognitive complexity have been examined. They propose an algorithm for honest language as follows: honesty = self-references + references to others + exclusion words - negative emotion words. They do the same when calculating cognitive complexity by summing exclusive words, tentative words, negation and discrepancy, and subtracting inclusive words. The results suggest that the most cognitively complex language is Cheney's and the least is Kerry's. Concerning honesty, the language of the vice-presidential candidates is found to be more authentic than that of the presidential candidates, with Cheney being the most honest of all four. Here, the positive relationship between cognitive complexity and honesty becomes clear [13].

Skillicorn & Little conduct a study on court testimonies which are considered non-emotional setting. They start out using the model developed by Newman et al. and end up modifying it in a way suitable for interrogative forms. In their revised model, deceptive messages are marked by obvious increase in all four linguistic cues to deception used in Newman et al.'s model [14].

A more recent study on deception in the political field is conducted by Skillicorn & Leuprecht. They analyze speeches given by candidates in the 2008 U.S. presidential election; namely, Barack Obama, Hillary Clinton and John McCain, employing Newman et al.'s model. The study also examines correlation between levels of deception and other factors: time, location, audience, content and polls. The results reveal that Obama's language is the most deceptive while Clinton's language is the least deceptive with McCain's in between. The level of deception varies across the time span of the elections which are divided into three time periods. Some factors affect the rises and falls in the levels of deception such as the time periods whether at the beginning or at the end. In addition, some national issues affect their deception such as the issue of economy which arose during the second period. Also, their confidence in the election results and polls plays a significant role in the level of deception adopted [15].

As mentioned, LIWC has been of great help in the domain of detecting deception in a variety of contexts. LIWC facilitates the processes of analyzing natural language and creates opportunities of discovering new untrodden areas. Nevertheless, much work has to be done in order to add to the previous research done and to transcend any limitations.

3 METHODOLOGY

A. Corpus Description

The Arabic corpora consist of selected speeches of the last five Egyptian presidents. The speeches are obtained from the State Information Service headquarters (SIS) as they are not available on its website. Not all the Arabic speeches have been transcribed, so there has been a process of transcription that preceded the analysis. The English corpora comprise selected speeches of the last five American presidents, and are retrieved from the American Presidency Project website (<http://www.presidency.ucsb.edu>). The total number of transcripts is 145. The criterion of selection is based on variety of the occasions of the speeches. In other words, the speeches selected for each president are separately given on special occasions of national distresses or problems, occasions of victory or achievement, and in times of stability. Moreover, the audience of all the selected speeches is the nation; which is considered a fixed variable.

B. Tool of Analysis

The primary language processing tool of analysis used in this study is LIWC 2015 developed by Pennebaker and his colleagues in 2001. As aforementioned, this tool is primarily developed in order to count and categorize words which reflect emotions, thinking styles as well as social issues and relations. 2015 version is capable of capturing around 85% of the words that people say or write. The words in any text are referred to as *target words* while the words incorporated in LIWC default dictionary are called *dictionary words* and they are 6400 words. The target words that are detected are matched with their dictionary counterparts, percentage of total words is calculated and the words are classified under one or more of LIWC's categories. At the same time, external user-defined dictionaries and categories can be uploaded to LIWC in order to fit certain different research objectives set by the users.

The output data include approximately 90 variables including four summary language variables, standard linguistic dimensions, categories of psychological construct and punctuation categories. The results are expressed as percentage of total words. Each dictionary entry may belong to more than one category. For example, the word *cried* if detected in the target words, it can be categorized under five categories: sadness, negative emotion, overall affect, verbs and past focus. Before running the files into the software, it is recommended to clean them from any unwanted texts. However, capitalization, grammar and sentence structure do not need to be corrected before the analysis since LIWC converts all files to lower case before processing them. LIWC dictionaries are translated into various languages including Spanish, Italian, German, Chinese and Arabic. More information about the program and testing it online is available at www.liwc.net/. Table 1 illustrates all the categories incorporated in LIWC [16]-[17].

The precise algorithms of how LIWC accurately works are not available due to commercial agreements. However, the Pennebaker Lab has published a number of papers that provide algorithms of how the four summary variables; namely, analytical thinking, clarity, authenticity and emotional tone are calculated. The one that is relevant to the present study is authenticity, and it is addressed in the aforementioned study of Newman et al. [7].

The authors of LIWC do not elaborate on how the software works with each language including Arabic. The author observes that the Arabic dictionary is similar to the English one to a degree. However, while the English dictionary is capable of capturing about 85% of what people say or write, the Arabic dictionary captures only around 40% of any text. The fact that some of the Arabic speeches contain some colloquial words contributes to the inability of the software to detect as many words as possible. The problem of undetectable Arabic words, both standard and colloquial, does not pose an obstacle on the analysis. The reason behind this is that this research mostly depends on function words in the analysis so the unrecognized words are not substantial. As per the problem of the colloquial function words, it is solved by manually adding a number of the common colloquial pronouns and negation words to the Arabic dictionary.

Another problem with the Arabic dictionary is that it does not include all the linguistic categories incorporated in the English one. Examples of these categories include positive and negative emotion words, verb tenses, common verbs, adverbs, motion words and cognitive words. Not all the categories are translated into Arabic, and the translated ones are based on LIWC default dictionary 2007 not the most recently updated one in 2015. Otherwise, LIWC process Arabic corpus using the Arabic dictionary in the same way it processes English corpus using the English dictionary.

The morphological analysis of pronouns does not pose an issue in Arabic corpus. The pronouns comprised in the Arabic dictionary include both detached personal pronouns such as "هو" و "نحن" و "أنا" as well as attached personal ones such as "ياء المتكلم" و "كاف الخطاب" و "هاء الغيبة". The latter set is attached to prepositions, and there are just over 1140 of attached and detached pronouns in the Arabic dictionary. Moreover, LIWC cannot be set to detect pronouns attached to verbs or nouns. Therefore, the analysis of pronouns in Arabic depends only on the pronouns that exist in the dictionary without posing any morphological obstacles to the analysis.

Concerning word sense disambiguation, it is taken into consideration to a high degree. There are some instances in English and Arabic that may cause ambiguities. Although LIWC is set to examine words and word stems, it may detect phrases such as "kind of" which is treated as a single word and categorized under the category 'tentative' instead of under the two categories 'adjective' and 'preposition'. A similar example is the phrase 'as well as'. However, hyphenated phrases such as 'this-or-that' are treated as three words because this phrase is not in LIWC's dictionary. Examples of common verb contractions such as 'I'll', 'she's' and 'you're' are embedded in the dictionary. However, examples like 'Jack's hat' are confusing because the contraction is considered verb contraction not possessive pronoun. Another problem is posed by common abbreviations such as 'U.S.' and 'Dr.' since they contain periods which signal the end of the sentence. In this case, the periods have to be removed without causing the creation of new words.

Concerning disambiguation in Arabic, it is observed that LIWC is sensitive to "الهمزة" which means the pronoun "أنا" is not categorized under first-person pronouns unless it is written with "الهمزة". This case is well dealt with in the built-in Arabic dictionary where all the words which contain "الهمزة" are included twice with and without it. In addition, this case is taken

into consideration while preparing the external Arabic dictionary. These examples of ambiguities in English and Arabic may lead to inaccurate and irrelevant results. However, such instances are very rare in the present study since the analysis mainly depends on function words which are not problematic regarding word senses. These scarce instances are taken care of while preparing the texts for analysis. Moreover, any irrelevant result detected is excluded from the total percentage.

C. Procedures

The Arabic and English speeches are analyzed in the same way. First, the researcher creates and uploads an external Arabic dictionary which includes a category that has not been present in LIWC Arabic dictionary but is there in the English dictionary; namely, “exclusive words”. This external dictionary also includes modifications and additions to a category that already exists, which is “negation”, because it has not been comprehensive enough. In addition, the dictionary comprises the category of “pronouns” that is already there in the Arabic dictionary, but it classifies the pronouns into first-person, second-person and third-person pronouns to guarantee specificity of results.

Second, each speech is uploaded onto LIWC to be analyzed. The target words which have matches with dictionary words, whether Arabic or English, are categorized under their respective categories. Third, after the texts are processed, LIWC outputs the results into percentages of the total words, revealing the set of categories relevant to the linguistic cues used to indicate veracity or deception in this study. A part of the analysis of Arabic data is manually done because the Arabic category of “conjunctions” lacks a substantial conjunction which is "ولو" in Arabic, so the researcher has to manually count it and add its percentage to the total percentage of “conjunction”. Fourth, some detailed results are added up to one another to limit the results as in compiling the results of the categories ‘time’, ‘place’ and ‘number’ into one category which is ‘details’. Then, the average use of each category by each president is calculated to draw a bigger picture of the use of each category. Finally, in order to get a simple and straightforward description of the veracity of each president, a formula is proposed based on the one developed by Slatcher et al. but has been modified to fit the data in question. It is made by summing the results of cognitive complexity, self-references and details, as well as subtracting the results of negation and other-references. The formula is as follows:

$$\text{Veracity} = \text{cognitive complexity (exclusive words + big words + conjunctions)} + \text{self-references} + \text{details (time + space + number)} - \text{(negation)} - \text{(other-references)}$$

Here, it is obvious how the proposed methodology integrates with LIWC. After the results are yielded in each category, given in percentages of the total words, the addition and subtraction follow in order to simplify the results into one percentage.

4 RESULTS and DISCUSSION

As mentioned earlier in this study, based on reviewed researches and experiments, a number of linguistic indices are highlighted in the analysis in order to draw attention to deception or veracity. They mainly include pronouns, negation, exclusive words, details, conjunctions and big words.

A. Cognitive Complexity

As a number of researchers state, low cognitive involvement positively correlates with deception. The calculation of cognitive complexity here is calculated by the sum of exclusive words, conjunctions and big words. Table 2 summarizes all the results of analyzing the Arabic and English data. It is clear that the language of both Morsi and Mubarak is more cognitively complex than that of Nasser, Sadat and El-Sisi. Moreover, the difference in language between presidents in each group is barely noticeable. However, Morsi uses conjunctions and exclusive words (18.17, 1.27 respectively) more than Mubarak does (16.73, 0.66). Actually, his use of these two categories surpasses the use of all the other Egyptian presidents. This entails that he tends to make a lot of distinctions to differentiate between what is in the category and what is not, or what happened and what did not happen. This reveals honesty and confers credibility of what he states.

TABLE I
LIWC CATEGORIES

Word count	Female references
Summary Language Variables	Male references
Analytical thinking	Cognitive processes
Clout	Insight
Authentic	Causation
Emotional tone	Discrepancy
Words/sentence	Tentative
Words > 6 letters	Certainty
Dictionary words	Differentiation
Linguistic Dimensions	Perceptual processes
Total function words	See
Total pronouns	Hear
Personal pronouns	Feel
1st pers singular	Biological processes
1st pers plural	Body
2nd person	Health
3rd pers singular	Sexual
3rd pers plural	Ingestion
Impersonal pronouns	Drives
Articles	Affiliation
Prepositions	Achievement
Auxiliary verbs	Power
Common Adverbs	Reward
Conjunctions	Risk
Negations	Time orientations
Other Grammar	Past focus
Common verbs	Present focus
Common adjectives	Future focus
Comparisons	Relativity
Interrogatives	Motion
Numbers	Space
Quantifiers	Time
Psychological Processes	Personal concerns
Affective processes	Work
Positive emotion	Leisure
Negative emotion	Home
Anxiety	Money
Anger	Religion
Sadness	Death
Social processes	Informal language
Family	Swear words
	Netspeak
	Assent
	Nonfluencies
	Fillers

Looking at the results of the English data, it is notable that they follow the same pattern of the Arabic results; there is no significant difference in results among the presidents. However, one may conclude that W. Bush's (Jr.) language is the most cognitively complex (32.85) while Obama's is the least (30.72). However, their usage of exclusive words suggests the opposite, since Obama's use of exclusive words(2.93) is the highest of them all, while W. Bush's (Jr.) the lowest (1.79).

It is also remarkable that the language of Egyptian presidents, as a whole, is more cognitively complex than that of the American presidents. However, more detailed results suggest an opposite pattern. For example, the American presidents' use of big words, in general, is higher than their use by the Egyptian presidents. The same applies to the use of exclusive words. However, significant increase marks the use of conjunctions by the Egyptian presidents. The researcher notices that using the conjunction "و" in Arabic represents a considerable part of conjunction usages as a whole; more than the half.

TABLE II
RESULTS OF THE ANALYSIS OF THE ENGLISH AND THE ARABIC SPEECHES

President	exclusive words	Big words	Conjunctions	Cognitive Complexity	self-references	details	negation	other-references
Abdel Nasser	1.07	17.61	17.21	35.89	.96	6.23	1.69	3.37
El-Saddat	1	17.46	17.5	35.96	.76	6.94	2.27	3.79
Mubarak	.66	21.28	16.73	38.67	.4	6.18	1.94	2.47
Morsi	1.27	18.62	18.17	38.69	.5	7.07	2.33	2.86
El-Sisi	1.14	18.67	15.58	35.39	1.09	5.47	2.37	3
Reagan	2.35	22.39	6.28	31.02	1.38	14.99	1.08	2.37
H. W. Bush (Sr.)	2.46	22.77	6.1	31.33	1.62	13.81	1.42	2.38
Clinton	2.4	23.14	6.49	32.03	1.66	14.16	1.07	2.12
W. Bush (Jr.)	1.79	23.37	6.513	32.85	1.4	12.89	.92	2.4
Obama	2.93	20.81	6.98	30.72	1.23	14.54	1.07	2.42

A. Pronouns

Although the Egyptian presidents' use self-references differently, the differences are not significant. However, it is noticeable that El-Sisi scores the highest use of first-person pronouns (1.09) compared to Mubarak whose use of these pronouns is remarkably the lowest (.4). Despite the observation that employing high rates of self-references indicates honesty, it has to be coupled with low rates of other-references, including second and third person pronouns. El-Sisi's use of other-references is relatively high (3) compared to their use by the other Egyptian presidents. This observation suggests that though his speeches are marked by veracity to a degree, he attempts to shift responsibility and dissociate himself from the situation.

The occurrences of other-references in the speeches of Abdel Nasser and El-Saddat are the highest (3.37, 3.79 respectively) although their use of self-references is within average (.96, .76 respectively). Obviously, they exhibit a lower degree of honesty than El-Sisi and a higher degree of others' involvement. Concerning Mubarak and Morsi, their use of both types of pronouns is the lowest suggesting a great deal of detachment.

The differences among the American presidents in their usage of self-references are similarly subtle highlighting Clinton as the highest user of self-references (1.66), while Obama the lowest (1.23). In addition, Clinton's lowest rate of using other-references (2.12) highlights his authenticity, followed by Reagan (2.37). Unlike Clinton, Obama's language exhibits the highest use of other-references (2.42) and lowest use of self-references (1.23). Reagan and W. Bush's use of both types of pronouns is moderate.

Comparing the results of the use of pronouns in the speeches of the Egyptian presidents and the American presidents, it is obvious that using self-references in the American speeches is higher than their usage in the Egyptian ones. On the contrary, using the other-references reveals the opposite observation. According to the use of pronouns as a linguistic indicator of deception, the language of the American presidents is relatively more authentic than the language of the Egyptian presidents.

B. Negation

The highest rate of using negation is found in El-Sisi's language (2.37), followed by Morsi (2.33), while Abdel Nasser's language is marked by the lowest rate of using negation (1.69). In the results of the American speeches, the highest use of negation appears in H. W. Bush's speeches (1.42) while the lowest mark W. Bush's speeches (.92). Negation is obviously more employed in the Egyptian speeches than in the American speeches. This note may suggest that Egyptian presidents tend to speak about what they do not do or experience.

C. Details

The calculation of details in this study includes the mention of time, space or place and numbers. Morsi's speeches feature the highest use of details (7.07), especially numbers, followed by El-Saddat (6.94). On the contrary, El-Sisi's use of details is the lowest (5.47) followed by Mubarak (6.18). Tracking this linguistic feature in the results of the analysis of

the American speeches reveals that the highest rate of using details occur in Reagan's speeches (14.99), followed by Obama (14.54), whereas the details in W. Bush's speeches are the lowest (12.89). Paying attention to the details in any speech suggests a lot about the president's specificity and straightforwardness.

According to the observations drawn from the results, the difference in word use among the presidents is quite observable. Describing the language of a person, consequently the person himself, as truthful or deceptive by no means depends on one linguistic feature because this may result into biased and incomplete results. For instance, one cannot state that a person is absolutely veracious because he/she uses self-references at high rates. Rather, combining a set of linguistic features together yields reliable and comprehensive results that can be generalized. In this study, the veracity of the language of each president is marked by high cognitive complexity and details, as well as low rate of negation and other-references.

The analysis of the results of President Clinton is an example to clarify how the process of analysis is conducted. LIWC separately yields results for the categories relevant to veracity. Compared to the other American presidents, Clinton's highest percentage of first-person pronouns and cognitive complexity, coupled with the lowest results of other-references suggest that he is the most authentic of them all. However, his use of negation and details is within average. In order to reach conclusive results concerning his veracity, the proposed formula is applied by summing up his results of cognitive complexity, self-references and details while subtracting the results of negation and other-references. The same methodology is followed relating to the other presidents. Table 3 illustrates the final results after calculating the veracity of each president descendingly arranged.

According to the results in table 3, neither the Egyptian presidents nor the American ones tremendously differ in their veracity. Among the Egyptian presidents, Mubarak is found to be the most veracious, closely followed by Morsi, whereas El-Saddat is the least veracious. Among the American presidents, Clinton is found to be the most authentic, followed by H. W. Bush while Obama is the least authentic. Generally speaking, the honesty of the language of the American presidents is higher than the honesty of the language of the Egyptian presidents.

Table III
RESULTS of VERACITY of EACH PRESIDENT

president	percentage of veracity	president	percentage of veracity
Mubarak	40.84	Clinton	43
Morsi	40.44	H. W. Bush	42.96
El-Sisi	38.5	W. Bush	42.64
Abdel Nasser	38.02	Reagan	42.56
El-Saddat	37.6	Obama	41.77

5 CONCLUSIONS

To conclude, this paper analyzes selected speeches of the last five Egyptian presidents and the last five American presidents. The corpora are analyzed using LIWC to detect the linguistic cues to veracity or deception. The study proposes a formula in order to simplify the final results of veracity. The results of the Egyptian speeches reveal that Mubarak's language is the most authentic, while El-Saddat's is the least. Among the American presidents, the analysis shows that Clinton's language exhibits the highest veracity, whereas Obama's language is characterized by the least veracity. The results also show that the American presidents are generally more honest than the Egyptian ones. Concerning Obama's results, they are somehow relevant to the previous study conducted by Skillicorn and Leuprecht on speeches delivered by presidential candidates in 2008. Obama's language is found to be the most deceptive of the three candidates, a result similar to the conclusion of the present study. Moreover, his language is marked by the low use of first-person pronouns, which is identical to the results of the present study though there are slight differences in the results of the use of negation and exclusive words.

The research in computerized analysis of deception and veracity is vast and growing. This study picks from where other previous studies have left, and it offers new insights into this matter for future work. One of the limitations of this study is that it cannot include the analysis of more linguistic features indicative of deception in Arabic because some categories, which are there in the English dictionary, do not exist in the Arabic one. Future studies may enhance the results through inserting the missing categories to the Arabic dictionary in LIWC. Analyzing more linguistic features

helps avoid bias and provides multi-faceted results. Moreover, the corpora used in this study can be expanded to include all the speeches of the presidents in question so that the results become more generalized. However, the study avoids bias as much as possible in many ways. First, as aforementioned, some additions, categories and dictionary words, are inserted into the Arabic dictionary as much as the resources needed are available. Second, Speeches delivered in a variety of occasions are selected in the corpus in order to provide a generalized account for the psycholinguistic features of each president regardless of the situational aspects.

To sum up, this study is an attempt to explore the area of veracity and deception from a psycholinguistic perspective. It widens the scope of the applicability of content analysis tools to extend to Egyptian presidential speeches, which is a relatively uncommon research quest. Moreover, it tests the efficiency of the computerized tool LIWC that has not been frequently used in Arabic exploring its strength points and weak points. The study also highlights some of the problems caused by the discrepancies between the English and Arabic versions of LIWC proposing solutions to some of them. It holds considerable promise for future work in this area and especially with a similar corpus. This is how this study is linked to language and processing and engineering.

REFERENCES

- [1] S. Freud, *Psychopathology of Everyday Life*, T. Fisher Unwin Ltd., 1914.
- [2] L. A. Gottschalk and G. C. Gleser, *The Measurement of Psychological States Through the Content Analysis of Verbal Behavior*, University of California Press, 1969.
- [3] P. Ekman, *Telling Lies: Cues to deception in the Marketplace, Politics and Marriage*, Norton, 2nd ed., 1992.
- [4] D. B. Buller, J. K. Burgoon, "Interpersonal Deception Theory," *Communication Theory*, vol. 6, no. 3, pp. 203-242, 1996.
- [5] S. H. Adams, "Communication under Stress: Indicators of Veracity and Deception in Written Narratives," Doctoral dissertation, the Faculty of the Virginia Polytechnic Institute and State University, Virginia, April 2002.
- [6] J. Bachenko, E. Fitzpatrick, and M. Schonwetter, "Verification and Implementation of Language-Based Deception Indicators in Civil and Criminal Narratives," in *Proc. of the 22nd International Conference on Computational Linguistics*, (COLING), pp. 41-48, Manchester, U.S.A, August, 2008.
- [7] M. L. Newman, J. W. Pennebaker, D. S. Berry, and J. M. Richards, "Lying Words: Predicting Deception from Linguistic Styles," *Personality and Social Psychology Bulletin*, vol. 29, no. 5, pp. 665-675, 2003.
- [8] M. L. Knapp, R. P. Hart, and H. S. Dennis, "An Exploration of Deception as a Communication Construct." *Human Communication Research*, vol.1, no. 1, pp. 15-29, 1974.
- [9] J. W. Pennebaker, *The Secret Life of Pronouns: What Our Words Say About Us*, Bloomsbury Press, 2011.
- [10] J. T. Hancock, L. E. Curry, S. Goorha, and M. Woodworth, "On Lying and Being Lied to: a Linguistic Analysis of Deception in Computer-Mediated Communication," *Discourse Processes*, vol. 45, no. 1, pp. 1-23, 2008.
- [11] S. H. Adams, J. P. Jarvis, "Indicators of Veracity and Deception: an Analysis of Written Statements Made to Police," *Speech, Language and the Law*, vol. 13, no. 1, pp. 1-22, 2006.
- [12] N. D. Duran, C. Hall, P. M. McCarthy, and D. S. Mcnamara, "The Linguistic Correlates of Conversational Deception: Comparing Natural Language Processing Technologies," *Applied Psycholinguistics*, vol.31, no. 3, pp. 439-462, 2010.
- [13] R. B. Slatcher, C. K. Chung, J. W. Pennebaker, and L. D. Stone, "Winning Words: Individual Differences in Linguistic Style among U.S. Presidential and Vice Presidential Candidates," *Journal of Research in Personality*, vol. 41, no. 1, pp. 63-75, 2007.
- [14] D. B. Skillicorn, and A. Little, "Patterns of Word Use for Deception in Testimony," *Security Informatics*, vol.9, 2010.
- [15] D. B. Skillicorn and C. Leuprecht, "Deception in Speeches of Candidates for Public Office," *Journal of Data Mining and Digital Humanities*, Episciences.org, pp. 1-41, 2015.
- [16] J. W. Pennebaker, R. L. Boyd, K. Jordan, and K. Blackburn, "The Development and Psychometric Properties of LIWC2015," Austin, TX: University of Texas at Austin, 2015.
- [17] J. W. Pennebaker, R. J. Booth, R. L. Boyd, and M. E. Francis, "Linguistic Inquiry and Word Count: LIWC 2015 Operator's Manual," Austin, TX: University of Texas at Austin, 2015.

BIOGRAPHY



Marina S. Badawi is a Teaching Assistant, Faculty of Al-Alsun, Ain Shams University. Research interests include cognitive linguistics, critical discourse analysis, corpus linguistics and psycholinguistics.



Dr. Khaled Elghamry is an Associate Professor of Computational Linguistics, Faculty of Alsun, Ain Shams University, Egypt. Dr. Elghamry obtained a Ph.D. in Computational Linguistics, 2004, Indiana University, USA. He was a visiting scholar, Department of Linguistics, University of Florida 2007-2010. Dr. Elghamry is the Co-Founder of the Midwest Computational Linguistics Forum, Indiana University, and also the Co-Founder of the Arabic Digital Content Statistical Database, and the Arabic Digital Content Foundation Report.



Dr. Radwa Kotait is a Lecturer of Linguistics and Translation, Faculty of Al-Alsun, Ain Shams University. She earned her Master's degree in Translation and Linguistics in 2007 on "Translating Conceptual Metaphors in the US speeches on the Middle East in the UN General Assembly (1990-present)". She also earned her PhD degree in Cognitive Translation Studies on "A Cognitive Approach to Translation Shifts: Using an LDC Arabic-English Parallel Corpus" in 2011. She is a freelance translator. She has taken part in translating several books and reports issued by international organizations such as the Organization of the Islamic Conference (OIC), the Euromed, as well as the International Service for Human Rights.

أساليب المكاشفة فى نماذج من خطاب رؤساء مصر (1956-2015) ورؤساء الولايات المتحدة الأمريكية (1981-2015): دراسة ذخائرية فى ضوء علم اللغة النفسى

مارينا بدوى¹، خالد الغمرى²، رضوى قطيط³

قسم اللغة الانجليزية، كلية الألسن، جامعة عين شمس

العباسية، القاهرة

¹marinasameh1990@gmail.com

²elghamryk@gmail.com

³radwa_kotait@alsun.asu.edu.eg

ملخص

يتلاقى علم اللغة مع علم النفس فى الكثير من النواحي، حيث يمكن الاعتماد على أحدهما فى فهم الآخر. إن كل كلمة ينطقها الإنسان تعد تعبيراً عن أنماط تفكيره، وحالاته النفسية، وقدراته الإدراكية، بالإضافة إلى علاقاته الاجتماعية. فخلال العقود القليلة الماضية، زاد الاهتمام بدراسة الأبعاد النفسية والاجتماعية للغة. لذا، تهدف هذه الدراسة إلى الكشف عن الملامح اللغوية المرتبطة بأساليب الصدق والخداع فى خطاب مختارة لآخر خمسة رؤساء مصريين فى الفترة ما بين 1956 وحتى 2015، وخطب مختارة لآخر خمسة رؤساء أمريكيين فى الفترة ما بين 1981 وحتى 2015. فى سبيل ذلك، تستخدم الدراسة برنامج *Linguistic Inquiry and Word Count (LIWC)* الذى يقوم بتحليل الكلمات فى النص إلى تصنيفات نفسية واجتماعية، وبيولوجية، وإدراكية مثبتة فى البرنامج لتحليل أى نص من هذا المنظور. تنتهج الدراسة المنهج الانتقائى فى التحليل حيث تعتمد على نموذج نيومان وزملائه (2003) بالإضافة إلى أجزاء من نماذج أخرى نتجت عن دراسات سابقة. يقوم تحليل النصوص بالكشف عن ملامح لغوية مثل: الضمائر، والنفى، و تفاصيل المكان والزمان، و كلمات الاستثناء.

الكلمات الدلالية: دراسة ذخائرية، الصدق والخداع، المنهج الانتقائى، خطاب رؤساء مصر والولايات المتحدة الأمريكية