Discourse Tagging of Political Speeches: A Corpus-based Study

Marwa Adel Abu El Wafa[*1], Sameh Alansary[**2], Shadia El Soussi[***3]

[*]Language and Translation Department, College of Language and Communication, Institute for Language Studies
Arab Academy for Science, Technology and Maritime Transport
Miami, Alexandria, Egypt
[1]marwa.adel.abuelwafa@aast.edu

[**]Phonetic and Linguistics Department, Faculty of Arts, University of Alexandria
ElShatby, Alexandria, Egypt
Bibliotheca Alexandrina, Alexandria, Egypt
[2]sameh.alansary@bibalex.org

[***]Institute of Applied Linguistics, Faculty of Arts, University of Alexandria
ElShatby, Alexandria, Egypt
[3]shadia.elsoussi@bibalex.org

Abstract—— **This paper discusses the creation of a tag set on the discourse level through tagging various rhetorical devices employed by both the American President Barack Obama in seven of his speeches and the African American leader Martin Luther King in seven of his speeches. This is done on the path of discourse tagging as a means of creating a discourse-based tag set of the devices and annotated corpus of political speeches. This tag set is meant to be fed into a concordance program namely MonoConc Pro 2.2. Once the speeches are manually annotated by the researcher, the tagged speeches are then analyzed by the concordance program searching for and counting the frequencies of the devices. The results help draw conclusions about the style of each character as well as the similarities and differences between each. This study might open the way for creating a discourse based corpus that can be used by other researchers experimenting in the same field.**

1. INTRODUCTION

The study is not a critical discourse analysis study, but rather a corpus-based study. This corpus-based study aims at building an annotated corpus of political speeches. The annotations are tags created by the researcher. This is from where the originality of the present study stems. To the researcher's knowledge, there are no studies on annotating a corpus of political speeches on the discourse level neither in English nor in Arabic. The created tag set is totally original as it is created by the researcher for the purpose of analyzing the present corpus. The study aims at creating an annotated corpus on the discourse level through analyzing speeches searching for the rhetorical devices used by the politicians to bond with their audience and evoke their emotions. Two speakers are selected for this study; the African American leader Martin Luther King and the American President Barack Obama. Through the rhetorical devices, politicians convince their audience with a certain frame work or view points or certain perspectives. In the present study the term rhetorical devices is the one used to refer to the devices which in other studies may be referred to as stylistic devices. The political speeches chosen serve as the corpus of the study.

2. STUDY SIGNIFICANCE

The originality of this study stems from the endeavor of creating an output of annotated corpus on the discourse level. A corpus of this nature has never been available before to help researchers portray the style of different writers or speakers. As a matter of fact, the shortage of annotated data for linguistic and language engineering research was a motive behind conducting this study. An annotated corpus is rich with linguistic data which can open the door to multiple linguistic and language engineering researches whose results open gates for language users in general and reveals secrets about language. This research could be one of the fewest resources in discourse tagging as there are few endeavors to tag on the discourse level. Moreover, the study aims at creating a discourse- based tag set. This tag set stands for the selected rhetorical devices for the study. This can enable other researchers to analyze texts in terms of the language used with a press of a button bringing out numbers of used rhetorical devices. Through annotating the texts using the assigned tags, the researcher can arrive at clear numbers of the occurrences of the devices. Consequently, conclusions can be drawn to

identify and describe the style of the different writers or speakers whether belonging to the political field or any other field.

3. DEFINING RHETORIC

Rhetoric has been defined as the art of speaking or writing effectively as a means of communication or persuasion. It is also a skill in the effective and creative use of speech and the use of language. Rhetoric is a tool that is used to enrich language in order to persuade, inform, express ideas and entertain. It is no surprise that the skill of persuasion is often in evidence with great politicians or religious leaders throughout history. Using rhetoric and its devices, a writer or speaker is capable of invading audiences' minds and changing or guiding their perspective. Rhetoric gives the power to communicate diverse messages through the use of powerful imagery or referring to reputable figures thus evoking emotions and creating the bond needed with the audience. Persuasion, although is present as an aim of any use of language, is viewed as one of four aims of using rhetoric. Informing is the second aim. Using rhetoric to inform may not appear as powerful as when it is used to persuade. Informing is clear in cases as teaching. A teacher uses the tools of rhetoric to bring ideas closer to the learners.

In rhetoric, a rhetorical device is any of the techniques that an author or speaker uses to convey to the audience a meaning with the goal of persuading him or her towards considering a topic or a number of topics or an ideology different from or similar to his or her from a different perspective. Not only do rhetorical devices evoke an emotional response in the audience and consequently bond them with their politicians, but also the main goal behind using them is to persuade the audience towards a particular frame of view, view point or a particular course of action. In this sense, appropriate rhetorical devices are used to shape the language that is designed both to make the audience receptive through emotional changes and to provide a rational argument for the frame of view, view point or course of action.

4. THE RHETORICAL DEVICES AND THEIR CATEGORIES

The selection of the devices was done in a very cautious manner. They are grouped according to their function into four classes. Each group or class encompasses devices that are employed for a certain purpose and a certain effect. The devices that belong to the first category are known to be used to present a strategy or point of view. The second group includes devices that give depth to the argument through stressing the ideas in a certain manner. The third group embraces devices that are used to organize the ideas. The fourth group includes devices that give a distinctive style to the writing. The four categories are presented as they conventionally appear in the literature. Devices that share the same or similar effect or purpose are grouped together under the same category. Neither the devices nor the categories are presented in priority order. Therefore, they could be alternating.

5. TAGGING THE DEVICES

Once the devices are selected, the phase of designing the tag set starts. These tags play the role of codes that stand for each device of the thirty five devices selected. These tags are used in the analysis stage and are annotated into the corpus selected. The tags assigned to the parts of speech are either one capital letter or three capital letters. For instance, verbs take the tag V, nouns N, prepositions P, adjectives ADJ and determiners DET. The second pattern is used by the researcher for the rhetorical devices chosen in the study. Three capital letters that resemble the device's pronunciation are given to each device as a tag (Bird & Liberman, 1999). Table I displays all the devices included in the study, the meaning of each and their tags.

A. Annotation

In the present study the term 'annotation' is used to refer to the process of adding interpretative linguistic information to the corpus (Bird &Liberman, 1999). Any act of corpus annotation is, by definition also an act of interpretation, either of the structure of the text or of its content. An unannotated corpus is simply a raw text where linguistic information and linguistic phenomena are hidden. On the other hand, an annotated corpus transforms texts into banks of linguistic information available for investigation and analysis. Annotating a corpus helps make the retrieval and extraction of linguistic information and the study of linguistic phenomena easier and faster thus enabling researchers to arrive at findings that would not have been feasible without the presence of an annotated corpus. Annotated corpora make up reusable resources for many researchers with multiple purposes. Hence, a linguistic database is available for analyses and studies can be compared and contrasted adding richness to the field. There are many levels of corpus annotations starting with the phonological moving to the morphological, then the lexical and finally the highest level which is the discourse level (Masahiro, Yukihiko, Takuya&Niimi, n.d.).

B. Leech's Annotation Maxims

The linguistic information that is added to a corpus is governed by the seven maxims of Leech. According to Leech, there should be flexibility in dealing with the annotated corpora. In other words, after annotation there should still be the possibility of recovering the corpus to its raw state. If the first maxim is the head of the coin, then Leech's second maxim is actually its tail. The first and the second maxims accentuate that on one hand the corpus can be regressed to its raw state without the annotations and on the other hand the annotations themselves can be solely extracted from the corpus. The first two maxims are put in such manner so as to ensure maximum flexibility for the manipulation of the corpus by the user. This totally applies to the corpus in the present study. In other words, the tags can be removed from the corpus and it can appear in its raw state once more. This is because the tags are not inserted into words and so removing them would destroy the words, but rather surround extracts. The third maxim is concerned with the end user and so stresses on the availability of clear guidelines for the annotation scheme adopted by the researcher. For this reason, a clear description of all the chosen rhetorical devices and their corresponding tags is given to ensure that other users can benefit from the present study in future research. The fourth maxim confirms that it should be made obvious how and by whom the annotation was performed. In the present study the corpus is manually annotated by the researcher. Manual annotation is one of the types of annotation which is highly valued for its accuracy.

Table I
The Rhetorical Devices and the Tag Set

Device	Description/Function	Tags
Allusion	A short reference to a famous person, event, history, Greek mythology, literature or reference to religion.	ALU
Understatement	A statement consciously weakened or expressed as less important than it actually is, either to soften the message for politeness and tact or to sound ironical.	UNS
Litotes	A figure of speech generated by denying the opposite or contrary of the word which otherwise would be used. It is a form of understatement. Litotes intensify the sentiment intended by the writer.	LTO
Antithesis	Opposition or contrast of ideas or words expressed often in parallel construction. It emphasizes the contrast between two ideas to draw the readers' attention directly to the contrast.	ANT
Hypophora	Question raised and then answered by the author / speaker.	HYP
Rhetorical question	Question without a direct answer. It is used for effect, emphasis, or provocation, or for drawing a conclusionary statement from the facts at hand.	RHQ
Procatalepsis	Allowing an argument to continue through anticipating an objection and answering it, putting into consideration points or reasons opposite to the train of thought.	PRO
Distinctio	Offering the meaning or meanings of a word in order to remove ambiguity.	DST
Simile	A direct comparison between two different things that resemble each other at least in one way, often by using the words' *like'* or *'as'*.	SIM
Analogy	Overlaps with similes Comparing two things with similarities in several aspects without adding *'like'* or *'as'*.	ANG
Metaphor	Comparing two totally different things by asserting that one thing *is* another thing.	MET
Eponym	A particular attribute of a famous person famous of such attribute.	EPM
Exemplum	Citing an example through offering an illustrative story.	EXM
Sententia	A means of quoting a wise saying or a statement of wisdom.	SNT
Anaphora	The same word or phrase is used to *begin* successive clauses or sentences. This draws the readers'/listeners' attention to the message of the sentence.	ANA

Epistrophe	The counterpart of anaphora where the repeated part comes at the end of successive phrases, clauses or sentences.	EPS
Symploce	Combining anaphora and epistrophe. This is displayed by repeating one word or phrase at the beginning and another is repeated at the end of successive phrases, clauses or sentences.	SYM
Personification	Metaphorically representing inanimate objects or animals or abstract terms as having human qualities.	PER
Amplification	Repeating a word or expression while offering more details as a means of emphasizing its importance.	AMP
Aporia	Expresses doubt about an idea or conclusion. It is a way to raise a number of choices without being obliged to any of them.	APR
Climax	Climax consists of arranging words, clauses, or sentences in an ascending order or the order of increasing importance for continuity and emphasis.	CLX
Parallelism	Similarly structuring successive clauses or sentences as a means to concentrate on the message to show that the ideas in the parallel structures are equal in importance as well as to create a musical effect.	PAR
Chiasmus	It is usually called 'reversed parallelism', because the second part of a grammatical construction is paralleled with the former but in reverse order.	CIA
Metabasis	A brief statement of what has been said before and what will follow. It acts as a sort of transitional summary to keep the discussion ordered and keep the audience focused.	MTA
Anadiplosis	The last word of one phrase, clause or sentence is being repeated at the beginning or very near to the beginning of the next.	AND
Conduplicatio	A key word is being repeated from a preceding phrase, clause or sentence to the beginning of the next.	CND
Apostrophe	Interrupting the discussion and directly addressing a person or personified entity either present or absent.	APS
Polysyndeton	The use of a conjunction between each word, phrase, or clause as an attempt to encompass something complex.	POL
Asyndeton	Omitting conjunctions between words, phrases, or clauses as an attempt to give the effect of multiplicity and spontaneity. It is the opposite of polysyndeton.	ASN
Zeugma	Zeugma includes grammatically linked parts of speech by another part of speech. This is done with two or more parts of speech.	ZGM
Synedoche	Any portion, section, or main feature stands for the whole itself or vice versa.	SYN
Metonymy	Another form of metaphor The thing chosen for the metaphorical image is closely associated with the subject with which it is compared.	MTN
Alliteration	Repetition of the *initial consonant sound* in neighboring words. Alliteration draws attention to the phrase and is often used for emphasis.	ALT
Expletive	A single word or short phrase, usually interrupting normal syntax, used to lend emphasis to the words immediately proximate to the expletive. The expletive can be placed at the beginning, middle or at the end. The words on each side are emphasized in order to maintain continuity of the thought.	EXP
Tricolon	A rhetorical term for a series of three parallel words, phrases, or clauses.	TRI

The fifth maxim sounds as an advice for the end user. This advice is concerned with clarifying that the annotation done in the corpus should not be viewed as a perfect and flawless production, but is a tool that can aid in future research. The sixth maxim stresses that any scheme used in the annotation process should be based on theory-neutral principles. That is, principles that are widely agreed upon by linguists and not controversial ones. The seventh maxim is both an advice for the annotator and the end user (Leech, 1993). The maxim emphasizes that no annotation scheme is to be considered as a standard. Standards are considered as such after general accord and this can happen only after the annotation scheme is practically applied. These maxims are taken very closely into consideration in the analysis of the corpus of the present study. The researcher focuses on meeting all the maxims of annotation so as to create an annotated corpus that not only would be of help to other researchers but also helps provide findings.

C. Different Types of Annotation

There are three types of corpus annotation; fully manually, fully automatically and semi- automatically (Bird &Liberman, 1999). All the three types have pros and cons. The fully manually annotated corpus has the virtue of being of highest quality, yet it is tremendously time consuming and still a human researcher's annotation is prone to error. Humans are of course more accurate than machines since they embrace the value of reasoning. This is the one used in the analysis of the corpus in the study. Annotation in this study plays the pivotal role in the analysis of the political speeches, the corpus of the study. The researcher analyses the speeches searching for the different instances or occurrences of the rhetorical devices to which the tags are assigned. The second type of annotation is the one automatically carried out (Bird & Liberman, 1999). Although this automatic type of annotation is quick, yet it is consistently full of errors. A computer program, no matter how suitable for the task, commits a high number of errors.

The third type of annotation is a mix between the first two types. This type entails automatic annotation with manual post annotation editing. Accordingly, the tags are annotated into the speeches to indicate the occurrence of the devices they stand for. Once an instance is spotted, the tag is placed at the beginning and at the end of the instance. An illustration of this is the following example taken from King's speech "I Have a Dream":
"<TRI> <PAR> <ALT> Life, Liberty <ALT/> and the pursuit of Happiness <TRI/> <PAR/>."

The above example shows several occurrences of several devices at the same time. Tricolon, alliteration and parallelism are assigned the tags <TRI>, <ALT> and <PAR> respectively. The tag is placed twice to surround the instance thus simplifying and clarifying the tracing of all the various instances. The brackets < TAG> surround the tag that opens at the beginning of the instance. The end of the instance is surrounded by the same brackets but includes an oblique <TAG/> to indicate that the instance has ended.

The study encompasses a corpus of approximately 40,000 words included in the fourteen speeches.

D. The Concordance Program: MonoConc Pro 2.2

The program chosen for the study is MonoConc Pro 2.2 for concordancing and corpus analysis (http://www.athel.com/mono.html.). The program's user interface makes the software easy to deal with. The program helps researchers upload a corpus and search. The search results appear in just a few seconds and are displayed in a very clear manner. It also offers expression searches and tag searches. This of course requires that the tag set is uploaded to the concordance program along with the annotated corpus. The program searches for word lists and frequency lists, for words and phrases, and also for collocates and collocations.

According to Barlow, MP 2.2 has newly added features such as highlighting the frequent collocates in a different color and they appear in the concordance result window. The results or the retrieved examples appear in a form of keywords that are highlighted and are shown in context. By clicking on the highlighted example appearing in the results window, the whole sentence where the word or occurrence lies appears in the context window. This helps identifying the data visually with utmost ease (Barlow, 2008). The originality of the present study stems partly from its distinctive tag set. Such tags are created by the researcher and they need a software program for the analysis. Most concordance programs analyze the part of speech tags, but for the present study the tags are discourse based. After the whole corpus is annotated, it is uploaded to the MonoConc. The tags are of course also added to the software to be able to spot them as needed for the researcher's purposes. One device is searched at a time, and the program displays all instances of the required tag search. Numbers of the occurrence of every device are displayed to the researcher, who then starts collecting the results to arrive at conclusions. The conclusions are related to the type of devices used by every speaker and the amount of usage. Once the search is done, the program spots the specified device and brings it forth to the researcher in the results window and other information is also displayed.

6. CHOOSING THE TWO POLITICAL FIGURES

The study has at its heart a corpus of political speeches. These speeches were given by two political figures described by many writers as two very eloquent orators. The first figure is Martin Luther King, Jr., the clergy man and the son of the African American Baptist church, who managed to change history through his eloquent speeches. He was a man driven by his dream of achieving equality for all of "God's children" as he always describes mankind in his speeches. The 4th of April, 2015 served as the 47th anniversary of King's assassination. Although Martin Luther King died at the age of 39, he had several contributions in various areas springing from his connections to the peace and social justice, humanist and civil rights movements of his time. He acted as a source of inspiration and a muse for a variety of the intellectual, cultural and political developments belonging to the twentieth century. King spent years of his life fighting to gain the dignity of the oppressed people all around the world and not only the blacks.

His oratory, infused with the experience he gathered from his readings in theology as well as his own insights, had a glowing effect on so many as was evident in his preaching activities. He joined and created so many associations and movements calling for the rights of the blacks. The Montgomery Improvement Association (MIA) which was formed by a number of notable Montgomery black leaders including Ralph Abernathy, his lifelong companion, is only one illustration of the many leaders who fought by his side. King took the role of the primary spokesperson of the year-long Montgomery bus boycott which he actually spoke about in his speeches. His oratory, deep beliefs in the equality of all human beings, theological background and Mohandas Gandhi's teachings of nonviolence of which King was an advocate, transformed him into a leader capable of expressing himself in memorable words thus mobilizing forces to fight by his side.

King's speeches are rich with the variety of rhetoric employed throughout. As a political and religious leader, King's aim is definitely to move and persuade thus leading to the major end desired from the listeners; to act. His speeches influenced masses of people belonging to his similar school of thought and others from different walks of life. In chapter three, the chapter responsible for the analysis, the language of the speeches is analyzed and the rhetoric of both speakers is put on display. Not only do King's speeches have linguistic richness between its lines that can help researchers arrive at theories and investigate language, but also King's speeches have many contributions and legacies in many areas of life. The year 2008 unfolded on the 45th anniversary of King's most famous speech "I Have A Dream" and in the same year Barack Obama became the first African American to accept the presidential nomination of a major political party at the 2008 Democratic National Convention. This is definitely the realization of one of King's dreams that all human beings are equal and that they should be assessed by "the content of their character and not the color of their skin". Consequently, King can be described as Obama's god father.

Assessed by the content of his character rather than the color of his skin or his African roots, Barack Obama is now the 44th Unites States President succeeding George W. Bush. Obama is the second political figure in this study whose political speeches serve as the other half of the study's corpus. Barack Obama became President at noon on January 20th, 2009 which is a date specified by the Twentieth Amendment of the Constitution. The Amendment requires that the president starts officially holding the office at noon on January 20 following the year of the presidential election. This day is known in America as the Inauguration Day thus marking the four-year term of both President and Vice President. Obama and King are highly connected for several reasons. Both speakers have African origins and both are always referred to as great eloquent orators who can stir and enchant audiences.

Obama's capability of stirring an audience is many linguists' area of research. He is described by many writers as having the ability to use simple words in his speeches, yet manage to elevate and inspire the audience through techniques that he uses (Assumndson, 2008). The speeches that Obama gave during his campaign running for presidential election are widely praised as master pieces which have inspired many writers to work on analyzing Obama's style. Many writers search for Obama's secret behind his ability to stir the crowd. Obama's election itself has its historical value and the way he uses his simple words to awaken, stir, inspire and stimulate the audience to revive their hopes that a better America in particular and a much better world in general is possible. Analyzing his words and looking deep into the stylistic or rhetorical devices used is the concern of the researcher of the present study.

7. ANALYSIS

Uploading the annotated corpus to the program, MonoConc Pro, was followed by the entering of the tag set designed by the researcher. The search takes place one tag at a time. Figure 1 displays an example of the occurrences of one of the devices in the uploaded corpus. The tag of the device shows in the middle of the window in blue surrounded by the fragments in which they occur.
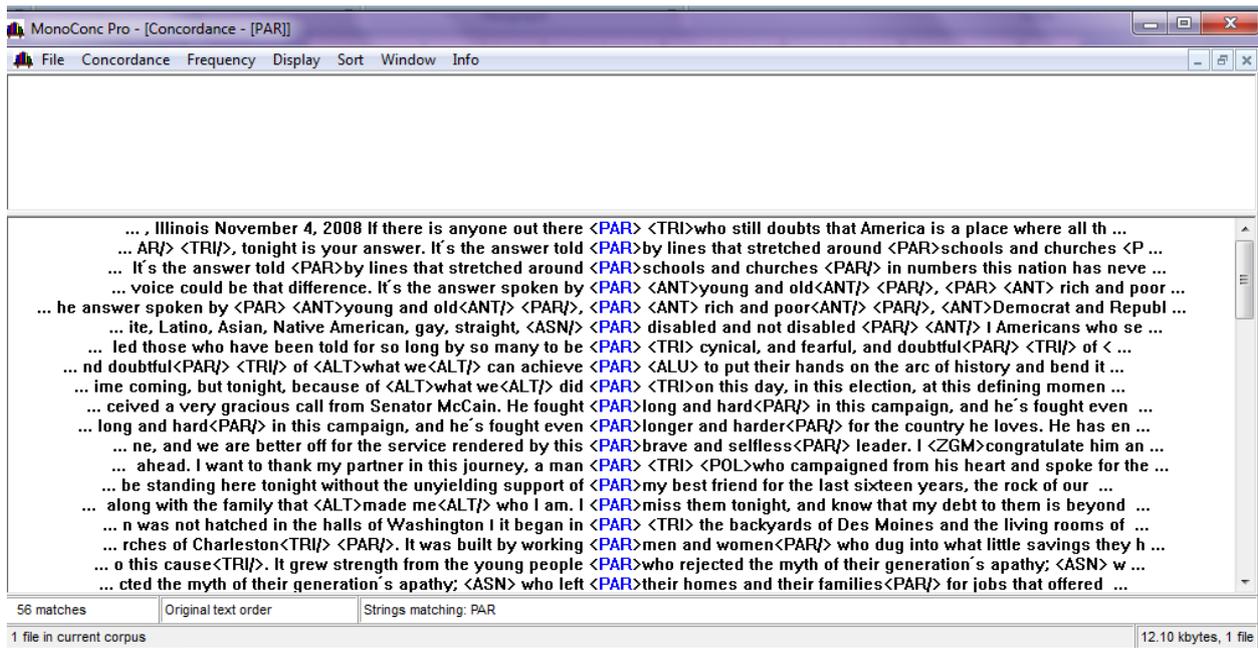
Figure 1: Occurrence of parallelism

By clicking on any of the highlighted occurrences, the whole instance of the device shows in an upper window. The results then appear in a double window where the researcher can clearly read the whole instance as a better way to understand the device in context. This is shown in Figure 2.
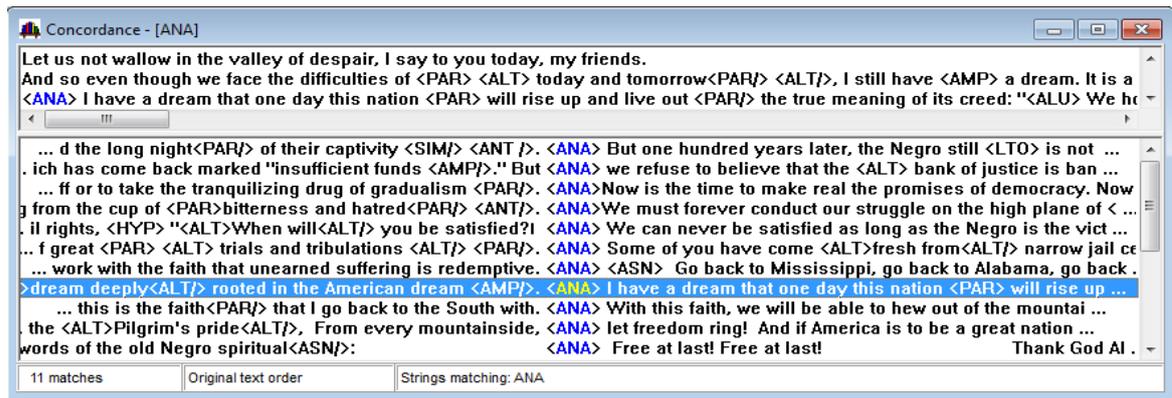

Figure 2: Search results of anaphora in two windows

A. The results of the analysis of King's speeches

"Give Us the Ballot", the first speech in this selection, does not show high numbers in the use of rhetorical devices. Parallelism, alliteration and allusion are the top three devices used occurring 55 times, 29 times and fourteen times respectively. This group is followed by antithesis and tricolon both used nine times. Anaphora is used six times, expletives five times with two similes in the speech. Understatement, sentential, amplification, conduplicatio, asyndeton and zeugma appear only once. The other devices have no occurrences at all.

The second speech in King's collection is The Great March on Detroit. Parallelism and alliteration occur 49 and 43 times respectively. These two numbers are followed by ones that are smaller, for example antithesis appears sixteen times. Anaphora occurs 12 times, allusion nine, tricolon eight and asyndeton seven times. Metaphor, personification and polysyndeton all occur four times. Symploce occur twice whereas rhetorical question and amplification appear only once.

The third speech is King's speech "I Have a Dream". The highest number of occurrences of a device goes for parallelism which occurs 38 times in the speech. This is followed by alliteration which is used for 35 times, antithesis occurring 21 times and anaphora 11 times. Some other devices are used in the speech but in little numbers, such as allusion which is

used six times and climax which occurred only once. On the other hand, some devices did not occur at all in the speech. These devices are rhetorical question, procatalepsis, distinctio, understatement, eponym, exemplum, sentential, epistrophe, personification, aporia, chiasmus, metabasis, anadiplosis, conduplicatio, apostrophe, zeugma, syndoche and expletive.

The fourth speech is the one that King gave in Oslo when he was receiving the Noble Prize. The numbers of occurrences of all devices in general in this speech are not as huge as its counterparts. The highest occurrence is of parallelism which occurs thirty seven times in the speech. This is followed by the second highest number of device occurrence which is alliteration. Tricolon occurs seven times in the speech and antithesis and anaphora occur five and four times respectively. Allusion appears three times and personification and conduplicatio both occur twice. Distinctio, simile, metaphor, symploce, amplification, zeugma and expletive each have a single occurrence only.

The following speech in this section is Our God Is Marching On. 70 occurrences of parallelism are found followed by 65 occurrences of alliteration. Following these two devices are allusion and anaphora appearing 14 and 13 times respectively. Antithesis occurred 11 times whereas tricolon and hypophora appeared nine and seven times respectively. Both personification and polysyndeton occur five times, while metaphor and amplification occur only twice. Each of exemplum, metonymy and expletive has only one occurrence.

The sixth speech in this collection is Beyond Vietnam. Numbers of occurrences of devices in this speech outnumber the first two speeches. Largest numbers of occurrences are scored by parallelism scoring 131 occurrences, followed by 79 occurrences of alliteration, thirty three occurrences of allusion and lastly thirty occurrences of tricolon. These huge numbers are followed by rhetorical question and antithesis appearing 25 and 24 times respectively. Litotes is used sixteen times, expletive used 12 and anaphora is used 10 the same as metonymy in the speech and personification is used nine times. There are six occurrences of asyndeton and five for metaphor. Procatalepsis, simile, symploce and zeugma all occur three times, while polysyndeton appears twice. Hypophora, epistrophe, amplification, metabasis and conduplicatio are all used only once. The other devices are not used in this speech at all.

I See the Promised Land is the last speech chosen for King. As usual, parallelism occupies the highest number of occurrences scoring fifty as shown in figure 13. This is followed by 38 occurrences of alliteration and 20 of allusion. Anaphora is used 12 times succeeded by rhetorical question appearing eight times and asyndeton six times. Both hypophora and symploce are employed five times in the speech while both antithesis and metonymy four times. Expletive and amplification appear twice and each of tricolon, distinctio, simile, personification, conduplicatio and apostrophe show only a single occurrence.

B. The results of the analysis of Obama's speeches

The first speech is the one he gave in South Carolina. The search results of the number of occurrences in the South Carolina speech show forty eight occurrences of parallelism, followed by fourteen occurrences of alliteration. Obama used antithesis eleven times in the current speech, and used tricolon ten times. These scores are followed by eight uses of polysyndeton. The three of anaphora, apostrophe and asyndeton are employed for five times. Metonymy is used for three times whereas amplification, conduplicatio and expletive are used only twice.

The second speech in Obama's selection is "Super Tuesday". The highest number of occurrences is scored by parallelism which occurs fifty times throughout the speech. The second highest number is shown through the occurrence of alliteration. Tricolon occurs 14 times and both expletive and asyndeton occur eight times. This is followed by polysyndeton and anaphora occurring seven times. Antithesis occurs six times whereas allusion four times. Each of rhetorical question, exemplum, amplification, conduplicatio and apostrophe occurs only once.

'Night Before the Election' is the speech that Obama gave one day before he was announced President of the United States of America. This speech is the fourth in the selection. Parallelism occurs 46 times in this speech followed by anaphora and asyndeton which occur 13 times and 11 times respectively. This is followed by tricolon and alliteration that occur nine times and eight times respectively. Metonymy occurs seven times whereas antithesis occurs six times. Polysyndeton, apostrophe and expletive occur in five, three and two times in that order. Finally, exemplum, amplification and climax each occurs only once.

The Election Night Victory Speech is the fourth speech in this selection. In this speech the highest number of occurrences of rhetorical devices goes to parallelism which occurs fifty six times. Alliteration follows parallelism occurring thirty six times. Tricolon makes the third highest number. These three high numbers are followed by nine occurrences of antithesis,

seven for apostrophe, six for allusion, four for anaphora and three for asyndeton. Rhetorical question, amplification and polysyndeton all occur twice. Distinctio, exemplum, zeugma and expletive each occurs only once.

The fifth speech is the Inaugural Speech which makes the first speech for Obama as President. Ninety four occurrences of parallelism are found in the Inaugural Speech, followed by 31 alliterations. Tricolon makes up eighteen occurrences, while anaphora occurs twelve times throughout the speech. One occurrence less than anaphora, antithesis occurs eleven times. This is followed by asyndeton, apostrophe and allusion occurring nine and eight and six times respectively. There are four occurrences of metonymy, three of amplification, and one for zeugma. Both symploce and climax occur twice.

The sixth speech in Obama's selection is the speech he gave in University of Cairo. This speech has a huge number of occurrences of both parallelism and alliteration occurring one hundred fifty four times and one hundred and five times respectively. These very huge numbers are followed by numbers that are close to each other. Tricolon and metonymy occurred thirty and twenty seven times respectively. Allusion occurs twenty times and antithesis, only two occurrences less, occurs eighteen times. Fifteen occurrences of expletives are found in this speech. Anaphora has eight occurrences and amplification occurs four times. Symploce and litotes and zeugma have three occurrences. Finally metaphor and epistrophe occur only once.

The last speech is the one that Obama gave after the Egyptian President Hosny Mubarak stepped down on the eleventh of February 2011. The speech does not contain a big number of devices used. Parallelism is used seventeen times followed by alliteration thirteen times. Tricolon is used for four times and allusion is used three times. Antithesis, simile and metonymy are used twice, while anaphora, amplification, polysyndeton, asyndeton and expletive are used one time.

Figure 3 is a bar chart showing the results of all the occurrences of all devices in the fourteen speeches of both King and Obama. The bars show the differences between the two speakers' usage of the devices belonging to the four categories. There is a similarity in the use of the devices that belong to the last two categories. These categories include devices that are used to organize the writing and the other includes devices that create certain structural pattern to add a distinctive style to the writing. In these two categories Obama shows a wider range of use of devices in both categories scoring 498 and 480 occurrences in the third and fourth categories respectively whereas King scored 438 and 459 occurrences respectively. King then shows higher numbers of occurrences than his counterpart with a difference of 182 occurrences in the first category and a difference of 51 in the second category.
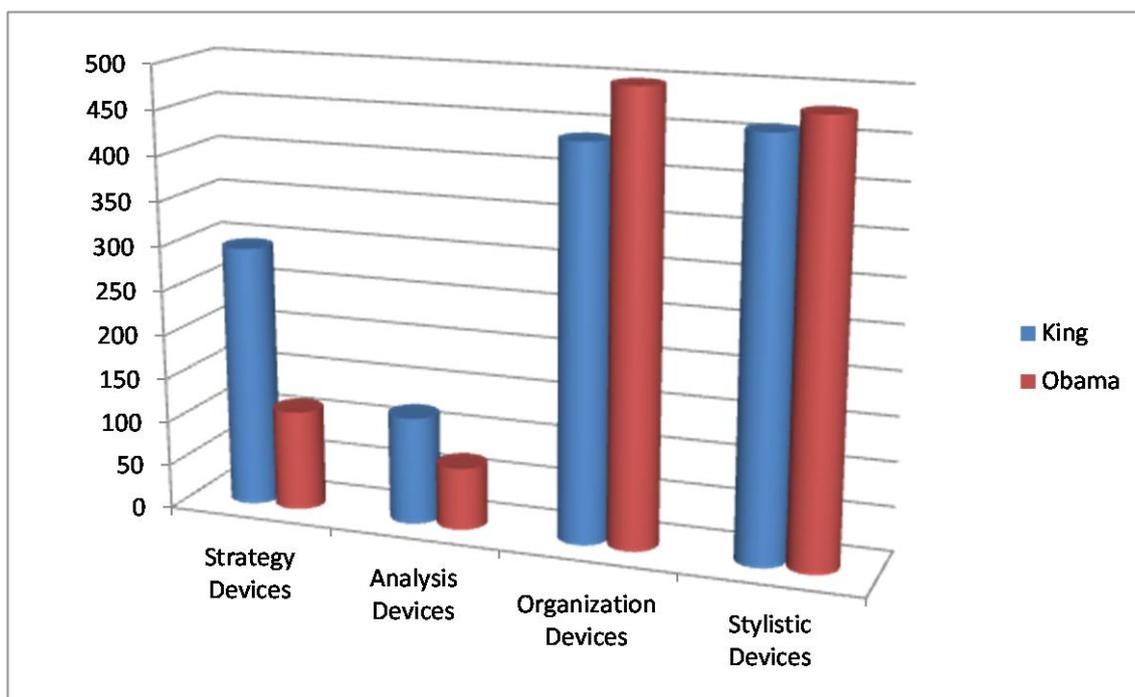

Figure 3: A Bar Chart Showing the Differences Between King and Obama's Use of the Devices in the Four Categories

8. CONCLUSION AND RECOMMENDATIONS

The present study is a corpus based study where the aim is to create an annotated corpus on the discourse level. Creating such output will facilitate the job of many other researchers trying to identify the style of speakers through analyzing the

language they use. Having such output available will enable them to come up with the findings with the press of a button. In the present study, part of the methodology is to create a tag set of the selected rhetorical devices. The tags were designed on the same lines of part of speech tagging. Three capital letters resembling the word being tagged were created for every device. The annotated corpus was then uploaded to the concordance program and the search began. The tags proved success as well as the choice of the computer program, the MonoConc Pro. Although many tags are sometimes placed in the same paragraph representing their occurrences, yet the search for the tags was not problematic at all. Once the search of a certain tag starts, the software separates the searched for tag from the others and the occurrences appear clearly. The program not only produces the occurrences of the tags, but also certain phrases.

The tagged corpus transformed the speeches from being an undiscovered creation into a living body of data that has tags within its lines. These annotations or tags help unfold the secrets behind stirring an audience and behind making them laugh or cry, behind the rise of a leader as a world leader and the fall of another. This annotated output enables researchers to investigate its language looking for the used rhetorical devices which will help know the style of the authors and speakers. They can search for any phenomenon they might be working on investigating. This annotated corpus enables researchers to find the power behind the language of political speeches. The investigations, analyses and results finally arrived at could not have been feasible except through a corpus as such.

The research questions that the study started with are answered through the analysis and investigation. A main aim of the study is to build a discourse- based corpus. Such output is the annotated corpus which embraces the designed tags corresponding to the chosen devices. The study proved that building an annotated corpus on the discourse level is possible. A second research question is concerned with the creation of the tags. The tags were created to represent the rhetorical devices and they follow the same pattern of the part of speech tags. The tags are placed in the corpus and uploaded to the computer program. The results of the search showed that the tag set worked successfully embracing the occurrences of the various devices. The chosen rhetorical devices are organized in categories based on the purpose of using them. Consequently, after the analysis of the tagged speeches using the concordance program, the search results clarified which devices are used and to which categories do these devices belong. This can enable researchers to both clarify the effect of the devices and also to identify the style of the politicians.

Leech's annotation maxims were a very good guide in the annotation process. The annotated corpus can be reverted to its original state through the removal of the tags. Equally, the annotations can be extracted by themselves from the corpus. Since the study is of good use to other researchers, a clear description of the annotation scheme is provided. This description also includes that the annotation was carried out by the researcher fully manually and of course such annotation scheme might be prone to error and is not presented as a standard but as the primary endeavor. The originality of the study stems from its discourse based corpus. Such output was never available before.

This output can definitely be enlarged in future research. If more speeches for the same speakers are annotated, this will enable researchers to arrive at more reliable conclusions about the speakers' styles. Annotating speeches that belong to different stages in the speakers' lives can also help trace the changes or spot the similarities in their styles as a means of arriving at a better understanding of their way of thinking. That is, this output can also be enlarged through using parallel corpora. Speeches in both Arabic and English can be annotated and the differences or the similarities be pointed out for further analysis and investigation. Enlarging the set of rhetorical devices will also add to the annotated output. Through a wide range of rhetorical devices, which will be assigned new tags of course, more reliable conclusions can be drawn about the speakers' styles. Researches on this path won't be possible before the presence of an annotated corpus as the one present in this study. With the press of a button the researcher can come up with numbers of the occurrences of the different devices, no matter how large these numbers might be. Such conclusions will allow the researcher to create theories about language.

REFERENCES

[1] A. Harris, Robert. (2003). *Writing with Clarity and Style: A Guide to Rhetorical Devices for Contemporary Writers.* Los Angeles: Pyrczak Publishing House.

[2] Aaarts, J. & Meijs, W. (Eds). (1984). *Corpus Linguistics: Recent Developments in the Use of Computer Corpora in English Research.* Amsterdam: Rodopi.

[3] Aaarts, J. & Meijs, W. (Eds). (1990). *Theory and Practice in Corpus Linguistics.* Amsterdam: Rodopi.

[4] Aarts, J., Haan, P. & Oostidik, N. (Eds). (1993). E*nglish Language Corpora: Design, Analysis and Exploitation.* Amesterdam: Rodopi.

[5] Abu El Wafa, Marwa. (2013). *Persuasion in Political Speeches: Discourse Tagging of Political Speeches.(*Unpublished Master's Thesis*)*. University of Alexandria, Alexandria, Egypt.

[6] Aijmer, K. & Altenberg, B. (Eds). (1991). *English Corpus Linguistics*. London: Longman.

[7] Allen, J., and Core, M. 1997. Draft of DAMSL: Dialog Act Markup in Several
Layers. http://www.cs.rochester.edu/research/cisd/resources/damsl/RevisedManual/.

[8] American Psychological Association. (2010). *Publication Manual of the American Psychological Association* (6th Edition, Second Printing).
Washington, DC: Author.

[9] Anonymous. (2008). Text Annotation for Political Science Research. *Journal of Information Technology & Politics*, 5(1).

[10] Araki, Masahiro, Kimura, Yukihiko, Nishimoto, Takuya & Niimi, Yasuhisa. (n.d). *Development of a Machine Learnable Discourse Tagging Tool.*
Department of Electronics and Information Science. Kyoto Institute of Technology.

[11] Assmundson, Mikael (2008). Persuading the Public: A Linguistic Analysis of Barack Obama's Speech on "Super Tuesday" Dalarna University,
School of Languages and Media Studies, English.

[12] Barlow, Michael. (2008). Parallel texts and corpus-based contrastive analysis, In: GÃ³mez GonzÃ¡lez, M., Mackenzie, L. and GonzÃ¡lez Alvarez,
E. (eds.), *Current Trends in Contrastive Linguistics: Functional and Cognitive Perspectives.,* Benjamins, 101-121.

[13] Barlow, Michael, and Bowker, L. (2008). A comparative evaluation of bilingual concordancers and translation memory systems, In: Yuste Rodrigo,
E. (ed.), *Topics in Language Resources for Translation and Localisation.*, Benjamins, 1-22.

[14] Barlow, Michael. (1999). MonoConc 1.5 and ParaConc. *International Journal of Corpus Linguistics.*

[15] Beaugrande, Robert de. (January, 2006). Critical Discourse Analysis: History, ideology,
methodology. *Journal of Studies in Language & Capitalism,* 29 – 56.

[16] Biber, Douglas. (2006). University Language: *A Corpus-based Study of Spoken and Written
Registers.* Amesterdam: John Benjamins.

[17] Biber, Douglas, Connor, Ulla & Upton, Thomas A.(2007). *Discourse on the Move: Using
Corpus Analysis to Describe Discourse Structure.* Amsterdam: John Benjamin.

[18] Biber, Douglas, Conrad, Susan & Reppen, Randi. (1998). *Corpus Linguistics: Investigating Language Structure, and Language Use.*
London:Cambridge University Press.

[19] Bird, Steven & Liberman, Mark. (March,1999). *Formal Framework for Linguistic Annotation.* Tech.Rep. MS-CIS-99-01, Department of
Computer and Information Science, University of Pennsylvania, Philadelphia, PA, 1999. http://xxx.lanl.gov/abs/cs.CL/9903003.

[20] Carlson, Lynn, Marcu, Daniel & Okurowski, Mary Ellen.(n.d.) *Building a Discourse-Tagged
Corpus in the Framework of Rhetorical Structure Theory.*

[21] Carlson, Lynn & Marcu, Daniel. (2001). Discourse Tagging Reference Manual. Technical Report, Information Science Institute.

[22] Carlson, Lynn & Daniel Marcu. (2012). Ecological Evaluation of Persuasive Messages Using   Google Ad Words. Journal: CoRR,
Vol.abs/1204.5369.

[23] Danieli, Morena. (1999). Discourse Tagging. Towards Tools and Standards for Discourse Tagging. (ACL-99 Workshop) June 22, 1999 University
of Maryland College Park, MD, USA. URL: http://www.mri.mq.edu.au/conf/acl99/html.

[24] Finegan, Edward. (1996). [Review of Text and Corpus Analysis: Computer-Assisted Studies of
Language and Culture, by Michael Stubbs].

[25] Garside,R Leech, G. & McEnery, Tony. (1997). *Corpus Annotation. Linguistic Information from Computer Text Corpora.* London: Longman.

[26] Halmari, Helena & Virtanen, Tuija (Eds). (2005). *Persuasion Across Genres.* Amsterdam:John Benjamin's Publishing Company.

[27] Hunston, Susan. (2002). *Corpora in Applied Linguistics.* United Kingdom: Cambridge University Press.

[28] J., Thomas & M., Short. (Eds.) (1996). *Using Corpora for Language Research.* London: Longman.

[29] Johansson, Stig & Oksefjell, Signe. (Eds) (1994). *Corpora and Cross-Linguistic Research: Theory, Method, and Case Studies.* USA: John
Benjamins Publishing Company.

[30] Kennedy, Graeme. (1998). *An Introduction to Corpus Linguistics.* New York: Longman Pearson Education Inc.

[31] Lapadat, Judith C. (2007). Discourse Devices used to Establish Community, Increase
Coherence, and Negotiate Agreement in an Online University Course. *Journal of Distance
Education*, Vol. 21(3), 59-92.

[32] Leech, Geoffrey. (1992). Corpora and Theories of Linguistic Performance. In Svartvik (1992)

[33] Leech, Geoffrey. (1993). Corpus Annotation Schemes. In Literary and Linguistic Computing, 8(4): 275-281.

[34] Nomoto, Tadshi & Matsumoto, Y. (1999). Learning Discourse Relations with Active Data Selection. In P. Fung and J. Zhou (eds.), Proceedings of
the 1999 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora, 158-167, University of
Maryland, MD.

[35] M., Stubs. (1996). *Text and Corpus Analysis*. Oxford:Blackwell.

[36] M., Oakes. (1998). *Statistics for Corpus Linguistics*. Edinburgh: Edinburgh University Press.

[37] Maurice De Schryver, Gilles. (2001). Corpus-based Activities versus Intuition-Based Compilations by Lexicographers, the Sepedi Lemma-Sign List as a Case in Point. Nordic Journal of African Studies 10(3): 374-398.

[38] McEnery, Tony. (2003).Corpus Linguistics. In Ruslan Mitkov (Ed.), *The Oxford Handbook of   Computational Linguistics* (pp.448-456). New York: Oxford University Press.

[39] McEnery, Tony & Wilson, Andrew. (1996). *Corpus Linguistics*. Edinburgh: Edinburgh University Press.

[40] McEnery, Tony & Wilson, Andrew. (2003). Corpus Linguistics. Retrieved from Essex.ac.UK:
http://www.essex.ac.uk/linguistics/external/clmt/w3c/corpus_ling/content/corpora/types/annotated.html.

[41] Meyer, Charles F. (2002). *English Corpus Linguistics: An Introduction*. United Kingdom: Cambridge University Press.

[42] Mitchell P., Marcus, Mary Ann, Marcinkiewicz & Santorini, Beatrice. (1993). Building a large annotated corpus of English: The Penn Treebank. Journal,19(2). 313–330. URL: http://portal.acm.org/citation.cfm?id=972475.

[43] Mladová, Lucie, Zikánová, Šárka & Hajičová, Eva. (2008).   From Sentence to Discourse:
Building an Annotation Scheme for Discourse Based on Prague Dependency Treebank. In
Proceedings of the 6th International Conference on Language Resources and Evaluation.

[44] Moliken, Paul, Grudzina, Douglas & McGuigan, Brendan. (2007). *Rhetorical Devices: A Handbook and Activities for Student Writers.*USA:Prestwick House, Inc.

[45] Nesselhauf, Nadja. (2005). Corpus Linguistics: A Practical Introduction. Amsterdam: Benjamins.

[46] Newman, John. (2008). Aiming low in linguistics: Low-level generalizations in corpus-based
research. Proceedings of the 11th International Symposium on Chinese Languages and
Linguistics (IsCLL-11), May 23-25 2008, National Chiao Tung University, Hsinchu,
Taiwan.

[47] Rapillo, Miriam. (n.d.). Corpus Linguistics: A General Introduction. Retrieved from:http:// www.givemeawhisper.tk.html.

[48] Schaffner, Christina. (1996). Editorial: Political Speeches and Discourse Analysis, Current Issues in Language and Society, 3: 3, 201 — 204.

[49] Schiffrin, Deborah, Tannen, Deborah & Hamilton, Heidi E. (Eds). (2001). *The Handbook of Discourse Analysis.* United Kingdom: Blackwell.

[50] Sherris, Ari. (n.d.). Corpus Linguistics as Agency for Quantitative and Qualitative Research. Retrieved from Find That PDF:
http://www.findthatpdf.com/search-9594609-hPDF/download-documents-essayoncorpuslinguistics.pdf.html.

[51] Sinclair, John & Carter, Ronald (Eds).(1991). *Corpus Concordance Collocation.* New York: Oxford University Press.

[52] Sinclair, John. (Eds). (1992). The Automatic Analysis of Corpora. New York: Routledge.

[53] Sinclair, John & Carter, Ronald (Eds). (2004). *Trust the Text: Language, Corpus and Discourse*. New York: Routledge.

[54] Teun A. van Dijk. (2001).Critical Discourse Analysis. In Schiffrin, Deborah, Tannen, Deborah & Hamilton, Heidi E. (Eds). *The Handbook of Discourse Analysis* (pp.103-136). United Kingdom: Blackwell.

[55] Teun A. van Dijk. (2001). Political Discourse and Political Cognition. In Schiffrin, Deborah, Tannen, Deborah & Hamilton, Heidi E. (Eds). *The Handbook of Discourse Analysis* (pp.203-233). United Kingdom: Blackwell.

[56] Urbanavičienė, Irena. (2004). Political Speeches: Exertion of Power through Linguistic Means
Retrieved from Ceeol: http:// www.ceeol.com.html.

[57] Walker, Marilyn. (1999). Towards Standards and Tools for Discourse Tagging. University of Maryland, USA: Association for Computational Linguistics.

[58] Wasowa, Thomas & Arnold, Jennifer. (2004). Intuitions in Linguistic Argumentation. Retrieved from www.sciencedirect.com:
DOI:10.1016/j.lingua.2004.07.001.

[59] Y. Kawaguchi, M. Minegishi & J. Durand (Eds.) (n.d.). *Corpus Analysis and Variation in Linguistics*. Amsterdam: John Benjamins.

[60] Yeates, Stuart& H. Witten, Ian. (n.d.). On Tag Insertion and its Complexity. Department of Computer Science,University of Waikato,Hamilton, New Zealand. Retrieved from s.yeates, i.witten @cs.waikato.ac.nz.

يناقش هذا البحث امكانية تصميم مجموعة من الأكواد على مستوى النص و ذلك من خلال تكويد أو عنونة أساليب بلاغية متنوعة تستخدم فى الخطب السياسية و غيرها من أنواع النصوص بغرض الاقناع و توصيل الأفكار بطريقة بلاغية و فى هذة الدراسة تم تكويد الأساليب البلاغية التى أستخدمها كلا من الرئيس الأمريكى باراك أوباما و القائد الأمريكى مارتن لوثر كنج فى سبع خطب لكل منهم. و يتم هذا على طريق عنونة النصوص كوسيلة لتصميم مجموعة أكواد خاصة بالباحث و استخدامها لتكويد أو عنونة الخطب الأربعة عشر التى تم اختيارها للبحث .كما تم اختيار برنامج **MonoConc Pro 2.2.** لرصد اعداد الأساليب البلاغية المختارة عن طريق تتبع الأكواد التى تم ادخالها على البرنامج من قبل الباحث. و قد تؤدى الأ رقام الناتجة عن الأعداد للوصول لأسلوب كل شخصية سياسية كما تبين ما يتشابهون فية و مايختلفون فية و قد تتيح هذة الدراسة الفرصة لباحثين اخرين من الاستفادة من الدراسة فى نفس المجال.

عنونة الخطب السياسية: دراسة تحليلية

مروة عادل أبو الوفا

قسم اللغة و الترجمة، كلية اللغة و الاعلام، معهد دراسات اللغات، الأكاديمية العربية للعلوم و التكنولوجيا و النقل البحرى

ميامى، الأسكندرية ، مصر

سامح الأنصارى

قسم الصوتيات و اللغويات، كلية الآداب، جامعة الأسكندرية

الشاطبى، الأسكندرية، مصر

مكتبة الأسكندرية، الأسكندرية، مصر

شادية السوسى

معهد اللغويات التطبيقية،، كلية الأداب، جامعة الأسكندرية

الشاطبى، الأسكندرية، مصر

Marwa Adel is an Assistant Lecturer at the Arab Academy for Science, Technology & Maritime Transport. She is originally a graduate of Faculty of Arts, University of Alexandria, English Department, Literature Section. She teaches Pragmatics, Syntax and Linguistics in College of Language and Communication, Language and Translation Department, and also teaches a variety of ESP courses in different colleges including College of Engineering, College of Maritime Transport, College of Logistics and College of Business Administration. She obtained her Applied Linguistics Diploma in 2008 with Very Good as a general grade. She then registered for the Masters Degree in 2009 in University of Alexandria in the field of Applied Linguistics working on a pioneering topic in the field of text tagging. She obtained the Degree in 2013 with a general grade Excellent. In the same year, in September 2013, she got enrolled in the PhD programme and registered the PhD thesis in September 2014 in the field of abstract meaning representation.

Her academic interests are in the fields of corpus- based studies, discourse tagging, discourse analysis and meaning representation.

Dr. Sameh Alansary is professor of computational linguistics in the Department of Phonetics and Linguistics and the head of Phonetics and Linguistics Department, Faculty of Arts, Alexandria University. He obtained his MA in Building Arabic Lexical Databases in 1996, and his PhD from Nijmegen University, the Netherlands in building a formal grammar for parsing Arabic structures in 2002. His main areas of interest are concerned with corpus work, morphological analysis and generation, and building formal grammars.

He is also the head of Arabic Computational Linguistics Center in Bibliotheca Alexandrina. He is supervising and managing the Universal Networking Language project in Library of Alexandria since 1-6-2005 till now.
Dr. Alansary is the co-founder of the Arabic Language Technology Center (ALTEC), an NGO aims at providing Arabic Language resources and building a road map for Arabic Language Technology in Egypt and in the Middle East. He Has many scientific works in Arabic Natural Language Processing published in international conferences and periodicals, and a member in many scientific organizations: (1) Egyptian Society of Language Engineering, Cairo, (2) Arabic Linguistic Society - USA, (3) Association of Computational Linguistics - USA – Europe, (4) Universal Networking Language foundation, United Nations, Geneva, Switzerland.

Shadia el-Soussi is a Lecturer at the Institute of Applied Linguistics and Translation, Faculty of Arts, University of Alexandria. She has extensive teaching experience in teaching and course management of English and Arabic as Foreign languages (TEFL &TAFL), English for Specific Purposes, Testing and Evaluation. She has supervised and co-supervised a large number of MA and PhD degree students.
Her academic interests are largely in Sociolinguistics, Discourse Analysis, Critical Discourse Analysis and Cultural Studies.
She has also been working a consultant for the Arabic Bibliographic Department, Library Sector, Bibliotheca Alexandrina, since 2009.