STATISTICAL ANALYSIS OF THE PRESIDENT'S SPEECHES DELIVERED FROM 2013 TO 2018

Ibrahima dit Bouran SIDIBE[a], Abdou FANE[a], Arouna SOUGANE[b], Imene DJELLOUL[c,d], Abdelhakim KHATAB[e,f]

[a]Centre de Formation et de Perfectionnement en Statistique (CFP-STAT)

[b]Institut National de la Statistique (INSTAT)

[c]Ecole Supérieure des Sciences Appliquées d'Alger (ESSA-Alger), Place des Martyres 16000 Alger, Algérie

[d]Manufacturing Engineering Laboratory of Tlemcen (MELT), Abou Bekr Belkaid University of Tlemcen,P.O. Box 230, 13000 Tlemcen, Algeria

[e]Laboratoire de Génie Industriel de Production et de Maintenance (LGIPM)

[f]National Engineering School of Metz, FRANCE, 57078 Metz cedex, 1 route d'Ars Laquenexy.

ABSTRACT

This paper investigates a corpus of 60 speeches that president produced during his first quinquennial between 2013 and 2018. Subsequently, a novel quantitative analysis based on descriptive and factorial methods are implemented on this corpus. These methods allow us to make an appraisal of president style of speaking, the vocabulary density, the diversity, the complexity and the lexical richness of sentences by using statistics and complexity indexes. Moreover, the readability and the intelligibility of the president speeches are analyzed for Malian people by use of statistical and graphical visualizing methods. Accordingly, this analysis measures the originality of the president's spoken words response to Malian social context characterized by the weakness of the government authority, the increase of insecurity and terrorism, the corruption of the public finance workers and the deterioration of the economic indicators such as the youth unemployment, the lack of valuable creation by service and production industries and their consequences on the Gross Domestic Product (GDP) and the poverty index. To perform these analyses, the corpus is therefore transformed into Term Document Matrix (TDM), which is a contingency table between speech and word counts. These analyses underline that the president usually delivered speeches with a high complexity (similar to Marcel Proust text) whose message content is unintelligible and incomprehensible for most Malian.

Keywords: Corpus analysis; Spoken style; Lexical richness; Term Document Matrix; Word-cloud.

1. INTRODUCTION

Text mining refers to knowledge discovery process when the source of data under consideration is a text document from speech, Internet or scientific article. Currently, text mining relies on important contributions from statistics, Artificial Intelligent (AI) and Machine Learning (ML) algorithms. Early, Correspondence Analysis (CA) was developed and applied to study theater text [1]. Escofier, with simultaneous representations on factorial axes, evaluated attractions between used words in theater "Phèdre" and characters [1].

In recent decades, text mining approach has received important support due to the increase of computational power and the growth of textual information from internet. Indeed, Text mining proposes a set of algorithms

which handles the detection of the pattern, the trend and the correlation in the spoken or/and written texts in order to efficiently extract relevant information to make text document more organized, classified, accurate to exploit. These algorithms are often applied on transformations of the text in which the text of speech is considered as a bag of words/terms or a matrix of words/terms. A great survey on the text mining subject is proposed in [2], [3]. The interested reader can consider the collection of article from [2] for a huge review in which several approaches were proposed by different authors to process texts and extract information. Indeed, these authors focused on the mathematics background of the text document processing. These researches mainly figure out the advantages of text mining methods in the document's classification, clustering and retrieval. A number of other interesting algorithms for email surveillance have been developed [3] to make email classification handy for filtering spam message. Most of these Algorithms are based on the content analysis through supervised machine learning approaches. The implementation of these approaches makes convenient the information exchange and reduces the number and the drawback of phishing and malicious viruses by emailing.

Nowadays, text mining is also used for an appraisal of the speech and its impact on the public opinion particularly from politicians. Speech is a communication or expression of thoughts in spoken words according to Merriam Webster (www.merriam-webster.com[1]). Presidents deliver periodically speeches to explain by combination of specific terms or words or substantives his agendas to citizenry in response of social and economic problems. Most presidents address periodically speeches to their citizenry as to relate to their political decisions along with the latter social and economic drawback on the population. In fact, text mining competes with the natural language processing approach for rational understanding of political speech. Early, Cohen [4] showed that the influence of speech depends on the resources and the public's receptivity. Cohen [4] carried out an analysis on the correlation between presidential rhetoric and the public agenda based on some US presidents' speeches. This work pointed out that "Increases in presidential attention to the economics, foreign, and civil rights policy lead to increases in public concern with those policies." Therefore, the persuasion of the public depends on the substantive that the president uses to convince. In addition, Wang [5] summarized the features of Obama speeches based on Halliday's Systematic Functional Grammar and some statistics. Wang figured out that Obama made his public more easily to understand his sentences regardless the speech topic due to his leadership and his ability to often use simple substantives and short sentences [5]. In the same way, the Sharndama in [6] and Osisanwo in [7] carried out a descriptive analysis of President Buhari's inaugural speech. Their analyses allowed identifying Buhari main political agenda for next years. Moreover, Osisanwo investigated the use of words from a pragmatic perspective in order to identify the pragmatic act involved and goals of this acts in [7]. These analyses made then an accurate appraisal of Buhari rhetoric and also a comparison and a balance between different agendas.

In the same way, our paper proposes an analysis of Malian president's (Ibrahima Boubacar Keita) speeches. Unlike some cited paper, our analysis is based on statistical methods in which two approaches are going to be used. First, a descriptive analysis is performed to evaluate the president's words density and diversity in sentences. The main purpose of this descriptive analysis is to define the lexical richness and the complexity of his sentences. Second, a factorial analysis is realized to evaluate the relation between words and speeches in order to figure out the similarity and dissimilarity in each space. We expect on this work to understand how president speeches can fit the Malian context defined by insecurity and strong poverty.

The remaining of this paper is organized as follows. The analyzed problem is introduced in Section 2 in which, the hypothesis of work is merely defined and described. Section 3 gives descriptive analysis of the speeches and words or terms in each speech. Section 4 makes a complexity and readability analysis of speeches. In addition, the corpus of speeches is transformed to matrix in order to make clustering and corresponding analysis. Section 5 concludes our text mining analysis.

[1] www.merriam-webster.com, (1st June 2018)

2. CONTEXT AND PROBLEM DESCRIPTION

Malian president Ibrahim Boubacar Keita (IBK), since his election to office by 2013, has often delivered speeches addresses on national issues and events. The delivered speeches allow interacting with citizenry on his agenda or trying to influence public expectation due to context. Mali was going through a deep crisis due to a complex combination of principal factors. In 2012, Mali underwent a multi-sectional crisis that translated in the collapse of the country's institution in favor of a military coup by March 2012. The actors of this coup justified their action by the former president's inability to face properly the rebellion in the North along with the degradation of working conditions in the regular Army. Consequently, the institutional vacuum this coup generated along with the international sanctions that ensued hampered the nation's capabilities to face the deployment, proliferation and progression of armed groups which ended up controlling three northern regions. This made the country unsafe for living and forced some people to move inside and abroad. The insecurity, due to armed groups, impacted negatively the economic development by idling the trade, goods and service and government tax revenues. It was under this explosive mixture that IBK won presidential election after a deal between state representatives, some armed groups under a supervision and financial aids of international institutions such as African Union, Economic Community of West African States (ECOWAS), and European Union. Therefore, president must often deliver speech with appropriated content in order to explain his strategies, his policy, and his performance and to convince people to join him on his current and future projects. To earn & preserve the people's support during his first term, the president delivered at least five permanent speeches per year. The first was the New Year speech and was delivered on January 1st in which president IBK discussed his achievements and developed his agenda for the next year. The second speech takes place on January 20th. As the chief of armed forces, president IBK celebrated army day. In this speech, president depicted the security context of the country and draws with the army staff the current and future challenges such as counter terrorism, insecurity and other threats. The third delivered speech by president took place on March 26th named Martyr's Day. President commemorates the blood sacrifices of those citizens who died for the advent of democracy in 1991. The fourth speech is delivered on May 1st which stands for Workers' Day. President appreciated the contribution of the nation's workforce to the development of the nation. The last speech is delivered on September 22nd and celebrates the Independence Day. Meanwhile, special speeches were made on emergency situations such as trouble related to insecurity and disasters which touch Malian people inside and abroad. Predicting the number of the president's speeches ahead of time is impossible because it is contingent to live of the nation (crisis).

3. CORPUS DESCRIPTION

To perform our analysis, a sample of the speeches is considered. These speeches were delivered between September 2013 and May 2018 and were available on the presidency web site[2] in 2018. A detail on each speech is provided in the table of annex. The table in annex gives detailed information on collected speeches in which the date and the context of each speech in corpus are available. This information is summed up in Table (1) with respect to the type (Permanent, Special) and the first concerned public (National public, International public) of each speech per year. Table (1) depicts that the corpus has 41 delivered to deal with casual events against 19 for permanent situation. This distribution remains the same for the nature of speech which focuses on the concerned public by the speech.

Based on this corpus, we are going to perform analysis in order to underline how these speeches fit the context of Mali during the period 2013 and 2018. The proposed analysis is non-linguistic. In this work, the purpose is to perform a statistical analysis of the president speech. For this study, some statistical approaches are going to be used on the corpus of speeches delivered by the president. In fact, each speech is going to be considered as a collection of terms or lemmatized words. Accordingly, the feature of used words is going to be studied with the sentence structure in order to evaluate in one way the complexity and in another way the nature of president response words to challenges of the country. To make descriptive analysis, each speech

2 http://www.koulouba.ml/discours-du-president/discours/ (2nd June 2018)

is going to be transformed to matrix in which the columns present the speeches while the rows stand for the terms/words used in the different speeches. This matrix is denoted Term Document Matrix (TDM). This matrix crosses document with used term through a contingency table. This table stands for a mathematical representation of documents in which each column stands for a speech and the $(i, j)^{th}$ entry represents the frequency of the term i in the document j.

Table 1: Speeches Distribution according to the Public and the Types

	Nature Of Speech		Type Of Speech		
Years	*National*	*International*	*Special*	*Permanent*	*Total*
2013	4	2	5	1	**6**
2014	9	3	7	5	**12**
2015	11	7	15	3	**18**
2016	6	2	5	3	**8**
2017	6	4	7	3	**10**
2018	5	1	2	4	**6**
Total	41	19	41	19	**60**

Next subsections make an exploratory analysis on the corpus of speeches. This analysis describes the nature of the president sentences based on the length, the diversity and the frequency of his usual words. The words are going to be counted, analyzed and compared per year in order to understand the natural outcome of president politics in relation to the current and future social and economic challenges characterized by the idleness of production activities and the increase of the insecurity and terrorism acts.

3.1 SENTENCE LENGTH

This section makes a descriptive analysis of speeches and sentences through an appraisal of word counts. The analyses focus therefore on the statistical description of the word counts distribution in each sentence as depicted in Tables (3).

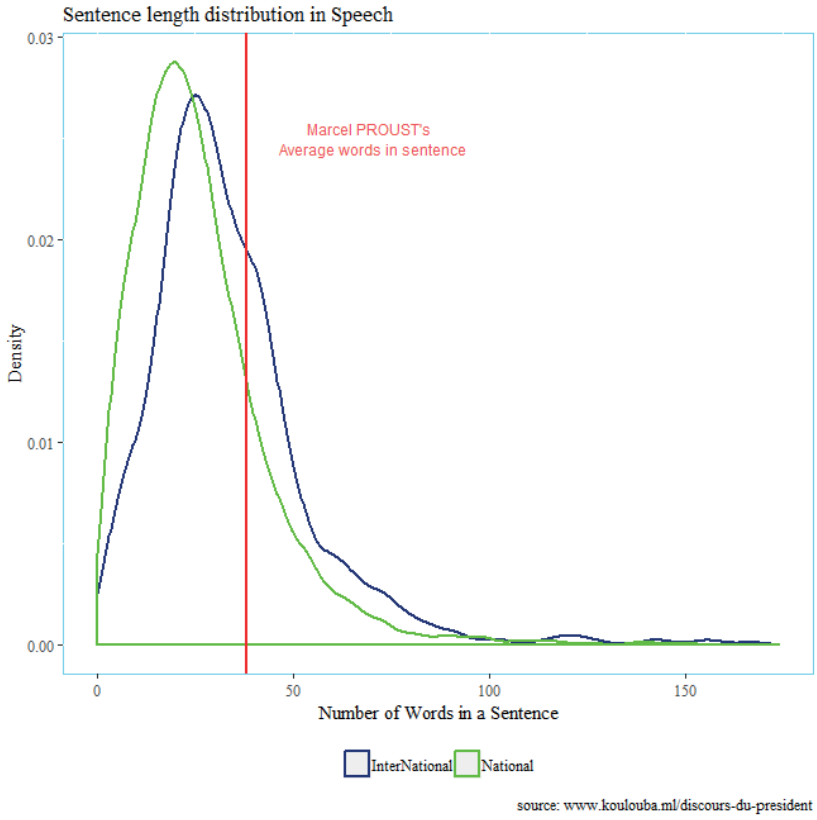
Table 3: Number of Words in sentences

		Number Of Words in Sentences					
		Min	*1st Quartile*	*Median*	*Mean*	*3rd Quartile*	*Max*
Nature	**International**	1.00	22.00	31.00	34.81	43.00	167.00
	National	1.00	15.00	24.00	27.41	35.00	191.00
Type	**Special**	1.00	17.00	27.00	31.20	39.00	191.00
	Permanent	1.00	15.00	24.00	26.99	35.00	152.00
	Total	**1.00**	**16.00**	**26.00**	**29.05**	**37.00**	**191.0**

To compute the statistics, we count 3655 sentences. From these Tables, we note that the recorded sentence lengths (words) vary from a minimum of 1 to a maximum of 191. A typical sentence has 29 words which stand for the average length of sentence. However, the length of sentence differs according to the type of speech. In international speeches, the president uses in general more words than in national ones Table (3). In the former, the average length of sentences is 35 words. This average is close to Proust's average 38 in his well-known book denoted in French "à la recherche du temps perdu". Table (3) depicts that the first quartile, the median and the third quartile are respectively 22, 31 and 43 words within sentences of international speeches. This respectively means that 75%, 50% and 25% percent of sentences, in international speech,

contain more than 22, 31 and 43 words. In contrast in national speeches, the first quartile, the median and the third quartile are respectively 15, 24 and 35 words per sentences. Figure (1) depicts the difference between both types of speech by drawing the curve of word counts in sentence relatively to the total words.

Figure 1: Sentence length distribution in words/terms



This figure highlights some differences in shape between both curves. These differences can relatively be statistically significant. To deal with this question, Welsch's statistical hypothesis testing comes in handy as we compare null and alternative hypotheses. Null hypothesis states that the average number of words in each speech is the same regardless the speech type while the alternative states that the average number of words in international speech is greater than the ones national speeches. The computation of Fisher's statistics of test gives a p-value = 0. This means that under null hypothesis, we have a null probability to observe this difference between both average words number. Therefore, we reject the null hypothesis and accept the alternative which states a significant difference between the average numbers of words in both types of speeches. The number of words remains a common measure of sentence length. According to Flesch [8], the sentence comprehension and readability may depend on its length in connection with to the reader or listener experience and ability to memorize. To fit a large audience, sentences of speech must be short in words and composed of words with a weak number of syllables [8]. In fact, Flesch proposed some standards for English reader, in [8, p. 38], based on the sentence average length in word. From *Art of Plain Talk* [8], Flesch defined a scale of difficulty with respect to the average length of sentence in speech in relation to the average reader assessment of the language. The scale of difficulty is defined on 7 levels (ranging from very easy to very difficult). For Flesch, the average length of sentence in a very easy speech would be 8 or less words while a standard, difficult and very difficult speech would have 17, 25 and 29 or more words. Based on Flesch analysis, a president's speech is considered difficult in general for a standard reader. Moreover, in international speech the difficulty increases. We note that Flesch's standard was deduced from English text. According to Richaudeau [9], the average length of sentence in word counts would not be the unique criteria to evaluate a speech readability. For Richaudeau, the readability depends more on sentence structure than sentence length particularly for French language. Richaudeau focused his analysis on the reader or listener memory span. The length of memory span determines then the reader memorization ability. The memory span varies between 8 and 20 words. Richaudeau realized 3 clusters of readers based on the average length

of memory span. From his paper Richaudeau [9], the average memory spans are approximately 8, 13 and 20 words respectively for moderately educated, enough educated and educated readers. The sentences have then to be divided to sub-sentences whose length in words fit readers memory span length in order that readers get enough chances to keep more information. Moreover, most important sub-sentence has to be put at the beginning of the sentence. However our analysis doesn't consider this aspect of sentences structures.

Next subsection is going to focus on the words of sentence. In which, speeches are considered as a bag of words in order to evaluate the lexical richness and figure out common words in president' speeches.

3.2 SENTENCE WORDS

This section analyzes the words of the president IBK. To do so, the speeches are considered as a collection of words regardless of the sentence structure. The main purpose remains identifying common words and their frequency in the speech in order to evaluate the main words which characterize the president's agenda. As a result, some words are lemmatized and grouped in the canonic form. Accordingly, the plural forms are transformed to singular for substantives. All of verbal forms are reduced to their infinitive. Figure (2) draws the bar plot of the most frequent words or terms. We note, among this shortlist, some words/terms such as "Mali, Pays, Paix, Gouvernement, President, Developpement, Securite, Etat". The most frequent 719 occurrences is "Mali". The word "Pays" means "Country" in English, counts 483 apparitions. Both words appear generally together. Jaccard's association level between both is given in Figure (3).

The association level stands for the ratio of simultaneous occurrences of two words. Jaccard's association takes its value from 0 to 1 as 0 describes a weak association rate and 1 one. Moreover association level, between the words "Mali" and "Pays", is equal to 0.75 Figure (3). In fact, these words constitute the foundations of the president addresses because of their high total apparition number in the texts, approximately (1202). Figure (3) denotes that the term "Mali" focuses the priority of president IBK' speeches according to a high association with some others terms. These terms are composed of words such as "Paix", "Developpement" and "Securite" which mean respectively "Peace", "Development" and "Security" in English. Accordingly, the president charges his government to organize the security and Development in order to contribute to the peace in the country. This is highlighted by the association value from Figure (3). In fact this figure shows that the association between the word "Government" and "Peace", "Development" and "Security" are respectively 0.56, 0.57 and 0.41. Among the cited actors, in this shortlist "Government" has the highest association with the priority words "Development" and "Security" Figure (3).

Figure 4: Association between common terms

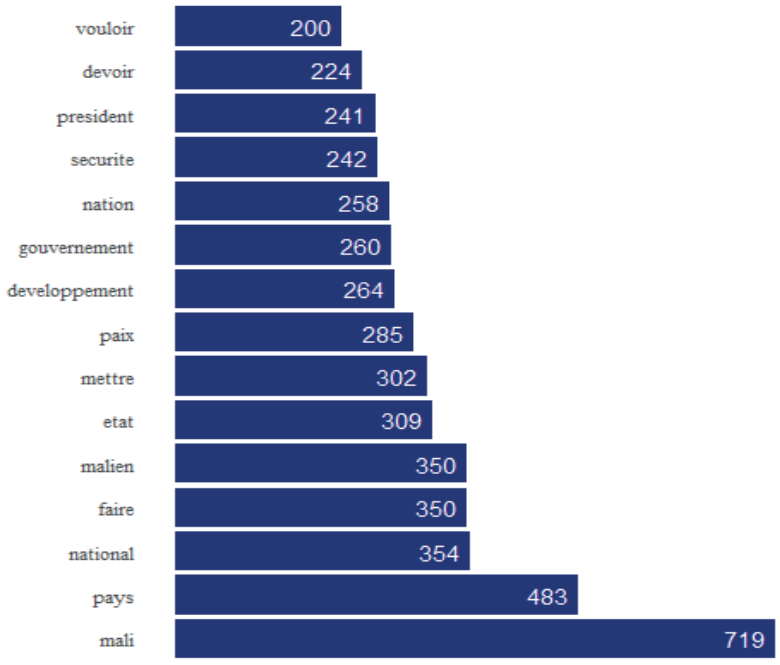


Figure 2: Most frequent words in the speeches

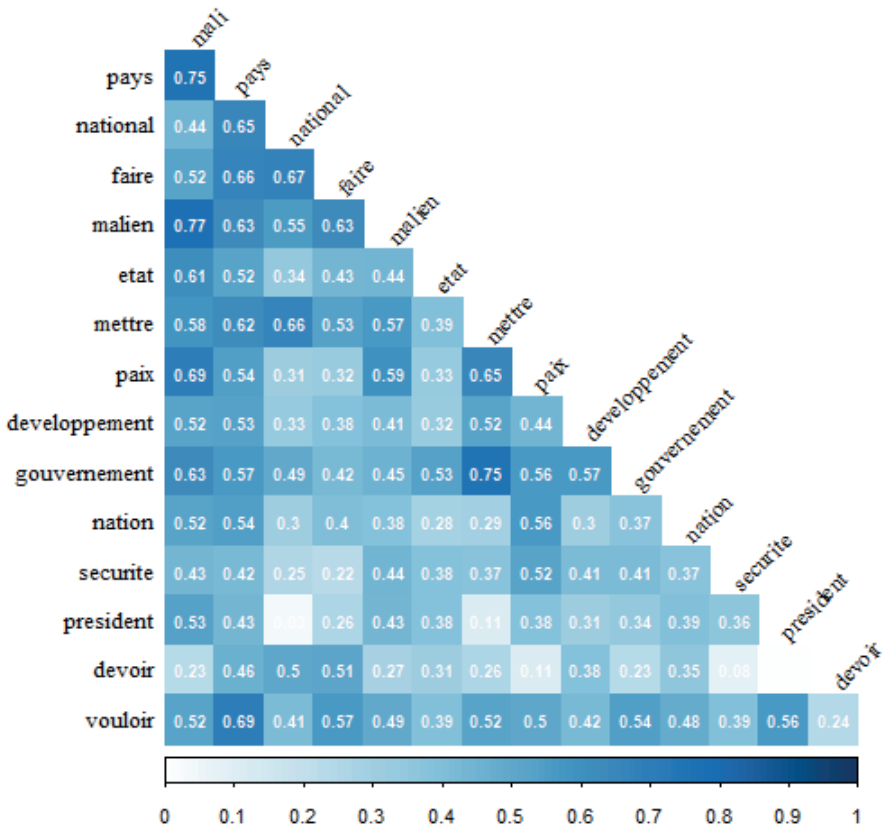


Figure 3: Association between terms

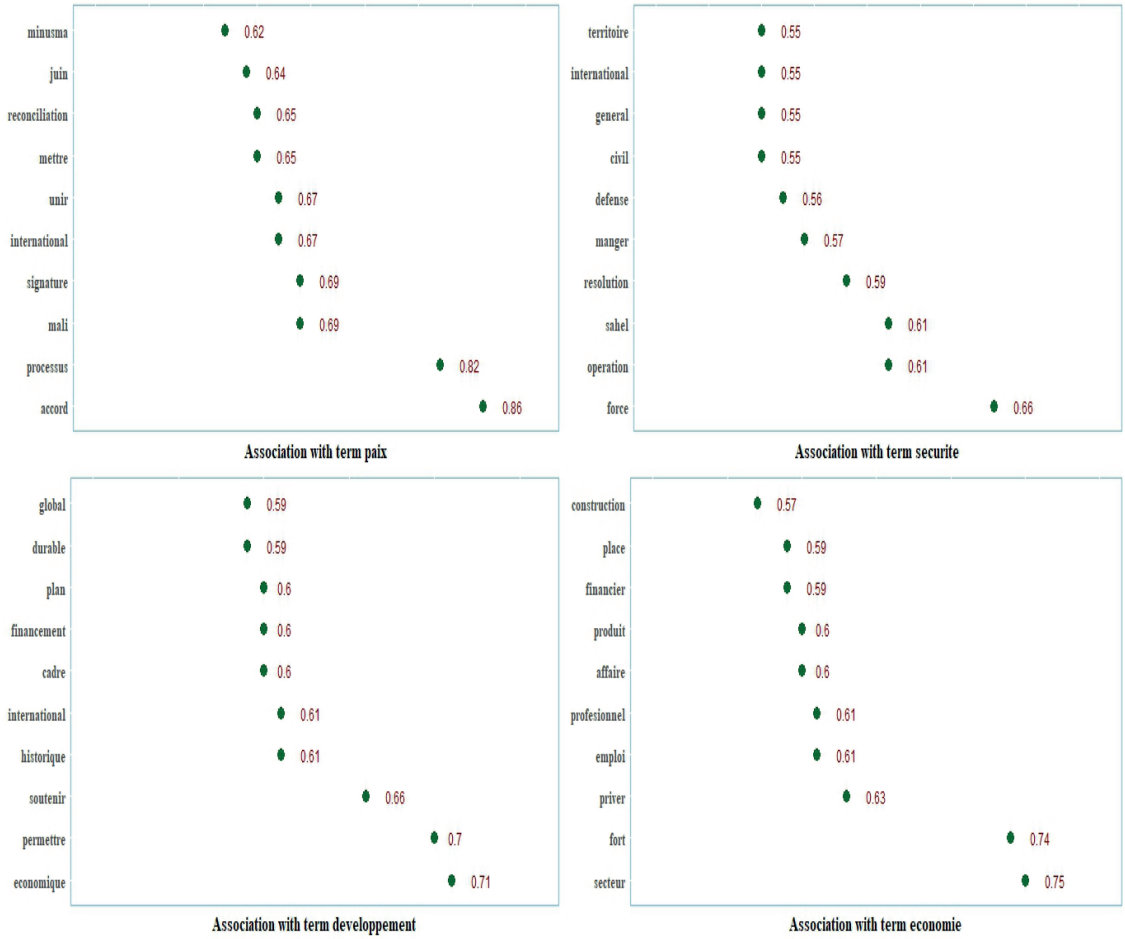


Figure (4) depicts the most associated words with "Security". According to this figure, the main threat comes from "Sahel" region and characterized by "Terrorists". To overcome this threat the president IBK counts on the military "Operations" conducted by Tchad and MINUSMA led "Operations" combined with a "Reform" of the Malian security sector. As such the national army needs to undergo major changes as to have a chance to deal with national securities issues. This translates, from the part of the president, little confidence on the national army's capabilities to face threats related to terrorism. Figure (4) shows also the association rate between word "Development" and others terms. "Development" is highly associated to the words "Economique" which stands for "Economic" in English. The speeches point out that the development must be economic and sustainably financed according to "Plan" and "Project". Moreover, most efforts have to be made to ensure economic development particularly for regions and population in North of the country. To overcome these challenges, the president relies mainly on tax receipt and mining resources. According to Figures (4), the speeches are characterized by an absence of economic terms. Economic aspects have weak consideration in his political agenda although this can be considered as the real cause or source of the insecurity in the North and social tension in the rest of country. The next section presents complexities analysis of the president's speeches in order to have another view on the contents of his speeches.

4. COMPLEXITY ANALYSIS

Herein, the analysis focuses on 5 usual indexes from literature review. These indexes are used to measure in general texts in speech features. The diversity of vocabulary in words is therefore measured by the index Type Token Ratio (TTR), Corrected Type Token Ratio (CTTR). Where TTR stands for the ratio between the numbers of unique words denoted Type (t) and the number of words in a speech denoted Token (n). In fact, repetition of types defines the token (words of speech). TTR gives us an idea of the number of different words used in a speech and the diversity of the vocabulary. Therefore, Equation (1) gives us the formula of TTR

$$TTR = \frac{\mathrm{t}}{n}, \tag{1}$$

and CTTR makes a correction of TTR to reduce the token dependence

$$CTTR = \frac{\mathrm{t}}{\sqrt{2n}} \tag{2}$$

In addition, the complexity, the intelligibility and the readability features are going to be analyzed by considering FLESCH, FOG index and Automated Readability Index (ARI). The formulas of these indexes are respectively

$$ARI = 4{,}71\,\frac{C}{W} + 0{,}5\,\frac{W}{S_{ari}} - 21{,}43, \tag{3}$$

where n and S_{flesch} stand respectively for the average number of words in sentence and the average number of syllables in words.

$$FLESCH = 206.835 - 1015n - 84{,}6S_{flesch}, \tag{4}$$

where S_{fox} represents the percentage of words whose number of syllables is greater than 3 in the speech. The formula of Automated Readability Index (ARI) is

$$FOX = 0{,}4(n - S_{fox}), \tag{5}$$

with C/W and W/S_{ari} represent respectively strokes per word or word length and words per sentence. For need of analysis, five indexes are computed for all speeches. Figure (5) shows Pearson correlation values between indexes. We note that FOG index and ARI are highly and positively correlated. This involves that both indexes vary in the same direction. However FOG index and ARI vary in opposite sense with FLESCH index Figure (5).

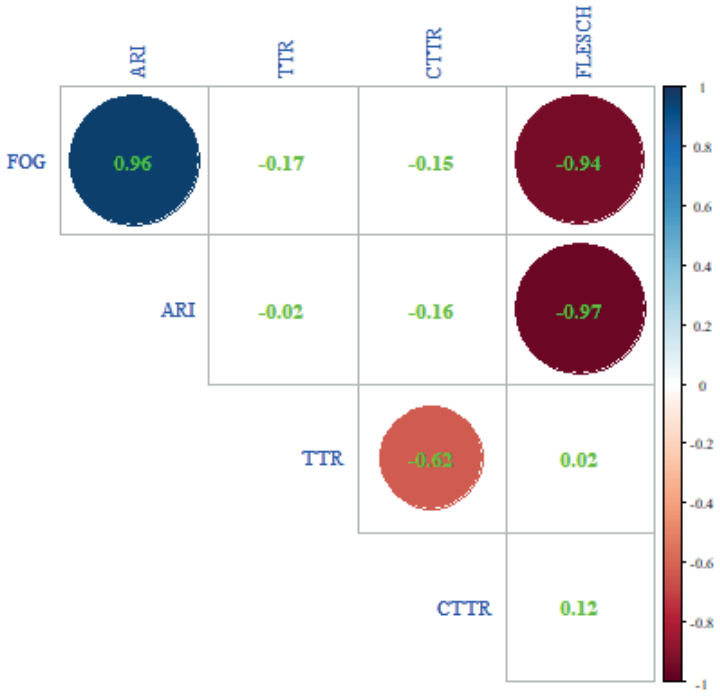


***Figure 5**: Correlation matrix between Indexes*

We observe a linear dependence between these indexes in president speeches. Another way for the diversity index, the correlation is −0.687. Therefore TTR decreases with CTTR. The diversity index has a weak linear dependency with the readability index as pointed out in Figure (5). Statistical tests are made on the nullity of these correlation values. The estimated p values of tests indicate that the correlations between TTR and FLESCH, ARI are not significantly different to zero Figure (5). This involves that an absence of linear relationship between TTR and FLESCH, ARI. Moreover, Figure 6 shows the results of principal component analysis on the indexes. We note both factorial axes explain 87% of total variability, in which, first axis highlights the readability and intelligibility characteristic of speech by opposing FOG and ARI to FLESCH index. Meanwhile the second axis opposes TTR & CTTR and denotes the lexical richness and diversity of speech. Accordingly, the analysis of all indexes is not necessary for our data. From correlation analysis, we consider TTR and FLESCH index instead of all due to the linear dependence inside the group of readability index and the lexical richness index. Figure (6) depicts this conclusion.

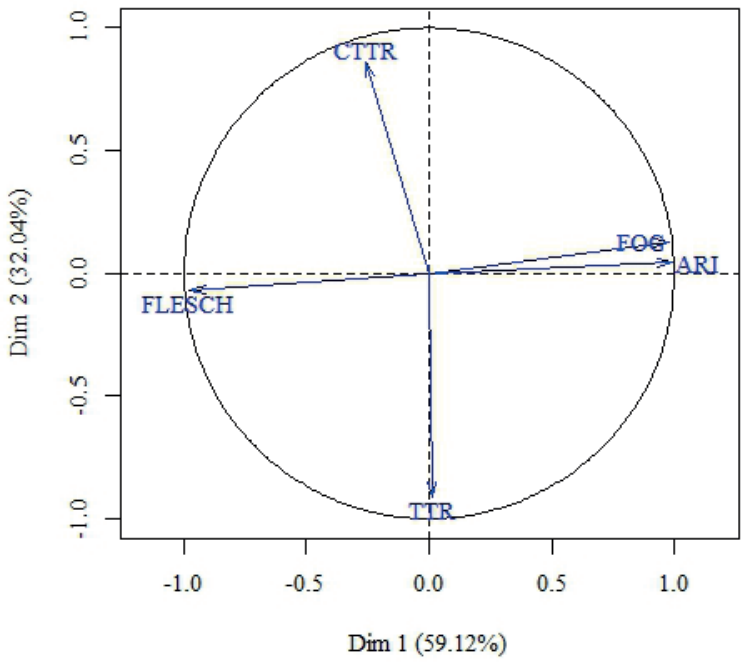


***Figure 6**: Principal component analysis of Index*

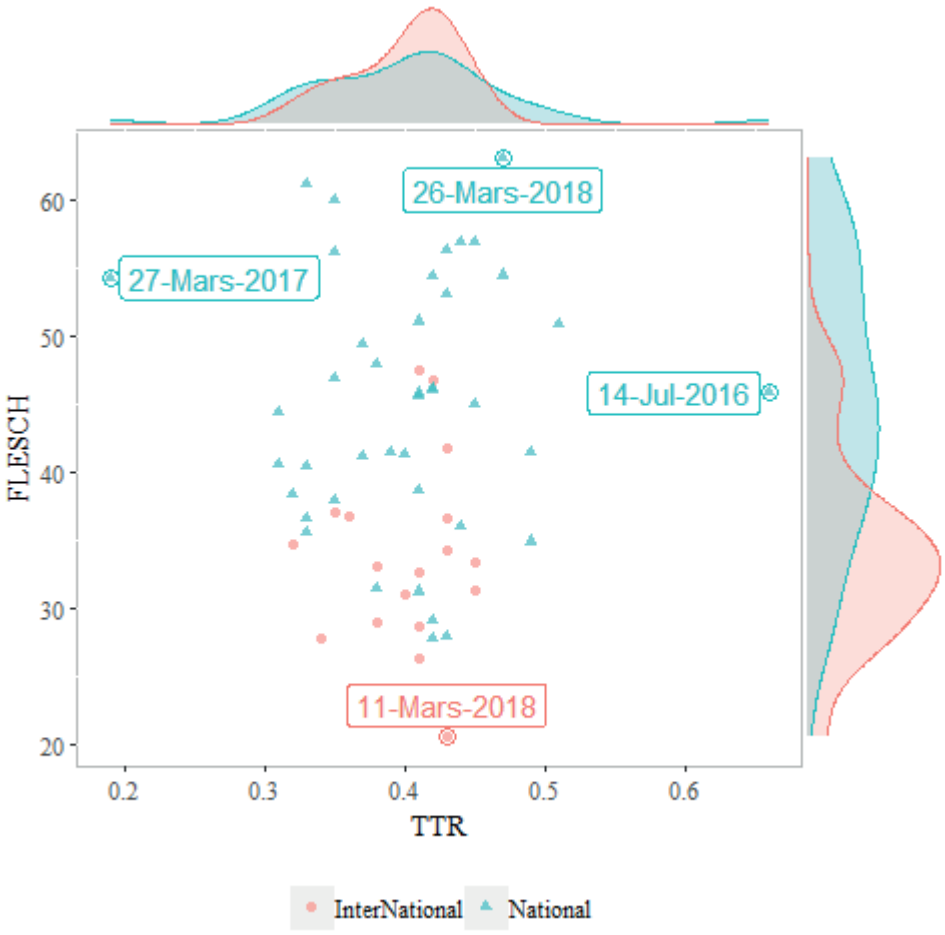


***Figure 7:** TTR and FLESCH scatter plot*

4.1 ANALYSIS OF TTR AND FLESCH INDEX

Figure (7) shows the scatter plot of TTR in function of FLESCH index for each recorded speech during observation period. TTR values are on the abscissa and FLESCH index on the ordinate. This shows that the minimal and maximal of TTR are 19% and 61% computed on the speech national speech on 27-March-2017 and 14-July-2016. In addition the average and standard deviation of TTR value are 40% and 6%. Subsequently, the computation figures out that in average the vocabulary diversity is the same regardless the type of speech. President international and national speeches have the same average value of the type token ratio (TTR). A statistical test on averages confirms this result with p-value = 0.97. Another way, a typical speech of the president IBK has 40.76 as FLESCH index with a standard of deviation equal 10.27. However, FLESCH index varies from 18.91 to 61.80. Unlike the TTR index, FLESCH depends on the type of speech. Accordingly, FLESCH index in National speeches is higher than International ones. This involves that the president IBK's national speeches contain in general less complex words and shorter sentences in comparison to international speeches. Therefore, international speeches average FLESCH score is 32.70 while national remains 44.21. In fact, an international speech is less readable than the national ones. To confirm the previous assumption a statistical test is made, on FESCH average index between both groups. The test concludes that the average readability index value in international speech is significantly less than his value in nation speech with a p-value = 0. Roughly speaking, we note that the vocabulary diversity or density in president speech seems steady around 40% and is not influenced by the type of audience (nation or international). Each speech is composed of 40% unique words. Nevertheless, an international speech contains in average more words with a high number of syllables, lengthened and complex structured sentences than a national speech. Therefore, international speeches have got weak

value in FLESCH scale index and are more difficult to read and understand than national speeches for standard educated audience. In addition, the averages FLESCH score in national and international speech get 32.70 and 44.21 scores correspond respectively to "difficult" and "Very difficult" level in FESCH scale. Accordingly, the president IBK's Speeches are as difficult as a scientific document. To understand them, audience shall be highly educated (at least 16 years at school). Meanwhile, the percentage of high educated people is weak with a higher education graduate rate around 1,7% according to RGPH2009[3]. This denotes that only a few percentage of the population have the requirements to understand presidential speeches. Accordingly, the president agenda remains unintelligible for most citizens. Indeed, president needs to work deeply on his speech texts in order to make his speeches with simple substantive in short sentences (around 17 words according to Flesch standard).

Next section proposes ideas identification and contents modeling in order to highlight the main priorities in his political agenda and to know how suitable his responses are to deal with citizenry social problems such as incivility, insecurity, unemployment, and social demands.

4.2. SPEECH IDEA DISCOVERY AND MODELING

Herein, the similarities in the words cloud are going to be analyzed. The purpose is to describe the distance in this cloud between different words and to aggregate some of them to compose homogenous groups. The similarity, in each homogenous group, characterizes the related issues or ideas supported. The identification of groups and their related ideas allows highlighting president main response to face the citizenry expectations. An unsupervised classification method is therefore applied to discover the similarities between groups of words. The hierarchical clustering is performed. The analysis focuses on the most frequent words which appear at least 150 times except words {mali, malien and grand}. To improve the accuracy of this approach, the hierarchical clustering is used on the results of corresponding analysis applied on the TDM. To reduce the number of column variables, the speeches are grouped per years. Indeed, the analyses allow underlining the main words and idea per years. Correspondence analysis is therefore performed on the Term Document Matrix, in which the speeches are grouped by year in column.

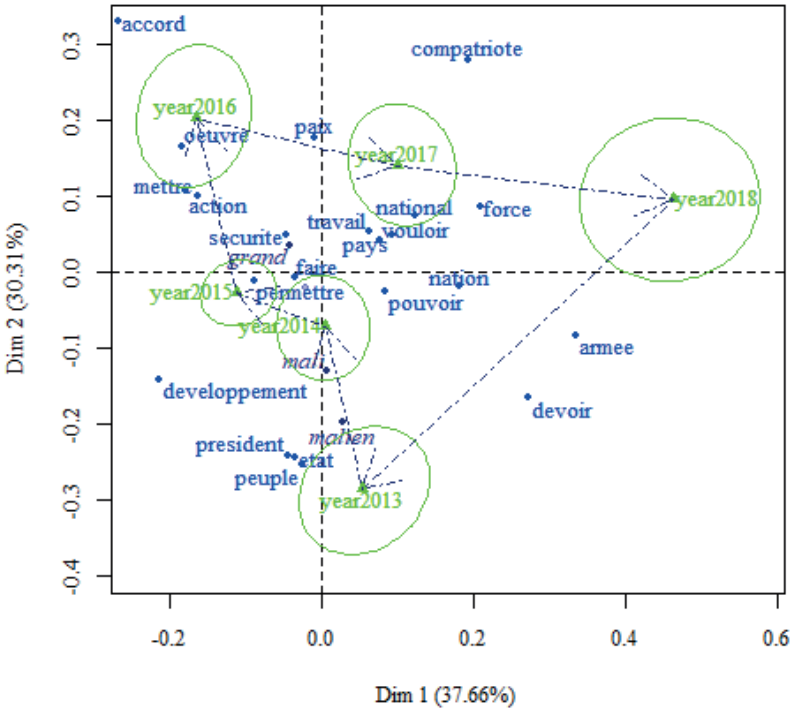


Figure 8: ***Corresponding analysis of words matrix and confidence ellipses at 95% around each year***

[3] The last census performed by « INStitut national de la Statistique du Mali (INSTAT). »

Figure (8) depicts simultaneously the delivered words per year. This figure depicts corresponding analysis and allows deducing information on the attraction between words and context of delivered years. From this analysis, we note that both factorial axes explain 67% of the total variability whose 37% and 30% are explained on the first and second axes. This figure underlines the evolution of the president words in his agenda year per year. Indeed, first factorial axis figures out the difference in the contents of speeches between 2016 and 2018.

In fact, 2016 was president first midterm. President focused his speech contents on setting up of the peace based on the agreement obtained with the armed group in Algiers, while 2018 was an election year and president had to prove his firmness by force face to the terrorist attack. Second axis opposes the contents of speeches deliverer in 2013 to 2016 and 2017. Accordingly, president focused firstly on the people by activating the patriotic leverage in 2013 but not for long time. After, a large place has been given to the research of peace according to the development policy and security without a great success. However, president changes his policy and underline the use of force in order to overcome the security not in the North but in the Center of country. Based on this analysis, a hierarchical clustering is performed and shows on Figure (9).

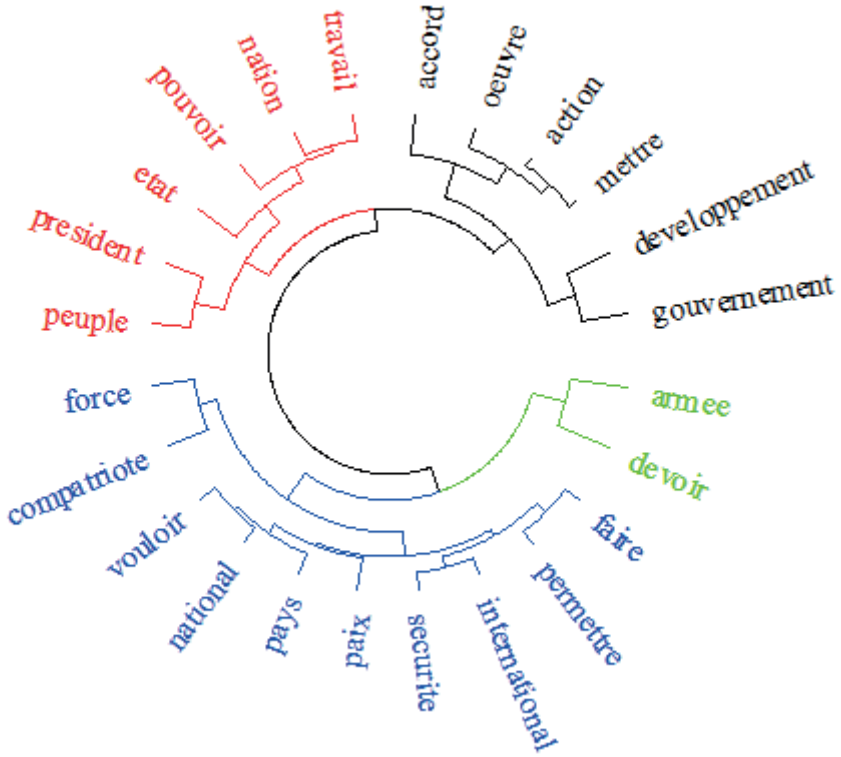


Figure 8: ***Words clustering and idea discovery in speeches***

The figure draws a circular cluster based on Ward criteria. We consider 4 main clusters to highlight the priority purpose during the five first years. Each cluster has a specific color. The first cluster has red color and contains eight words which were highlighted between 2013 & 2014 just after his election. On Figure (9), this cluster defines the set of words whose "President" is in charge. Second cluster underlines the president priority in 2015 and 2016. This date priority was "Development" and the realization of allowed "Action". Accordingly, the development in Mali must be conceptualized and rationalized by Government in order that People can take their fate to overcome the social problems. The third cluster is composed of two words ("Armee" and "Devoir") with green color. Both words are naturally evoked in president speeches and particularly in 2013 and 2018. The last cluster has blue color and underlines the basement of the president agenda just before 2018 (election year). This diversity of words in this denotes an inability to president to well understand the complexity of context and to propose suitable solutions to overcome the consequence of the crisis.

5. CONCLUSION

This paper investigated the usual words that the president IBK used in a corpus of speeches delivered between from 2013 to 2018. To perform a considered analysis, the corpus of speeches is transformed to Terms document Matrix (TDM). Term Document Matrix allowed us to compute some ratios and to evaluate and extract relevant information on the lexical richness and the complexity. Therefore, the president speeches are composed of 40% of unique words in average according to the type and token analysis. This average is almost the same regardless of the public. Meanwhile, the complexity evaluated by the Flesch criteria shows that speeches delivered remain difficult for standard readers. Indeed, his speeches are not understandable by the vast majority of citizenry. With a complexity close to Proust, the president delivers unreadable speeches in which citizenry cannot generally decipher the real message. Subsequently, it is difficult to know if his speeches have any impact or influence on the citizenry behavior. For the next, the president IBK has to work on his speeches in order to propose short sentences with simple clear substantives. This will allow him to have a suitable speech for the vast majority of Malian people. To complete this analysis, clustering and correspondence analysis methods are used on the TDM. The clustering allows building groups of similar words/terms in order to extract some idea while the corresponding analysis leads to evaluate the evolution of the president priority year after year during observation period. With clustering, we extract four main ideas in the speeches such as patriotism for the country, respect, peace and development. However president expects on the economic development but this is not very concise in their speeches. Therefore, the economics word is not supported by suitable economic terms. Indeed, president wishes economic development without clear appraisal of problems and without a real identification of leverage variables. This explains how president has a problem to clearly define his economic policy and find solution to deal with the economical aspect which can be perhaps the cause of homeland insecurity. In addition, the corresponding analysis leads to identification the president priority year per year. In fact, corresponding analysis shows that president began his quinquennial by working on the patriotism leverage of citizenry. His speeches were structured around homeland security without real responses during midterm.

With this work, statistical methods were applied in order to understand the speech of Malian president. The purpose of this analysis was to make an evaluation of speech vocabulary richness and its complexity in order to know if these are suitable for citizenry. Our analysis identifies some weakness in his speech in relation to the complexity and the lack of appropriated economical terms. For the future, we hope on the machine learning methods to deeply evaluate text in order to analyze words in sentences structure. This allows us to mix linguistics and statistics approaches together and to figure out more relevant information from president speeches texts.

ANNEX: CORPUS OF SPEECHES

Table 4: Selected Speeches in corpus

Date	Speech	Date	Speech		Date	Speech
04 Sept. 2013	Investiture speech	**01 Jan. 2014**	New year		**29 Dec. 2014**	Greetings to the Armed Forces for new year
22 Sept. 2013	Independence day	**20 Jan. 2014**	Army day		**29 Dec 2014**	Greetings to the government for yew year
27 Sept. 2013	UN speech at 68th Conference	**31 Jan. 2014**	African Union speech		**01 Jan. 2014**	New year
02 Oct. 2013	Speech to the nation after 68th Conference at UN	**26 Mars 2014**	Martyr's Day		**20 Jan. 2015**	Army day
01 Nov. 2013	National conference on the North of Mali	**01 May 2014**	Workforce Day		**19 Fev. 2015**	Forum of Bamako
09 Dec. 2013	Strasbourg parliament speech	**05 May 2014**	5th councils of agriculture		**13 Mar. 2015**	Conference on Eco-Development in Cairo at Egypt
		19 May 2014	After the event Kidal due to PM visit		**15 Apr. 2015**	CILSS
		22 Sept. 2014	Independence day		**06 June 2015**	FAO
		04 Nov. 2014	Official opening of the university		**20 June 2015**	Peace Accord signature process
		27 Nov. 2014	UN speech 69th conf.		**26 June 2015**	Presentation of VGAL report
		01 Dec. 2014	AIDS day		**11 Sept. 2015**	CILSS
		09 Dec. 2014	At national School of Administration (ENA)		**22 Sept 2015**	Independence day
					26 Sept 2015	Tribune of UN 70th session
					28 Sept 2015	Session of General discussion at UN 70th
					30 Sept 2015	MINA event during pilgrimage of mecca
					29 Oct. 2015	India and Africa summit
					14 Nov 2015	Open day of Institutions
					20 Nov 2015	After the attack of Radisson Blu

Table 4 bis: Selected Speeches in corpus

Date	Speech	Date	Speech	Date	Speech
01 Jan. 2016	New year	**01 Jan. 2017**	New year	**01 Jan. 2018**	New year
20 Jan. 2016	Army day	**20 Jan. 2017**	Army day	**05 Jan. 2018**	1th ministers council with S.B. Maiga
10 Apr. 2016	Islamic summit in Istanbul reading By A. DIOP	**27 Mar. 2017**	National conference (Opening)	**20 Jan. 2018**	New year
07 June 2016	In Response to VEGAL report of 2015	**27 Mar. 2017**	National conference (Closing)	**11 Mar. 2018**	New Delhi, International Solar Alliance
17 June 2016	National conference on the North first anniversary	**08 Apr. 2017**	1th councils of ministers with A.I. Maiga	**26 Mar. 2018**	Martyr's day
19 June 2016	Forum IDENEUF 2	**19 Sept. 2017**	UN conference	**01 May 2018**	Workforce day
14 Jul. 2016	Gao tragic event	**22 Sept. 2017**	Independence day		
22 Sept. 2016	Independence day	**16 Nov. 201**	Dubai Forum		
		07 Dec. 2017	Investor Forum		
		19 Dec. 2017	G5-sahel summit		

RÉFÉRENCES

[1] B. Escofier, 'Analyse factorielle des correspondances', Université de Renne, Renne, 1965.

[2] M. W. Berry and M. CASTELLANOS, Eds., *Survey of Text Mining II: Clustering, Classification, and Retrieval*. London: Springer-Verlag, 2008. doi: 10.1007/978-1-84800-046-9.

[3] M. W. Berry and J. Kogan, *Text Mining: Applications and Theory*. John Wiley & Sons, 2010.

[4] J. E. Cohen, 'Presidential Rhetoric and the Public Agenda', *Am. J. Polit. Sci.*, vol. 39, no. 1, p. 87.

[5] J. Wang, 'A Critical Discourse Analysis of Barack Obama's Speeches', *J. Lang. Teach. Res.*, vol. 1, no. 3, pp. 254–261, May 2010, doi: 10.4304/jltr.1.3.254-261.

[6] E. C. Sharndama, 'Political discourse: A critical discourse analysis of president Muhammadu Buhari's inaugural speech', p. 13, 2015.

[7] A. Osisanwo, 'I Belong to Everybody yet to Nobody: Pragmatic Acts in President Muhammadu Buhari's Inaugural Speech', *Athens J. Mass Media Commun.*, vol. 3, no. 4, pp. 297–320, Sep. 2017, doi: 10.30958/ajmmc/3.4.2.

[8] R. Flesch, *Art of Plain Talk*, First Edition. New York; London: HarperCollins, 2000.

[9] F. Richaudeau, 'Les structures des phrases', *Commun. Lang.*, vol. 1, no. 1, pp. 19–25, 1969, doi: 10.3406/colan.1969.3706.