



## Discourse Markers in Sudanese Spoken English: A Corpus-Based Study

**Abdulaziz B Sanosi \***

Prince Sattam Bin Abdulaziz University, Saudi Arabia

### Abstract

The present study investigates the use of Discourse Markers (DMs) in Sudanese non-native spoken English. Two corpora were compiled to typify Sudanese and British spoken English. Each corpus comprises 108600 words constituting over fifty TEDx Talks presented by Sudanese and British speakers in the last seven years. The study focusses on the frequency and placement of three DMs, namely: *you know*, *well*, and *okay*, in the Sudanese corpus by adopting data from the British native speakers' corpus as a measuring criterion. The frequency of use of the DMs was calculated using AntConc software while the data were manually analysed using Key Word in Context (KWIC) lists to study the positioning patterns of the DMs in the two corpora. The results of the study point to the higher frequency of DMs in the British corpus compared to the Sudanese one. Non-native speakers, nevertheless, used the three DMs in all in-sentence positions i.e. initial, medial, and final just as their native counterparts did. The results also show that *you know* was the most used DM in the two corpora and it consistently occurs in the medial position. It was also found that Sudanese speakers overused the DM *okay*. A number of implications of the results are discussed and relevant recommendations are presented.

**Keyword:** discourse markers, ted talk genre, corpus linguistics, discourse analysis

### INTRODUCTION

One basic feature of discourse as a specific linguistic structure is that it should be coherent and logically presented. Speakers tend to use different linguistic units or particles which are termed Discourse Markers (DMs) to maintain the coherence and logic flow of the discourse. Accordingly, it can be argued that efficient use of DMs can help in achieving the ultimate goal of any discourse i.e. mutual understanding and effective communication. Absence or shortage of using such linguistic units, on the other hand, may result in pragmatic failure.

These characteristics of DMs attracted the attention of researchers who investigated issues such as their functions in different genres, the volume of their use by specific speech communities, the best methods to help EFL learners master them, and the use of them across various languages. Therefore, DMs were the topic of a large body of research that addressed various inquiries. An interesting question attempted by some of these

studies was whether DMs are culture or L1-specific. The argument dealt with regarding this question was supported by investigating situations such as the different techniques used by EFL writers of various L1 backgrounds. Such variations were attributed to the cultural and metalinguistic features of different languages which provoke using writing techniques including DMs in diversified ways. The Contrastive Rhetoric Hypothesis by Kaplan (1966) suggested that “rhetorical patterns of language are unique to each language and culture” (Connor, Nagelhout, & Rozycki, 2008, p. 1) and thus implied that writers transfer their spoken language behaviour to writing since culture and social aspects of language are more likely to be connected to oral discourse. Using different numbers and types of DMs was perceived as one feature of this claimed variation.

Research issues regarding DMs were dealt with by utilizing different methodologies. One of the trending approaches in investigating DMs is corpus-based studies. The current research applied such an approach to investigate the DMs used by Sudanese non-native English speakers in contrast with British native speakers. It is envisaged that the findings of this research will have theoretical and pedagogical implications regarding the amount and methods of teaching DMs since English in Sudan is acquired mainly through classroom instruction.

## **BACKGROUND OF THE STUDY**

### **Discourse Markers**

Although DMs were abundantly studied in the recent decades, there is no consensus on their definition or functions. Even the term DMs is not agreed upon since the concept is also referred to by many other terms such as Sentence Connectives (Halliday & Hasan, 1985), Pragmatic Markers (Fraser, 1996), Discourse Particles (Aijmer, 2002), and Cohesion Markers (Gee, 2018) to name few. The term DMs, however, is the most preferred by researchers and scholars.

Researchers also suggested different definitions of DMs most of which were from a functional perspective. They are considered as elements that facilitate the addressees’ interpretation of the discourse according to the surrounding context (Aijmer, 2002) and mark different aspects of participants’ intentions (Guo, 2015). Thus, it is noted that the occurrence of these “seemingly empty expressions found in oral discourse” (Brinton, 1996, p. 29) normally signals a change in the development of the discourse (Jabeen, Rai, & Arif, 2011) whether by the speaker or the addressee. Some other researchers e.g. (Fraser, 1996) discussed the wide range of the classes of DMs. In the English language, for instance, a DM can be an adverb of time (*now*), an adjective (*okay*), a conjunction (*however*), a prepositional phrase (*by the way*), a phrase (*I mean*), or an interjection (*oh*). The focus on the functional approach in defining DMs indicates that although they are important in facilitating interpretation of the discourse, DMs nevertheless have no semantic value and may not denote their exact meaning when used as cohesion devices. Correspondingly, a specific DM can have different meaning according to the context or the speaker.

The main issue emerged as a consequence of the various classes and the different perceived functions of DMs is reflected in the typology of them. The dispute related to

such an issue is whether to classify DMs according to their syntactic categories or to their role in the discourse (Asik & Cephe, 2013). This is basically because it is difficult sometimes to decide if the word *actually* in a sentence such as “he is actually sick” is a discourse marker implies emphasis or an adverb qualifies the adjective sick. For other words such as the interjections *ah*, *oh* or the gap fillers *hum*, *aha* there is a more consensus to consider them as DMs because they have no literal connotation to be confused with their pragmatic function.

In fact, the notion of pragmatic function was taken by many researchers to classify DMs accordingly. For Example, (Farser, 2006) suggested four classes of what he termed Pragmatic Markers which were: Basic, Commentary, and Parallel Pragmatic Markers. He termed the fourth group Discourse Markers and defined them as those elements which connect adjacent discourse segments. It is noted that Fraser (2006) used the term DMs only to point to those elements that connect the upcoming unit of discourse to the previous one, a concept that was used earlier to define DMs by many scholars.

A more generic classification subdivided DMs into two categories. In this regard, (Brinton, 1996) admitted that DMs cover various pragmatic functions. He, therefore, suggested that DMs can best be classified according to the functions they serve which can be either textual or interpersonal. The textual category includes those elements which are used, according to context, to maintain the coherence of the discourse. This group includes DMs that are used to launch and close discourse, to indicate old and new information, to mark turn-taking, and to signal the relevance of discourse units. In short, this group contains elements that related to the context (Asik & Cephe, 2013) and devotes language structures to achieve this goal. Interpersonal category, on the other hand, includes those DMs which mark “the expression of the speaker's attitudes, evaluations, judgments, expectations, and demands, as *well* as of the nature of the social exchange, the role of the speaker and the role assigned to the hearer” (Brinton, 1996, p. 38). Thus, interpersonal DMs are related to social interaction and feature the role of speaker and hearer in the discourse.

The aforementioned discussion reveals the uncertainty about DMs definition, terminology and classification. However, an identification of the characteristics of what can be termed a DM can help in standardizing the concept at least for research purpose. In this regard, Holker (1991) reported four characteristics as distinguishing criteria of DMs. According to (Juker, 1993), these criteria were:

(1) They do not affect the truth conditions of an utterance; (2) They do not add anything to the propositional content of an utterance; (3) They are related to the speech situation and not to the situation talked about; and (4) They have an emotive, expressive function rather than a referential, denotative, or cognitive function (p. 436).

Brinton (1996) added that DMs markers are normally short elements that are used more generally in oral discourse and form different tone group whether rising-falling or falling-rising. He also noted that they more frequently occur at the sentence-initial position either inside or outside the syntactic structure.

The flexible and multifunctional nature of DMs makes them one of the most used classes in English. (Aijmer, 2002) noted that DMs outrank basic grammatical category of prepositions, adverbs, determiners, conjunctions, and adjectives" (p. 2) in the frequency of occurrence of 50000-word sample from the London-Lund Corpus of Spoken English. This importance of DMs in controlling the flow of discourse inspire researchers to study them in regard to gender, age, and linguistic background (Guo, 2015) in different settings and using different methodologies a relatively new trend of which is corpus linguistics.

### **Corpus Studies**

Since its emergence in the 1960s, corpus studies have represented a significant methodological development in linguistics. It can be simply defined as "the study of language based on examples of real language use" (McEnery & Wilson, 1996, p. 1). It thus marks an outstanding shift from the previous approaches which tended to describe the structure of language and generalize the findings to portray how language is used. The focus of corpus-based studies, on the contrary, is on the functional description of the language and therefore based on a real sample of used language. Moreover, it is crucial that this sample of a used language is not gathered in an observed or structured setting as this may violate its natural quality. It is rather concerned with the "language which has occurred under circumstances in which the speaker was known to be doing something more than demonstrate how the system works" (Timmis, 2015, p. 2).

The novelty feature is then the backbone of corpora with which it would be safer to generalize research findings and implications since these findings are based on frequent regularities. This highlights the common objective of most of the recent corpus studies which is to investigate a specific linguistic structure and to present the generated information about its frequency, its various forms and its communicative capabilities (Meyer, 2002). These features and potentials of corpus-based studies have recently turned the approach into one of the preferred methodologies of investigation in modern applied linguistics.

The approach of using corpora to analyze language-in-use is termed Corpus Linguistics. The term, although widely used and accepted, raised some argument regarding the question of whether corpus linguistics is a branch of linguistics such as semantics and syntax. It is apparent that corpus linguistics is a methodology that is used by linguists to study other branches of linguistics. Being this tightly-related to other linguistic fields is a sufficient reason, some scholars believe, to consider it a branch of linguistics. According to (McEnery & Wilson, 1996) "while corpus linguistics is not an area of linguistic inquiry it does at least allow us to discriminate between methodological approaches taken to the same area of inquiry by different groups, individuals or studies" (p. 2). It can be convincing to consider corpus linguistics a branch of applied linguistics equal to psycholinguistics and sociolinguistics for instance (Aswini & Srinivasan, 2016) thus it can yield two instances of each linguistic approach i.e. corpus-based syntax versus theory-based syntax, corpus-based semantics and non-corpus-based semantics ...etc.

Using corpus linguistics approach for linguistic analysis was old. Early studies of child language e.g. (Brown, 1973) on morpheme acquisition were based on longitudinal data

based on child utterance. Lexicographers collected diaries about Key Word in Context (KWIC) to decide on the frequency of vocabulary use. However, the execution of these methods was rudimentary and laborious. Recently, the vast technological advance has driven corpus studies into new levels. Corpus linguistics is now associated with computer-readable texts that are analyzed with sophisticated software packages (Timmis, 2015) and stored on the cloud for broad and flexible access. This development has not changed the core of the methods followed by early corpus studies. Concordancing was used before many decades in linguistic analysis but not with such an efficacy that makes corpus studies an ideal option for modern applied linguistics research.

Corpus linguistics is now used to study both written and spoken discourse for studying historical linguistics, sociolinguistics, pragmatics, and language acquisition to present few examples. The word *text* referred to both written and spoken discourse in corpus linguistics literature (McEnery & Wilson, 1996). However, corpus-based studies that deal with the spoken discourse are relatively less. This can be attributed to, firstly, the nature of the written discourse which has clear-cut boundaries of linguistic units whereas spoken discourse is marked with pauses, fillers, and hesitations which may harden the analysis process of some linguistic features. Furthermore, upon building their own spoken corpora, researchers need to exert more effort because “there is a need both to make and to transcribe the recordings, which is time-consuming” (Timmis, 2015, p. 15). Accordingly, it can be argued that compiling a spoken corpus should be devoted to studying linguistic features that are more likely to occur in the oral discourse. Following this would make it more cost-effective and accordant with the research variables. Since DMs are normally a feature of spoken discourse as reported above, analysing their occurrence in various spoken genres, including the newly emerging ones, can provide more implications about discourse structure. One of the relatively new genres in this regard is what is used by TED Talks speakers.

### **TED Talks**

TED conferences started in 1984 aiming at publishing knowledge and new insights about Technology, Entertainment, and Design the three domains from which its label TED was coined. In 2006, TED launched its websites which accommodates videos of TED conferences and TEDx events that are organized all over the world by independent communities. Various themes are presented in TED talks since then including, but not limited to, entrepreneurship, self-development, education, biographies, economics, and public health.

The popularity of TED talks attracted the attention of researchers who suggested that it is “eventually becoming a new spoken web-based genre” (Scotto di Carlo, 2014, p. 122) that has distinctive discourse features. Researchers found that TED talks have common characteristics that mark them from other conference speeches (Ludewig, 2017). It is apparent that TED talks are marked with a conversation informal style that often includes some degree of comic and humour though it deals with formal issues. However, studies in this domain are still relatively new and therefore no consensus has been reached regarding the specific type or features of this genre.

The philosophy of TED talks that is symbolized by TED slogan Ideas worth spreading is behind choosing such style. It is a philosophy that aims to publicize knowledge to common people and bring elite scholars from many fields to present speeches that can be understood by the average audience. It requires no previous knowledge of the subject to enjoy the presentation aiming ultimately to entertain and inform in not more than eighteen minutes for a presentation.

The fact that TED presenters are related to different social, educational, and professional backgrounds allows TED talk-based corpora to have diverse discourse by diverse speakers. When it comes to TED presenters who speak in L2, their presentation might be considered more insightful as they represent elite people with high intellectual potentials. Therefore, they are likely to represent the highest linguistic levels that might be reached by their peers. Also, the rhetorical strategies followed by TED presenters imposed using various cohesive techniques including DMs. However, no much research was carried out to investigate this discourse. The present study attempts to contribute to filling this gap. A review of the trends that followed by most of the previous studies about DMs is presented in the next section.

### **Previous studies**

Mostly, corpus-based studies that investigated DMs are contrastive, functional-oriented, and pedagogic. They get the advantages of corpora in that they compare sizeable data of using DMs and compare the findings with existing native-speaker corpora. They also investigated the functions, positioning and frequency of DMs and revealed suggested implications for raising second language learning and teaching. Due to the recency of the approach, however, there are some gaps to be filled. Below, some studies that are similar in the scope and approach of the current study are reviewed.

(Fung & Carter, 2007) conducted a corpus-based study to contrast the use of DMs by non-native English speakers to the British use. They utilized a researcher-developed corpus of classroom discourse and compared it to a subcorpus of CANCODE British English corpus. Their findings revealed that non-native speakers used referential functional DMs more frequently. They nevertheless, showed a restricted use of other DMs. As it was presumed, native speakers were found to use DMs for a wider variety of pragmatic functions. This represented a great source of implications on how to familiarize EFL learners to DMs to make them more competent speakers.

Far from pedagogic setting, a study by (Jabeen, Rai, & Arif, 2011) investigated the frequency and setting of Pakistani English as spoken by media figures. Their YouTube-based corpus was then compared to ICLE-GB British English corpus to investigate the frequency, function and placement of eight English DMs by Pakistani English speakers taking British native as a measuring criterion. Repeatedly, more DMs were revealed to be used by native speakers, however, Pakistani speakers used DMs for a variety of functions and in all positions within the sentence i.e. initial, medial, and final. The authors did not consider this as a mark of improper use of English, rather, they adopted it as an evidence of the existence of a special variety of Pakistani English. They hinted to the importance of

raising such a variety and that Pakistani teachers and students should not preoccupy themselves to achieve native-like competence.

Another corpus-based analysis was conducted by (Polat, 2011) aiming to investigate the acquisition of DMs by learners. Using a longitudinal approach, the study was based on a developmental learner corpus compiled along a period of one year to trace the acquisition of *you know*, *like*, and *well*. The subject showed a dynamic method of acquisition of the first two DMs with overuse at the beginning of the period, a notable drop in the middle, and a proper use at the end of the period. The DM *well* was not used by the subject. The result indicated a method of acquisition that is similar to the U-shaped mode of acquisition followed by most second language learners. Such Implications can be used to design lessons and teaching plans regarding DMs instruction.

To compare the use of DMs by Norwegian learners of English to the native British English speakers, (Sandal, 2016) conducted a corpus-based study. The Norwegian corpus LINDSEI-NO corpus was the source of the data of advanced learners spoken English. Such data were compared to native speaker spoken English data drawn from the LOCNESS. The results revealed an unsystematic underuse of DMs by the Norwegian advanced English learners. Nonetheless, it was exposed that a variety of function was covered by this use. Sandal (2016) attributed such a misuse to both lacks of input in schools and L1 interference factors. Other cultural and personal factors were also acknowledged as causal factors for such improper use. The researcher recognized DMs as a pivotal factor that enhances communication and called for drawing the attention of their instruction.

(Ozer & Okan, 2018) investigated the DMs used by teachers in EFL classroom. It was based on two corpora of native and non-native teachers of English and contained a transcription of class recordings delivered by the participants. Like the case with EFL learners, EFL non-native teachers were found to use the majority of DMs less than native ones. The researchers suggested that such low-level use of DMs might be reflected on the learners' competence of DMs as another reason implied improving teaching syllabi on teacher education level.

It is apparent that all the reviewed studies agreed on the low level of non-speakers of English in using DMS. Improper use was represented in the low frequency of use and improper contextualization of DMs. It was found that this inadequate use to be caused by a number of linguistic, cultural, and instructional reasons. Because of the relative paucity of corpus-based studies that investigate spoken discourse, there is still a research gap to bridge. More studies that utilized genres which incorporate diverse speakers from the non-native and native background are needed. Also, conducting studies that include diverse participants from different linguistic backgrounds can contribute to the body of the research and present more evidence to its hypotheses. The present study is an attempt in this regard.

## THE PRESENT STUDY

The main objective of this study is to investigate the use of three selected DMs by Sudanese English speakers at a number of TEDx events that were held in Sudan. The focus of the study is on the frequency and the placement of the used DMs by Sudanese speakers

compared to native speakers' use. The DMs *you know*, *well*, and *okay* were selected on the bases that they are common in spoken English and they occurred frequently on the two corpora that were designed for the research purpose.

## METHOD

### Research design

The current study followed a quantitative descriptive research design. It is quantitative in the sense that it computed the frequency of occurrences of DMs in both Sudanese and British corpora of TEDx speakers. The frequency and percentage of use by each group of speakers were analyzed. Numbers of occurrences in the initial, medial, and final position were also calculated and analyzed.

### Material

The researcher compiled two corpora for the purpose of the research. The two corpora were extracted from the TEDx YouTube channel following specific procedures to maintain the appropriateness of the data and the equality of both corpora.

#### *Sudanese TEDx Spoken English (STSE)*

The data for STSE is composed of TEDx talks presented in Sudan in the period between April-2011 to July 2018. The talks were sampled from a number of events held in Sudan such as TEDx Khartoum, TEDx University of Khartoum, TEDx Soba Women, TEDx Nile Street and TEDx Mogran. Few videos were presented by distinguished Sudanese speakers at TEDx events outside Sudan e.g. (Dafaa-Alla, 2014) which was presented at TEDx MidAtlantic, USA. STSE incorporates transcriptions of 67 English talks offered by female and male speakers of different ages (from 13 to 50 years old). It contains 108621 words and they were delivered in a total of 12.8 hours. The talks covered various topics such as self-development, entrepreneurship, human rights, science, social issues, and education. The average length of the talks was 1638.2 words.

#### *British TEDx Spoken English (BTSE)*

A comparable native speaker corpus of the same genre was needed to measure the non-native speaker use of DMs against. For this purpose, a corpus comprising of 54 TEDx talks presented by British speakers was developed. The talks were presented at TEDx London, TEDx Birmingham, TEDx Manchester and TEDx Liverpool in the period from November 2011 to July 2018. The total words of the corpus were 108618 and it contained talks on similar topics as STSE corpus. Moreover, certain procedures for selecting the talks were followed to maintain the diversity of speakers by incorporating male and female speakers from different age levels. It was noted that, however, the average length of these talks was greater than that of the STSE speakers i.e. it was 2128.5 for a talk. The procedures adopted in selecting, transcribing and storing the data are detailed below.

### Procedures

All the data for the two corpora were extracted from the TEDx YouTube channel. For the STSE corpus, the researcher first searched for the uploaded videos of the events. Since



TEDx talks in Sudan are presented in English and Arabic, Only the talks that were presented in the English language were selected. Moreover, the researcher referred to the biographies of the speakers to select only those who have spent at least part of their lives in Sudan and/or have attended schools or college there. In the cases where there was uncertainty regarding that, the concerned talks were excluded.

To convert the video talks in a text format, the researcher used DIY caption extracting tool by Ridgway (2015). With a browser plugin, the tool provides the subtitle of the video in a *txt* format. The researcher manually pasted the transcription of each video in an independent file and code it as [STS>SAMPLE NUMBER]. There were two types of subtitles. Some talks were subtitled manually by TED volunteers while other were subtitled automatically by the YouTube subtitle and transcription service. For the first type, no further procedures were conducted. For the auto-generated subtitles, however, the researcher checked the transcription against the original video to correct any junk data that might be caused by sound quality, environment or non-English words. The final step of compiling the corpus was to create a metadata file containing information about the presenter, the topic, the file properties and the YouTube URL of each file. The researcher has reviewed the licensing considerations of TEDx from their website and email communication with the TED team before performing the aforementioned procedures.

An equal size corpus was drawn from TEDx talks delivered in Britain. Only the talks that were presented by British native speakers were elected. The above-mentioned measures for extracting and verifying auto-generated transcriptions of the videos were followed and the files were coded with the prefix [BTS>SAMPLE NO]. The files were stored in an independent folder with an independent metadata file giving information about the corpus.

## Data Analysis

The researcher adopted a multi-method approach in analysing the data. Both automated and manual analysis processes were followed in concordancing the corpora and analysing the data. A quantitative approach is followed in investigating the frequency of DMs occurrence in the two corpora. For this purpose, the researcher utilized AntConc software (Anthony, 2018). The generated results were then checked manually to exclude non-discourse markers usage of the targeted DMs. The filtered data were processed by an *Excel* formula to count the percentage of occurrences of the DMs in both corpora and compare the results. Further, the researcher used the KWIC lists generated earlier to determine the placements of the three DMs in sentences in the two corpora.

## RESULTS

This section reports the results of the use of the Sudanese and British TEDx speakers of three DMs i.e. *you know*, *well*, and *okay*. As the three elements were sometimes used by the speakers of the two groups to denote their actual semantic connotations, the first phase of the analysis took account of the cases where the non-DM use of them was observed. Table 1 details the overall uses of the three elements as DMs or non-DMs.

**Table 1.** Occurrences of *you know*, *well*, and *okay* in the two corpora.

Phrase/word	BTSE			STSE		
	Overall	As non-DM	%	Overall	As Non-DM	%
<i>you know</i>	214	16	7.5	177	35	19.8
<i>well</i>	180	70	38.9	121	50	41.3
<i>okay</i>	42	14	33.3	139	21	15.1

The non-DM use of the three elements across the two corpora implies that *well* was the most-used as non-DM among the three elements. The uses of the other elements, however, were varied. After filtering the results, frequencies of occurrence of the selected DMs in both corpora were calculated. The results pertaining to BTSE are reported in table 2.

**Table 2.** The frequency of the selected DMs in BTSE corpus (total words = 108618)

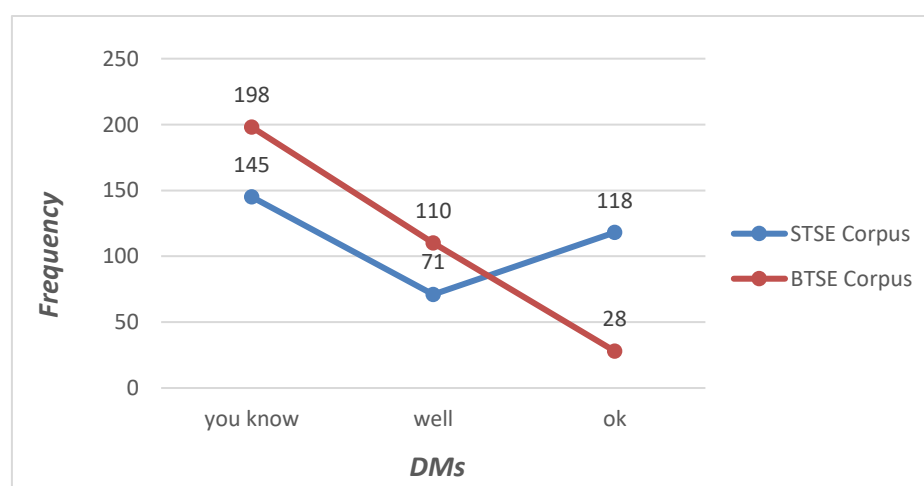
DMs	Frequency	Percentage
<i>you know</i>	198	0.18 %
<i>well</i>	110	0.1 %
<i>okay</i>	28	0.03 %

The results obtained from BTSE data analysis were the standards against which Sudanese use of DMs was compared. The frequency of use of the selected DMs by Sudanese non-native speakers is displayed in table 3 as follows.

**Table 3.** The frequency of the selected DMs in STSE corpus (total words = 108621)

DMs	Frequency	Percentage
<i>you know</i>	145	0.13 %
<i>well</i>	71	0.07 %
<i>okay</i>	118	0.11 %

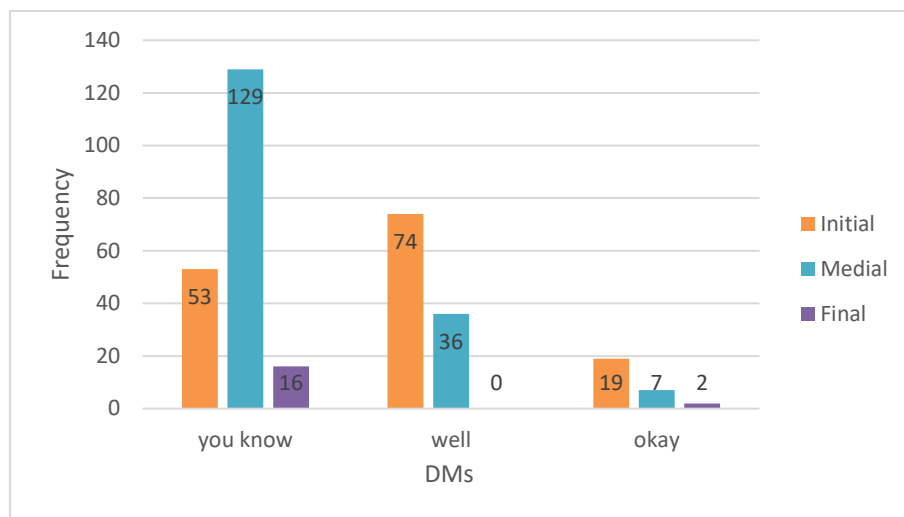
The use of the selected DMs by the speakers of the two corpora is illustrated in figure 1 below:

**Figure 1.** Frequency of occurrence of *you know*, *well*, and *okay* in BTSE and STSE corpora

A quick glance at figure 1 reveals that unexpectedly Sudanese speakers surpassed their British counterparts in using the DM *okay*. Apart from that, the results cope with the

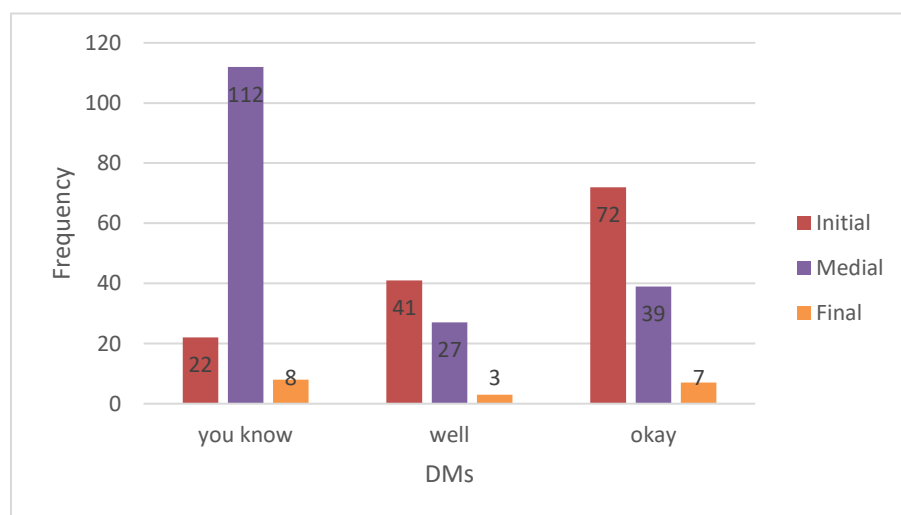
previous studies in assuming native speakers to outperform non-native ones in using DMs.

Regarding the placements of DMs markers, British speakers tended to use the DMs *well* and *okay* more in the initial position of the sentence however they normally use *you know* in the medial position as shown in figure 2 below:



**Figure 2.** In-sentence placement of the DMs *you know*, *well*, and *okay* in BTSE corpus

It is apparent that the DM *okay* was hardly used in the final position of the sentence while the DM *well* was never used as a sentence ending in BTSE corpus. On the other hand, a comparable placement pattern was shown by the Sudanese speakers as illustrated in figure 3.



**Figure 3.** In-sentence placement of the DMs *you know*, *well*, and *okay* in STSE corpus

Although the overall pattern of placing the DMs are identical in the two corpora, the non-native speakers used DMs in the final position less more than the natives did.

## DISCUSSION

The pragmatic nature of the selected elements imposes that their uses as DMs are controlled by the context in which they occur. Accordingly, considerations were to be taken to analyze the cases where the studied elements were used as DMs or mere

semantic units. In this regard, it was noted that you know was mostly used as a DM by both native and non-native speakers with a notable supremacy in favour of native speakers. In other words, Sudanese speakers used the term you know in its literal sense more than their British counterparts. It was used in this sense to ask audiences about their previous knowledge of a topic or to confirm such knowledge. Examples from STSE can include "I don't know what yet, if you know please contact me." (Osman, 2011), "you know what a cart is, right?" (Mahjoub, 2013)). This implies less competence in using this specific DM as the native speakers used it mainly as a pause filler when they ran short of suitable words, for instance, "...he's like many teenagers in London spending lots of time hanging around, you know, and I wanted him to really understand ..." (Gratton, 2012).

The case was different for the DM *well*. Speakers of both groups used it as a non-DM at around 40% of the situations. Bearing in mind the wide range of using the word *well* in spoken English, this finding seems sensible. The use of *well* by the non-native group was proven to be the closest to the native among the three regarding the distinction between DM and non-DM usage. The majority of the cases of non-DMs usage of *well* were as an adjective (*well* + past participle) e.g. "It's a novel story, *well* written and *well* presented." (Dafaa-Alla, 2014), "Who am I as a dropout to judge someone on how *well*-qualified they are." (Barlett, 2016); or as an adverb, for example, "we have to choose a set of customers that this product works *well* for them." (Yousif, 2015), "a very curious thing that very often doesn't match up *well* with reality." (Shrimban, 2017); and as a conjunction e.g. 'it is some sort of judgment as *well* as an attitude that we have towards ourselves self-esteem" (Hussein, 2013).

The third word i.e. *okay*, was mostly used as an adjective by the British speakers. Around 64 % of occurrences of *okay* were used in its semantic sense. Examples from BTSE can be "you're going to sleep and you're *okay* with that" (Owen, 2016) and "they weren't great they were *okay* though I had ..." (Bailey, 2016). On the other hand, more cases of DMs usage of *okay* in the STSE were observed. Sudanese speakers manifested a very high frequency of using this word as a DM; however, this was likely caused by the high frequency of using the word itself. Examples from STSE can include "*Okay*, thank you, hello everybody my name is Hiba Salih, I'm a physician from Sudan" (Salih, 2012) and "just now let's go to the science fiction, *okay*? Let's go to the science fiction, just be open-minded with me. *Okay* now ..." (Hydar, 2015).

The high frequency of using *okay* by the non-native speakers makes the findings of the present research inconsistent with those of previous studies. Sudanese speakers used *okay* 4 times as their British counterparts did. According to the researcher's viewpoint, this is more likely to be caused by first language interference factor. In Arabic, many words that can be translated to *okay* are used as DMs in formal and informal settings. Not only that but also the word *okay* itself is used as a borrowed term by Arabic speakers of different educational and social background in their everyday speech. This would make the DM *okay* the closest candidate for Arabic speakers to use as a frame marker, a pause filler, or a feedback stimulator. This justification, nevertheless, should not be taken as an indicator to the properness of this overuse. Alternatively, it should be an implication to assist DMs instruction at all levels. Effort should be directed to promote the use of various

DMs in different situations and avoid repetition and overuse of a specific DM. Taking the aforementioned discussion into account, it can be argued that as it was previously suggested by the researchers, native speakers surpassed non-native ones in the quantity and quality of DM usage. The high frequency of using *okay* by the non-native subjects of the current study is to be interpreted as an overuse rather than supremacy in using this marker.

Although non-native speakers used DMs less than their native counterparts, they showed a similar pattern of use regarding the DMs positions within sentences. Members of both groups rarely used DMs in the final position of sentences. The nature of TED talk may explain this as they are delivered using a genre that is more oratory than conversational. Accordingly, presenters rarely stop for stimulating feedback or checking for agreement, or follow-up, the cases which require using final DMs.

Respecting the initial and medial positions, it was found that the speakers tended to use each DM more in specific positions. Native speakers used the DM *you know* in the medial position in 65 % of the occurrences while non-native speakers used it more in this position i.e. in 79 % of the occurrences. It can be inferred that TED speakers use *you know* to fill pauses rather than to introduce a new topic or to check that their audiences share the same knowledge. Non-native speaker subjects of the current research exhibited similar trends indicating that fewer difficulties were encountered by them regarding the use of this DM. The same was true for the final position which was found the least to host the DM *you know* in both corpora i.e. 8% in BTSE versus 6% in STSE.

The third discourse marker, *well*, was more likely to occur in the initial position of sentences. Both native and non-native speakers used it more in that position i.e. 65 % of the occurrences of *well* in the BTSE and 58 % in the STSE were in the initial position. The interpretation of this is that both native and non-native speakers used this DM as a frame marker and a new topic indicator. The DM *well* was also used in the medial position by both groups implying that TED speakers also use this DM when they run out of words or want to sustain the audience attention. On the other hand, while native speakers did not use *well* in the final position, few cases in STSE occurred in the sentence tail. This suggests the sole improper placement of DMs by non-native speakers.

Results pertaining to DMs in-sentence positioning reveal that Sudanese non-native speakers encountered no difficulties in placing DMs. This finding had been suggested by many previous studies which found no variation in DMs placement by native and non-native speakers similar to the variation in their frequency of occurrence. It can be implied from this that DM instruction should focus on frequency and diversity of DMs more than on their placement, especially in formal speech and presentation. Considering the high potentials that are presupposed to be possessed by TED speakers in general, it is expected that more problems would reveal if a similar study was conducted on average Sudanese EFL learners.

## CONCLUSION AND RECOMMENDATIONS

The present study examined the use of three DMs by Sudanese non-native speakers of English compared to British spoken English. A corpus-based method was utilized and a

multi-method analysis approach was followed to analyse the two developed corpora. The results reveal that, in general, Sudanese non-native speakers used the three selected DMs less than British native speakers. Notwithstanding, an overused pattern was exhibited by the Sudanese learners regarding the DM *okay*. It was also revealed that the DMs *well* and *okay* were used mostly in the initial positions as a frame marker while the phrase *you know* was generally used as a DM in the medial positions by native and non-native speakers. Moreover, no significant differences were found regarding the distribution of the DMs within sentences.

The obtained results can be used to provide insights regarding DM instruction. Focusing on the diversity of DMs and their functions and distribution is a recommendation to consider. For further research, investigating the use of more DMs and their functions can provide more beneficial indications. As a considerable part of this research was conducted manually, a pertinent suggestion is that future research should also consider adopting more developed software that can handle automatic recognition of sentence boundaries to facilitate the process of detecting DMs positions and functions. The STSE corpus also should be developed to include more samples of Sudanese spoken English extracted from different genres. The aim of these suggestions is to validate the findings of this study and hence to present more insightful implication for DM instruction.

## REFERENCES

- Aijmer, K. (2002). *English Discourse Particles: Evidence from corpus*. Amsterdam: John Benjamins.
- Anthony, L. (2018). *AntConc (Version 3.5.7)*. Computer Software. Tokyo, Japan: Waseda University. Retrieved from <http://www.laurenceanthony.net/software>
- Asik, A., & Cephe, P. T. (2013). Discourse Markers and Spoken English: Non-native Use in the Turkish EFL Setting. *English Language Teaching*, 6(12), 144-155.
- Aswini, P., & Srinivasan, R. (2016). Corpus-based studies: Some perspectives. *International Journal of Applied Engineering Research*, 11(4), 2340-2342.
- Bailey, C. (2016, September 6). A more human approach to productivity. *Tedx Liverpool*. Retrieved from <https://www.youtube.com/watch?v=REFh5TMOoS0&t=310s>
- Barlett, S. (2016, July 14). Entrepreneurs should follow their instincts. *Tedx London*. Retrieved from [https://www.youtube.com/watch?v=Dri\\_W3dvd6Y&t=309s](https://www.youtube.com/watch?v=Dri_W3dvd6Y&t=309s)
- Brinton, L. (1996). *Pragmatic markers in English: Grammaticalization and discourse functions*. Berlin: Mouton de Gruyter.
- Connor, U., Nagelhout, E., & Rozycki, W. (Eds.). (2008). *Contrastive Rhetoric: Reaching to Intercultural Rhetoric*. Amsterdam: John Benjamins.
- Dafaa-Alla, A. (2014, April 27). How TEDx is bringing change to Sudan amidst political crackdown. *TEDxMidAtlantic*. Retrieved from <https://www.youtube.com/watch?v=XLOkrJx9vr4&t=56s>
- Farser, B. (2006). Towards a theory of discourse markers. *Approaches to discourse particles*, 1, 189-204.
- Fraser, B. (1996). Pragmatic markers. *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)*, 6(2), 167-190.
- Fung, L., & Carter, R. (2007). Discourse markers and spoken English: Native and learner Use in the pedagogic setting. *Applied Linguistics*, 28(3), 410-430.
- Gee, J. P. (2018). *Introducing Discourse Analysis*. New York: Routledge.

- Gratton, L. (2012, Jun 7). How to be ready for your future, now. TEDx London Business School. Retrieved from <https://www.youtube.com/watch?v=VbZ3eKbFi3g&t=378s>
- Guo, F. (2015). A review of discourse Markers from the functional perspective. *Journal of Arts and Humanities*, 69-75.
- Halliday, M., & Hasan, R. (1985). *Language, context and text: aspects of language in a social-semiotic perspective*. Oxford: Oxford University Press.
- Hussein, A. A. (2013, December 31). Tackling low self-esteem. TEDx University of Khartoum. Retrieved from <https://www.youtube.com/watch?v=JL4lOc5vjV4&t=111s>
- Hydar, A. (2015, January 16). From human to human. TEDxGeziraUniversity. Retrieved from <https://www.youtube.com/watch?v=NgbtRG-RWCc&t=388s>
- Jabeen, F., Rai, A., & Arif, S. (2011). A corpus-based study of discourse markers in British and Pakistani speech. *International Journal of Language Studies*, 69-86.
- Jucker, A. H. (1993). The discourse marker *well*: A relevance-theoretical account. *Journal of Pragmatics*, 19(5), 435-452.
- Ludewig, J. (2017). TED Talks as an emergent genre. *CLCWeb: Comparative Literature and Culture*, 19(1). doi:10.7771/1481-4374.2946
- Mahjoub, O. H. (2013, January 2). Unheard Voices. TEDxYouth@Khartoum. Retrieved from <https://www.youtube.com/watch?v=CcZI7uinekk&t=83s>
- McEnery, T., & Wilson, A. (1996). *Corpus Linguistics: An introduction*. Edinburgh: Edinburgh University Press.
- Meyer. (2002). *English corpus linguistics: An introduction*. Cambridge: Cambridge University Press.
- Osman, W. I. (2011, December 15). Stopping Wars. TEDxYouth@Khartoum. Retrieved from <https://www.youtube.com/watch?v=c2tDEZjPJYY>
- Owen, J. (2016, March 28). Good Kids, Bad Kids, & Weird Kids. TEDx Birmingham. Retrieved from <https://www.youtube.com/watch?v=mcybIn2V5QY&t=373s>
- Ozer, H., & Okan, Z. (2018). Discourse markers in EFL classroom: A corpus-driven research. *Journal of Language and Linguistics Studies*, 14(1), 50-66.
- Polat, B. (2011). Investigating acquisition of discourse markers through a developmental learner corpus. *Journal of Pragmatics*, 43, 3745-3756.
- Salih, H. (2012, May 21). Untraditionally traditional. TEDx College Hill. Retrieved from <https://www.youtube.com/watch?v=mXv1IIYSesA&t=174s>
- Scotto di Carlo, G. (2014). New trends in knowledge dissemination: TED Talks. *Acta Scientiarum. Language and Culture*, 36(2), 121-130.
- Shrimban, M. (2017, Jun 27). Is confidence ever reasonable? Tedx London. Retrieved from <https://www.youtube.com/watch?v=WAQ44MLby5I&t=4s>
- Timmis, I. (2015). *Corpus linguistics for ELT*. New York: Routledge.
- Yousif, M. (2015, 12 14). Gut Feeling. TEDx UMST. Retrieved from [https://www.youtube.com/watch?v=Nu1r8\\_tiTiM&t=42s](https://www.youtube.com/watch?v=Nu1r8_tiTiM&t=42s)