

Discourse markers *so* and *well* in Zimbabwean English: A corpus-based comparative analysis

Faith Chiedza Chapwanya¹  | Joanine Hester Nel² 

¹Department of Afrikaans, University of Pretoria, South Africa

²Department of Afrikaans, University of Pretoria, South Africa

Correspondence

Faith Chiedza Chapwanya, Department of Afrikaans, University of Pretoria, South Africa.
Email: u12167160@tuks.co.za

Funding information

National Research Foundation South Africa; University of Pretoria

Abstract

This analysis of discourse markers *so* and *well* in Zimbabwean English (ZE) and British English was carried out to determine possible statistically significant variations in their occurrence and function frequencies in spoken and written registers, and in different genres to ascertain if they are used in the same manner in both languages and in different language use contexts. The ZE corpus and the International Corpus of English-Great Britain (ICE-GB) were compared to reveal statistically significant variations in some registers and genres. Regarding *so*, the marking implied result and the sequential *so* functions were more frequent in the ICE-GB compared to the ZE corpus. *Well* occurred more in the ICE-GB compared to the ZE corpus. Searching for the right phrase, rephrasing or correcting, move to main story, indirect answer, contributing an opinion, direct answer, continuing an opinion, and evaluating a previous statement occurred more in the ICE-GB compared to the ZE corpus.

1 | INTRODUCTION

The objective of this article is to investigate the use of the discourse markers (DMs) *so* and *well* in the ZE corpus and the ICE-GB¹ to determine whether there are statistically significant variations between the two corpora and to ascertain if they are used in the same manner in both languages and in different language use contexts. Comparisons were made between ZE and British English (BrE) because English was introduced in Zimbabwe through British colonialism and has subsequently played a vital role in language use in Zimbabwe. Although there may be influences from other English varieties on ZE, BrE mostly provides for the norms, being used in teaching and learning as well as in

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *World Englishes* published by John Wiley & Sons Ltd.

print media, because newspapers and publishing houses follow the BrE conventions (Kadenge, 2009). The discourse marker (DM) *so* was chosen because it was the most frequent DM in the ZE corpus whilst *well* was chosen because it is among the most frequently studied DMs in second language (L2) Englishes. Previous research has shown that there are variations in the frequencies and uses of *so* and *well* in L2 varieties of English (De Klerk, 2005; Müller, 2005; Lam, 2009a, 2009b; Unuabonah, 2019; Algouzi, 2021). Examining the variation of the use of DMs in ZE will enable comparisons with results from other L2 varieties of English in future research.

Following a pluricentric approach to world Englishes, the term 'Zimbabwean English' is used to refer to the English that is spoken by Shona mother tongue speakers in Zimbabwe (Kadenge, 2009; Marungudzi, 2016). English is one of the 16 official languages of Zimbabwe and is used as a language of teaching and learning, and in most government departments (Kadenge, 2009). The other official languages, namely, Barwe, Chewa, Kalanga, Khoisan, Nambya, Ndau, Ndebele, Shangani, Shona, Sotho, Tonga, Tswana, Venda, Xhosa, and Zimbabwean Sign Language are rarely used in high-status and formal domains of society such as education, industry, government, and parliamentary business (Kadenge, 2010). Regarding L2 English varieties such as ZE, Kachru (1992) shows that the varieties emerging in different countries are a sign of linguistic innovation, accommodation, and adaptation, language-change features which form part of world Englishes and occur as a result of language contact. Kachru (1985, 1988) visualizes the spread of English through the Three Circles Model namely the Inner Circle, the Outer Circle, and the Expanding Circle. Countries such as Canada, Australia, the United Kingdom, the United States, and New Zealand where English is the first language (L1) belong to the Inner Circle. Countries where English is used as a second or official language and former British colonies such as Kenya, Zimbabwe, India, and Nigeria are found in the Outer Circle. Countries like Israel and Japan, where English is used as a foreign language, are classified under the Expanding Circle.

In section 2, a brief overview of the definitions, terminology, and theoretical framework of DMs is provided. Section 2.1 discusses the use of DMs in L1 varieties of English while DMs in L2 Englishes are discussed in section 2.2. The functions of *so* and *well* from previous research are reported in section 2.3 while section 3 focuses on the research methodology utilized in this study. Results from this study are reported in section 4. Finally, section 5 discusses the results and the paper concludes in section 6.

2 | DISCOURSE MARKERS ACROSS ENGLISHES

Broadly, DMs are defined as singular words or multi-word phrases which aid in managing the flow and structure of discourses that is either spoken or written. DMs have been approached and analyzed from different perspectives across different varieties of English resulting in a diverse use of terminology, definitions, and theoretical frameworks associated with the creation and maintenance of cohesion and coherence in discourses. Regarding the varying terminology in the field, terms like 'discourse markers' (Schiffrin, 1987; Schourup, 1999; Blakemore, 2002; Müller, 2005), 'discourse connectives' (Blakemore, 2002), 'discourse particles' (Aijmer, 2002; Lam, 2009a, 2009b), and 'pragmatic markers' (Brinton, 2010; Oladipupo & Unuabonah, 2020) are used as synonyms for the same concept; while some scholars make a distinction between DMs and discourse particles by including DMs as part of a broad category called discourse particles. For instance, Aijmer (2002, p. 2) noted that 'discourse particles seem to be dispensable elements functioning as signposts in the communication facilitating the hearer's interpretation of the utterance on the basis of various contextual clues.' This definition fits in with other definitions provided for DMs. The difference is that Aijmer (2002) classifies DMs into a broad category of discourse particles consisting of interactional signals and DMs. This study employs the term 'discourse markers' because it shows that linguistic items such as *so* and *well* show the relationship between discourse segments.

As outlined above, there seems to be no consensus on the definition of DMs as evidenced by the several definitions provided by researchers. For instance, Schiffrin (1987, p. 31) regards DMs as 'sequentially dependent elements which bracket units of talk' whilst Aijmer's (2002) definition focuses on the fact that DMs are dispensable elements whose function is to enable the hearer to interpret utterances based on the context. In this article, DMs are conceptualized as free morphemes, made up of items from different grammatical categories, which signal a specific message, connect

discourse units, and have a procedural core meaning (Fraser, 1993; Schourup, 1999). Different theoretical frameworks have been proposed to analyze and account for DMs such as relevance theory, which analyzes the contribution made by DMs in text processing and comprehension (Blakemore, 2002) and coherence theory, which concentrates on indexical functions of DMs (Schiffrin, 1987). Given the indexical nature of DMs, as elements whose reference can shift from context to context but also link circumlocuted references or phrases, in this study, coherence theory is used as a basis for analysis because DMs are examined in the contexts in which they appear to check their relationship with other words. The discussion above has highlighted that the study of DMs is marked by diversity in definitions, terminology, and theoretical frameworks, leading to variations in the way DMs are conceptualized in research. The next section focuses on DMs in L1 varieties of English.

2.1 | Discourse markers in L1 varieties of English

DMs have been studied extensively in L1 varieties of English (Schiffrin, 1987; Fraser, 1993, 2009; Schourup, 1999; Blakemore, 2002). Some of the studies are descriptive, like Schiffrin's (1987) analysis from a sociolinguistic perspective using coherence theory. Another descriptive account of DMs was given by Blakemore (2002) who outlines that DMs ought to be examined according to their contribution to the cognitive processes that are fundamental in language processing. Using a pragmatic approach, on the other hand, Fraser (1993, 2009) discusses the characteristics and functions of DMs and states that DMs have procedural meanings aiding in discourse coherence. The three classes of DMs suggested by Fraser include elaborative DMs, which show how information in the first sentence is elaborated in the second sentence, contrastive DMs, which show direct or indirect contrast between the first sentence and the second sentence, and inferential DMs, which show that the first sentence provides a basis for inferring in the second sentence.

Other studies on L1 varieties are comparative, like Aijmer's (2013) study that utilized a variational pragmatics approach to analyze the DMs *well*, *in fact*, and *actually* in different contexts in BrE. The study also considered New Zealand-, American-, and Australian English. Regarding *well*, the author reported that contextual factors like the agenda, audience, turn-taking organization, medium, social roles, and discourse organization play a role in the use of *well*. Aijmer (2013) noted that there were no differences in the use of *in fact* in BrE and American English. In addition, *in fact* and *actually* are reported to occur in similar text types but with differing frequencies. The next section discusses the use of DMs in L2 varieties of English.

2.2 | Discourse markers in L2 varieties of English

Previous research on DMs in L2 varieties of English shows variability in their use (De Klerk, 2005; Müller, 2005; Lam, 2009a, 2009b; Unuabonah, 2019; Mohr, 2021). For example, De Klerk (2005) explored the procedural meaning of *well* in Xhosa English, an L2 variety of English and reported the diversified and complex use of *well* by Xhosa English speakers. In addition, De Klerk noted that *well* was less frequent in Xhosa English compared to specific L1 English corpora namely the New Zealand English corpus and the London Lund corpus. De Klerk (2005) attributes the limited frequency of DMs in the Xhosa English corpus to the context in which Xhosa English is acquired. Müller (2005) studied the use of DMs by German L2 English speakers and American L1 English speakers in a movie-telling experiment and noted that the former has more functions of DMs than the latter. Regarding the frequency of functions of DMs, Müller (2005) noted that the textual functions of *so* were used more by female American participants than males, while the German data showed that the use of interactional functions of *so* was higher in female participants than males. In Müller's study, the youngest speakers were reported to utilize some functions of DMs more than other age groups.

The focal point of work by Lam (2009b) outlines how text type influences the use of DMs. The author studied *so* in different contexts in Hong Kong English to determine its different functions. Results reveal that the frequency and functions of DMs in Hong Kong English vary according to text type. Lam (2009b) mentioned that different contexts

in which DMs are used, should be considered when reporting on their functions. Unuabonah (2019) investigated stylistic variability of DMs in Nigerian English, an L2 English variety. Results from the study showed that DMs in Nigerian English and BrE varied significantly in terms of frequency and style. Unuabonah (2019) attributed the variations to the fact that Nigerian English is an L2 and BrE is an L1. Similarities were also noted in the frequency of elaborative DMs in both varieties of English, which were attributed to the influence of BrE on Nigerian English.

Other studies on L2 varieties of English include one by Diskin-Holdaway (2021), who analyzed the use of *you know* and *like* in Polish and Chinese L2 speakers of English in Ireland and Australia. Results from the study show that there are no significant differences in the use of *you know* between L1 and L2 English speakers. In addition, *you know* was used more by Polish L1 speakers compared to other groups. The author noted that L2 English speakers who migrated to Ireland did not achieve similar levels as L1 English speakers in their use of clause-final *like*. The discussion above has shown that most scholars of L2 Englishes employ a comparative analysis between L1 and L2 Englishes. As can be gleaned from the aforementioned studies, similarities and differences exist in the use of DMs in L1 and L2 English varieties. The following section will outline the functions of the DMs *so* and *well*.

2.3 | Functions of *so* and *well*

A range of functions are fulfilled by the DM *so* and researchers focus on a varied list of functions. For instance, the function of *so* as a marker of result is well documented (Müller, 2005; Lam, 2009b; Buysse, 2012; Vickov & Jakupčević, 2017; Algouzi, 2021), as exemplified below.

(1) <le> yeah and they had a shortage of actors. So they had some actors playing several parts. (Buysse, 2012, p. 1769)

In (1), *so* shows that the fact that some actors played various parts is a consequence of the scarcity of actors. Another function that has received considerable attention is the main idea marker (Schiffrin, 1987; Müller, 2005; Algouzi, 2021). In this case, *so* is used to return to the main idea after a digression as shown in example (2) where speaker A digresses from the question by explaining the reasons for liking the film and then returning to the main idea starting with *so*.

(2) (B): Did you like the film?

(A): Yeah I did like it because it was very funny an' entertaining and um in general I like Chaplain. So I enjoyed the movie. (Müller, 2005, p. 72)

The sequential *so* is another function reported in literature (Müller, 2005; Buysse, 2012; Algouzi, 2021). In example (3), *so* marks a coherent transition from one event to another.

(3) <le> . . . erm she's not happy with it because he's painted her as she is she obviously wants to look a bit more glamorous . . . er so he repaints the picture and she hangs it up very impressed er to show all her friends . . . who per who perhaps aren't quite as impressed as her. (Buysse, 2012, p. 1773)

Regarding the summary or rewording function, example (4) shows *so* being used to rephrase what was said in the first part of the sentence.

(4) (T4): . . . and think of your own examples, so don't use examples from the book. (Vickov & Jakupčević, 2017, p. 664)

Another function of *so* is marker of transition relevance place as shown in example (5), where *so* marks a changeover of speakers from speaker B to speaker A.

- (5) (B): When they're talking, they—
 (A): Right?
 (B): They were moving their mouths.
 (A): OK.
 (B): uh,
 (A): So everything was over-emphasized, what you were saying a little earlier there. (Müller, 2005, p. 86)

Well as a DM has several functions (Schiffrin, 1987; De Klerk, 2005; Müller, 2005; Lam, 2009a). One of the functions is searching for the right phrase as shown in the sentence *and then he just ends up on the streets of well I guess New York* (Müller, 2005, p. 110). *Well* is also used to rephrase or correct a statement as is outlined in (6).

- (6) Charlie Chaplin um ... decides to shuffle, well is chosen to shuffle. (Müller, 2005, p. 112)

The move to main story function shows change in the topic under discussion or a move back to the previous topic as in (7) where there is a move from instructions to the start of narratives.

- (7) (B): Well, when you left they went into America.
 (A): OK, um, well, the film starts on a ship. (Müller, 2005, p. 116)

Müller (2005, pp. 108–135) also highlighted other textual functions of *well* such as the quotative *well*, introducing the next scene, and the conclusive *well*, and interactional functions namely indirect answer, direct answer, response to self-raised expectations, contributing an opinion, continuing an opinion or answer, and evaluating a previous statement. For the purpose of this study, the analyses of the functions of *so* and *well* in the ZE corpus and the ICE-GB were based on the functions outlined by Müller (2005). In the next section, the research methodology employed in this study is explained.

3 | METHODOLOGY

In this study, the data stemmed from two sources, namely the ZE corpus and the ICE-GB. The ZE corpus consists of 206 007 words compiled by the researchers as well as part of Marungudzi's (2016) corpus consisting of 72 samples totalling 150 000 words, which was added.² For this study and for future research, the researchers collected their own data to add to Marungudzi's (2016) corpus to make the ZE corpus bigger and to have a wider representation of texts. The study outlined here, forms part of a larger study on the use of articles, modal verbs, and selected discourse markers in Zimbabwean English. In total, the ZE corpus consists of about 356 007 words. Regarding part of the ZE corpus compiled by the researchers, ethical clearance (number: GW20181012HS) was applied for and obtained from the faculty of Humanities at the University of Pretoria. The procedures followed were in accordance with the Declaration of Helsinki 1975, as revised in 2008. Each participant from whom spoken and written language data was elicited signed an informed consent form. To preserve anonymity and confidentiality throughout data collection and subsequent reporting, pseudonyms were used to identify participants.

Data was gathered in different urban and rural areas in Zimbabwe, namely Harare, Masvingo, Gweru, and Mutare. Private semi-scripted dialogues and private dialogues were recorded from participants whose socio-cultural and linguistic backgrounds varied. Participants were Shona L1 speakers who used English as an L2 and were aged at least 18 years. Data for private dialogues stem from 10 pairs of participants whilst five participants provided data for business letters. Samples for private semi-scripted dialogues came from 45 participants. Online newspaper articles (editorials and newspaper reportage) were also added. A mixed method sampling technique encompassing three non-probability sampling methods which included snowball, convenience, and purposive sampling, was used (Leedy & Ormrod, 2021).

TABLE 1 Word count for text categories in ZE corpus and ICE-GB

Register	Text type	Words in ZE corpus	Words in ICE-GB
Spoken	Private dialogues	131 961	185 208
	Public scripted monologues	4 193	43 061
	Public dialogues	98 760	171 062
Written	Creative writing	17 890	42 646
	Popular writing	2 759	21 199
	Academic writing: Examination	2 077	21 225
	Editorials	26 342	20 719
	Newspaper reportage	48 939	41 539
	Social letters	5 008	31 085
	Business letters	18 078	30 491
Total		356 007	608 235

The international corpus of English conventions of corpus design and annotation (at textual mark-up level) were followed (Nelson, 2002a, 2002b) as the same conventions were used in the ICE-GB, enabling comparison between the corpora.

The second corpus employed in this study is part of the ICE-GB Release 2. In total, the ICE-GB has 1 million words but, 608 235 words were used because this study only used genres which occurred in both the ZE corpus and the ICE-GB. Regarding comparability, the ZE corpus consists of samples from the 1990s (Marungudzi, 2016), which is the same timeframe as the compilation of the ICE-GB. For instance, the spoken texts include public dialogues such as discussions, interviews, and public scripted monologues such as speeches, radio and television news reports that were available from archives at radio and television stations. The written samples of the ZE corpus include business letters, editorials, and private business letters written in the late 1990s. The word count for each text type in the ZE corpus and the ICE-GB is presented in Table 1.

A corpus-based approach was used in this study (Biber et al., 1998). It incorporated quantitative data analysis in terms of frequency counts, after which a qualitative analysis of the functions of *so* and *well* was done. Sketch Engine³ was used to analyze the ZE corpus and the ICE-GB and concordances were generated in order to examine the key words in context. The task of assigning specific functions to DMs is not clear-cut due to the multi-functional nature of DMs (Lam, 2009b). The different functions of *so* and *well* were counted by checking the concordance lines where the DMs occurred in Sketch Engine. A qualitative analysis in terms of outlining the meanings of the different functions as semantic interpretations helped with assigning primary functions to each occurrence. Using the Sketch Engine Manual annotation (Skema) tool, concordance lines of *so* and *well* were searched and annotated for their primary functions.

The ICE-GB is annotated for DMs and in the ZE corpus, instances of *so* and *well* that did not serve the DM functions, were excluded using the Skema tool by manually checking and annotating the corpus. The examples given below illustrate some of the uses of *so* that were not discussed. In (8), *so* functions as a conjunction whilst in (9), *so* functions as an adverb.

- (8) We produce results so that when they are ready to distribute funds for the next round, Zimbabwe is in a favourable position. (ZE: Newspaper reportage 1 – from current study data)
- (9) And her kids were so happy. (ZE: Private semi-scripted dialogue 43 – from current study data)

Examples (10) and (11) were excluded from the discussion because *well* did not serve as a DM in the sentences but rather as adverbs.

TABLE 2 Frequency of *so* across registers (normalized per 10 000 words)

Register	ZE corpus		ICE-GB		Log likelihood
	Absolute frequency	Normalized frequency	Absolute frequency	Normalized frequency	
Spoken	1 484	63.2	1 556	39	+174.99****
Written	104	8.6	196	9.4	-0.54
Total	1 588	44.6	1 752	29	+156.95****

(10) We would like to thank them for a job well done. (ZE corpus: Editorials 5 – from current study data)

(11) They performed very well. (ZE corpus: Private semi-scripted dialogue 26 – from current study data)

The occurrence and function frequencies of *so* and *well* were normalized per 10 000 words. To determine whether the observed variations were statistically significant, a log likelihood calculation was performed (Rayson et al., 2004) using an online calculator created by Rayson (2023).⁴ One asterisk (*) indicates that the log likelihood was > 3.84 where $p < 0.05$. Two asterisks (**) show that log likelihood was > 6.63 at level $p < 0.01$, whilst three asterisks (***) indicate that log likelihood was > 10.83 where $p < 0.001$. Four asterisks (****) symbolize that log likelihood was > 15.13 where $p < 0.0001$. Regarding the log likelihood values, if the higher value occurred in the ZE corpus, this direction is indicated with a plus sign before the log likelihood value and if the higher value occurred in the ICE-GB, the direction is indicated with a minus sign before the log likelihood value. The next section focuses on the results from the study.

4 | RESULTS

In this section, occurrence frequencies of *so* and *well* are reported per register, and then per genre. Afterwards, the function frequencies of *so* and *well* are reported and a qualitative functional analysis of the two DMs is provided.

4.1 | Occurrence frequencies of *so* across registers

The frequency of *so* in the spoken and written registers and log likelihood values are given in Table 2.

In Table 2, there are no variations in the frequency of *so* (per 10 000 words) in the written register because in ZE, *so* occurs 8.6 times whilst in BrE *so* occurs 9.4 times. The picture is different in the spoken register because there are variations in the normalized frequency of *so*, which is higher in ZE (63.2 per 10 000 words), compared to 39 occurrences per 10 000 words in BrE.

4.2 | Occurrence frequencies of *well* across registers

A good overview of the frequency of *well* in the spoken and written registers can be gleaned from Table 3.

Results above show that there are no variations in the normalized frequency in the written register in ZE and BrE. There are significant variations in the spoken register, which shows that BrE has more instances of *well* per 10 000 words (47.3 compared to 8.8 in ZE).

TABLE 3 Frequency of *well* across registers (normalized per 10 000 words)

Register	ZE corpus		ICE-GB		Log likelihood
	Absolute frequency	Normalized frequency	Absolute frequency	Normalized frequency	
Spoken	206	8.8	1 888	47.3	-809.71****
Written	25	2.1	61	2.9	-2.23
Total	231	6.5	1 949	32	-782.81****

TABLE 4 Frequency of *so* across genres (normalized per 10 000 words)

Genre	ZE corpus		ICE-GB		Log likelihood
	Absolute frequency	Normalized frequency	Absolute frequency	Normalized frequency	
Private dialogues	1136	86.1	902	48.7	+164.51****
Public scripted monologues	14	33.4	39	9.1	+13.87***
Public dialogues	334	33.8	615	36	-0.81
Creative writing	22	12.3	23	5.4	+7.39**
Popular writing	1	3.6	15	7.1	-0.51
Academic writing: Examination	0	0	26	12.2	-4.84*
Editorials	8	3	6	2.9	+0.01
Newspaper reportage	23	4.7	12	2.9	+1.95
Social letters	29	57.9	97	31.2	+7.59**
Business letters	21	11.6	17	5.6	+5.08*
Total	1 588	44.6	1 752	29	+156.95****

4.3 | Occurrence frequencies of *so* across genres

In this section, the frequencies of *so* in different genres are reported for the ZE corpus and the ICE-GB. Table 4 shows the frequencies of *so* in different genres.

If the 10 genres are considered, significant differences can be observed in private dialogues, public scripted monologues, creative writing, and social letters in the two corpora. The highest variation is in private dialogues where *so* occurs more frequently in the ZE corpus (86.1 times per 10 000 words) compared to the ICE-GB (48.7 times). There are no variations in the frequencies in public dialogues, popular writing, newspaper reportage, and editorials.

4.4 | Occurrence frequencies of *well* across genres

The frequencies of *well* across genres are indicated in Table 5 below.

The normalized frequency of the DM *well* shows some variations between ZE and BrE. This is evident in private dialogues and public dialogues where *well* is more prevalent in BrE than in ZE (69.5 and 32.6 occurrences per 10 000 words compared to 7.4 and 10.5 respectively). The newspaper reportage genre shows significant differences, with ZE recording two occurrences per 10 000 words compared to 0.5 occurrences per 10 000 words for the ICE-GB. There are no variations in public scripted monologues, creative writing, business letters, and social letters. Another observation

TABLE 5 Frequency of *well* across genres (normalized per 10 000 words)

Genre	ZE corpus		ICE-GB		Log likelihood
	Absolute frequency	Normalized frequency	Absolute frequency	Normalized frequency	
Private dialogues	98	7.4	1287	69.5	-848.58****
Public scripted monologues	4	9.5	44	10.2	-0.02
Public dialogues	104	10.5	557	32.6	-141.37****
Creative writing	7	3.9	14	3.3	+0.14
Popular writing	0	0	0	0	0
Academic writing: Examination	0	0	0	0	0
Editorials	0	0	0	0	0
Newspaper reportage	10	2	2	0.5	+4.59*
Social letters	6	12	45	14.5	-0.20
Business letters	2	1.1	0	0	+2.95
Total	231	6.5	1949	32	-782.81****

is that both corpora do not have any instances of the DM *well* in popular writing, academic writing, and editorials. The frequencies of different functions of *so* are reported next.

4.5 | Function frequency of *so*

Although the task of assigning specific functions to DMs is challenging because of the multifunctional nature of the DMs under investigation, searching for *so* and *well* as key words in context and examining the contexts in which they appear, and using the prosodic clues helps with determining the primary functions of DMs (Holmes, 1984). That is why some prosodic features such as filled and unfilled pauses and non-speech sounds such as laughing and coughing were transcribed in this study. The transcription conventions are given in Appendix 1. Although the researchers did not have access to the ICE-GB recordings, the ICE-GB followed the same ICE corpus design and annotation as the ZE corpus. This enabled analysis of the same prosodic cues that were transcribed in the ZE corpus. In the ZE corpus, there were 30 unclassified instances, and in the ICE-GB there were 34 unclassified instances where *so* did not fit into the functional categories. Some instances of *so* in spoken texts could not be classified because they occurred in contexts where speech was inaudible or interrupted. Regarding written texts, some words were deleted or incomplete. Repetitions of *so* were also put in the unclassified instance group if they did not have a different function from that of the adjoining *so*. Examples of such instances are shown in examples (12) to (14) which stem from the current study.

(12) <\$T>: So so I never, in my life I, that was my first time. (ZE corpus: Private semi-scripted dialogue 20)

(13) <#> So I, so I asked the university to to let me do part-time work. (ZE corpus: Private semi-scripted dialogue 38)

(14) But they do don't they don't due to lack of opportunities. So <&>incomplete sentence</&>. (ZE corpus: Private semi-scripted dialogue 7)

In Table 6, the marking result or consequence function of *so* occurred more in ZE, with a frequency of 26.2 per 10 000 words compared to BrE (13 per 10 000 words). The second most significant difference is observed in the sequential *so* function where the ICE-GB had more occurrences of *so* compared to the ZE corpus. Interestingly, ZE speakers use so

TABLE 6 Frequencies of functions of *so* (normalized per 10 000 words)

Function	ZE corpus		ICE-GB		Log likelihood
	Absolute frequency	Normalized frequency	Absolute frequency	Normalized frequency	
Marking result or consequence	934	26.2	793	13	+209.45****
Main idea unit marker	162	4.6	123	2	+46.44****
Summarizing/ rewording/ giving an example	132	3.7	86	1.4	+49.87****
Sequential <i>so</i>	20	0.6	272	4.5	-144.68****
Boundary marker	0	0	0	0	0
Speech act marker – question or request	189	5.3	214	3.5	+16.73****
Speech act marker – opinion	52	1.5	91	1.5	-0.02
Marking implied result	12	0.3	45	0.7	-6.71**
Marker of a transition relevance place	57	1.6	94	1.5	+0.04
Unclassified - Unfinished sentences, repetitions	30	0.8	34	0.6	+2.64
Total	1 588	44.6	1 752	29	+156.95****

to mark result or consequence, to summarize, as a main idea unit marker and as a speech act marker for a question or a request more than BrE speakers.

4.6 | Function frequency of *well*

Table 7 shows the different functions of *well* and how they are attested in the two corpora. The unclassified category included the use of *well* in repetitions, incomplete sentences, interrupted speech, and inaudible sentences as exemplified in (15) and (16) from the current study.

(15) <\$Z> <#> Well well well, it was when I talked to my district nursing officer. (ZE corpus: Private semi-scripted dialogue 26)

(16) <\$E> <#>How are you?

<\$A>: Very well <,> thank you.

<\$E>: Right eeh well when <&>incomplete sentence</&>

<\$A>: Yes go ahead Sir <#> <„> <#> Seems we have lost that call. (ZE corpus: Public dialogue 47)

Significant differences in the frequencies of functions of *well* can be seen in Table 7. BrE recorded more frequencies of *well*, with searching for the right phrase recording the highest frequency of 10.6 per 10 000 words compared to 1.6 for ZE. The option of using a direct answer is the second highest statistically significant function of *well* in BrE with a normalized frequency of nine whilst ZE has 1.6. Other functions that show significant differences between the two corpora are options such as (i) rephrasing or correcting, (ii) using the option of a move to the main story, (iii) choosing an indirect answer, and (iv) continuing an opinion or answer. The quotative *well* function was used more in the ZE corpus compared to the ICE-GB. The following section evaluates the different functions of *so*.

TABLE 7 Frequencies of functions of *well* (normalized per 10 000 words)

Function	ZE corpus		ICE-GB		Log likelihood
	Absolute frequency	Normalized frequency	Absolute frequency	Normalized frequency	
Searching for the right phrase	56	1.6	644	10.6	-314.80****
Rephrasing/ correcting	17	0.5	231	3.8	-122.82****
Quotative <i>well</i>	4	0.1	1	0.02	+3.89*
Move to the main story	13	0.4	129	2.1	-57.85****
Introducing the next scene	18	0.5	33	0.5	-0.06
Conclusive <i>well</i>	4	0.1	12	0.2	-1.04
Indirect answer	23	0.6	126	2.1	-33.75****
Direct answer	57	1.6	549	9	-241.59****
Response to self-raised expectations	0	0	0	0	0
Contributing an opinion	8	0.2	43	0.7	-11.26***
Continuing an opinion/ answer	13	0.4	117	1.9	-49.21****
Evaluating a previous statement	2	0.1	20	0.3	-9.01**
Unclassified - Unfinished sentences such as repetitions	16	0.4	44	0.7	-2.84
Total	231	6.5	1 949	32	-782.81****

4.7 | Qualitative functional analysis of *so*

The following section provides a qualitative functional analysis of *so* while the discussion is subdivided by each function in which a comparison occurs between the ZE corpus and the ICE-GB.

4.7.1 | *So* as marking result or consequence

In example (17), *so* indicates that the writer's responsibility for two offices is a consequence or result of the absence of the principal and the secretary. In example (18), it can be deduced that *so* marks the fact that the writer is living in someone else's residence as a result of not securing accommodation (*res*). By using *so*, the writer shows the chronological sequence of events that led to the writer not having accommodation.

- (17) The principal is now in Australia but her Secretary's going to be away too so I'll be running between two offices. (ICE-GB: W1B-001 #185)
- (18) I failed to secure a Res, so I'm squatting with my former Res-mate in New Complex 5, and I will contribute towards the payment of the accommodation fee of \$8000. (ZE corpus: Social letters 1)

4.7.2 | *So* as the main idea unit marker

The main idea unit marker is illustrated below.

- (19) (A): Is that an irritation when you have a vague feeling you've lent a book to somebody and you can't quite figure it out.
 (A): It's not there.
 (E): If it's a paperback no.
 (E): If it's a hardback.
 (E): And on occasion I have lent hardbacks and not got them back.
 (E) So, I just don't lend hardbacks to anyone now. (ICE-GB: S1A-013 #092-097)

In (19), speaker (E) first answers the question posed by speaker A with an affirmative statement that shows they do not have a problem with lending a *paperback* but not a *hardback* and then digresses to explain what has happened in the past regarding lending hardbacks. Speaker (E) then repeats the main idea with explanations that they no longer loan hardbacks. Therefore, a return to the main idea is marked by the DM *so*. Let us consider an example from the ZE corpus below.

- (20) <\$LL>: Uh I got my, when I got my PhD. I was, I think I was sixty years old. I had been I have been lecturing for more than twenty years with only master's degree. Then I decide to go to enrol for PhD and most people laughed at me saying I was wasting my time to study. But I graduated after six years. I was proud of myself. Uh it was tough but I didn't, did not quit. <„> So graduating with a, with a, with PhD sixty years made me proud. <#>My children were proud of me. (ZE corpus: Private semi-structured dialogue 38)

In (20), speaker <\$LL> is answering a question regarding the day they felt proud by responding that it was the day *when I got my PhD*. Afterwards the speaker digresses by explaining the number of years they had been lecturing, how long it took to complete the PhD and the criticism from other people. Finally, *so* is used to come back to the main idea that getting a PhD at 60 years was satisfying. It is interesting to note that the return to the main idea is marked by an unfilled pause which is longer than a single syllable. The use of an unfilled pause can be viewed as a focusing device by speaker <\$LL> to go back to the main idea.

4.7.3 | *So* used for summarizing/ rewording/ giving an example

Let us consider the following examples.

- (21) I could do what I like in that respect so there was no restriction there uhm which was very unusual, very uncommon. (ICE-GB: S1A-072 #144)
 (22) <\$G>: I just know the Lord is my shepherd, I shall not want. So, you need to trust in God coz she is our savior. (ZE corpus: Private semi-scripted dialogue 8)

In (21), *so* is used to put what was said before *so* in different words. The utterance following *so* expresses the same propositional idea as the utterance that precedes *so*. In this instance, the rewording is expressed in one sentence. Example (22) indicates *so* as a marker of summary. Speaker <\$G> starts by stating the verse in the first sentence. In the second sentence, speaker <\$G> then uses *so* to summarize it immediately afterwards.

4.7.4 | The sequential *so*

In example (23), speaker (A) describes the first sequence of events that led to partial destruction of the tower. The second sequence of events, which is the building something on as a way of *thanksgiving* is preceded by *so*. In this

case, *so* serves the sequential function because it shows a coherent transition from one event to another (Müller, 2005).

- (23) (A): There'd been some subsidence or some terrible storm that'd uhm uh demolished part of the tower.
 (A): The original.
 (A): So then he built this on as a sort of thanksgiving. (ICE-GB: S1A-094 #275-277)

In the ZE corpus, the sequential function of *so* is attested in the corpus as shown below.

- (24) <\$F>: So, Chombo came and I was uh junior councilor. So, you would give a speech, right. So, when you give a speech, you will just uh, you were expected to give a slogan at the end. So unfortunately, I was representing children in parliament. So, I was not obliged to give a slogan. So, when I sat down, his guys were like uh no councilor, you left something, a slogan. So, then I stood up and I tried to do the slogan of which I did not know how to. (ZE corpus: Private semi-scripted dialogue 6)

A sequential transition from one event to another is exemplified in (24). In the first sequence, speaker <\$F> first explains that when they were a junior councilor, they were required to give a speech and do a slogan but did not. The speaker further explains that after being reminded to do so, they tried to do the slogan. In the second sequence marked by the last *so*, speaker <\$F> transitions to the act of doing the slogan. An interesting observation about *so* in (24) is that it is used extremely repetitively. This can be accounted for by the fact that the speaker is chronicling a series of events and uses *so* to connect, organize, and manage the discourse.

4.7.5 | *So* as a boundary marker

According to Müller (2005), *so* as a boundary marker is prevalent between types of talk such as between instructions and at the start of narratives. The boundary marker function does not seem to be attested in either the ZE corpus or the ICE-GB.

4.7.6 | *So* as a speech act marker for questions or requests

Regarding the function of speech act marker for question or a request, Müller (2005) categorized all instances where a question or a request is preceded by *so*, as shown in (25) and (26).

- (25) So why am I going? (ICE-GB: W1B-010 #067)
 (26) <\$A>: Eat you are eating a pollutant. Then when you eat a pollutant you are going to be affected you know healthwise. And then from there you are in trouble.
 <\$B>: Uhm, uhm, mm. So, what can be done? Is there anything that's been done? (ZE corpus: Public dialogue 20)

In (26), *so* is preceded by hesitation markers *uhm, uhm, mm*, which can be viewed as a strategy by speaker <\$B> to process the information given by the doctor regarding pollution and to think about what to say next. After the hesitation markers, speaker <\$B> asks a question preceded by *so*.

4.7.7 | So as a speech act marker for opinion

Müller (2005) describes the use of *so* as a speech act marker for opinion and notes that sometimes the words *I assume*, *I think*, and *I guess* can be used to show an opinion. The following examples indicate *so* as a marker of opinion.

- (27) And the way I'm going about it is uhm I use a lot of turps and rub the color in into the grey areas, so I suppose that's one of the reasons why it appears luminous. (ICE-GB: SB1-008 #113)
- (28) <\$L>: Ok. From your narration, thank you. From your narration I think you said you are a teacher by profession, and you are also a family member. Eh so I assume that you've got children who go to school. (ZE corpus: Private dialogue 6)

Example (27) shows opinion about the color appearing luminous by using *I suppose* and in example (28), *so* prefaces an opinion about the hearer having school going children. *So* is preceded by a hesitation marker *eh*, an indicator that speaker <\$L> is thinking about what to say next.

4.7.8 | So marking implied result

The marker of implied result function of *so* in the ICE-GB is illustrated below.

- (29) (C): Have you got his album.
 (B): Yeah.
 (C): I'd really love to tape it from you if you didn't mind.
 (B): Yeah. If you give me a tape I've got a tape to tape and I can run it off.
 (C): Oh great that'd be
 (C): Well I mean cos actually thinking about it I've not got a uhm record player or anything.
 (C): So yeah <„>. (ICE-GB: S1A-042 #127-136)

In (29), speaker (C) is asking speaker (B) for an album and then when speaker (B) offers to help, speaker (C) is excited but later goes on to reveal that they do not have a *record player*. Afterwards, speaker (C) ends the conversation with *so yeah*. Here, *so* indicates that speaker (C) wants speaker (B) to infer what is implied, and in this case, since speaker (C) mentioned that they do not have a *record player*, then speaker (B) can infer that the tape will not be provided. The function of *so* to mark implied result in the ZE corpus is shown below.

- (30) <\$C>: Ah actually an operation like this has been done in Zimbabwe in nineteen eighty-three eighty-four at Harare Hospital. But as you know uh every set of conjoined twins is different from the next one. So <„> you, we can't say because I've done one set before therefore, I can do the other one. So, uh yeah. (ZE corpus: Public dialogue 27)

After talking about a similar operation being done in the past, speaker <\$C> reveals that *every set of conjoined twins is different*. Speaker <\$C> uses *so* as a cue to indicate the assumption that the hearer can deduce the importance of treating every operation differently. The filled pause after *so* is used as a cue for the hearer to infer what speaker <\$C> is referring to.

4.7.9 | So as a marker of a transition relevance place

At interactional level, *so* can show a changeover of speakers initiated by either the speaker or the hearer (Müller, 2005). In this case, *so* is preceded by a filled or unfilled pause before another speaker takes over. In (31), transition from one speaker to another is marked by the use of *so*. Speaker (D) has been talking about a journey that involves getting the bus and the train. After an unfilled pause lasting more than a single syllable, speaker (D) uses *so* to turn the floor over to another speaker. This is a cue that speaker (E) takes and then asks a question, which is responded to by another speaker.

- (31) (A): You know it's all right if you've got a car but
 (D): I mean,
 (D): ... you have to go to Cambridge and get the village bus out.
 (D): And it took over three and a half hours to get there and over three and a half hour to get back and uh probably four all told.
 (D): Involves getting the tube.
 (D): <„> so.
 (E): Not worth it, is it?
 (C): No. (ICE-GB: S1A-019 #366-377)

Example (32) shows transition relevance place where speaker <\$C> has finished speaking and there is an unfilled pause indicating that the floor is open for speaker <\$A>. Speaker <\$A> does not talk immediately but pauses before beginning the sentence with *so*. In this instance, *so* marks the transition from speaker <\$C> to speaker <\$A>.

- (32) <\$C>: We are going to be going to schools like Chinhoyi<#> We know in Chinhoyi <,> they've got a faculty that focuses on fashion design and all that <,> so we're going there as well Bulawayo <,> we be focusing on that as well <„>.
 <\$A>: <„> So having said that <,> let's ah take you through eeh what I like to call quick fire questions as we get to know better <,> as as Eve Ruoko<,>. (ZE corpus: Public dialogue 34)

4.8 | Qualitative functional analysis of *well*

The following section provides a qualitative functional analysis of the DM *well* while the discussion is subdivided by each function in which a comparison occurs between the ZE corpus and the ICE-GB.

4.8.1 | | *Well* as used for searching for the right phrase

According to Müller (2005, p. 109), the use of *well* to search for the right phrase is usually 'combined with other means of expressing deliberation', for example by using (i) filled and unfilled pauses, (ii) truncated words and intonation units, (iii) repetition, or (iv) other markers. This can be observed in the example from the ICE-GB below.

- (33) (A): What are they about.
 (B): Sorry.
 (A): What are they about I mean.
 (B): Uhm well uh the one I'm sort of trying to finish at the moment is a play for a company called Quicksilver.
 (ICE-GB: S1A-096 #023-026)

In (33), *well* is surrounded by filled pauses and repetition of *uhm* where speaker (B) is searching for the right phrase to answer the question. In the ZE corpus, the function of *well* to search for the right phrase is given below.

- (34) <\$A>: Pig farmers they say they don't have a market for their cattle uhm for their pigs and uh uh they say buyers, traditional buyers such as Colcom refuse to buy from small scale farmers <#> What's your comment on that?
<\$B>: We <,> well uhm l'm l'm l'm surprised because we see we see pork in the supermarket. (ZE corpus: Public dialogue 19)

Example (34) shows an unfilled pause before *well* and a filled pause and repetition after *well*. The DM *well* and the hesitation markers can be viewed as delay strategies used by speaker <\$B> in search of the right phrase.

4.8.2 | *Well* used for rephrasing/ correcting

The use of *well* to correct an utterance or to rephrase it is also highlighted by Schiffrin (1987). In (35), speaker (B) uses *well* to modify or correct the first utterance, which mentions annual dinners and corrects it to an invitation to the *Christmas departmental dinner*.

- (35) (B): I might have to do the uh after dinner speech at our annual well not annual, our Christmas departmental dinner. (ICE-GB: S1A-030 #072)

In the ZE context, *well* is used for rephrasing or correcting in (36). Speaker (B) first refers to the program as *brand new* in the first sentence. Thereafter, there is an unfilled pause and then *well* marks the correction of the utterance to say that the program is in its fourth week. The unfilled pause can be regarded as a strategy used by the speaker to re-evaluate the first utterance, which is then corrected in the second utterance.

- (36) <\$B>: Stand by for a brand-new program. <„> Well it's no longer brand new. It's now four weeks into the program. (ZE corpus: Public dialogue 30)

4.8.3 | The quotative *well*

According to Müller (2005), *well* is used to indicate the starting point of direct speech and the most common quotative construction is BE + *like* with other possibilities being *go*, *say*, *ask*, and *think*. For instance, in (37) from the written register, *well* signals the start of direct speech, as indicated by the use of quotation marks before *well* and at the end of the quoted sentence.

- (37) I asked for a hard piece and she said, 'A really hard one?' So I said 'Well why not?' So I'm given Harriet Shelley's Mad Suicide speech when she's in Hyde Park at midnight. (ICE-GB: W1B-010 #060)

The use of *well* to mark the start of direct speech is manifested in (38), which is an extract from the written register. In writing, quotation marks indicate the use of direct speech (Müller, 2005). All the four examples of the quotative *well* in the ZE corpus were from the written register where direct speech was quoted.

- (38) The Old Man finishes his eating, hands his plate to Betty, washes his hands and after a brief respectable pause says: 'Well, boy, see you when you come back.' (ZE corpus: Creative writing 3)

4.8.4 | *Well* used to move to the main story

The use of *well* to move to the main story is evident below.

- (39) (B): Uhm I like to watch sport
 (B): ... and I enjoy any kind of physical exercise that is you know stretching or uhm working out but it sort of stops there uhm.
 (B): With working now in movement and dance you can use that
 (B): You need some strength.
 (B): Oh this is terrible sorry.
 (A): Right well
 (A): Uh ok.
 (A): Uhm why do you think physiotherapy is really important? (ICE-GB: S1A-003 #007 017)

In (39), after speaker (B) discusses how they enjoy watching sport in detail, speaker (A) uses *well* to move back to the main story, *physiotherapy*. In (40) from the ZE corpus, speaker (B) starts by explaining the effects of smoking when someone is pregnant. Speaker (B) shifts the conversation by including information about the diseases affecting rural communities. Speaker (A) uses *well* to move back to the main story about the effects of smoking on pregnancy and speaker (B) reverts back to it. The use of an unfilled pause is indicative of an attempt to refocus the discussion.

- (40) <\$B>: And Anna I would want to point that there is eh the tendency of this smoking habit aah rising especially among teenagers. Um those that get pregnant it would not be surprising to then get ah premature deliveries or even still births in these pregnancies. So that is why it is important to know that ah smoking ah and pregnancy is quite hazardous. There are also diseases such as cholera and tuberculosis that are increasing in a rural areas. Our teams have been going there to help educate communities on how to prevent cholera. A lot of them face challenges accessing clean drinking water.
 <\$A>: Alright <,> well you mentioned about aahm you mention the issue of aahm <,> smoking actually having an effect before you actually get pregnant. Right. Highlight that for us.
 <\$B>: Yes, you are very right ah ah Anna. Smoking makes it harder for a woman to get pregnant. (ZE corpus: Public dialogue 2)

4.8.5 | *Well* used for introducing the next scene

The use of *well* to introduce the next scene serves to show a transition in the focus of a narrative or when there is the introduction of a new character or to show a change from one scene to another (Müller, 2005). In this instance, in the ICE-GB, the first scene is when speaker (A) discusses in detail a picture showing the presence of officials on the Nile Valley, a prisoner being punched and the tomb of Horum-Heb. Speaker (A), then uses *well* to transition to the next scene. In this case, speaker (A) wants to transition and to focus on discussing *Egyptian* art for the period in question as shown below.

- (41) (A): There's a scene of the officials on their horses <,> relative new comers on the Nile Valley.
 (A): An interesting detail such as the prisoner then receiving a punch on the jaw from the Egyptian official.
 (A): And the tomb of Horum-Heb is one of the masterpieces of the uh Memphite necropolis.
 (A): Right.

(A): Well after that preamble we can now launch into the evidence of the Egyptian art for this period. (ICE-GB: S2A-052 #056-061)

The use of *well* for the purpose of introducing the next scene in the ZE corpus is represented in (42) where the first scene about the graduation is introduced. Afterwards, *well* is used to transition to another scene where they went to *Mega Two* pub. In this instance, the speaker changes the focus from the graduation to the pub.

(42) <\$GG> <#> And when we finally wore our caps to indicate that we were now graduates, hah that was the most important time in my life. Well, then we went to to this uhm pub called uhm it's Mega Two where we had some drinks and braai. (ZE corpus: Private semi-scripted dialogue 33)

4.8.6 | The conclusive *well*

Müller (2005, p. 120) suggests that 'many of the instances of conclusive *well* carried a notion of "I have said enough" about this scene or topic.' This can be observed in the example from the ICE-GB below.

(43) (B): There you go.
 (A): It's a very nice shirt by the way, well.
 (B): It's a very old shirt.
 (A): It's very nice. Thank you. (ICE-GB: S1A-038 #002-005)

In (43), speaker (A) concludes the sentence with *well*. Speaker (B) takes *well* as a sign that speaker (A) has finished talking and responds. In (44), speaker <\$A> uses *well* to give a cue that they are done talking and speaker <\$B> responds with *ok*. In this instance, *well* is followed by an unfilled pause, which is an indicator that the floor is open. Hence, *well* in this case can be considered as aiding in turn-taking among speakers.

(44) <\$A>: In March nineteen ninety-three he was posted to all Arms battle school as the commander and where he served until thirty-one December nineteen ninety-three and well <„>.
 <\$B>: Ok. And would you say you learnt from him and the work that he did? (ZE corpus: Public dialogue 4)

4.8.7 | *Well* occurring as an indirect answer

Insufficient answers, answers that do not directly supply the information required, replies with delayed answers, and replies that provide partial answers are categories of indirect answers (Müller, 2005). An example is given below.

(45) (C): How many legs would you think we'd be talking about with ply?
 (B): Well the ends will be OK uhm.
 (B): If we put in ply, as a shelf I'd put one centre and then one centred in the two halves. (ICE-GB: S1B-073 #286-288)

In (45), speaker (B) does not directly tell speaker (C) the number of legs to be put on the table. Instead, speaker (B) starts by telling speaker (C) that *the ends will be ok* and then moves on to explain how the legs will be fitted. The answer to the question comes at the end where speaker (B) does not provide the actual number of legs, but speaker (C) has to deduce that it is two legs by adding one plus one. An example of indirect answer in ZE is shown below.

- (46) <\$B>: Alright thank you very much Doctor Mbuyeyesango <#> They were sharing one liver. So ha... what did you do?
 <\$C>: Well they they had eh a liver that was ah joined but each one had eh a separate biliary system. So ><, >< eh eh you know the liver makes eh eh bile ><, >< or gall and each had has a separate system. So, what we had to do was to split the liver so that each of the eh patients had a liver with them with a biliary system. (ZE corpus: Public dialogue 25)

After speaker <\$B> asks speaker <\$C> about what they did, the doctor responds indirectly, starting the sentence with *well* and first explains that the conjoined twins shared a liver but had a *separate binary system*. Thereafter, speaker <\$C> answers the question by saying that they *split the liver*. Example (46) has filled and unfilled pauses. This is a strategy by speaker <\$C> to recall the process that was involved in the separation.

4.8.8 | *Well* used as a direct answer

The function of *well* to indicate direct answer occurs when the answer given directly supplies the information needed (Müller, 2005), as illustrated below.

- (47) (A): Uhm can you describe to me if it's possible uhm a typical day in your home when you were a boy of less than fourteen. What do you remember?
 (B): Mm well I remember I remember sort of there were very hot summers uh for some reason. (ICE-GB: S1A-073 #066-067)

In (47), speaker (A) asks a *wh*-question and then speaker (B) provides a direct answer which supplies the information needed by speaker (A). The response starts with a hesitation marker *mm*. This could be because speaker (A) is asking about past events that speaker (B) needs to recall. Therefore, speaker (B) uses *mm* as a delay strategy while thinking back to the time frame asked. *Well* then precedes the direct answer that Speaker (B) recalls that the summers were hot. The direct answer in the ZE corpus is highlighted below.

- (48) <\$A>: Right fantastic stuff but of course locally the premiership continues. What's going on at Luveve?
 <\$B>: Well at Luveve Chiredzi is ah being entertained by Chicken Inn. (ZE corpus: Public dialogue 42)

Considering example (48), speaker <\$B> responds to the *wh*-question regarding what is happening at Luveve posed by speaker <\$A> by supplying a direct answer which provides the information needed by speaker <\$A>. The answer is neither incomplete nor preceded by other information that is not relevant to the question asked. Speaker <\$B> has been asked what is happening at Luveve and answers directly by saying *well at Luveve Chiredzi is ah being entertained by Chicken Inn*.

4.8.9 | *Well* as a response to self-raised expectations

According to Schiffrin (1987, p. 123), in responses to self-raised expectations, 'speakers are treating their own prior talk as something to be responded to.' Response to self-raised expectations does not seem to be attested in either the ZE corpus or the ICE-GB.

4.8.10 | *Well* used in contributing an opinion

In example (49), after asking speaker (B) a question and getting a response, speaker (A) agrees with speaker (B), and then adds an opinion by starting with *well*. *Well* is preceded by a hesitation marker, indicating that speaker (A) thought about what to say next.

- (49) (A) What is the difference for you between experiencing it and feeling that you've made it up.
 (B): Because if I made it up because it had a purpose then I should be able to stop it.
 (A): And will keep you in control.
 (A): Mm. Well also that means that's a bit like carry on taking the tablets you know. (ICE-GB: S1A-062 #036-041)

In (50), *well* has a function of marking where speaker <\$L> begins to contribute an opinion regarding whether it will rain or not.

- (50) <\$L> <#>Now people are not farming enough because of poor rains. <#>I hope that it rains, well but I don't think it will. (ZE corpus: Private semi-scripted dialogue 12)

4.8.11 | *Well* used to continue an opinion/ answer

Well is used to continue expressing an opinion on a subject matter as done in (51), which shows speaker (A) talking about their opinion regarding *pictures I saw at the exhibition*. The continuation of opinion is preceded by *well* and then speaker (A) goes on to share an opinion about the quality of a painting at the exhibition. The hesitation markers *uh* and *uhm* help the speaker to think of a way to continue the opinion.

- (51) (A): Well my impression of the pictures I saw at the exhibition was that he wasn't filtered uh by an actual filter.
 (A): But he was just painting what he wanted to paint.
 (A): And the quality was variable.
 (A): And there was no one standing around saying to him you should junk this.
 (A): And well you know. Uhm and
 (A): Uh so some of it was not very good quality. (ICE-GB: S1A-015 #091-097)

In (52), speaker <\$D> is discussing about people cutting down trees in the area where they live. There is a continuation of an answer after *well* with the speaker further explaining why people cut down trees and the consequences of cutting down trees.

- (52) <\$D> Actually people most of the people are are now buying firewood from nearby farms because the the farms have uh a lot of trees. The area that we live people have actually cut down a lot of trees and well they cut down a lot of trees and build houses there. So, we have very small trees and we cannot find firewood there. (ZE corpus: Private semi-scripted dialogue 4)

4.8.12 | *Well* used as evaluating a previous statement

In (53), speaker (A) first expresses that they do not like *Eastenders*. Speaker (A) then evaluates the statement and changes it to say they just saw the ending. Speakers (A) and (B) go on to discuss that Clyde won despite looking roughed up.

- (53) (B): Did you watch Eastenders?
 (A): Uh uh I don't really go a bundle on it.
 (A): I just saw well just near the end.
 (B): Yeah.
 (B): Clyde won.
 (A) Yeah that's . . . He looked a bit <unclear word> to me.
 (B): He looked to me as if he'd been a bit roughed up you know.
 (A): Yeah he did rather.
 (B): It shouldn't. It looked like it should've been an RSF there didn't it but uh. But obviously he carried on and won.
 (A): Yeah.
 (A): Yeah well all credit to the lad. (ICE-GB: (S1A-095 #197-210)

Example (54) shows *well* being used to evaluate a statement. Speaker <\$A> first highlights the lack of action from the newly inaugurated government. Afterwards, speaker <\$A> uses *well* to evaluate the statement by noting the achievements of the new government such as the economic revival. The unfilled pauses before and after *well* indicate a delay in which speaker <\$A> is assessing the statement. Afterwards, speaker <\$A> uses *well* to change the previous statement.

- (54) <\$A>: I think there is a question there about what is ZANU PF doing since the ah ah inauguration of the new government. I mean we have not seen any particular action. <,> Well <,> but that's not actually true. The reality is that we have produced a turnaround document which is ZIMASSET which is aimed at eh reviving the economy. (ZE corpus: Public dialogue 13)

The next section is a discussion of the results.

5 | DISCUSSION

This study has revealed that there are differences in the frequency of *so*, which occurred more in the spoken register compared to written register in both corpora. This finding is in line with previous studies which report that DMs are mainly found in the spoken discourse (Lam, 2009a; Brinton, 2010). Interestingly, *so* was used more in the spoken register in the ZE corpus compared to the ICE-GB. This may be because Shona has an equivalent DM *saka* which is translated as 'so' in English. Therefore, speakers may be using their knowledge of Shona during conversations, underscoring the influence of Shona on the use of DMs in ZE. In addition, the ZE corpus had 29 occurrences of *saka*. These instances occurred as DMs and in some instances preceded *so* as shown in (55).

- (55) <\$KK>: Why I am saying *so* is because some of the times you you end up in a different career because of the situation or sometimes because of eh your background. Saka so so I can say ih what I was supposed to do personally was to study very well. (ZE corpus: Private semi-scripted dialogue 37)

In (55), *saka* occurs before the DM *so* showing that speaker <\$KK> first thought of the Shona equivalent of *so* and said it before remembering that the utterance is in English and then switching back to English. This can be attributed to the effects of language contact between Shona and English. With regard to *well*, variations were observed in the spoken register, where BrE has more instances compared to ZE. A possible motivation for this variation is the different environments in which English is learned and used. BrE is an L1 learned and used both at home and at school whilst ZE is an L2, and is usually learned at school and rarely used at home (Kadenge, 2010).

Regarding genres, *so* was more frequent in the ZE corpus compared to the ICE-GB in private dialogues, public scripted monologues, creative writing, business letters, and social letters. *Well* occurred more frequently in the ICE-GB compared to the ZE corpus in private dialogues and public dialogues. These results support the claim that contextual domains and text type influences the frequency of DMs (Lam, 2009b). No differences were observed in the frequency of *so* in public dialogues, popular writing, editorials, and newspaper reportage. This may be due to the influence of BrE on ZE since BrE is mainly the norm-providing variety in teaching and learning and in print media (Kadenge, 2009).

Regarding the function frequencies of *so*, variations were evident because the ZE corpus recorded higher frequencies than the ICE-GB in terms of (i) marking result or consequence, (ii) using the main idea unit marker, (iii) summarizing or rewording or giving an example, and (iv) using speech act markers for questions or requests. As discussed above, the possible motivation for the high frequency of these functions in the ZE corpus is the presence of the Shona DM *saka* which is equivalent to 'so' in English. The marking of the implied result and the use of the sequential *so* functions were more frequent in the ICE-GB compared to the ZE corpus. No variations were recorded in the speech act marker for opinion and the marker of a transition relevance place functions. Considering the functions of *well*, the highest variations were recorded in (i) searching for the right phrase, (ii) rephrasing or correcting, (iii) moving to the main story, (iv) choosing the indirect answer, (v) contributing an opinion, and (vi) choosing the direct answer. These functions occurred more in the ICE-GB compared to the ZE corpus. A motivation for these variations is the fact that BrE is used as an L1 whilst ZE is used as an L2. The functions of *well*, namely introducing the next scene and conclusive *well*, did not show variations between the two corpora. The motivation for the functions of *so* and *well* that did not show variations between BrE and ZE is that although there may be influences from other English varieties, ZE is greatly influenced by BrE, which was introduced through colonialism, resulting in similarities between the two varieties (Kadenge, 2009). The following section concludes the study.

6 | CONCLUSION

This paper analyzed the use of discourse markers *so* and *well* in Zimbabwean English and British English to determine whether there are statistically significant variations in their occurrence and function frequencies in spoken and written registers, and in different genres in order to determine if these DMs are used in the same manner in both languages and in different language use contexts. Through an analysis of the ZE corpus and the ICE-GB, this study has provided a comparative description of the use of DMs *so* and *well*. Findings from the study have shown that the DMs occur more in the spoken register compared to the written register. In addition, text type played a role in the frequency of the DMs. The relationship between Shona and English in Zimbabwe was highlighted as a possible motivation for the high frequency of *so* in the spoken register in ZE compared to BrE, showing the impact of language contact on these two languages. In terms of discourse functions, contextual cues helped in assigning primary functions of *so* and *well*. This study has shown that the following functions of *so* occurred more in the ZE corpus compared to the ICE-GB, namely (i) marking result or consequence, (ii) using the main idea unit marker, (iii) summarizing or rewording or giving an example, and (iv) using the speech act marker for questions or requests. The option of using the marking of the implied result use of the sequential *so* functions occurred more in the ICE-GB compared to the ZE corpus. Regarding the functions of *well*, the following options occurred more in the ICE-BG compared to the ZE corpus, namely, (i) searching for the right phrase, (ii) rephrasing or correcting, (iii) moving to the main story, (iv) choosing the indirect answer, (v) contributing an opinion, and (vi) choosing the direct answer occurred more in the ICE-GB compared to the ZE corpus whilst the quotative *well* function was more prevalent in the ZE corpus than in the ICE-GB. These statistically significant combinations of variations of the frequencies and functions in different registers and genres point to the fact that ZE speakers and BrE speakers do not use some of the functions of *so* and *well* in a similar way.

ACKNOWLEDGEMENTS

This project was supported by funding from the National Research Foundation (NRF) of South Africa and the University of Pretoria.

CONFLICT OF INTEREST

The authors state that there are no conflicts of interest.

ORCID

Faith Chiedza Chapwanya  <https://orcid.org/0000-0001-8298-9606>

Joanine Hester Nel  <https://orcid.org/0000-0001-7618-6014>

NOTES

¹ <https://www.ucl.ac.uk/english-usage/projects/ice-gb/>

² Although permission was given to use Marungudzi's corpus from North-West University, access to only 150 000 words out of the 390 000 words was provided because he passed away in January 2018 before handing the corpus over to the university. This meant that the data collection process had to be adapted to collect more samples than originally planned.

³ <https://www.sketchengine.eu/>

⁴ <https://ucrel.lancs.ac.uk/llwizard.html>

REFERENCES

- Aijmer, K. (2002). *English discourse particles: Evidence from a corpus*. Amsterdam/Philadelphia: John Benjamins.
- Aijmer, K. (2013). *Understanding pragmatic markers: A variational pragmatic approach*. Edinburgh: Edinburgh University Press.
- Algouzi, S. (2021). Functions of the discourse marker *so* in the LINDSEI-AR corpus. *Cogent Arts & Humanities*, 8, 1–14.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Blakemore, D. (2002). *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers* (Vol. 99). Cambridge: Cambridge University Press.
- Brinton, L. J. (2010). *Pragmatic markers in English: Grammaticalization and discourse functions*. Berlin: Mouton de Gruyter.
- Buysse, L. (2012). *So* as a multifunctional discourse marker in native and learner speech. *Journal of Pragmatics*, 44, 1764–1782.
- De Klerk, V. (2005). Procedural meanings of *well* in a corpus of Xhosa English. *Journal of Pragmatics*, 37, 1183–1205.
- Diskin-Holdaway, C. (2021). *You know* and *like* among migrants in Ireland and Australia. *World Englishes*, 40, 562–577.
- Fraser, B. (1993). Discourse markers across language. *Pragmatics and Language Learning Monograph Series*, 4, 1–16.
- Fraser, B. (2009). An account of discourse markers. *International Review of Pragmatics*, 1, 293–320.
- Holmes, J. (1984). Hedging your bets and sitting on the fence: Some evidence for hedges as support structure. *Te Reo*, 27, 47–62.
- International Corpus of English-Great Britain (ICE-GB). Retrieved from <https://www.ucl.ac.uk/english-usage/projects/ice-gb/>
- Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: The English language in the outer circle. In R. Quirk & H. G. Widdowson (Eds.), *English in the world: Teaching and learning the language and literatures English in the world* (pp. 11–30). Cambridge: Cambridge University Press.
- Kachru, B. B. (1988). The sacred cows of English. *English Today*, 4(4), 3–8.
- Kachru, B. B. (1992). Models for non-native Englishes. In B. B., Kachru (Ed.), *The other tongue: English across cultures* (pp. 48–74). Urbana and Chicago: University of Illinois Press.
- Kadenge, M. (2009). African Englishes: The indigenization of English vowels by Zimbabwean native Shona speakers. *The Journal of Pan African Studies*, 3, 156–173.
- Kadenge, M. (2010). Zimbabwean English: A sociophonological exploration. *South African Journal of African Languages*, 30, 35–51.
- Lam, P. W. Y. (2009a). Discourse particles in corpus data and textbooks: The case of *well*. *Applied Linguistics*, 31, 260–281.
- Lam, P. W. Y. (2009b). The effect of text type on the use of *so* as a discourse particle. *Discourse Studies*, 11, 353–372.
- Leedy, P. D., & Ormrod, J. E. (2021). *Practical research: Planning and design*. Essex: Pearson Education Limited.
- Marungudzi, T. (2016). *Some morphosyntactic features and variety status of Zimbabwean English: A corpus-based study*. [Doctoral dissertation, North-West University, South Africa].
- Mohr, S. (2021). *You know* and *I think* in English(es) in Zanzibar. *World Englishes*, 40, 534–547.
- Müller, S. (2005). *Discourse markers in native and non-native English discourse*. Amsterdam: John Benjamins.

- Nelson, G. (2002a). International Corpus of English: Markup manual for spoken texts. Retrieved from ice-corpora.net/ICE/spoken.doc
- Nelson, G. (2002b). International Corpus of English: Markup manual for written texts. Retrieved from ice-corpora.net/ICE/spoken.doc
- Oladipupo, R. O., & Unuabonah, F. O. (2020). Extended discourse-pragmatic usage of now in Nigerian English. *World Englishes*, 40, 371–389.
- Rayson, P. (2023). Log likelihood calculator. Retrieved from <http://ucrel.lancs.ac.uk/llwizard.html>
- Rayson, P., Berridge, D., & Francis, B. (2004). Extending the Cochran rule for the comparison of word frequencies between corpora. In G. Purnelle, C. Fairon, & A. Dister (Eds.), *Le poids des mots: Proceedings of the 7th International Conference on Statistical analysis of textual data* (pp. 926–936). Louvain-la-Neuve: Presses Universitaires de Louvain.
- Schiffrin, D. (1987). *Discourse markers*. Cambridge: Cambridge University Press.
- Schourup, L. (1999). Discourse markers. *Lingua*, 107, 227–265.
- Unuabonah, F. O. (2019). Frequency and stylistic variability of discourse markers in Nigerian English. *Corpus Pragmatics*, 3, 249–271.
- Vickov, G., & Jakupcevic, E. (2017). Discourse markers in non-native EFL teacher talk. *Studies in Second Language Learning and Teaching*, 7, 649–671.

How to cite this article: Chapwanya, F. C., & Nel, J. H. (2023). Discourse markers *so and well* in Zimbabwean English: A corpus-based comparative analysis. *World Englishes*, 1–24. <https://doi.org/10.1111/weng.12626>

APPENDIX 1

TRANSCRIPTION CONVENTIONS

Transcription based on Nelson (2002a, 2002b).

<\$A>, <\$PP>, etc.: Speaker identification

<&>: Incomplete sentence

<|>...</|>: Subtext marker

</.>: Incomplete word

<,>: Short pause

<„>: Long pause

<[>: Overlapping string

<@>...</@>: Changed name or word

<indig>...</indig>: Indigenous word(s)

<unclear>...</unclear>: Unclear word

<O>...</O>: Marks utterances such as laughter, coughs, tongue clicks, and sneezes.

<&>background noise</&>: Background noise

<&>background conversation</&>: Background conversation

<#>: Text unit marker

↔...</>: Normative deletion

<+>...</>: Normative insertion

<=>...</>: Original normalization

<p>: Paragraph marker

<h>...</h>: Heading

<&>...</&>: Editorial comment