

A corpus-based investigation of the Given before New principle in Tanzanian English

Sondos Hassan Ibrahim

Northumbria University

Abstract

This paper investigates the information-packaging structures of Tanzanian English in order to evaluate the universality of the given before new (GBN) principle. Since Halliday (1967) observed that familiar information tends to precede new information, GBN has been accepted as a ‘linguistic truism’ (Birner & Ward, 2006 p. 291) and rarely challenged. However, recent cross-linguistic studies suggest that L2 learners of English prefer a new before given (NBG) structure (e.g. Park, 2011, p. 109), calling into question GBN’s universality. As a region where English largely functions as a second language, Tanzania is a worthy domain for further investigation of this kind. In this context, I analyse the personal columns category of the Tanzanian component of ICE-EA. I compare the frequency of GBN and NBG structures in this corpus category, evaluating the contexts in which these structures occur. My findings reveal that, although NBG is more prevalent in Tanzanian English than in standard British English, GBN remains a dominant feature in this English variety. The goal of this research is to use corpus-based methods to scrutinise the accuracy of this principle to describe non-standard varieties of English.

Key words: *Tanzanian English; Standard British English; Given Before New (GBN) principle; New Before Given (NBG) structure*

Introduction

This paper documents an investigation into the information-packaging structures of Tanzanian English, conducted using ICE-EA, which is the East-African component of the International Corpus of English. Compiled in the early 1990s, ICE-EA contains just over one million words of written and spoken language from Kenya and Tanzania; it is one of the only Tanzanian English language corpora in present circulation (Schmied & Hudson-Ettle, 1999). The recent emergence of new ICE-corpora has witnessed a surge of interest in grammatical variation across World Englishes, facilitating the scrutiny of generic assumptions about English syntax (Kortmann, 2006, p. 604).

One idea that has recently been challenged is the assumption that all varieties of English follow Michael Halliday's given before new principle (GBN). Though precise definitions of 'given' and 'new' differ, GBN broadly posits that, in discourse situations, familiar knowledge normally precedes unfamiliar information (Halliday, 1967, p.213). Though some grammarians such as Birner & Ward (2006) assert that GBN is a 'linguistic truism' of all languages (p. 291), recent studies have found that Polish, Korean and Swahili may prefer a new before given structure (NBG) (Mithun, 1992; Park, 2011; Vitale, 1981). The question that arises, then, is whether language contact has any impact on the packaging structure of English varieties.

In this context, I question the extent to which GBN is a feature of written Tanzanian English. Tanzania's sociolinguistic situation, where English (GBN) and Swahili (NBG) co-exist as joint official languages, marks out this variety for an investigation of this kind. Beginning with a brief review of the critical literature on GBN and on Tanzania's unique language context, I outline the design, challenges and results of the corpus experiment. Each clause from the 20,125 word Tanzanian personal columns subsection of ICE-EA is considered for the alternation between NBG and GBN structures. Deviations from GBN are analysed in particular detail in terms of the effect on intelligibility and journalistic style. Therefore, a consideration of GBN's prevalence in a non-standard English variety is here used as a vehicle to explore the pragmatic effects of this principle. The goal of this investigation is to use corpus-based methods to test my working hypothesis that Tanzanian English exhibits a reduced preference for GBN than standard British English due to Swahili/English contact. In so doing, this paper hopes to pave the way for further research into the impact of language contact on the ordering of information in texts.

Definitions and scope: What is ‘given’? What is ‘new’?

In the critical literature on information packaging, there is little theoretical consensus on the exact scope of the terms ‘given’ and ‘new’ (Halliday 1967, Ozón 2006). This section defines these terms for the purposes of the corpus-based study and highlights the results of major studies on GBN and NBG in English and other languages.

The earliest reference to information packaging came from Halliday who broadly defined given information as any knowledge which is judged to be retrievable by the hearer or addressee, either ‘situationally or anaphorically’ (Halliday, 1967, p. 204). Conversely, new information is defined as any unfamiliar content. This distinction is illustrated in the following examples:

- (1) Mary paid Peter so he bought himself a chocolate bar

G N

- (2) Sue was having a picnic. The beer was cold because that’s how she likes it.

N G

The use of the pronoun ‘he’ in (1) refers to the named individual Peter and would therefore be judged as anaphorically given. By contrast, ‘a chocolate bar’ constitutes previously unnamed information and is therefore syntactically new. ‘The beer’ is both situationally and anaphorically irretrievable and must, necessarily, be new. The principle of GBN states that English (like many other languages) is more likely to adopt a structure similar to (1) than (2)¹.

Whilst notional definitions of GBN and NBG remain uncontroversial, challenges arise when attempting to identify exact syntactic criteria for givenness and newness. The issue of scope remains problematic. Whilst Ozón’s (2006) and De Cuypere & Verbeke’s (2013) GBN investigations have been limited to a consideration of dative alternation in English (e.g. ‘I gave the book to Joe’ vs. ‘I gave Joe the book’), more experimental studies have worked with a significantly broader range of syntactic criteria. For instance, Di Tullio (2006) has named cleft-construction as examples of NBG clauses (p.483). By contrast, Sityaev’s (2000) corpus-

¹ Linguists disagree on the exact labels ascribed to givenness and newness, with some referring to ‘familiarity’ (Prince, 1981, p. 254) and others preferring ‘retrievability’ (Lambrecht, 1994, p. 84), however, these are relatively minor terminological differences.

based study distinguished between givenness and newness on the basis of different intonation patterns and word stress (p.285).

Returning to examples (1) and (2), some general features of GBN and NBG clauses can be observed. For instance, it is clear that the identification of information as given and new only applies to noun phrases (NPs) in non-canonical clauses (Birner & Ward, 2006, p. 291; Ozón, 2006). For this reason, the sentence ‘Sue was having a picnic’ cannot be analysed as GBN or NBG, as its ‘subject + predicate structure’ constitutes a canonical clause (Ward & Birner, 2002). As my 20,125 word corpus sample must be analysed manually, it is preferable to work with a broader set of criteria from givenness and newness. These criteria include (but are not limited to): clefting, personal and demonstrative pronouns and dative alternation. For reasons outlined above, my investigation must exclude all canonical clauses where the alternation between given and new information is non-existent.

It is only in the last decade that researchers have begun to use corpora to investigate the packaging structures of English language varieties, acknowledging GBN variation across World Englishes. However, neither Ozón’s (2006) study on GBN in British English, nor DeCuypere & Verbeke’s (2013) investigation on Indian English have offered an explanation for this phenomenon. The uniqueness of the present study lies in its foregrounding of language contact as a significant factor in influencing the word order of English language varieties. It is for this reason that my own study focuses on a variety of English that has emerged in close proximity to a known NBG language, Swahili.

Tanzania: Language Situation, Context and Genre

Just as there are various syntactic reasons to suggest that the NBG preference of Swahili has some effect on the structure of Tanzanian English, many sociolinguistic factors mark out this variety as interesting. In this section, I outline the sociolinguistic context of Tanzania, with a focus on the relationship between the country’s two official languages: Swahili and English. In section 3.2, I discuss the reasons behind my chosen corpus subsection: Tanzanian personal columns.

Tanzanian English and English in Tanzania

English first came to Tanzania shortly after World War One, when the country was divided into two regions: Zanzibar and Tanganyika (Kipacha, 2006, p. 502). Whilst these regions were British colonies until 1964, the British governors did not side-line Swahili, and the language continued to thrive in churches, schools and in public administration (Schmied, 2006, p. 190). Thus, Tanzanian English emerged in a different environment to other New Englishes in colonised countries (Platt et al., 1984, p. 17). Although the 1967 Tanzanian constitution explicitly names English and Swahili as joint official languages, their societal functions are very different. Marten (2006) points out that Swahili is the majority language of Tanzania's 36 million inhabitants (p.502), whilst Schmied (2006) has called it the country's 'true national language' (p.191). English, on the other hand, serves a largely public role and is a clear marker of an individual's education and social standing. Although Tanzania, Kenya and Uganda are often grouped together as 'East African Englishes', Tanzanians exhibit significantly reduced levels of English fluency in comparison to their Kenyan and Ugandan neighbours: just 5% of all Tanzanians identify as English language speakers (Schmied, 1991, p. 81). With this in mind, it would seem that the readership of the personal columns selected for my corpus study represents a very small (and privileged) section of Tanzanian society.

Attitudes towards English in Tanzania vary significantly, and recent studies have recognised that the relationship between English and Swahili is not a harmonious one (Rubanza, 1995). This is due, in part, to language teaching in Tanzanian schools. It is a bizarre feature of the education system that the language of instruction in primary schools is Swahili whilst secondary schools teach exclusively in English. Moreover, only 10% of primary school graduates proceed to secondary school and just 2% of these ever attend university (Rubanza, 1996, p. 84)². Many pro-Swahili campaigners have called for a change in these policies, arguing that Swahili and English should be taught together at all educational levels. It has been argued that the current system damagingly privileges English over Swahili, as its teaching at secondary and tertiary level is invariably reserved for the wealthiest children in society (Rubanza, 1996, p. 17)³. Nonetheless, it is important to remember that even for

² These startling statistics are broadly supported by evidence from the most recent Tanzanian Census (2012).

³ As of February 2015, the Tanzanian government has announced a radical change to these policies. For the first time in the country's history, Swahili will replace English as the sole medium of instruction in all Tanzanian primary and secondary schools. Although it is too early to review the effects of this policy change, the announcement sparked significant national and international debate (Global Voices Online, 2015).

those Tanzanians privileged with an English education, the preferred medium of communication in non-institutional settings will normally be Swahili (Marten, 2006).

Tanzanian Personal Columns

Although only a tiny section of Tanzanian society would willingly read an English newspaper, it is worth remembering that this figure is considerably greater than the number who would hold an informal conversation in English. Indeed, Schmied and Hudson-Ettle (2007) have pointed out that East African English newspapers carry significant linguistic influence over their readers (p.103). Personal columns are explicitly named in the ICE-EA manual as playing a prominent ‘social and linguistic role’ in Tanzanian society (Schmied & Hudson-Ettle, 1999, p. 18). From a social perspective, this is because the columnists are important public figures, exercising considerable influence in daily life. From a linguistic perspective, the significance lies in the frequent use of Swahili proverbs, which are very often left un-translated for the wide appeal of their readership. Unlike Kenya, which boasts eight daily English newspapers, Tanzania lays claim to just one: The Daily News (Schmied & Hudson-Ettle, p. 104).

The data for my experiment are all drawn from a single newspaper, allowing my investigation to confidently draw a conclusion about Tanzanian personal columns as a whole, without the need to consider a particular newspaper’s house-style.

Corpus Experiment

This theoretical and contextual outline has highlighted the importance of my research question. In questioning the extent to which GBN is a feature of Tanzanian English, I am researching a variety that has emerged in a sociolinguistic setting that might be expected to prefer the NBG structure of Swahili. There is a clear gap in research of this kind and ICE-EA seems to be an appropriate corpus with which to carry out this investigation. The present section outlines the methodology, reviews the data and analyses the results from the corpus experiment.

Experiment Design

In order to assess the extent to which GBN in Tanzanian English differs from British English, the methodology for this experiment draws heavily on Ozón's (2006) study using ICE-GB. In this instance, Ozón's findings are used as a baseline against which to measure deviations from British English in the Tanzanian English corpus. However, the specialist parsing and tagging software, ICE-CUP, that allowed Ozón to analyse language data from ICE-GB is not yet available for ICE-EA. Instead my own analysis is manual and therefore draws on a subcorpus approximately fifty times smaller than Ozón's (2006). The resulting experimental design implications are that Ozón's data for British English and my own for Tanzanian English are not strictly comparable. Nonetheless, these issues are somewhat mitigated by the broader definitions for 'givenness' discussed in section 2. Table 1 draws on definitions from the critical literature to outline the criteria for tagging a clause in the corpus as 'given'.

Table 1: Instructions for tagging a clause in the corpus as 'given'.

Tag as <i>Given</i>	Examples from ICE-EA
The personal pronouns <i>he, she and they</i>. Only where the referent has explicitly been mentioned by name previously.	<u>Kibogovo</u>...<u>he</u> transformed me into a chatterbox. (ICE-EA, W2E019T).
The personal pronouns <i>I and you</i>. Assume that <i>I</i> refers to the columnist, and is therefore situationally given. Likewise, assume that <i>you</i> is also situationally given and is referring to the reader. This does not apply to any quotations or idioms.	<u>I</u> have quarreled with my stove. (ICE-EA, W2E011T).
Repeated proper nouns, common nouns and other noun phrases.	This guy had foresight...<u>this guy</u> is going to make business. (ICE-EA, W2E012T).
The second part of a cleft construction. (i.e. the cleft part will always be tagged as 'new').	It's the teachers <u>who will pay</u>. (ICE-EA, W2E020T)

It was decided that the simplest means of representing given and new information was to work with a colour-coded system. All given data was marked in pink and all new information was marked in green. The benefit of this system was that it was easy to see at a glance whether any significant clumping of GBN structures tends to occur, aiding the qualitative analysis of results. Unclear clauses were marked in yellow, including those in which Swahili words or proverbs were used. Figure 1 shows a small sample of the colour-coded analysis. For reasons already outlined in section 2, canonical clauses were excluded from the analysis, except in the case of canonical clauses that included one of the personal pronouns mentioned in Table 1.

Figure 1: Sample colour-coded analysis, displaying given (green), new (pink) and unknown (yellow).

```

<X_></X></I>
W2E013T
<I> <h> </>Darubini </h>
<h> </><ca/>Mageuzi' in our homes </h>
<#/>IN the process of pondering the issue which the writer intends to exchange views on with his
fans and critics, the title he initially opted for was - 'when wives become family breadwinners'.
<#/>Interest in the subject was spurred by reading a document containing research findings on
the pattern of income generation in Dabangan urban families, and its attendant social problems.
<#/>The smart social worker behind this project also happens to lecture at one of the nation's
high institutions of learning and used Dalia city as a case study.
<#/>Given that workers in government, parastatal institutions, and private companies constitute a
sizeable percentage of Dalia city's population, the researcher had to pay much attention to this
social group.
<#/>Among families visited during the research exercise was one of Mr Mashaka. <#/>He works
for the government as an office supervisor and earns about 8,000/- per month, a good salary by
Dabangan government remuneration standards.
<#/>You may as well note that Mashaka heads a family of eight people who include a wife and
six children.
<#/>According to findings of this research, Mashaka needs about 800/- a day to provide his
family with poor quality food, leaving requirements like clothing, school fees etc. unattended.
<#/>The first thing you are likely to think of in this case is that a family whose head earns 8,000/-
a month when his minimum food budget during this very span stands at 24,000,- must be
starving.
<#/>Surely, the family would be starving, had it not been for <ca/>Mama Zawadi's determined
plunge into the food industry. <#/>By the way, <ca/>Mama Zawadi is Mashaka's legal wife.

```

The next stage was to sort through this tagged sample in order to tabulate the number of GBN, NBG and unclear clauses in each 2,000-word text. The results were then reorganized, taking into account the topic of the column. Finally, the clauses were considered in terms of different types of GBN and NBG structure, in order to ascertain the preferred packaging structures of Tanzanian English and in this particular genre.

Exclusions

Idioms, Swahili words and column headlines presented a particular challenge in my analysis and were ultimately excluded from the final results. This section will outline the issues posed by these three categories and the challenges involved in tagging them.

Farsi (2013: 4) has noted that Swahili is a particularly idiom-rich language, so the presence of some Swahili proverbs or idiomatic expressions was unsurprising. Around forty translated Swahili proverbs were identified in the corpus sample, including the following examples that were repeated several times:

- (3) Fire is raging in this family. (ICE-EA, W2E013T)
- (4) Havens of peace. (ICE-EA, W2E011T)
- (5) His eyes betray untruthfulness. (ICE-EA, W2E015T)

Although it was possible to ascertain the meaning of these idioms from reference books, tagging them was not straightforward. In an example like (3), the clause seems to behave as an NBG structure with ‘this family’ referring to anaphorically given information and ‘fire’ being new. However, from an idiomatic perspective, it seemed to make more sense to take the entire idiom as one unit, replicating the trend in some corpus tagging systems to ditto-tag entire phrases (Denison, 2007).

The problem remained as to whether to treat this entire unit as given or new. Either choice would be problematic. Whilst it seems sensible to tag a unit as ‘given’ if it has not been mentioned before, it is important to consider the rationale behind a columnist’s choice to translate a popular Swahili proverb. By definition, an idiom is deducible not by an understanding of its individual parts but by a familiarity with its use as a whole. If given information is defined as anything which the speaker ‘assumes the addressee to know’, then it seems that idioms should always be tagged as ‘given’, particularly those idioms which have been translated (see section 2). Ultimately, idioms seemed to occupy a fuzzy position in the corpus sample, necessitating their exclusion from the results.

Similar challenges arose when working with non-English lexis. Whilst ICE-corpora very often include non-English words, this poses a challenge to the tagging process. A significant number of Swahili words are included in the corpus-sample, as exemplified in (6) – (8). In keeping with ICE-EA conventions, Swahili words are italicised and tagged <ea/>.

These examples have been translated using the OED and Awde's (2000) Swahili-English dictionary.

- (6) we have a very quiet, <ea/>*staarabu* manner. (ICE-EA, W2E020T)

English: we have a very quiet, civilised manner.

- (7) here in the <ea/> *Babaangu* and <ea/> *Mamaangu* land.

(ICE-EA, W2E010T)

English: here in the father and motherland.

- (8) those chunks of <ea/> *ugali* and <ea/> *ubwabwu* dishes.

(ICE-EA, W2E100T)

English: those chunks of maize and porridge dishes.

It was difficult to determine whether these Swahili words should be treated as the absolute synonyms of their English counterparts. For instance, although it would be appealing to tag (7) as given following the mention of 'fatherland' a few lines earlier, this seems to be a slippery slope, potentially leading to every single synonym pair in the sample being tagged as given (e.g. raising the question as to whether 'allies' should be tagged as given because 'friends' has already been mentioned). As with idioms, it seemed more sensible to exclude Swahili words in the first phase of corpus analysis.

Column headlines were the last to be excluded. As the catchy, opening gambit of any newspaper column, a headline is rarely a full non-canonical clause. Issues arose when parts of the column made some reference to the headline in a way that could be interpreted as anaphorically given, although in general the knowledge was 'new'. Attention-grabbing headlines such as 'That parking business!' use deictic markers in such a way that they seem to refer to experiences with which the reader is called upon to empathise. It is possible to read these as situationally given but overall this seemed to complicate my analysis. Therefore, although headlines are a key part of my chosen genre, excluding headlines from the analysis seemed the most reasonable course of action.

The dataset

After taking into account the exclusions outlined in section 4.2, the 20,125-word corpus sample yielded around 1,036 distinct clauses. Having excluded all canonical clauses, I was left with 747 clauses to analyse. Thirty individual columns by seven columnists were

(10) I instructed my brain.

G N

= *situationally given*.

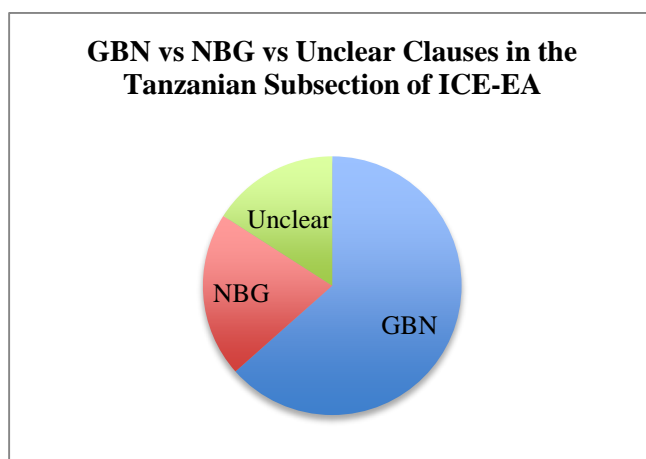
(ICE-EA, W2E0150T)

The results are displayed in Table 3, whilst Figure 2 graphically displays the proportion of GBN, NBG and unclear clauses from the first phase of the experiment. Figures are correct to 1 decimal place.

Table 3: Results from the analysis of the total 747 clauses.

GBN	NBG	Unclear
474	154	119
=63%	= 21%	=16%

Figure 2: Proportions of GBN, NBG and Unclear Clauses in the corpus sample



NBG structures accounted for over one-fifth of the corpus sample, as in the following example:

(11) Ugly creatures and silly they seem to me. (ICE-EA, W2E0150T)

In (11), taken from a column describing the alcohol consumption of government officials, ‘ugly creatures’ clearly places extra emphasis on the columnist’s judgement. The information could very easily be re-packaged with any of the following constructions:

(11a) They seem to be ugly and silly creatures to me

(11b) To me they seem to be ugly and silly creatures.

The original example (11) is by no means unintelligible, but it does seem to display a certain journalistic flourish. It would seem from this case that flouting the GBN principle does not result in unintelligibility but it does shift the focus of the clause.

A significant number of unclear cases arose in the corpus analysis, accounting for 16% of the total results. Many of these were the opening lines of the column, where deictic markers such as ‘this’ or ‘that’ seemed to indicate a sense of givenness, but the information was generally irretrievable, situationally and anaphorically. This is illustrated in the following use of the word ‘this’ as the opening of the column:

(12) **This** was supposed to be a secret between **me, I and myself** (ICE-EA, W2E011T)

Other unclear cases arose due to the journalistic tendency to replicate the colloquial style of general conversation.

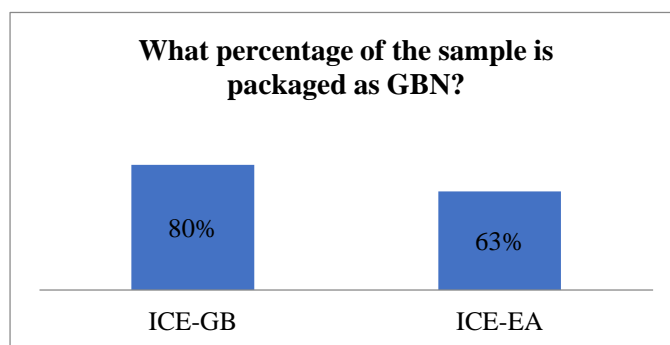
(13) Should **I** take it from **my country, my beloved country?** (ICE-EA, W2E012T)

Example (13) raises the problem of repetition. It is unclear whether ‘my beloved country’ should be treated as a separate unit from ‘my country’ or if they should all be tagged as the same noun phrase. This sentence is conversational in its tone and would not be out of place in a sample of spoken language. As with the examples of NBG constructions, cases of repetition do not result in unintelligibility but instead serve as a stylistic marker of the columnist’s journalistic aptitude and their confidence in manipulating conventional packaging structures.

Finding that GBN dominates Tanzanian English personal columns did not adequately address my research question. It was thus necessary to draw some comparison between Tanzanian English and British English to assess the extent to which Tanzanian English deviates from packaging norms. Taking into account methodological differences,

approximately 80% of ICE-GB was identified as GBN, in comparison to just 63% of my own sample from ICE-EA. This shows a reduced preference of 17% (Figure 3)⁵.

Figure 3: Comparison between my own results and those of Ozón's (2006) work with ICE-GB.



Elaboration: The role of language contact

In order to test my hypothesis that Swahili/English contact is responsible for reduced GBN tendencies, it is necessary to return to the Swahili words in the sample. Although those clauses containing Swahili lexis were initially excluded from the analysis, it seemed important to ascertain their packaging structures in isolation from the rest of the sample. This was judged to be the only conclusive way of evaluating the role of language contact.

In total, there are 144 Swahili words in the corpus, with many of these repeated several times over. Examples include:

(14) The <ea/>babaangu and <ea/>mamaangus here were rumoured to have been planning it.

N

G

(ICE-EA, W2E011T)

English: The mothers and fathers here were rumoured to have been planning it.

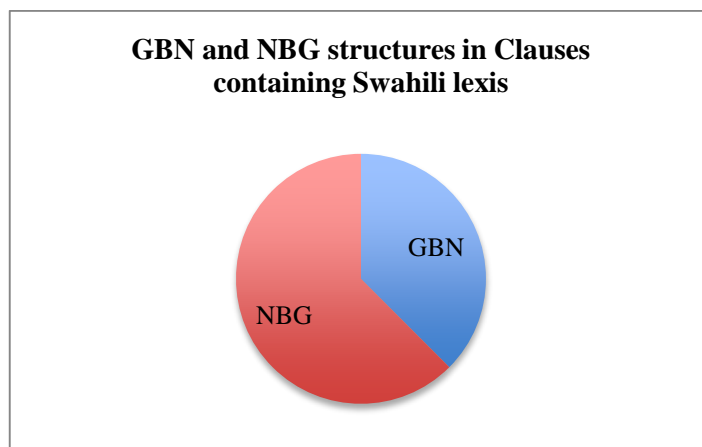
Putting the problems of synonymy to one side, 90 clauses (62.5%), which contain Swahili words, exhibit an NBG structure whilst 54 (37.5%) are tagged as GBN.

This is a stark comparison to the 21% NBG clauses in the entire sample. The results from the final stage of the experiment are displayed in Figure 4. The chart in Figure 5

⁵ These calculations are based on Ozón's (2006) observations that 'the DO...shows a marked preference for new information (approximately 80%)' in DOC constructions (p. 255).

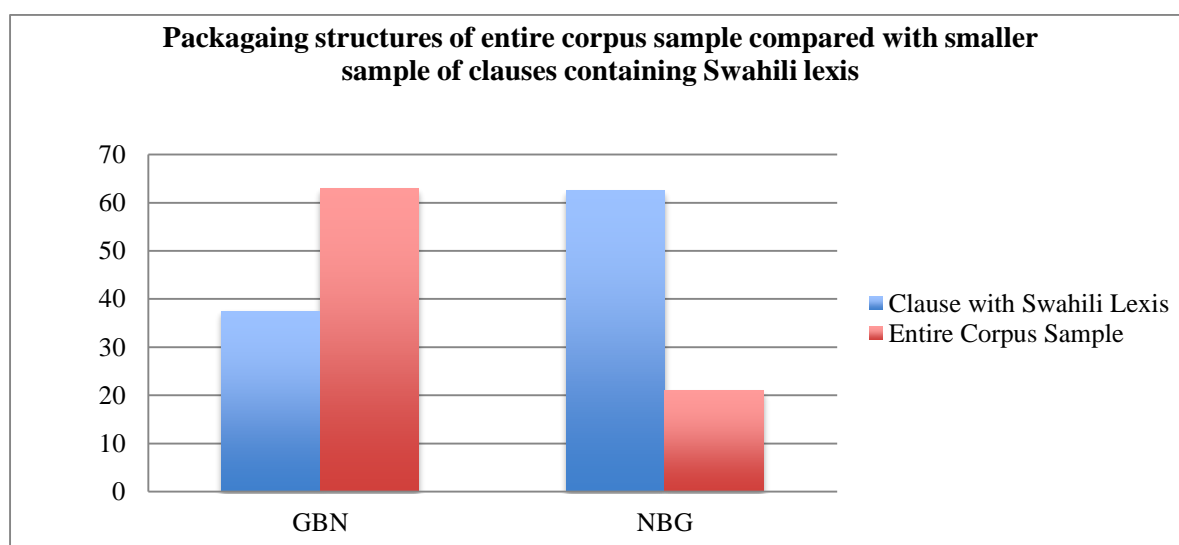
compares the packaging structures of the entire corpus sample to the results from the clauses containing just Swahili lexis.

Figure 4: Comparison of GBN and NBG structures in clauses containing Swahili lexis.



Given these significant differences, it is reasonable to suggest that language contact is at least partially responsible for the different packaging structures of Tanzanian English. However, it is impossible to conclusively prove whether other factors, such as the genre of personal columns and the high esteem of Swahili in Tanzanian society, also carry some influence.

Figure 1: Chart showing that NBG is significantly more prevalent in those clauses containing Swahili lexis.



Conclusions and future research

This paper began by questioning the extent to which GBN is a feature of written Tanzanian English. Generally, it would seem that GBN does dominate this variety supporting previous scholarship that has found that English varieties tend to prefer this packaging structure. On the question of whether Tanzanian English shows a reduced preference for GBN than British English, it would seem that the answer is a very tentative yes, but with some important caveats (see section 4.6). On the whole, I have argued that defining and delimiting the terms ‘given’ and ‘new’ so that they take into account shared, extra-linguistic knowledge as well as syntactic retrievability is essential. In this way, I prefer Quirk et al.’s (1985) use of the term ‘information processing’ as it places the onus on the addressee’s interpretation of knowledge; pragmatic considerations of how language is understood in context are key. In this case, working manually afforded me significant freedoms and allowed me to carefully refine the criteria for givenness and newness. Moreover, the challenging cases of idioms and headlines could be excluded with relative ease. On balance, the benefits of manual data analysis outweighed the difficulties.

My investigation has concluded that language contact has a significant part to play in influencing the reduced preference for GBN in Tanzanian English, although I concede that this may be one of many factors. Future researchers may wish to apply the framework outlined in this experiment to a corpus sample of a different English variety. There is also potential to work with a different written section of ICE-EA, perhaps Kenyan personal columns, to evaluate the role of social factors such as higher English fluency and better education on packaging structures. So, whilst GBN is certainly not ‘a linguistic truism’ of all languages, it would seem that, in the case of Tanzanian personal columns, it remains a dominant feature.

Caveats

This paper has outlined the methodological difficulties of working with an untagged corpus. Without computer software to automatically judge what is retrievable and irretrievable, remaining consistent in the analysis was very challenging. My work with Tanzanian personal columns shows that NBG constructions very often mark out

a columnist's stylistic confidence instead of a lack of English competence. Indeed, it is impossible to separate my results from an awareness of the journalistic tendency to flout standard grammatical rules. For instance, Weir (2009) has identified the prevalence of subject-dropping in headlines as one of many examples where newspapers, and other news sources, willingly break syntactic conventions. Indeed, it is possible that my results say less about language contact in Tanzania than they do about the packaging structures of newspaper columnists across English varieties. This poses a compelling counter-argument to my overall thesis. If Tanzanian personal columns show a reduced preference for GBN than British English, it could be suggested that the personal columns genre is the reason for this difference. However, taking into account the results shown in Figure 5, it would appear that Swahili lexis has a significant part to play in altering the packaging structures of a clause. Language contact seems to be a very likely explanation of these results.

In spite of this, it must be made clear that my results are not easily transferable to other genres of written Tanzanian English, especially those which are free from the grammatical idiosyncrasies of journalism. My study does not provide a solution to this challenge of applicability and future researchers in this area should be aware of the socio-pragmatic norms that govern their chosen text type before attempting to replicate a study of this kind.

References

- Awde, N. (2000). *Swahili-English Practical Dictionary*. New York: Hippocrene Books.
- Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, 8 (4), 243-257.
- Birner, B. & Ward, G. (2006). Information and Structure. In B. Aarts & A. McMahon (Eds.). *The Handbook of English Linguistics* (pp. 291-317). Oxford: Blackwell.
- Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. Li (Ed.) *Subject and topic* (p.25-33). New York: Academic Press.
- Census Results 2012, Tanzanian National Bureau of Statistics for Development*. Retrieved April 4, 2014, from <http://www.nbs.go.tz/>.
- Clifton, C. & Frazier, L. (2004). Should given information come before new? Yes and no. *Memory and Cognition*, 32 (6). 886-895.

- Columnists, Daily News Online Edition*. Retrieved April 21, 2014 from <http://www.dailynews.co.tz/index.php/columnists/>.
- Davies, M. (2013). *Corpus of Global Web-based English (GloWbE)*. Brigham Young University, Utah. Retrieved March 30, 2014, from <http://corpus2.byu.edu/glowbe/>.
- De Cuypere, L. & Verbeke, S. J. (2013). Dative alternation in Indian English: A corpus based analysis. *World Englishes* (32) 2, 169-185.
- Denison, David. (2007). Playing Tag with category boundaries. In A. Meurman-Solin & A. Nurmi (Eds.). *Annotating Variation and Change*, vol.1
- Di Tullio, A. (2006). Clefting in Spoken Discourse. In K. Brown (Ed.). *The Elsevier Encyclopedia of Language and Linguistics, second edition* (p.483-491). Amsterdam: Elsevier.
- Farsi, S.S. (2013). *Swahili Idioms*. Nairobi: Kenya Literature Bureau. *Global Voices Online, 2015*. Retrieved April 6, 2015 from: <http://globalvoicesonline.org/2015/02/24/tanzanian-schools-will-teach-students-in-swahili-not-english/>
- Halliday, M. (1967). Notes on transitivity and theme in English, Part 2. *Journal of Linguistics*, 3, 199-244.
- Kortmann, B. (2006). Syntactic Variation in English: A Global Perspective. In B. Aarts & A. McMahon (Eds.), *The Handbook of English Linguistics* (p.604-624). Oxford: Blackwell.
- Kipacha, A. (2006). Tanzania: Language Situation. In K. Brown (Ed.), *The Elsevier Encyclopedia of Language and Linguistics, second edition* (p.502-505). Amsterdam: Elsevier.
- Lambrecht, K. (1994). *Information structure and sentence form*. Cambridge: Cambridge University Press.
- Marten, L. (2006). Swahili. In K. Brown (Ed.) *The Elsevier Encyclopedia of Language and Linguistics, second edition* (p.502-505). Amsterdam: Elsevier.
- McEnery, T., Xio, R & Tono, Y. 2006. *Corpus-Based Language Studies: An advanced resource book for students*. London: Routledge.
- Miller, J. (2006). Focus. In K. Brown (ed.) *The Elsevier Encyclopedia of Language and Linguistics, second edition* (p.511-518). Amsterdam: Elsevier.511-518.
- Mithun, M. (1992). Is basic word order universal? In D. Payne (Ed.) *Pragmatics of Word Flexibility* (p.15-62). Amsterdam: John Benjamins

- 'Ndugu',n. *OED Online*. June (2003). Oxford: Oxford University Press. Retrieved on March 17, 2014 from <http://www.oed.com/view/Entry/125523>.
- Ozón, G. (2006). *The Given Before New* principle and textual retrievability: A corpus-based study. In A. Renouf & A. Kehoe (Eds.), *The changing face of corpus linguistics* (p.243-262). Amsterdam: Radopi.
- Park, K. (2011). Information Structure and Dative Word Order in Adult L2 Learners. In J. Herschensohn & D. Tanner (Eds.) *Proceedings of the 11th Generative Approaches to Second Language Acquisition Conference* (p.101-109). Massachusetts: Cascadilla Press.
- Platt, J, Weber, H. and Lian H. (1984). *The New Englishes*. London: Routledge.
- Prince, E. (1981). Toward a Taxonomy of Given-Before-New Information. In P. Cole (Ed.) *Radical Pragmatics* (p.223-255). Michigan: University of Michigan Press.
- Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. (1985). *A Comprehensive Grammar of the English language*. London: Longman.
- Rubanza, Y. (1995). The use of English and Kiswahili in Tanzania: complement or conflict? *Journal of Linguistics and Language in Education Department of Foreign Languages and Linguistics* (1), 23-42.
- Rubanza, Y. (1996). Can a tri-partite language policy work in Tanzania? A new perspective. *Ufahamu: A Journal of African Studies* (24) 1, 1-16.
- Schmied, J. (1991). *English in Africa: An Introduction*. London: Longman.
- Schmied, Josef. (2006). East African Englishes. In B. Kachru, Y. Kachru & C. L. Nelson (Eds.) *The Handbook of World Englishes*. (p.188-202). Oxford: Wiley-Blackwell.
- Schmied, J. & Hudson-Ettle, D. (1999). *Manual to accompany The East African Component of The International Corpus of English (ICE EA)*. Retrieved on January 12, 2014 from http://khnt.hit.uib.no/icame/manuals/ice_ea.pdf.
- Schmied, J. & Hudson-Ettle, D. (1999). International Corpus of English, East African Component (ICE-EA). Chemnitz: University of Chemnitz.
- Schmied, J. & Hudson-Ettle, D. (2007). Analyzing the style of East African newspapers in English. *World Englishes*, 15 (1) 103-113.
- Sityaev, D. (2000). The relationship between accentuation and information status of discourse referents: A corpus-based study. *UCL Working Papers in Linguistics* (12) 1, 285-304.

- Stefanowitsch, A. & Gries, G. (2002). Givenness and Linear Precedence: A construction-grammar approach. *2nd International Construction Grammar Conference*. Retrieved on March 21, 2014 from <http://www.eng.helsinki.fi/janola/iccg2.htm>.
- Van Peer, W. (2006). Foregrounding. In K. Brown (Ed.) *The Elsevier Encyclopedia of Language and Linguistics, second edition* (p.546-551). Amsterdam: Elsevier.
- Vitale, Anthony. (1981). *Swahili Syntax*. Berlin: Walter de Gruyter.
- Ward, G., Birner, B & Huddleston, R. (2002). Information packaging. In G. Pullum and R. Huddleston (Eds.) *The Cambridge grammar of the English language* (p. 1363-1447). Cambridge: Cambridge University Press.
- Ward, G. & Birner, B. (1988). *Information Status and Non-Canonical Word Order in English*. Amsterdam: John Benjamins.
- Weir, A. (2009). *Article drop in English headlines*. MA Dissertation. University College London. Retrieved on April 9, 2014 from <http://people.umass.edu/aweir/papers.html>.