# ANALYTICITY AND SYNTHETICITY IN EAST AFRICAN ENGLISH AND BRITISH ENGLISH: A REGISTER COMPARISON

B.A.-Arbeit

zur

Erlangung des akademischen Grades

"Bachelor of Arts"

der Philologischen, Philosophischen und

Wirtschafts- und Verhaltenswissenschaftlichen Fakultät der

Albert-Ludwigs-Universität

Freiburg i.Br.

Vorgelegt von

Katharina Luisa Ehret

aus Freiburg i.Br.

SS08

English and American Studies

**Table of contents**

## 1. Introduction

In this paper East African, i.e. Kenyan and Tanzanian, English registers are compared to British English registers in terms of their *analyticity* and *syntheticity* using quantitative data from two different corpora namely, the International Corpus of English for East Africa (ICE-EA) and the British National Corpus (BNC).

Firstly, a brief outline of the historic development of the terms used in this paper as well as a literature review on the topic will be provided. I will then move on to give details on the data sources, retrieval and analysis method used for the current study. In section 2 the data will be analysed and the results interpreted. Lastly, I will give a short summary of the issues discussed and conclude with a final interpretation of my findings.

### 1.1. Literature review and historical overview

As the terms *synthetic* and *analytic* are used in a wide range of different meanings (cf. Anttila 1986: 15, Danchev 1992: 26), a brief overview of the historic development of the terms used here and their meaning in the context of this paper will be given.

Traditionally, Danchev writes in his paper "The evidence for analytic and synthetic developments in English" (1992), syntheticity and analyticity are used in a morphological context. He notes that it is important to maintain "the distinction between formal and semantic definitions of analyticity and syntheticity" (Danchev 1992: 26). He further explains that the formal definition – which is referred to in this paper – of analyticity implies two or more language units whereas the definition of syntheticity implies one unit (cf. Danchev 1992: 26). This means analytic constructions are manifested in at least two free morphemes whereas synthetic constructions are manifested in one bound morpheme.

Generally, morphological typology distinguishes three types of languages, namely *isolating*, *agglutinative* and *fusional* (cf. Comrie 1989: 42). Isolating describes languages in which morphology is practically non-existent as there is a one-to-one correspondence of words and morphemes, i.e. a word consists of only one morpheme. Agglutinative, on the other hand, involves words with more than one morpheme in which there is a clearly recognisable morpheme boundary; in fusional types, this

boundary is not clear-cut. Occasionally, Comrie writes, the fourth category *polysynthetic* or *incorporating*, describing a language type in which a large amount of morphemes is combined into a single word, is added to the three basic types (cf. Comrie 1989: 42-43).

The term 'synthetic' derives from this context of morphological typology and allows categorising languages on a scale from synthetic (isolating) to analytic (polysynthetic), whereas the index of synthesis basically indicates the number of morphemes per word.

Historically, as early as the eighteenth century, attempts at classifying types of languages have been made and in order to categorise languages into different types classification schemes like the one mentioned above were necessary.

According to Anttila (cf. Anttila 1989), Friedrich von Schlegel introduced a two-type classification dividing languages into either flective or affixive types, which were however, quite ambiguous and the terms were not used in the generally accepted way as flective described not only root modification but also proper inflection (including affixation). The term 'affixive' on the other hand included both particles and proper affixation. Von Schlegel understood these types as only an "outward sign of a fundamental distinction between organic form (inflection) and merely mechanical form" (Greenberg 1974: 37), i.e. affixation.

Later von Schlegel's brother added a third category, 'languages without grammatical structure' (cf. Anttila 1989: 310). This term corresponds to, and was later replaced by the term 'isolating'. He further split the inflectional type into the sub-categories synthetic and analytic. However, this differentiation merely represented two stages in the development of a given language, synthetic being the earlier stage and analytic the later language stage (cf. Greenberg 1974: 38).

Wilhelm von Humboldt extended Schlegel's threefold scheme for categorising languages with a forth category which he called *incorporating*. But his importance for the early typology of languages had been his system of language philosophy which saw languages as the manifestation of the spirit displaying different degrees of perfection, i.e. each language was seen as a realisation of this spirit referring to the "inner form" of a language (cf. Greenberg 1974: 38).

The model of a three-way division of languages has remained, the label of the categories, however, had been changed over time into isolating, agglutinating and flectional. Additionally to these classification categories Max Müller introduced the

terms synthetic and analytic describing the segmentability of morphological units which have been used simultaneously with the terms mentioned above. The same is valid for W.D.Whitney's term polysynthetic which he introduced for difficult cases in terms of segmentability (cf. Anttila 1989: 311).

Still, the criteria for morphological typology in the nineteenth century were not flexible enough and could not reliably classify language types as firstly there was a lack of reliable data and secondly languages had to be exclusively assigned to one category only which was basically impossible as a language can represent one or two of the features applied at the same time.

Edward Sapir developed a more flexible system for categorising languages into types by employing three independent parameters, each of these including several sub-categories, which could be combined with each other and were thus not mutually exclusive. He introduced four concepts as the first parameter: I Basic (concrete) concepts, II Derivational concepts, III Concrete relational concepts, IV Pure relational concepts, which are represented in four fundamental types of languages while the concepts I and IV have to be present in all of these types whereas the types II and III are optional. According to Greenberg, Sapir considered these concepts of his classification model "much deeper and more significant" (Greenberg 1974: 41) than the other two criteria because they would describe the "expression or non-expression of certain classes of concepts" (Greenberg 1974: 41). This point of view is related to von Humboldt's inner form of languages and his theory that a language is the expression of a people's spirit (cf. Greenberg 1974: 41).

In addition to these fundamental types, i.e. concepts, Sapir further established his second parameter, a degree of complexity or synthesis using the labels analytic, synthetic and polysynthetic in ascending order of complexity. The third parameter Sapir set up is technique, by which he means the way of construction (agglutinative, fusional, symbolic) or in the case of isolating, relation of words and morphemes (cf. Anttila: 312-315).

This model which is based on three parameters represented an improvement in categorising languages as they did not have to be exclusively assigned to one single category anymore, but allowed arranging languages on a comparative scale according to certain properties. Greenberg criticises Sapir's model because the method of placing

languages on the scale was unreliable and there was “no effective procedure” (Greenberg 1974: 23) for doing so.

Sapir’s model was extended to five criteria by Joseph Greenberg who gives a comprehensive description of his typological approach in his paper “A Quantitative Approach to the Morphological Typology of Language” (1960). Greenberg used roughly the same classification criteria as Sapir did, but he established (several) numerical indices for each of these criteria. Each feature was calculated by a ratio of two units based on the relative text frequency of these two units over a passage of 100 words. The degree of synthesis can thus be calculated as the ratio of morpheme/word. Greenberg set the lowest level of syntheticity at 1.00 as each word has to consist of at least one morpheme (cf. Greenberg 1960: 185). Greenberg, basing his typological research on these indices, chose to analyse languages which have frequently been cited as examples of certain language types in order to prove that his quantitative approach agrees with traditional nonquantitative judgements (cf. Greenberg 1960: 194). Applying his indices Greenberg showed how language types change over time, for example from a more synthetic to a more analytic language type in the case of Sanskrit and Anglo-Saxon which developed into modern languages such as, for instance, English. He concedes though that his results need to be confirmed by further data and that more languages should be analysed.

The indices used in this paper are roughly based on the above model by Greenberg and have been modified by Szmrecsanyi and Kortmann according to the purposes of their paper “Between simplification and complexification. Non-standard varieties around the world” (to appear). They used Greenberg’s method in order to investigate “morphosyntactic complexity variance” (Szmrecsanyi & Kortmann to appear: 2) in Englishes around the world and showed that variety type is the best predictor of a variety’s complexity by measuring four linguistic notions of complexity, among these grammaticity and redundancy for which they established the following three indices also used in the present study, which are based on a set of 76 linguistic features (Szmrecsanyi & Kortmann to appear: 2-4).

A *syntheticity index* representing the proportion of bound grammatical morphemes/markers, an *analyticity index* representing the number of free grammatical morphemes/markers and a *grammaticity index* which consists of the former two indices. All three indices are calculated per 1000 word tokens (cf. Szmrecsanyi & Kortmann to

appear: 8). They based their study on survey data of 46 non-standard varieties using the World Atlas of Morphosyntactic Variation in English and naturalistic data of 15 spoken varieties retrieved from corpora. More details concerning this study are given in section 2.3 where the grammaticity and complexity of spoken East African English in comparison to the varieties investigated by Szmrecsanyi and Kortmann will be analysed.

In another study by Szmrecsanyi the same indices have been applied to data – retrieved from the BNC – for the purpose of investigating intra-linguistic register variation in British English and aimed at analysing the connection between analyticity and syntheticity with different notions of complexity. The indices were interpreted in terms of complexity of the analysed registers as follows: the higher the analyticity index the more explicit – and to some extent easier – the grammatical structure becomes, i.e. the lower the reader/hearer complexity gets. On the other hand, the higher the syntheticity index, the higher reader/hearer complexity but at the same time the efforts for the speaker/writer are reduced. In short, the higher the syntheticity index the higher speaker/writer economy (cf. Szmrecsanyi submitted).

In the current paper I will investigate spoken and written registers of the ICE-EA and compare them to the corresponding registers of the BNC applying the indices based on Greenberg's typological classification model in the modified version from Szmrecsanyi and Kortmann to my corpus data. I will further discuss and seek to give an answer to the following research questions:

1. How the East African English registers under investigation differ in terms of their analyticity and syntheticity.
2. Whether and how East African English registers differ from the corresponding British English registers in terms of analyticity and syntheticity.
3. How spoken East African English fits in with Szmrecsanyi and Kortmann's comparison of Englishes around the world and where it is positioned within the other varieties of English in terms of analyticity and syntheticity as well as the related concepts of grammaticity and complexity.

### 1.2. Method and data

The data used in this study has been sampled from the East African sub-corpus of the International Corpus of English (ICE-EA). The East African sub-corpus comprises written and spoken material from two countries, Kenya and Tanzania, which has been collected between 1991 and 1996. Due to the special language situation in Tanzania, where written English is used far less frequently than in Kenya, the corpus contains two parallel text collections (of 400,000 tokens each), one for Kenyan and one for Tanzanian written English in order to represent both varieties appropriately (cf. Hudson-Ettle & Schmied 1999: 5-8). The written parts contain 200 texts each and are made up of 50 non-printed and 150 printed texts according to the ICE stipulations. However, minor modifications concerning the content of some written categories had to be made due to a lack of text availability, e.g. in Tanzania social letters are not written in English and could therefore not be included (cf. Hudson-Ettle & Schmied 1999: 8).

This difference in the use of English certainly derives, at least to some extent, from the colonial situation of the two countries: Kenya was a British protectorate and colony for more than half a century whereas Tanzania belonged to the German Empire and only later, in 1920, became a British mandate. But even then, due to its uncertain legal status, the British influence and thus the impact on and development of English in Tanzania was not as strong as in Kenya (cf. Schmied 1985: 28-29). The influence of this time on the use of English today can also be seen from the fact that Swahili is the only official language in Tanzania whereas both Swahili and English have the status of official languages in Kenya (cf. Schmied 1990: 218).

The spoken material of the East African corpus has been collected in a joint corpus and set up in accordance with the ICE guidelines as closely as possible. It is, however, reduced in size for lack of data on informal and other conversation categories; 120 monologues are also included, yet they are scripted and not truly spoken material (cf. Hudson-Ettle & Schmied 1999: 6-7).

Another special property of the spoken East African corpus is the category ‘written as spoken’ which contains, for instance, written recordings of court procedures or parliamental records as it was impossible to sample truly spoken material (cf. Hudson-Ettle & Schmied 1999: 8).

For the current study, which investigates 11 registers of the ICE-EA, three spoken and nine written registers, the written sub-corpora have been merged and equal samples from both varieties have been taken. These registers have been matched with corresponding registers of the BNC (for details on the name of registers, their content and composition see table 1 and 2).

| **Registers** | **ICE-EA** | **BNC** |
|---|---|---|
| Private | conversation, interviews oral narratives (s1a) | conversations, interview, interview oral history |
| Public | class lessons, broadcast discussion & interviews (s1b) | parliament, courtroom, classroom, broadcast discussion, public debate |
| Scripted | broadcast news & talks, speeches and lectures (s2b) | broadcast news & documentary, speech scripted |

Table 1: Corresponding spoken registers and their content in ICE-EA and BNC

The results from the BNC, as used here, have been established by Szmrecsanyi in his paper on "Typological parameters of intralingual variability: grammatical analyticity vs. syntheticity in varieties of English" (submitted). The private, public and scripted registers are spoken material, all the other registers under investigation are written.

| **Registers** | **ICE-EA** | **BNC** |
|---|---|---|
| Student Writing | timed & untimed essays (w1a) | university & school essays |
| Letters | social & business letters (w1b) | private & professional letters |
| Academic | academic writing & papers (w2a) | academic writing |
| Popular | popular writing (w2b) | popular writing |
| Reportage | newspaper articles (w2c) | newspaper articles |
| Instructional | administrative & regulatory (w2d) | instructional |
| Persuasive | institutional & personal columns (w2e) | institutional |
| Creative | novels & short stories (w2f) | fictional |

Table 2: Corresponding written registers and their content in ICE-EA and BNC

Methodically, the data under investigation was retrieved by an automated algorithm which randomly selected 1000 tokens of the ICE-EA per register so that the total data set contains 11 registers with 1000 tokens each, adding up to a set of 11,000 tokens of East African English.

The data has then been tagged manually using the BNC coding scheme for POS-tags (cf. Aston & Burnard 1998: 230-233) plus four additional tags, distinguishing between auxiliary use and use as full verb of the verbs *have*, *be* and *do*. Below, examples from the corpus illustrate the auxiliary use (1a-3a) of the verbs *have*, *be* and *do* as well as their use as full verbs (1b-3b).

(1) a I have forgotten (ICE-EA, s1b)

b We have cults (ICE-EA, s1b)

(2) a The child is encouraged (ICE-EA, s1b)

b There is suspicion of murder (ICE-EA, s1b)

(3) a A farmer does not forget (ICE-EA, w2b)

b In determining what unions do (ICE-EA, w2b)

The BNC POS-tags relevant for this study have been categorised as either being analytic (table 3) or synthetic (table 4) or, in some cases, both analytic and synthetic (table 5) so that the sources of variation in analyticity and syntheticity can be traced. Purely lexical tags, for instance tags marking proper nouns, have been ignored as they are not relevant for this analysis.

| **Analytic tags** | **Part of Speech** | **Example** |
|---|---|---|
| AT0 | Article | *the, a, an , no* |
| AVQ | Wh-adverb | *when, how, why* |
| CJC | Coordinating conjunction | *and, or, but* |
| CJS | Subordinating conjunction | *because, when* |
| CJT | Subordinating conjunction *that* | |
| DPS | Possessive determiner-pronoun | *your, their* |
| DT0 | General determiner-pronoun | *this, that* |
| DTQ | Wh-determiner-pronoun | *which, what* |
| EX0 | Existential *there* | *there is, there are* |
| PNI | Indefinite pronoun | *none, nobody* |

| | | |
|---|---|---|
| PNP | Personal pronoun | *he, you* |
| PNQ | Wh-pronoun | *who, whom* |
| PRF | Preposition *of* | |
| PRP | Preposition (except for *of*) | *about, in, on* |
| TO0 | Infinitive marker *to* | |
| AVB* | Auxiliary *be* | *be* |
| AVD* | Auxiliary *do* | *do* |
| AVH* | Auxiliary *have* | *have, 've* |
| XX0 | Negative particle *not* or *n't* | |

Table 3: Analytic POS-tags

| **Synthetic tags** | **Part of Speech** | **Examples** |
|---|---|---|
| AJC | Comparative adjective | *better, older* |
| AJS | Superlative adjective | *best, oldest* |
| NN2 | Plural common noun | *pencils, geese* |
| POS | Possessive/genitive marker *'s* or *'* | |
| VBB | Present tense forms of the verb *be* (except for *is, 's*) | *am, 'm, are, be* (subjunctive or imperative) |
| VBD | Past tense forms of the verb *be* | *were, was* |
| VBG | -ing form of the verb *be* | *being* |
| VBN | Past participle form of the verb *be* | *been* |
| VBZ | -s form of the verb *be* | *is, 's* |
| VHD | Past tense form of the verb *have* | *had, 'd* |
| VHG | -ing form of the verb *have* | *having* |
| VHN | Past participle form of the verb *have* | *had* |
| VHZ | -s form of the verb *have* | *has, 's* |
| VVD | Past tense form of lexical verbs | *sang, forgot* |
| VVG | -ing form of lexical verbs | *singing, forgetting* |
| VVN | Past participle form of lexical verbs | *sung, forgotten* |
| VVZ | -s form of lexical verbs | *sings, forgets* |

Table 4: Synthetic POS-tags

| Analytic-synthetic tags | Part of speech | Example |
|---|---|---|
| PNX | Reflexive pronouns, plural form | *myself, ourselves* |
| AVBD | Past tense forms of the verb *be*, auxiliary use | *was, were* |
| AVBG | -ing form of the verb *be*, auxiliary use | *being* |
| AVBN | Past participle form of the verb *be*, auxiliary use | *been* |
| AVBZ | -s form of the verb *be*, auxiliary use | *is, 's* |
| AVDD | Past tense form of the verb *do*, auxiliary use | *did* |
| AVDG | -ing form of the verb *do*, auxiliary use | *doing* |
| AVDN | Past participle form of the verb *do*, auxiliary use | *done* |
| AVDZ | -s form of the verb *do*, auxiliary use | *does, 's* |
| AVHD | Past tense form of the verb *have*, auxiliary use | *had, 'd* |
| AVHG | -ing form of the verb *have*, auxiliary use | *having* |
| AVHN | Past participle form of the verb *have*, auxiliary use | *had* |
| AVHZ | -s form of the verb *have*, auxiliary use | *has, 's* |
| VM0 | Modal auxiliary verb | *will, could, can, 'll, 'd* |

Table 5: Analytic-synthetic POS-tags

The tagged ICE-EA registers have then been subjected to a computerised analysis – implemented with the perl programming language – which retrieved the text frequencies of the above-mentioned POS-tags and returned values for the following three indices: syntheticity (being the ratio of bound grammatical markers per 1000 words), analyticity (the ratio of free grammatical markers and function words per 1000 words) and grammaticity which represents the sum of the syntheticity and analyticity indices (cf. Szmrecsanyi & Kortmann to appear: 8). An example for a bound grammatical marker would be the genitive marker *'s*, whereas the preposition *of* as used in genitive constructions would be a free grammatical marker. Function words include prepositions, pronouns, conjunctions, auxiliary verbs and existential *there*.

In a final step, the indices obtained for each of the ICE-EA registers have been manually compared with one another as well as to each of the indices of the BNC registers. Hereby, the results were subjected to statistical analysis of significance conducting chi-square tests with an online chi-square calculator (cf. Preacher 2001) for all of the indices. The threshold for statistically significant results has been set to a p-value of $p \leq 0.05$ for the intervariety comparison, i.e. the comparison of East African with British English registers, and a p-value of $p \leq 0.1$ for the comparison of the various ICE-EA registers, as the generally accepted threshold value for p would otherwise have led to a lack of significant differences in this dataset.

## 2. Analyticity and syntheticity: A comparison of East African and British English

### 2.1. Analyticity and syntheticity variance in East African English registers

Comparing East African registers with one another in regard to analyticity and syntheticity (for details see table 6), one can observe that the written creative register is the most analytic of all the registers whereas the written register reportage is the least analytic; the spoken registers private and public can also be counted among the most analytic registers. In terms of syntheticity, it is exactly the other way around: the creative and private registers are the least synthetic whereas reportage is the most synthetic register.

The other registers, all of them written except for the scripted register, are to be placed in a mid-range position for both analyticity and syntheticity.

| ICE-EA Registers | Analyticity | Syntheticity | Grammaticity |
|---|---|---|---|
| Private | 490 | 129 | 619 |
| Public | 491 | 152 | 643 |
| Scripted | 440 | 141 | 581 |
| Student Writing | 445 | 142 | 587 |
| Letters | 426 | 120 | 546 |
| Academic | 428 | 145 | 573 |
| Popular | 421 | 149 | 570 |
| Reportage | 408 | 183 | 591 |
| Instructional | 417 | 128 | 545 |
| Persuasive | 458 | 143 | 601 |
| Creative | 506 | 137 | 643 |

Table 6: Analyticity, syntheticity and grammaticity according to register

The creative register is, along with the public and private registers, the most analytic register (see figure 1). It has the highest frequency of analytic markers and is the most grammatically complex of the written registers under investigation. This is, however, not surprising as written language generally becomes more grammatically complex and latinised the more formal the register gets. What is surprising though, is that the academic register displays, at a p-value of 0.04, a significantly lower index for analyticity and a marginally significantly lower index for the overall grammaticity ($p$ = 0.1) than the creative register: normally, one would expect academic writing to show

more grammatical marking due to its required formality and scientific standards. On the other hand, creative writing is the most elaborate and detailed form of writing and does not suffer from restrictions such as, for instance, a limitation of the number of characters as might be required by scientific standards.

This difference in analyticity between the creative and academic register can be ascribed to several sources of which the use of pronouns is at $p = 0$ certainly the most significant one. Indefinite, personal, reflexive and wh-pronouns are six times more frequent in the creative register than in the academic one. Other sources for analyticity variation are the less frequent use of the auxiliary verb *have* and *more/most*, negative particles and conjunctions in descending order of significance in the academic register.

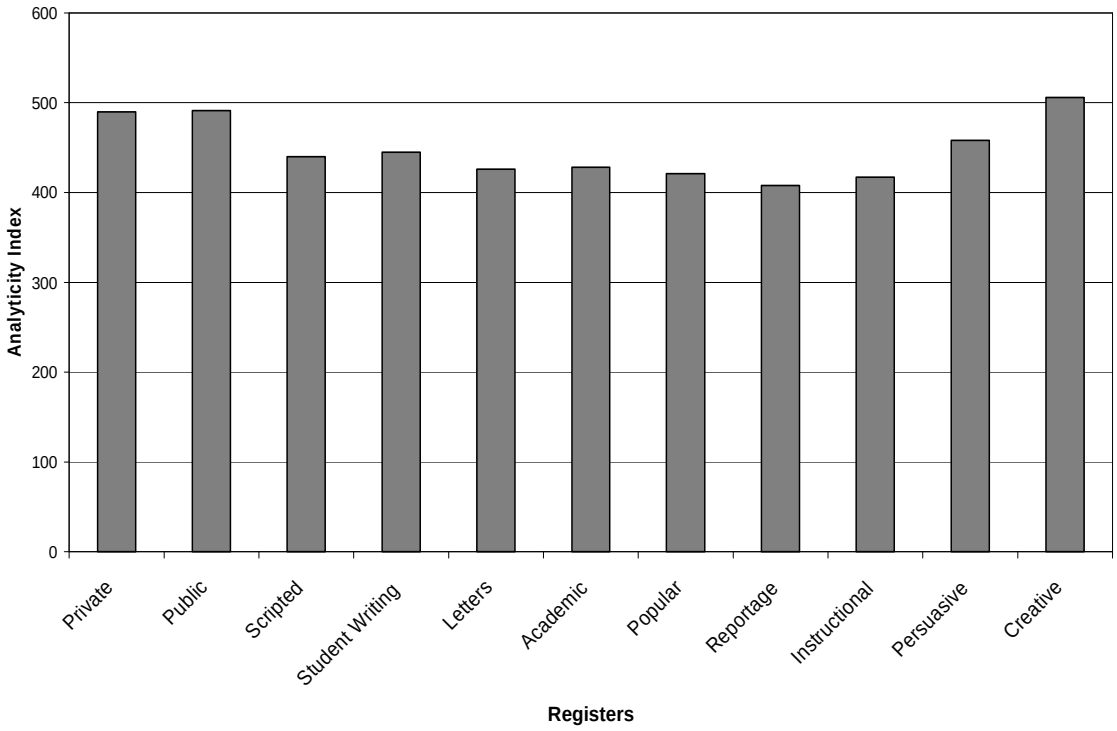


Figure 1: Analyticity by ICE-EA register

Another interesting observation is that the spoken registers public and private are also highly analytic and even significantly more analytic than academic writing at a p-value of 0.09 or, for example the written register reportage at $p = 0.02$. However, considering the fact that East African English is a non-native L2 variety, the findings appear logical: in a non-native language, especially in its spoken registers, speakers may be likely to avoid synthetic constructions and opt for analytic markers instead, in order to make the

language and the constructions more transparent to the benefit of the listener. This ties in neatly with what Danchev writes in his article on analytic and synthetic developments in English: “Analyticity facilitates comprehension, i.e. the listener [comprehension], whereas syntheticity enhances economy and expressivity, i.e. it serves the speaker” (Danchev 1992: 36).

The L2 status of East African English and the relation between overall grammaticity and language complexity would also explain why the grammaticity index is – along with the creative register – the highest in these registers: While the absolute complexity, i.e. the purely grammatical complexity as the sum of the synthetic and analytic indices, rises, pragmatic hearer-centred complexity drops (cf. Szmrecsanyi submitted). This means that the more grammatical markers, the more obvious and explicit the relations between the parts of the sentences become and the easier it is for the hearer to process the information.

Yet, this explanation does not serve to explain the high grammaticity index in the creative register, as one should be able to assume that readers of fiction are proficient in the language they read at their leisure and do not rely on easy comprehension. Literary standards may, however, require writers to use analytic markers instead of synthetic ones.

The most striking difference in the use of analytic markers can be observed between the creative register and reportage: at a p-value of 0.007 reportage is highly significantly less analytic due to the much more frequent use of indefinite, personal, reflexive and wh-pronouns as well as the auxiliary verb *have* and the existential *there* in the creative register.

All the other registers, except for the persuasive register, have a significantly lower analyticity index than the creative, private and public registers, yet the differences are not as striking as the ones mentioned above.

The creative register, i.e. the register with the highest analyticity index, has at the same time one of the lowest syntheticity indices and the register reportage with the lowest analyticity index has the highest syntheticity index. This is surprising as I would have expected – according to Szmrecsanyi and Kortmann’s paper (to appear) – that, in analogy to their finding that “a variety which is comparatively analytic is also comparatively synthetic” (Szmrecsanyi & Kortmann to appear: 10), the same could be

said for different registers. However, this assumption is not borne out by my data and is apparently not valid for different registers of the same variety.

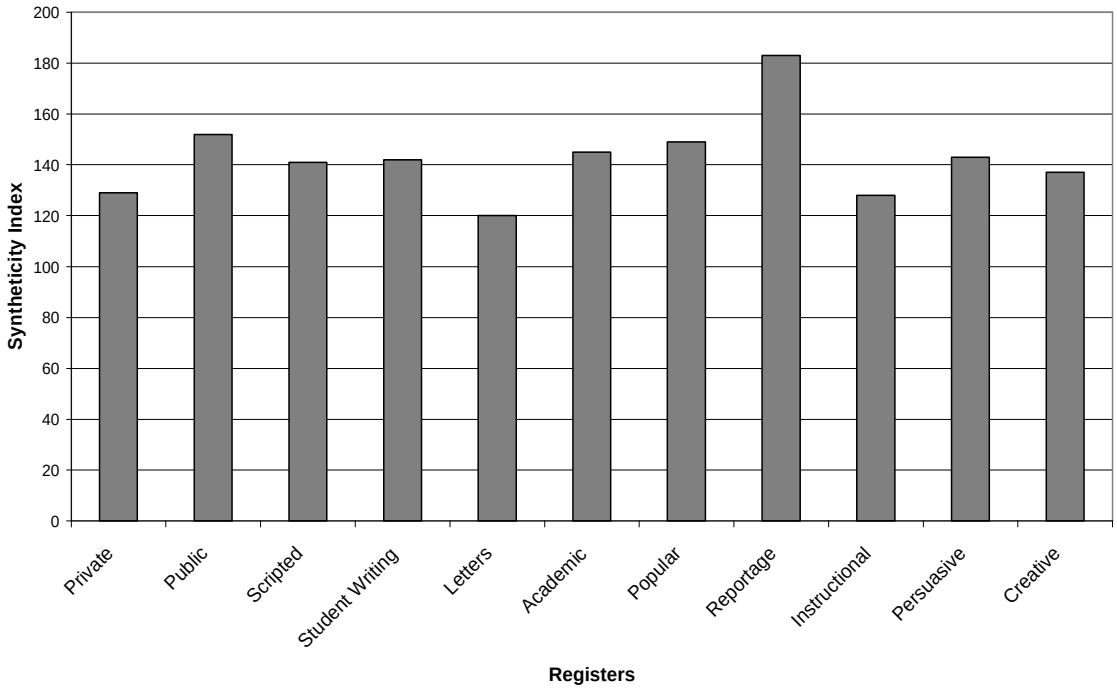


Figure 2: Syntheticity by ICE-EA register

Reportage is, as stated above, the register with the highest value for syntheticity, probably because the number of characters is heavily restricted in journalistic writing and a lot of synthetic constructions are used for writer economy.

The most significant difference (at $p = 0$) in terms of syntheticity can be found between the registers reportage and letters, which has the lowest syntheticity index (see figure 2 for visualisation). This might be due to the informal character of the language used in letters which is closer to spoken language than any other written register. This would also explain why analytic markers are relatively high in the register letters. It is also interesting to note that the analyticity indices for letters and academic writing are almost identical.

The syntheticity variation can be ascribed to the less frequent use of plural noun markers and verb inflection in the register letters: according to the data, fourteen times more past tense forms of the verb *be* and three times more past tense forms of lexical verbs occur in reportage than in letters. This difference influences the overall syntheticity of both registers so that reportage is more synthetic – despite the fact that

letters is more synthetic in regard to its use of superlative and comparative adjectives than reportage.

The private and instructional registers are also very significantly less synthetic than reportage: the private register at a p-value of 0.004 and the instructional register at a p-value of 0.003. In the case of the private register this difference is due to the fact that at a p-value of $p$ = 0.00002 very significantly less plural noun markers occur than in reportage. The private register also uses less s-genitive markers as well as less verb inflections and comparative/superlative forms of adjectives, however, these differences are not as significant as the difference in plural noun marking.

In the instructional register the main source of variation is not plural noun marking but verb inflection: reportage uses 62% more verb inflections than the instructional register. The use of 19 times more past tense forms of lexical verbs and almost five times the number of past tense forms of the verb *be* can be made responsible for this variation.

The registers creative at $p$ = 0.2, scripted and student writing at $p$ = 0.03 also use significantly less synthetic markers than the register reportage. The creative register uses significantly less plural noun marking and comparative/superlative adjectives than reportage. In the scripted register the difference in syntheticity clearly derives from the variation in verb inflection: the number of past tense forms of *be* and lexical verbs as well as the number of –ing forms of lexical verbs is two to three times higher in reportage. In contrast to the other registers, no major source can be made responsible for the variation in student writing but all four synthetic categories, s-genitive, plural noun markers, verb inflection and comparative/superlative adjectives contribute to the difference in syntheticity.

As for the overall grammaticity (for reference see figure 3), the creative and public registers occupy the top end of the scale, followed by the private register.

The other written registers as well as the scripted register show overall less grammatical marking, which again I find unexpected because normally written language tends to be a lot more grammatically complex in terms of markers. Generally, I would expect spoken language to be more negligent in regard to grammatical marking than written language which is here – apart from the creative register – obviously not the case as all the other written registers show a substantially lower index of grammaticity than the two spoken registers private and public whereas the written creative register is an exception.

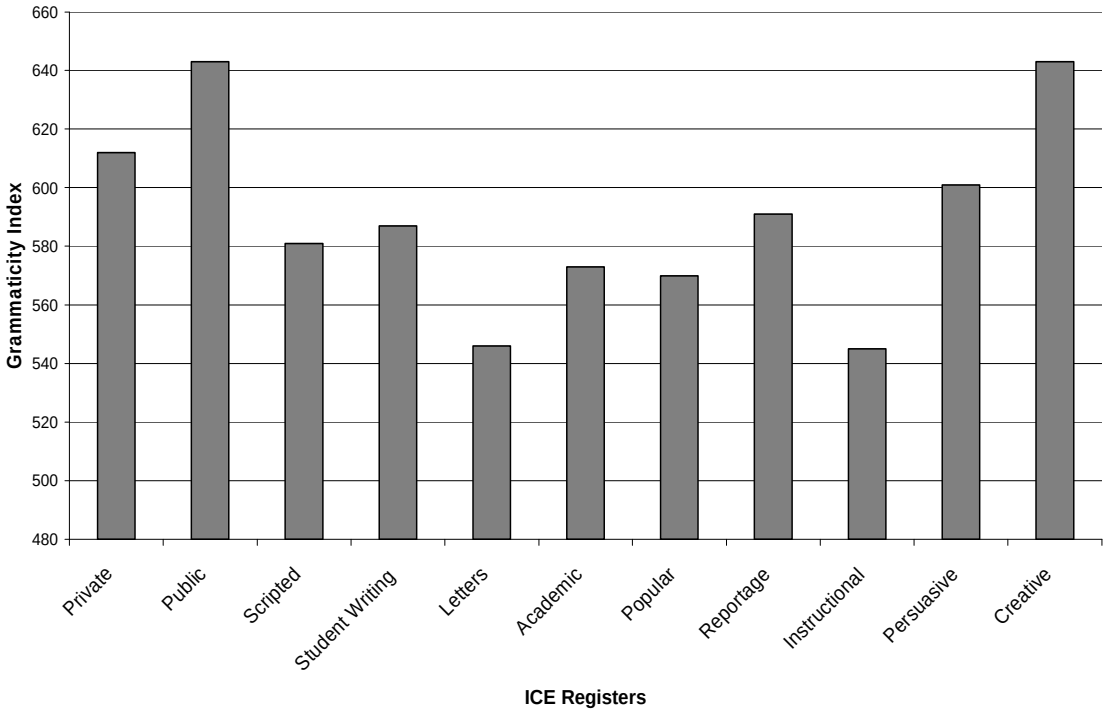


Figure 3: Grammaticity per ICE-EA register

This could, again, be explained by the status of English in Kenya and Tanzania, which is that of a second language and not a mother tongue: more grammatical markers are used in discourse in order to facilitate listening.

Having a closer look at the spoken registers it is not surprising that the public register has the highest grammaticity index because public language is obviously used in more official situations and therefore required to be generally more precise and correct in the use of grammatical markers. This formality is not required in everyday private conversations, thus the relatively lower index of grammaticity in the private register.

Both registers, letters and instructional, have the lowest indices for overall grammaticity; the register letters with a p-value of $p = 0.03$ and the instructional register at $p = 0.02$. These differences to the creative and public registers represent the most significant variation of grammaticity.

2.2. Comparison of registers across varieties in terms of analyticity and syntheticity

In this section the differences in analyticity and syntheticity between the corresponding registers across the two varieties, East African and British English, are analysed (for details see table 7).

Generally, one can say that the two varieties differ less than one would expect them to; there is hardly any difference in regard to analyticity (figure 5) and overall grammaticity (appendix B, figure 7). In terms of syntheticity (figure 4) only in six out of eleven registers – all of them written – a statistically significant difference can be observed. In the following I shall discuss these variances between the corresponding registers in descending order of significance.

| **Register** | **Analyticity** | | **Syntheticity** | | **Grammaticity** | |
|---|---|---|---|---|---|---|
| | ICE-EA | BNC | ICE-EA | BNC | ICE-EA | BNC |
| Private | 490 | 500 | 129 | 150 | 619 | 650 |
| Public | 491 | 485 | 152 | 143 | 643 | 628 |
| Scripted | 440 | 477 | 141 | 155 | 581 | 632 |
| Student Writing | 445 | 455 | 142 | 189 | 587 | 645 |
| Letters | 426 | 444 | 120 | 145 | 546 | 589 |
| Academic | 428 | 437 | 145 | 179 | 573 | 617 |
| Popular | 421 | 414 | 149 | 172 | 570 | 587 |
| Reportage | 408 | 399 | 183 | 182 | 591 | 582 |
| Instructional | 417 | 417 | 128 | 162 | 545 | 579 |
| Persuasive | 458 | 408 | 143 | 193 | 601 | 600 |
| Creative | 506 | 481 | 137 | 188 | 643 | 668 |

Table 7: Analyticity, Syntheticity, Grammaticity by register across varieties

The most conspicuous and eye-catching observation can be made within the written register creative. The East African variety is less synthetic than the British variety: at $p = 0.0005$ British English uses very significantly more synthetic markers than East African English. There is no significant difference in analyticity and overall grammaticity. However, East African English tends to be slightly less grammatically complex than British English which suggests, as the analyticity index is not significantly higher in East African English, that zero marking might be preferred to synthetic markers.

At $p$ = 0.0004 the major difference in syntheticity can be attributed to verb inflection: 185% less s-forms and 68% less past participle forms occur in East African English. S-genitive and plural noun markers are also slightly less frequent in East African English therefore British English is overall more synthetic – although comparative/superlative adjectives are about four times more frequent in East African than in British English.

In the persuasive register, East African English is at $p$ = 0.0008 far less synthetic than the British variety. At the same time, the persuasive East African register – as the only register with a statistically significantly higher analyticity index – uses at a p-value of $p$ = 0.04 significantly more analytic markers meaning that the overall grammaticity of the two varieties is the same. It is thus likely that in East African English, due to its status as L2 variety, analytic markers are preferred to synthetic markers in order to enhance reader comprehension. Six out of ten analytic POS-tags (for reference see table 8 and corpus examples 4-9 below) show significantly higher values in the East African persuasive register. The main source for the higher analyticity is at a p-value of $p$ = 0 the extensive use of indefinite, personal, reflexive and wh-pronouns in East African English. Other statistically significant sources are wh-adverbs which occur 29% more often in East African English as well as negative particles with 18% and the infinitive marker *to* with 59% higher frequencies.

| **Analytic POS-tags** | **ICE-EA** | **BNC** |
|---|---|---|
| AVQ | 172 | 133 |
| EX0 | 3 | 2 |
| PNI/PNP/PNQ/PNX | 49 | 15 |
| TO0 | 25 | 16 |
| VM0 | 18 | 14 |
| XX0 | 8 | 3 |

Table 8: Normalised values of analyticity POS-tags in the persuasive register

(only tags which show significant differences are shown)

(4) Why-AVQ has Ford taken so long to call (ICE-EA w2e)

(5) There-EX0 is the clamour for power (ICE-EA w2e)

(6) One-PNI wonders why (ICE-EA w2e)

(7) No one bothers to-TO0 put coins in the metres (ICE-EA w2e)

(8) Pilau may-VM0 be sweet (ICE-EA w2e)

(9) We did not-XX0 because it was for free (ICE-EA w2e)

The difference in syntheticity can mainly be ascribed to plural noun marking: almost 80% more markers are used in British English than in the corresponding East African register. Comparative/superlative adjectives are the second major source for the analyticity variance: British English uses almost five times as many as East African English in which comparative/superlative adjectives cannot be attested for at all. There are also minor differences in the use of s-genitive markers and verb inflection, yet these are not statistically significant.

In the register student writing, no significant differences in analyticity or overall grammaticity can be observed. However, East African English is at $p = 0.001$, like most of the other written registers, significantly less synthetic than British English. Although, three out of four synthetic POS-tags (for reference see table 9 and corpus examples10-13 below) are lower in East African English, only the difference in verb inflection is of statistic significance and can thus clearly be made responsible as the main source for this variation in syntheticity: in East African English at $p = 0.001$ very significantly less past tense forms are used than in British English. The different frequencies of past participle, -s and –ing forms in East African and British English are not of statistic significance.

| **Synthetic POS-tags** | **ICE-EA** | **BNC** |
|---|---|---|
| POS | 2 | 2 |
| AJS/AJC | 2 | 4 |
| NN2 | 53 | 64 |
| inflVERBS | 71 | 92 |

Table 9: Normalised values of syntheticity POS-tags in the register student writing

(10) Civilization is attained by man's-POS endeavour (ICE-EA w1a)

(11) One of the earliest-AJS empires (ICE-EA w1a)

(12) The totality of the authorities-NN2 (ICE-EA w1a)

(13) This practice poses-inflVERBS a threat (ICE-EA w1a)

In the instructional register another interesting observation can be made: the analyticity indices of the two varieties are identical. This observation along with the significantly lower syntheticity index and the slightly lower grammaticity index in East African English leads to the conclusion that in the instructional register, writers do not opt for the more transparent analytic markers but zero marking might be preferred instead.

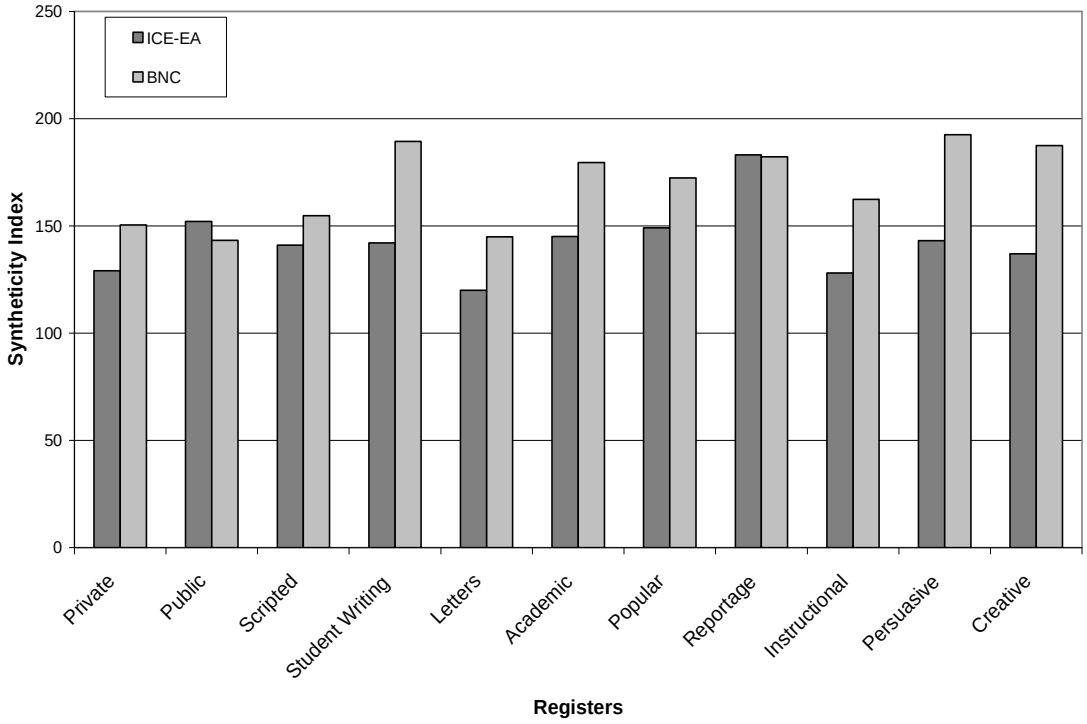


Figure 4: Syntheticity in British and East African English registers

At $p$ = 0.01 the instructional register in East African English is less synthetic due to significantly more frequent verb inflection in the corresponding British English register: British English uses 60.5% more s-forms than East African English. The value for past tense inflection is also slightly higher in the British variety, so that British English is overall more synthetic, although East African English shows higher frequencies of past participle and –ing forms. Other sources for the syntheticity variance are plural noun markers, comparative/superlative adjectives and s-genitive markers which occur slightly more often in the British English variety than in the East African one.

The East African academic register displays at $p$ = 0.01 a significantly lower index for syntheticity than the corresponding British register. There are no statistically significant

differences in analyticity and grammaticity to be observed, but the East African academic register is slightly less analytic and less grammatically complex. The lower syntheticity in East African English can be ascribed to one major source: verb inflection. At $p$ = 0.00004 East African English uses highly significantly less verb inflection than British English. The main difference is in the use of past tense markers which occur 139% more often in British English. Inflections for past participle forms are 54% and s-forms are 56% more frequent in British English. In the academic register East African English shows a slightly higher frequency of plural noun markers than the English variety, although this variance does not influence the overall syntheticity, it is nevertheless interesting to note as in all the other written registers plural noun marking was substantially lower in East African English.

In the register letters, the least difference in syntheticity can be observed; East African English is at a p-value of $p$ = 0.05 significantly less synthetic than British English. As in most of the other registers, the main sources for the syntheticity variance are verb inflection and plural noun marking. East African English letters use 29% less plural noun markers and 18% less verb inflections than British English. The difference in verb inflection can mainly be ascribed to the less frequent use of past tense markers and s-forms in East African English. Interestingly, East African English uses more comparative/superlative adjectives in this register, however, this difference is not significant enough to influence the overall syntheticity.

In the popular register an only marginally significant difference in syntheticity can be observed thus that East African English is less synthetic than the corresponding British English register. This is due to the higher frequency of verb inflection in British English. The values for analyticity and grammaticity are slightly lower in the East African register than in the BNC.

The register reportage is the only written register with no significant variation in syntheticity and above this the only register in which the indices for syntheticity are identical across the two varieties. This can be explained with the close connection between syntheticity and writer economy: in journalistic writing the number of characters is heavily restricted thus synthetic markers, which contribute to writer economy, are predominant in both varieties. The indices for analyticity and grammaticity do not show any significant differences.

So far, only the differences between the corresponding written registers have been discussed. The main reason for this is that there are no statistically significant observations to be made between the spoken registers. Nevertheless, a brief discussion of the spoken registers shall be given.

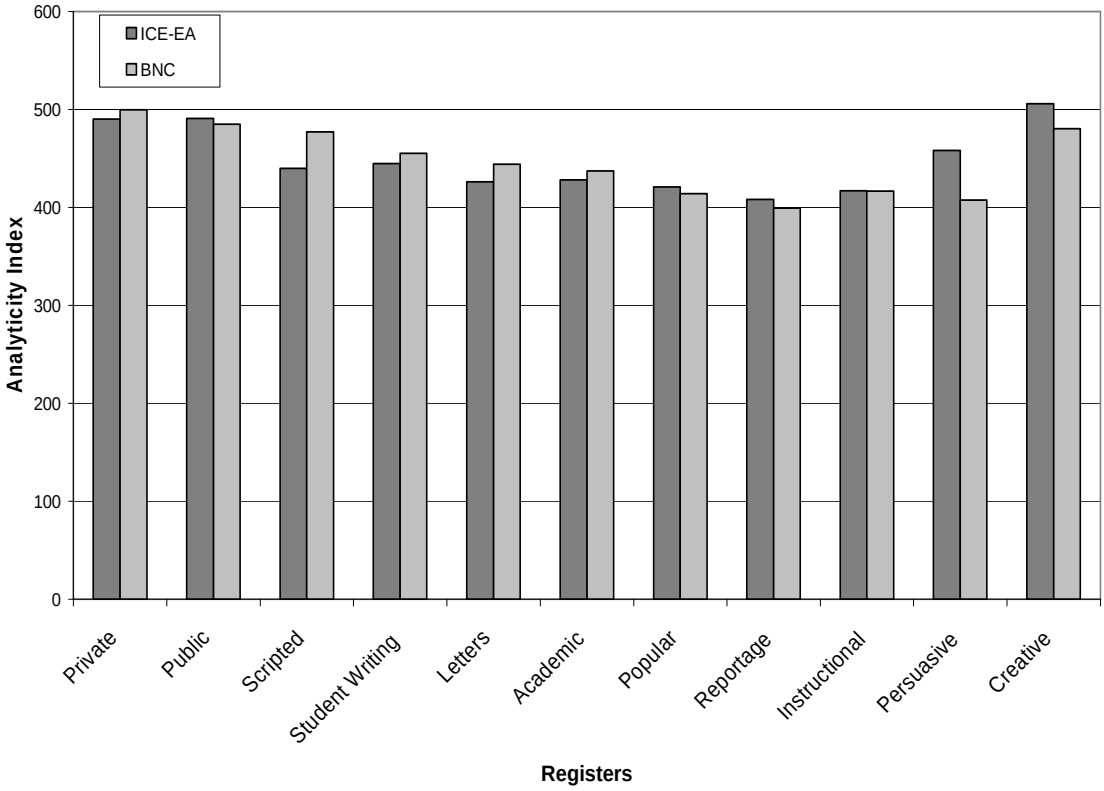


Figure 5: Analyticity in British and East African English registers

The most interesting of the spoken registers are the public and private registers in which overall no major differences can be observed. But in the public register the values for the BNC are consistently, although not considerably, lower than for the ICE-EA. The only reason for this could be that the formality in public discourse in terms of grammar is the same in both varieties.

In the private register it is the other way around: the values for the ICE-EA are consistently lower than for the BNC. This can be explained by the status of the two varieties. As British English is a native variety there is no need for an explicit and transparent structure as listener comprehension can be taken for granted in native communication, but in East African English, zero marking might be preferred in order to facilitate comprehension thus the overall lower indices.

In the scripted register, overall less grammatical markers are used in the East African variety, though, the grammaticity index is only marginally significantly lower. This is probably due to the lower indices for syntheticity and analyticity in East African English compared to British English.

## 2.3. Grammaticity and complexity variance in spoken Englishes

This section is based on the paper "Between simplification and complexification. Non-standard varieties around the world" by Szmrecsanyi and Kortmann (to appear) and is concerned with the relation between syntheticity and analyticity in regard to a variety's grammatical complexity. Spoken East African English will be compared to the 15 spoken varieties analysed by Szmrecsanyi and Kortmann. For the purpose of this comparison of spoken English varieties, the average values of the two truly spoken registers, public and private, from the East African corpus have been used. Table 10 lists the 15 varieties investigated by Szmrecsanyi and Kortmann plus spoken East African English, detailing type and the corresponding indices for syntheticity and analyticity:

| Variety | Type | Analyticity | Syntheticity |
|---|---|---|---|
| collAmE | high-contact L1 | 453 | 91 |
| collBrE | high-contact L1 | 471 | 114 |
| EAE | L2 | 491 | 141 |
| HKE | L2 | 414 | 67 |
| IndE | L2 | 462 | 99 |
| JamE | L2 | 454 | 95 |
| Mid | traditional L1 | 474 | 137 |
| N | traditional L1 | 473 | 139 |
| NIrE | high-contact L1 | 461 | 110 |
| NZE | high-contact L1 | 445 | 122 |
| PhilE | L2 | 459 | 102 |
| ScH | traditional L1 | 472 | 135 |
| SE & EA | traditional L1 | 517 | 120 |
| SgE | L2 | 457 | 107 |
| SW | traditional L1 | 457 | 103 |
| WelE | high-contact L1 | 464 | 131 |

Table 10: Spoken English varieties their type, analyticity and syntheticity

Szmrecsanyi and Kortmann are interested in, as already mentioned in the introduction, in how far variety type is related to morphosyntactic complexity. For their research they distinguish four types of spoken varieties of which the following three are relevant for this paper:

1. Traditional, low-contact L1
2. High-contact L1
3. Non-native L2

In order to investigate the relation between grammaticity and complexity several notions of complexity have been established. However, I will for reasons of relevance only focus on one of these notions, namely grammaticity and redundancy. Grammaticity is the overall frequency of grammatical markers whereby synthetic and analytic markers are distinguished. Szmrecsanyi and Kortmann equate this concept of grammaticity with redundancy, i.e. the repetition of grammatical information.

From the analysis of 15 spoken varieties, a hierarchical structure arises in regard to the overall grammaticity level: traditional L1 varieties have been found to be the most verbose type, i.e. most grammatically redundant, whereas non-native L2 varieties are the least verbose. High-contact L1 varieties are to be placed in the middle (cf. Szmrecsanyi & Kortmann to appear: 9).

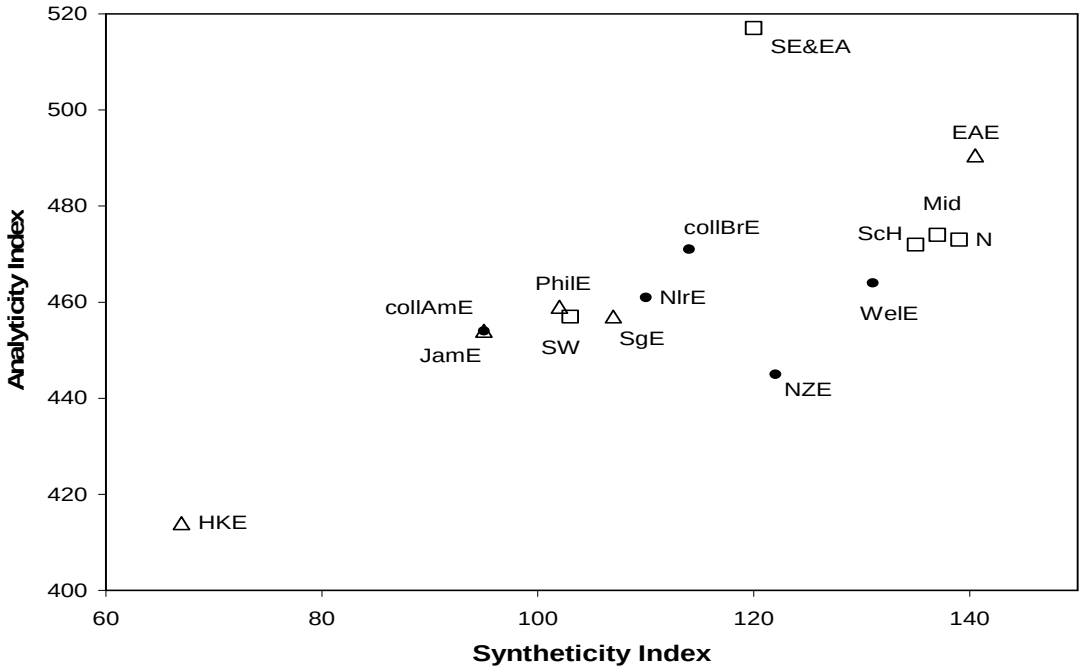


Figure 6: Analyticity by Syntheticity

Figure 6 shows the distribution of the varieties investigated by syntheticity and analyticity. East African English being positioned in the top right corner of the scatter plot shows L2-atypical behaviour: all L2 varieties, apart from Hong Kong English, are in the lower middle section of the plot and are neither particularly synthetic nor analytic. However, East African English and Hong Kong English could be seen as positioned on a virtual diagonal line, though on opposite ends, both varieties marking extreme cases: Hong Kong English is the L2 variety with the lowest indices for both syntheticity and analyticity which means it also has the lowest overall grammaticity index and is the least redundant variety. East African English, on the contrary, has the highest syntheticity index of all varieties including traditional and high-contact L1 varieties. Its analyticity index is also one of the highest, only Southeast/East Anglia English is more analytic, so overall, it is extremely verbose, i.e. grammatically redundant.

Szmrecsanyi and Kortmann showed that low-contact, traditional L1 varieties are more complex than high-contact L1 varieties proving that contact seems indeed to have a great impact on the development of a variety's complexity (cf. Szmrecsanyi & Kortmann 2007: 13). The more language contact the less grammatically redundant a variety becomes as superfluous grammatical information seems to get eliminated or at least reduced. This is why the majority of L2 varieties are neither particularly synthetic nor analytic.

The concept of language contact and reduced redundancy certainly explains why Hong Kong English is to be found in the bottom left corner of the scatter plot because Hong Kong was a trading stronghold – with Macao the most important trading point in Asia (cf. Endacott 1964: 4) – which means contact between different languages must have been extremely high. In addition to that, Hong Kong was a British colony for 155 years and even today English has the status of an official language (cf. Luke 1982: 47-55). Yet, this leaves the question why East African English is the overall most redundant L2 variety, unanswered.

Perhaps the type and status of a language, with which a given variety is in contact, is crucial for the development of this given variety's complexity. The English in Hong Kong was in contact with Chinese, a language based on a complex sound system with a monosyllabic structure of words, i.e. most words consist of one syllable only (cf. Haarmann 2001a: 97). In addition to this, Chinese uses hardly any inflections and marking, but analytic particles and word order are used to convey grammatical and

syntactic relations instead (cf. Haarmann 2001a: 97). By contrast Kiswahili, the main contact language for English in East Africa, uses a lot more inflections than English and is a highly synthetic language (cf. Haarmann 2001b: 370). This could explain why East African English with a syntheticity index of 140.5 is at $p$ = 0.000001 very highly significantly more synthetic than Hong Kong English with a syntheticity index of 67 and substantially more synthetic than any other variety of English. So it might be possible that the complexity of English developed into opposite directions according to the overall grammaticity of the contact language in these two countries.

Another possible explanation might be the different status of English in these two countries; in Kenya and Tanzania, English is, even today, a prestige language (cf. Hudson-Ettle & Schmied 1999: 4) mainly spoken by the educated and upper classes so that language contact might not play such an important factor and might not have as much impact on the development of complexity than in other L2 varieties. Moreover, Kiswahili was the main trading language in East Africa (cf. Haarmann 2001b: 370) and must have taken over the function and position of English as the major communication language, thus further reducing the impact and importance of English in Kenya and Tanzania. On the contrary, in Hong Kong, English was and still is the major language of trade and communication. This means it was probably spoken across all social classes so that frequent and extensive language contact was given and had a very strong impact on the complexity development of Hong Kong English.

## 3. Summary and conclusion

In this paper, I investigated the differences between East African English and British English registers in regard to syntheticity and analyticity. Starting off with these two grammatical dimensions, further connections with overall grammaticity and complexity have been made.

Firstly, the variance in syntheticity and analyticity of the different East African English registers has been compared.

In the spoken registers, private and public, as well as in the written register creative, the highest values for analyticity and overall grammaticity can be observed whereas the written register reportage has the lowest index for analyticity. All the other registers show mid-range values for analyticity. Reportage, having the lowest analyticity index, has at the same time the highest index for syntheticity. The other registers, except for the written registers letters and instructional which show the overall lowest indices for syntheticity, can be placed in a mid-range position for syntheticity.

The high frequency of analytic markers in the spoken registers private and public and the low frequency of analytic markers in the written reportage can be explained by the different focus of communication. In the spoken registers the focus is clearly on the hearer and in the written register reportage on the writer: so the current study confirms that analyticity facilitates listener comprehension whereas syntheticity serves writer economy. In the case of the written register creative, this explanation does obviously not serve. However, the high frequency of analytic markers as well as the high value for overall grammaticity derive from the fact that creative writing is the most elaborate and detailed form of writing which is not subjected to any kind of restrictions or limitations in terms of grammar and number of characters. It is thus not surprising that this register is rich in grammatical markers. Yet, it is very surprising that the written register academic, which one would expect to be very formal and an overall latinised form of writing, does show a significantly lower index for analyticity as well as overall grammaticity.

The written registers letters and instructional show the lowest values for syntheticity and overall grammaticity which is not particularly surprising as far as the register letters is concerned. The style of writing used in letters is likely to resemble that of (informal) spoken language, i.e. there is no need for formal grammatical marking.

Secondly, it has been investigated how East African English registers differ from the corresponding British English registers in terms of analyticity and syntheticity. Generally, the variance between the corresponding registers is less than one would expect it to be. There is, surprisingly, hardly any difference in analyticity at all as the values across all registers, with the exception of the persuasive register, show no statistically significant variation. Generally, British English is more analytic than East African English with the exception of two registers, creative and persuasive, in which the East African variety is more analytic.

In regard to syntheticity, in only six out of eleven registers a statistically significant difference can be observed whereas the East African English registers in which there is a significant syntheticity variance are without exception always less synthetic than their British English counterparts.

As mentioned above, the written register persuasive is the only register in which a statistically significant difference in analyticity can be observed: in East African English the frequency of analytic markers is higher than in British English. The sources for this variation are indefinite, personal, reflexive and wh-pronouns. At the same time, the East African persuasive register is less synthetic than the corresponding British register. As both varieties have identical indices for overall grammaticity, it can be assumed that the variation in analyticity is due to the L2 status of East African English which requires more transparent and explicit marking than the British native variety.

The most significant difference in syntheticity occurs in the creative register. East African English uses far less synthetic markers than British English.

Another interesting finding is that in the written register reportage the values for syntheticity are identical across the two varieties. This can be explained by journalistic standards which seem to be the same in both varieties as well as with the fact that synthetic markers enhance writer economy.

In all registers, the variation in syntheticity can mainly be ascribed to less frequent verb inflection, mainly past tense forms, in East African English. In the registers letters and persuasive, plural noun markers are the second major source for the variation in syntheticity.

Thirdly, it has been analysed how East African spoken English fits in with Szmrecsanyi and Kortmann's comparison of Englishes around the world in terms of analyticity and syntheticity as well as the related concepts of grammaticity and complexity.

Surprisingly, East African English does not behave like the other L2 varieties investigated by Szmrecsanyi and Kortmann and is therefore L2-atypical because it is not, as expected, in the mid-range like the majority of L2 varieties, but is extremely verbose, i.e. grammatically redundant. This might be due to the type and status of the contact language in Kenya and Tanzania: Kiswahili is an extremely synthetic and an overall highly grammatically complex language (cf. Haarmann: 2001b: 370) and might have influenced the complexity development such that East African English is the most synthetic and a highly verbose L2 variety. Another explanation could be that the status of English as prestige language reduced the impact of language contact as a factor on the complexity development of East African English.

Further research on the reasons for the L2-atypical behaviour of East African English would certainly be worthwhile and interesting as the above mentioned explanations are purely speculative. It might also be of interest to compare written material of East African English and the L2 varieties investigated by Szmrecsanyi and Kortmann in order to investigate whether the difference in complexity and grammaticity can be confirmed or whether this difference is only a phenomenon of spoken language.

## Appendix A: Zusammenfassung

In dieser Arbeit werden 11 Englische Register aus Ostafrika, d.h. Kenia und Tansania, mit entsprechenden britisch-englischen Registern hinsichtlich ihrer *Analytizität* und *Synthetizität* verglichen. Diese Analyse basiert auf quantitativen Daten von zwei Korpora, dem International Corpus of English for East Africa (ICE-EA) und dem British National Corpus (BNC).

In Greenbergs Abhandlung "A Quantitative Approach to the Morphological Typology of Language" (1960), etabliert er numerische Indizes für verschiedene grammatikalische Charakteristika, die eine exakte Berechnung der Analytizität und Synthetizität einer Sprachvarietät zulassen. Diese Charakteristika wurden mit einem Quotienten zweier Einheiten berechnet, welcher auf der relativen Textfrequenz dieser zwei Einheiten in einer Textpassage von 100 Wörtern basiert. Der Grad von Synthese kann demnach als dem Quotienten von Morphem/Wort berechnet werden, wobei der minimale Wert für Synthetizität auf 1,00 festgelegt wurde, da ein Wort mindestens aus einem Morphem bestehen muss (vgl. Greenberg 1960: 185).

Die Indizes, die in dieser Studie verwendet werden, basieren auf dem obigen Model von Greenberg und wurden von Szmrecsanyi und Kortmann hinsichtlich ihrer Abhandlung "Between simplification and complexification. Non-standard varieties around the world" (to appear) modifiziert. In ihrer Studie haben sie folgende drei Indizes etabliert:

Ein *Synthetizitätsindex*, das Verhältnis von gebundenen Morphemen darstellend, ein *Analytizitätsindex*, die Anzahl an freien grammatischen Morphemen darstellend, und ein *Grammatizitätsindex*, welcher aus den zwei vorherigen Indizes besteht. Alle drei Indizes wurden per 1000 Wortzeichen berechnet (vgl. Szmrecsanyi & Kortmann to appear: 8).

Die Korpusdaten, welche der vorliegenden Studie zugrunde liegen, wurden aus dem Subkorpus für Ostafrika des International Corpus of English entnommen und manuell mit BNC Wortklassenannotationen (POS-tags) versehen, welche in synthetische und analytische tags eingeteilt wurden. Mit Hilfe eines computerisierten Algorithmus wurden dann zufällig 1000 Wortzeichen per ICE-EA Register ausgewählt, sodass das gesamte Datenset 11 Register mit jeweils 1000 Wörtern, insgesamt also 11.000 Wörter

ostafrikanischen Englisch zählt. Die Textfrequenz der POS-tags wurde anschließend mit einem Computerprogramm erfasst, welches die Indizes für Synthetizität, Analytizität und Grammatizität berechnete, und manuell ausgewertet.

Die Daten der entsprechenden Register aus dem BNC, welche hier verwendet werden, wurden von Szmrecsanyi in seinem Artikel “Typological parameters of intralingual variability: grammatical analyticity vs. syntheticity in varieties of English” (submitted) etabliert und unverändert übernommen.

Anhand dieser Indizes wurde versucht auf folgende drei Forschungsfragen eine Antwort zu finden:

1. Inwieweit sich die verschiedenen ostafrikanischen Register im Hinblick auf Analytizität und Synthetizität untereinander unterscheiden.

Hierbei fällt auf, dass die gesprochenen Register private und public, sowie das geschriebene Register creative die höchsten Werte für Analytizität und Grammatizität aufweisen. Reportage, welches zugleich den höchsten Synthetizitätsindex besitzt, hat den niedrigsten Index für Analytizität. Bis auf die geschriebenen Register letters und instructional, die die niedrigsten Werte für Synthetizität aufweisen, können alle Register in eine mittlere Position sowohl für Synthetizität als auch Analytizität platziert werden.

2. Inwiefern sich die entsprechenden Register im ICE-EA und BNC hinsichtlich ihrer Analytizität und Synthetizität unterscheiden.

Überraschenderweise finden sich weniger Unterschiede als man annehmen würde: Es lässt sich kaum eine Abweichung in Analytizität feststellen und nur in einem Register, persuasive liegt ein statistisch signifikanter Unterschied vor: Ostafrikanisches Englisch ist analytischer als britisches Englisch. Von dieser Ausnahme abgesehen, sind die britisch-englischen Register allgemein analytischer als ihr ostafrikanisches Gegenstück. In Bezug auf Synthetizität kann in nur sechs von elf Registern ein statistisch signifikanter Unterschied beobachtet werden, wobei die ostafrikanischen Register ohne Ausnahme weniger synthetisch sind als die britische Varietät.

3. Wie ostafrikanisches Englisch im Vergleich mit den von Szmrecsanyi und Kortmann untersuchten Varietäten hinsichtlich ihrer Grammatizität und Komplexität einzuordnen ist.

Überraschenderweise verhält sich Ostafrikanisches Englisch nicht wie andere L2 Varietäten, welche normalerweise Mittelwerte für Verbosität aufzeigen, sondern ist eine

sehr verbose Varietät. Dies könnte zum einen am Status des Englischen in Kenia und Tansania liegen: auch heute noch ist Englisch die Sprache der gebildeten Elite, sodass der Einfluss und die Auswirkung des Sprachkontaktes auf die Komplexitätsentwicklung deutlich geringer sein muss als zum Beispiel für Hong Kong Englisch, welches die am wenigsten verbose L2 Varietät ist. Zum anderen mag der Typ der Sprache mit welcher Englisch in Kontakt war bzw. ist dessen Komplexitätsentwicklung beeinflussen: Kisuaheli ist eine sehr synthetische und grammatikalisch komplexe Sprache (vgl. Haarmann 2001b: 370), sodass sich Englisch in Ostafrika zu einer dementsprechend verbosen und komplexen Varietät entwickelt hat.

Es wäre sicherlich lohnenswert, die Gründe für das L2-atypische Verhalten von ostafrikanischem Englisch näher zu untersuchen, da die obigen Thesen rein spekulativen Charakters sind. Interessant wäre auch zusätzlich zu sprachlichem Material geschriebenes Englisch aus Ostafrika mit weiteren L2-Varietäten zu vergleichen um zu sehen, ob sich die Unterschiede in Grammatizität und Komplexität bestätigen lassen oder nur ein Phänomen der gesprochenen Sprache sind.

## Appendix B: Grammaticity in British and East African English registers

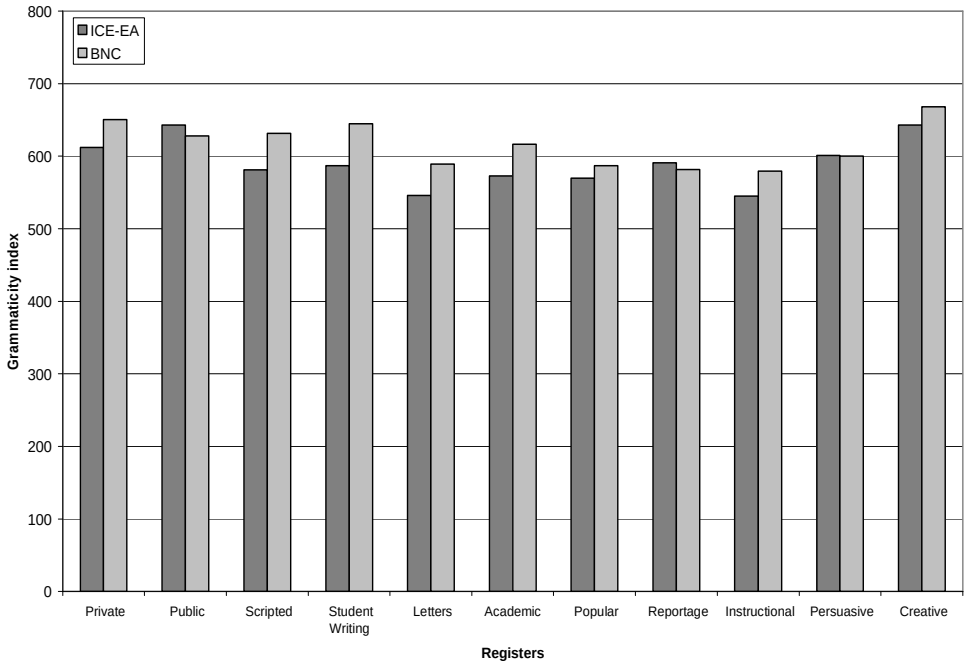


Figure 7: Grammaticity in British and East African English registers

**Bibliography**


Anttila, Raimo. 1989. *Historical and Comparative Linguistics*. Amsterdam, Philadelphia, Benjamins.

Aston, Guy & Lou Burnard. 1998. *The BNC Handbook*. Edinburgh, University Press: 230-233.

Comrie, Bernard. 1989. *Language Universals and Linguistic Typology*. Oxford UK, Blackwell.

Danchev, Andrei. 1992. "The evidence for analytic and synthetic developments in English." In: M.Rissanen, O.Ihalainen, T. Nevalainen and I. Taavitsainen (eds). *History of Englishes: New Methods and Interpretations in Historical Linguistics.* Berlin, New York, Mouton de Gruyter: 25-41.

Endacott, G.B. 1964. *A History of Hong Kong*. 2nd edition. Oxford University Press: London.

Greenberg, Joseph H. 1960. "A Quantitative Approach to the Morphological Typology of Language." *International Journal of American Linguistics* 26(3): 178-194.

Greenberg, Joseph. 1974. *Language Typology: A Historical and Analytic Overview*. The Hague, Mouton.

Haarmann, Harald. 2001a. "Chinesisch". *Kleines Lexikon der Sprache*. München, C.H. Beck: 95-100.

Haarmann, Harald. 2001b. "Swahili". *Kleines Lexikon der Sprache*. München, C.H. Beck: 369-372.

Hudson-Ettle, Diana & Josef Schmied. 1999. "Manual to accompany The East African Component of The International Corpus of English".

Luke, Kwang-kwong, J.C. Richards. 1982. "English in Hong Kong: function and status". *English World-Wide: A Journal of Varieties of English* 3:1: 47-64.

Preacher, K. J. (2001, April). Calculation for the chi-square test: An interactive

calculation tool for chi-square tests of goodness of fit and independence [Computer software]. Available from http://www.quantpsy.org. Website last visited: 19.03.2008

Schmied, Josef. 1985. *Englisch in Tansania*. Heidelberg: Julius Groos Verlag.

Schmied, Josef. 1990. "Language use, attitudes, performance and sociolinguistic background: a study of English in Kenya, Tanzania, and Zambia". *English World-Wide: A Journal of Varieties of English* 2: 217-138.

Szmrecsanyi, Benedikt. Submitted. "Typological parameters of intralingual variability: grammatical analyticity vs. syntheticity in varieties of English".

Szmrecsanyi, Benedikt & Bernd Kortmann. To appear. "Between simplification and complexification: non-standard varieties of English around the world". In: Geoffrey Sampson, David Gil & Peter Trudgill (eds). *Language Complexity as a Variable Concept*. Oxford: Oxford University Press.