

PROTEST AMERICAN ENGLISH INFLUENCE OR PROTEST AGAINST IT? CHANGING PREPOSITIONS IN NIGERIAN TWITTER ENGLISH

Matthias Hofmann

Abstract

According to Alo and Mesthrie (2008), Nigerian English (NigE) becomes increasingly more influenced by American English (AmE), due to contact with American-trained professionals among other factors (cf. Gut 2008, Jowitt 1991). The online micro-blogging service Twitter offers potential communication with a vast number of English natives around the globe, using English in a vernacular usage domain, among other domains (or genres such as a news tweet vs a private tweet). With its foundation in 2006, Twitter is a new communication technology, which may indicate that it is used predominantly by “younger” urban people, and which may influence linguistic choices. The question I attempt to answer is whether Twitter influences NigE such that the British English (BrE) heritage of the country is contested by AmE influence. In this paper, I focus on the usage of prepositions and orthographic realizations of lemmata ending in -o(u)r, which can be categorized as BrE and AmE origin, respectively, in a NigE Twitter Corpus compiled in 2016-17 (13 mill. words). These features’ frequencies are contrasted with those of the Nigerian component of GloWbE (Davies 2013). Results from chi-squared tests suggest that AmE prepositions increasingly enter NigE Twitter discourse. Differences in spelling tend towards American English, but are not statistically significant. The only exception is the lemma *labour*, which is more often used in its British English spelling variant ($\chi^2 = 26.30$; $df = 1$; $p_{\text{one-tailed}} < 0.001$).

Keywords

Nigerian Twitter English, youth language, prepositions, chi-square test, American English influence, specialized discourse, orthography

1 Introduction

Few scholars have used Twitter as a resource for the investigation of urban youth English usage in African countries. Even Nigeria has received little attention in this respect, although it is the African country with most relative internet users (Schmied 2015). As I have shown elsewhere (Hofmann in prep.), Twitter is more than suitable to investigate a very important aspect of youth language, which is its creativeness when used in an online micro-blog such as Twitter. Although its findings might not necessarily be generalizable to the whole speech community of urban youth in Nigeria, the paper lays the groundwork for

a written-to-be-spoken text type in such a variety of English based on a much bigger corpus than that in Hofmann (in prep.).

The speech community of Nigeria is not homogeneous. Traditionally, Hausa English spoken in the North of Nigeria is much more standard than Yoruba English is in the south of the country. Its standard is British English due to its socio-historical development (e.g. Alo & Mesthrie 2008). And yet, some authors have found an increasing influence of American English on several levels of language such as lexis and grammar. They hypothesized that this influence could be due to American broadcasts in Nigeria and contact with American-influenced professionals, which would reduce a differentiation of linguistic features on a sub-national level (Hofmann in prep., also cf. Jowitt 1991, Alo & Mesthrie 2008, Gut 2008).

In the next section of this paper, I narrow down the terminology that I find suitable for the present investigation and outline the suitability of Twitter English data for the investigation of youth language within Nigeria, before I introduce the features under investigation regarding the influence of American English through broadcasts. In Section 3, I describe the corpus compilation via a tool that has been developed by Albrecht and Schmied (2015), briefly summarize data cleaning and annotation procedures I used to build the corpus, and explain the method of investigation. In section 4, I present and discuss the results of prepositional usage in Nigeria, compare the findings in the Twitter data to those of the Corpus of Global Web-Based English (GloWbE; Davies 2013) for Nigeria in order to provide a context for the Twitter data. In the last part of section 4, I provide additional empirical support for the influence of American English on orthographic features in Nigeria. The last section concludes the analysis and provides a brief outlook for further investigations.

2 New sources of corpus linguistic data: Urban tweets in Nigeria

2.1 Relevant terminologies

Linguists from various fields have engaged in investigations of data from the web, using various terminologies for situating each of the studies in their respective domains. This development has released a variety of terms to choose from when it comes to denote the domain, such as Web as Corpus (WaC), Web for Corpus building (cf. Hundt & Nesselhauf 2007), web linguistics, internet linguistics, etc. I am not concerned with defining concepts and data sources exhaustively in such a way that each of the terms is clear-cut and separable from the next by definitions. Rather, I consider this web/internet-linguistics paper to be of corpus-linguistic nature and investigating a highly specialized corpus, namely a corpus comprised of Nigerian English tweets.

Other terms have been around to denote the variety of English under investigation, such as Internet English or Cyber(space) English (cf. Mair 2013, Heyd & Mair 2014). As outlined elsewhere (Hofmann in prep.), I use the term ‘Twitter English’ as hyponym to loosely include the meanings of the above-mentioned terms, but more importantly to capture and reflect the linguistic peculiarities of tweets that are caused by the limited number of characters allowed per tweet (140 before and 280 after 2017). The difference to English SMS talk, which has a character limitation as well, is obvious in the strong ties or embeddedness of tweet content to or on the internet. That is, unlike the content of conventional text messages, tweets include predominantly pictures, links, ads for homepages, and linguistically valuable text in the form of a spread of imminent and short-lived information. Character limitation sets Twitter English also apart from Facebook posts (e.g. cf. Burghardt 2015), although the content and aim of both are similar (cf. Schmied 2012). The variety under investigation in this study can hence be termed *Nigerian Twitter English (NiTE)*.

2.2 Twitter and the youth

Twitter is a rather recent online news and social networking tool that has been implemented in our daily lives in manifold ways. Nevertheless, it has also gained momentum in much more specialized discourses as well. Today, it is not only used for staying up-to-date on opinion-sharing among peers and sometimes even friends or for following a celebrity, but also for communicating with colleagues in academia and with university students in a formalized setting (cf. Schmied 2012 for more details). Although Twitter originally started out as an online news service attracting about 30 per cent older adult users three years after its foundation (cf. Lipsman 2009, Kelly 2009), eight years later the service has been established as a standard tool among the youth, as most users world-wide are aged between 18 and 29 (Hutchinson 2017). This may partially be because increasingly more teenager-idol celebrities have joined the social networking service, as a quick-and-dirty look at, for example, Katy Perry’s Twitter profile shows 106 million followers compared to Christina Aguilera’s 16.6 million followers – a more experienced celebrity – in November 2017.

Another reason for predominantly young people using Twitter may be its genres: 37.6 per cent of tweet content is conversational and 40.1 per cent is “Pointless Babble” (Kelly 2009: 5). Pointless babble refers to “‘I am eating a sandwich now’ tweets” (ibid.: 4) that I have denoted as imminent and short-lived information above. Mainstream news made only a marginal portion of 3.6 per cent of tweet content in the study (ibid.). Since Kelly’s study is rather dated, the content of tweets may have shifted somewhat towards more formal styles, such

as mainstream news or tweets by politicians. Unfortunately, more recent, reliable studies on Twitter are not available.

I would consider online conversing and babbling an unnatural linguistic behavior in informal styles that requires some acquisition during formative years. That is, the more Tweeters grow up having conversations online on Twitter, the more such online conversing and babbling becomes natural linguistic behavior for them. Peers of the Twitter founders would still have to get used to the service in terms of substituting natural face-to-face situations in which they learned to converse and babble with such anonymous online situations. That is, people older than 29 have acquired conversation and babbling skills in face-to-face situations and may consequently experience conversing and babbling in anonymous online situations as “different” or “unnatural” (cf. Tagg & Seargeant 2015). Due to the contents of most tweets, the style is rather informal.

In summary, the language used in Twitter is connected speech by predominantly young people. Although it is written, I consider the language used in tweets likely to be more similar to written-to-be-spoken text types than to formal writing (Halliday 1989, cf. Tagg & Seargeant 2015), considering the prominent topics of the majority of tweets. Their limited character number reduces sentence-length induced language complexity.

2.3 Twitter and Nigeria

Although Africa’s countries do not first come to mind when thinking about the internet and from where its users contribute to it, some of them are home to quite many internet users. As, for example, Schmied (2015: 190) points out, Nigeria’s large population causes the absolute number of Nigerian internet users to be higher than that in the United Kingdom and with 70 million users to be highest on the African continent. Nigerian scholars have valued and continue to value the internet as a vital means of communication in Nigeria by contributing to the international academic discussion of digital communication (e.g. cf. Taiwo 2010) and by investigating language and content of discourse data available online (e.g. cf. Opeibi 2011, Sousa & Ivanova 2012). While demographics for Nigerian Twitter usage are scarcely available, Schmied (2015: 189) maintains that Twitter is as much an integral part of Nigerian internet usage as Facebook, blogs and other internet fora. As such, Twitter is a rich source of collecting Nigerian English data. Complaints about internet reliability, affordability, and power failures in Nigeria – which by no means reduces the popularity of electronic media – may skew the use of Twitter and other services across the country towards the urban areas. For this reason, I have narrowed the data collection site down to include urban areas only, such as Lagos and Kano. A differentiation into Hausa, Igbo,

and Yoruba English (cf. Jowitt 1991) is, however, not possible with this means of data collection, as social meta data is scarcely filled out reliably in Twitter's user profiles (e.g. cf. Nguyen et al. 2013, Bamman et al. 2014).

2.4 American English influence and the role of English in Nigeria

As in other former British colonies, English is a high-function foreign language with about 4-20 per cent speakers/users in Nigeria (Jowitt 1995). Since 1998, English is the medium of instruction starting grade four in Primary School (Gut 2008). Today, especially among younger speakers, English usage is on the rise in all contexts of daily life except at home, so that its purely elite status (Jowitt 1995) and a somewhat exclusive use of Nigerian Pidgin for domestic functions (Osisanwo 2015) become blurred. More and more speakers of English value the former language of colonialism highly as a potential for material and social gain, a symbol of modernization, and a means of success and mobility (Gut 2008).

The structure of Nigerian English has been well described on all levels of language, e.g. phonology (e.g. Gut 2008, Adedeji 2015), morphosyntax, and lexis (e.g. Alo & Mesthrie 2008), and as a World English variety more generally (e.g. Schmied 1995, Schneider 2007, Mesthrie & Bhatt 2008). Approximately a decade ago, Alo and Mesthrie (2008: 338)¹ have pointed out that Nigerian English becomes increasingly more influenced by American English, allegedly due to American broadcasts (but cf. Stuart-Smith et al. 2013 on the influence of broadcasting media on varieties), music, cinema and contact with American-trained professionals (also cf. Jowitt 1991, Gut 2008: 38, 40). This influence is apparent on several levels of language, including pronunciation, idioms, and lexis. One example of American English influence on Nigerian English is the use of “[...] business terms like *Monday through Friday*” (Alo & Mesthrie 2008: 338), rather than British English *Monday to Friday*. Another example is the use of “[...] verb sub-categorizations like *to protest* + [direct object], rather than *to protest against* (Alo & Mesthrie 2008: 338).

I am concerned with the analysis of American English prepositions as used, for example, in phrasal verbs or fixed phrases/idioms. The motivation behind the investigation of prepositions in addition to lexical features is that the latter are usually not as good an indicator of systemic changes in the English language as grammatical or phonetic features are. Lexical features are quite salient to users of a language, which becomes apparent, for example, in the fast adoption of slang terms or buzzwords among younger groups of people. Likewise, these same terms disappear fast. None of them really makes it into the register² of a speech community. Hence, lexical items alone are not necessarily good features for

defining a variety or the influence of another variety. The grammatical features under investigation in the Nigerian English Twitter data set are randomly selected and outlined in Table 1:

Feature No.	British English	American English
1	fill in a form	fill out a form
2	[*Weekday] to [*Weekday]	[*Weekday] through/thru [*Weekday]
3	get on (with somebody)	get along (with somebody)
4	at the weekend	on the weekend
5	haven't had anything for years	haven't had anything in years
6	in [a street name]	on [a street name]
7	to protest against [+ DO]	to protest [+ DO] (direct object)

Table 1: Tabular overview of randomly selected British English and American English prepositions.

Features 1 to 4 are standard examples of British versus American prepositional usage that are cited frequently for illustration and indicate prepositional use in phrasal verbs (*fill in/fill out*, *get on/get along*), business terms (e.g. *Monday to Friday/Monday through Friday*), and fixed phrases (*at the weekend/on the weekend*). For feature 4, however, it might be the case that the American English variant of the prepositional usage is not derived from American English influence, but cognitive processes of creolization. On its own, it thus may not allow for measuring a direct influence of American English, but as a complement to the other features, it may do so if its distributional patterns are similar to those of the other features. Additionally, the potential origin through processes of creolization of this prepositional usage in NiTE is only confounding, if its frequencies of occurrence outnumber those of the other features by far.

Features 5 to 6 are prepositions that do not fall into the above-mentioned categorization. Feature 5 can be generalized as use of preposition *in* in a certain grammatical environment, namely after a negative construction (*have not had*), feature 6 as preposition used before street names, and feature 7 as prepositional verb. The latter three features are certainly used less often when outlining differences between the two native varieties of English in educational settings.

I assume that NiTE neither exclusively uses prepositions from British nor from American English. Rather, it is likely to use prepositions from both, which can also be viewed as empirical support for the question of whether American English influences Nigerian English. If American English does not influence Nigerian English, none of the American English prepositions should be in use by Nigerians (except feature 4). However, if both native Englishes' prepositions

are in use, a mix of the two should be observable in the Twitter data rather than a dominance of American English variants. Since the literature suggests that an increasing influence of American English should be visible, I consequently interpret this influence to be represented by a mixed use of prepositions denoting the same concept, rather than a dominance of American English prepositions.

3 Methodology

3.1 Corpus compilation

Using *TwitCollect* (Albrecht & Schmied 2015), a tool based on *Twython* (Schettler n.y.), tweets were collected from 2016 to 2017. The tool uses Twitter's API to download tweets, without capturing sociolinguistic meta-data about the authors of the tweets and saves them into a .csv-file by default. Twitter limits the number of tweets downloadable in this fashion to approximately 1 per cent of all tweets (Murthy et al. 2016: 34), yielding an average amount of approximately 1 Megabyte-sized csv-files per 24 hours. This size roughly corresponds to 1,000 tweets per file. In terms of corpus size, NiTE corpus totals 12,775,082 words. Frequency data will be shown in units of per million words (pmw).

3.2 Data cleaning and corpus annotation

Although software is available, “boilerplate” removal (cf. Bernardini et al. 2006: 20, Kilgarriff 2007) from corpus data off the web is typically offered for websites. An example of a discontinued tool as part of a PhD project is *jusText* (Pomikálek 2014), which is designed to remove html code, including navigation bars, headers, and footers, as well as duplicate information from online documents. Such tools are unsuitable to use for tweets, because their content is relatively free of html code, except for hyperlinks. To preserve corpus linguistic data in tweets, hashtag-topics (#category) and addressee-tags (@twitteruser) should be removed in addition to links. However, some tweets contain extremely few characters so that such a removal of text would render the content unintelligible (Kilgarriff 2007). To avoid this, I cleaned the corpus data manually in *MS Excel* to maintain boilerplate when crucial to interpret the semantic meaning of the tweets' content.

Tokenization, POS annotation and lemmatization of the Twitter corpus posed many problems, even after data cleaning. Tweets typically contain relatively great shares of non-standard spelling and a high density of neologisms and acronyms. For this reason, Bernardini et al. (2006: 23) suggest that “POS taggers should be re-trained on Web data”. Their suggestion is, however, based on the assumption that the web itself serves as a corpus that needs to be annotated, i.e.

a billions-of-words corpus (WaCky corpus). The Twitter corpus is too small to allow for general patterns of investigation that would require POS tagging, and the frequency with which the prepositions outlined in Table 1 are used can be measured without grammatical categorization of the lexical items in the corpus. Consequently, NiTE was not POS-tagged for this analysis.

3.3 Method of investigation

I used *Antconc* (Anthony 2018), Version 3.5.7, for the analysis of prepositional usage and spelling patterns in NiTE, GloWbE's online search interface for the Nigerian sub-component (Davies 2013), and R version 3.5.0 for the statistical assessment of the results (R Core Team 2018). Each preposition and orthographic realization was searched for using lemmata of the concordances, e.g. *fill in*, *filled in*, *filling in*, *fills in* are all categorized as instances of *fill in*. The search terms included wild cards (e.g. asterisks) and POS tags as necessary. The search terms for feature 5 in Table 1 in GloWbE's search interface, for example, included up to four slots for words between the negative and the prepositional phrase *in/for years*: *[v*] [xx*] [v*] * * * * in years*. The first POS-tag consisting of the small letter *v* plus the asterisk refers to all auxiliary verbs in the corpus, the second tag to negative constructions, and the third tag to auxiliary verbs again. The search for negative constructions followed by *in years* started with one word after the negative and continued with up to four words after the negative. I stopped at a maximum of four because the frequencies of occurrence became too small to include instances with more than four words after the negative construction. As mentioned above, the frequencies of occurrence are shown in units of per million words in the respective tables. All the corpus searches were done using a stoplist, which predominantly included single-letter morphemes among others, abbreviating words like *okay* (short: *k*).

For the statistical assessment, I drew on a chi-squared (χ^2) goodness-of-fit test (cf. Gries 2013) for the comparison of British versus American English origin of prepositions and spellings. The probability-of-error values, *p*, are calculated for a one-tailed test, as the alternative hypothesis is directional: The frequency of occurrence of the British English preposition is higher than that of the American English preposition.

The assumptions of a chi-squared test are, first, that all observations are independent of one another, second, that 80 per cent of the expected frequencies are greater than or equal to five, and third, that all expected frequencies are greater than one (cf. Gries 2013: 166). Although the first assumption is difficult to assure for every single tweet, I made sure that each of the prepositions and spelling variants under statistical assessment was taken from a different tweet. In

addition, tweets occurring more than once in NiTE were deleted. Prepositions and spelling variants with expected frequencies smaller than five were disregarded in the analyses and marked with *not tested* in the respective tables. It should also be noted that the total amount of tweets was too small to include genre as a variable for statistical assessment.

4 Analysis and discussion

4.1 British and American prepositions in NiTE

The results for the investigation of preposition usage patterns in NiTE are summarized in Table 2:

Feature No.	Variants	Occ. of British prep.	Occ. of American prep.
1	<i>fill in : fill out</i>	2.27	0.47
2	[*day] <i>to : through/thru</i> [*day]	1.33	0.08
3	<i>get on (with) : get along (with)</i>	0.63	1.72
4	<i>at the weekend : on the weekend</i>	0.31	2.66
5	[negative] + <i>for years</i> : [negative] + <i>in years</i>	1.25	2.43
6	<i>in : on</i> [street names]	0.94	6.11
7	<i>to protest against : to protest</i> [+ DO]	3.52	4.31

Table 2: Normalized frequencies (pmw) of British versus American English prepositional occurrence in NiTE.

A quick look at the numbers shows that for all features investigated the frequencies are quite low, especially the American English variants in features 1 and 2, suggesting that the corpus is quite small despite the approximately 13 million words in total. The same is true for the British English variants in features 3 and 4. Likewise, the British-English occurrences of features 5 and 6 are rather small. Feature 7 is different from the other features, as it shows the highest frequencies for both geographical variants, and yet features 3 to 7 all clearly show a tendency towards the American English version of the respective prepositions (marked in bold in the table). Regarding the remaining two features, a minor trend towards British English phrasal verbs (feature 1) and business terms (feature 2) is visible and contrasted starkly by a probably significant influence of the American English prepositional usage of *on* before street names.

A valid interpretation seems to be that both native varieties influence Southern NiTE or have left traces in it. That is, a claim of exclusive use of prepositions from British or American English would be much too strong, even for the British English business term *Monday to Friday*. Although too marginal to generalize,

the American English business term *Monday through Friday* is virtually never used in tweets of NiTE. Almost the same is, however, true for British English *at the weekend*, so that such an exclusive use of a feature is balanced out across all features considered. In total, the American English variants are used more often than their British English counterparts, with five out of seven features occurring more often in their American English version. A closer look at the differences is provided in Table 3 below, which shows the result of their statistical assessment.

Feature No.	Variants	χ^2	$p_{\text{one-tailed}}$ -value
1	<i>fill in</i> : <i>fill out</i>	15.11	<0.001
2	[*day] <i>to</i> : <i>through/thru</i> [*day]	14.22	<0.001
3	<i>get on (with)</i> : <i>get along (with)</i>	6.53	0.011
4	<i>at the weekend</i> : <i>on the weekend</i>	23.68	<0.001
5	[negative] + <i>for years</i> : [negative] + <i>in years</i>	4.79	0.029
6	<i>in</i> : <i>on</i> [street names]	48.40	<0.001
7	<i>to protest against</i> : <i>to protest</i> [+ DO]	1.00	(0.317)

Table 3: Chi-squared and probability-of-error values for the difference between British English and American English prepositional usage in NiTE.

As can be seen, all the preferences in the usage patterns of the investigated grammatical features are significantly different from the expected frequencies, except for feature 7. The British English variants in features 1 and 2 occur significantly more often than expected and the American English variants in features 3 to 6 occur significantly more often than expected. This is a clear influence of American English variants – at least for features 3 to 6 – in line with what is suggested by the literature (e.g. Alo & Mesthrie 2008).

In summary, it is important to stress that both native varieties influence the English used in NiTE. The question that may arise from this observation is whether there is a general tendency of young Nigerians to use American English in written language more often or exclusively in certain domains, which will be addressed below.

4.2 Using GloWbE as a reference for prepositional usage in Nigerian English

Although the texts used in GloWbE (Davies 2013) belong to a similar genre as the ones in the Twitter corpus, the age of their authors remains unknown. Having stated that, GloWbE consists of texts taken from businesses' and companies' websites. People older than 29 usually have websites for professional reasons, if at all, and 18- to 25-year-olds usually represent a minority in businesses and companies. Likewise, although GloWbE allows us to track difference in

frequencies of usage between web-based texts from different countries, it does not allow us to differentiate Northern Nigerian usage patterns from Southern Nigerian usage patterns. GloWbE can, however, serve as a reference corpus when it comes to answering the question of whether the British English prepositions under investigation in this paper are exclusively used or not and of how often they occur. The GloWbE results might further help – at least in part – to decide whether the findings in the Twitter corpus are data-specific or not. For these reasons, I use the Nigerian component of GloWbE to investigate whether there is a general tendency of Nigerians to use American English in written text.

Feature No.	Variants	Occ. of British prep.	Occ. of American prep.
1	<i>fill in : fill out</i>	5.60	2.39
2	<i>[*day] to : through/thru [*day]</i>	7.10	0.33
3	<i>get on (with) : get along (with)</i>	3.31	4.20
4	<i>at the weekend : on the weekend</i>	7.62	1.22
5	<i>[negative] + for years : [negative] + in years</i>	0.45	0.73
6	<i>in : on [street names]</i>	0.82	2.06
7	<i>to protest against : to protest [+ DO]</i>	1.92	0.12

Table 4: Frequencies (pmw) of British versus American English prepositional occurrence in the Nigerian sub-corpus of GloWbE.

Most apparently, the numbers are similar in size to the ones reported earlier for the Twitter corpus, i.e. none exceeds 10 occurrences per million words. Likewise, it is apparent that British and American English prepositions are mixed in the variety of Nigerian English represented by the 42.6 million words component of GloWbE, i.e. none of the two prepositional usages is exclusive.

When we compare the results from GloWbE with those from NiTE, we find that features 1 and 2 are similarly distributed. Both features show a clear preference for the British English preposition. The same clear preference can be found for features 3, 5, and 6, although in these cases the American English alternative is preferred, a distribution also shared with NiTE. While the difference for feature 3 seems marginal, the use of preposition *on* before street names seems pervasive in GloWbE's Nigerian component (feature 6). For features 4, and 7, the results in the NiTE corpus are contrary to those in the Nigerian sub-corpus of GloWbE. While feature 4 virtually never occurs as British English variant in NiTE, it is the feature with the highest frequency in the British variant in the Nigerian component of GloWbE. Regarding feature 7, a strong preference for the British English usage can be seen in Table 4, contradicting a slight preference for the American English usage in Table 2.

Table 5 shows that, except for feature 5, these preferences for British English or American English prepositional usage patterns in GloWbE are significantly different from the expected frequencies. In case of features 1, 2, 4, and 7 the observed frequencies of British English prepositions are significantly higher and in case of features 3 and 6 the observed frequencies of American English prepositions are significantly higher than the expected frequencies. The lack of significance of the difference between the British and the American variant in feature 5 means that both variants occur as often as expected.

Feature No.	Variants	χ^2	$p_{\text{one-tailed}}$ value
1	<i>fill in</i> : <i>fill out</i>	55.04	<0.001
2	[*day] <i>to</i> : <i>through/thru</i> [*day]	263.47	<0.001
3	<i>get on (with)</i> : <i>get along (with)</i>	4.51	0.034
4	<i>at the weekend</i> : <i>on the weekend</i>	197.69	<0.001
5	[negative] + <i>for years</i> : [negative] + <i>in years</i>	2.88	(0.090)
6	<i>in</i> : <i>on</i> [street names]	22.84	<0.001
7	<i>to protest against</i> : <i>to protest</i> [+ DO]	68.15	<0.001

Table 5: Chi-squared and probability-of-error values for the difference between British English and American English prepositional usage in the Nigerian sub-corpus of GloWbE.

Despite the prevalence of using the British prepositions significantly more often in the Nigerian component of GloWbE, it becomes obvious that American English prepositions are used in Nigerian web-based English and that the American variants *get along (with)* and *on* + [street name] even occur significantly more often statistically.

In general, it seems fair to resolve that both the British and the American variants are used in Nigerian Web-Based English. It is further fair to conclude that the preference for the usage of a certain preposition is a function of region and genre. As mentioned in section 2.2, the content of tweets is predominantly babble of the sort “I am eating a sandwich now”. In other words, most tweets are intended to inform the online community or imagined target audience about what one is up to – be it for purely informative reasons or for inviting other people to join in or to share their opinion. This genre-specificity thus helps to explain the numbers of occurrence of the verb *protest* in the Twitter data from Nigeria (feature 7), as it makes Twitter the perfect means to invite a potentially large number of random people to join the protests or to raise an awareness of the fact that people are protesting something publicly (cf. Schmied & Opeibi 2017). At the same time, it may be the limited amount of 140/280 characters per tweet that gives rise to the occurrence of *protest* + [direct object], omitting the preposition *against*.

Having stated that, it must be kept in mind that the frequencies for all features in GloWbE are not representing the prepositional usage patterns of the country's youth. They rather represent the usage patterns of the country's businesses' and companies' websites. Eighteen-to-25-year olds usually do not have websites or represent a minority in businesses and companies. I consider these usage patterns in NiTE contrasted by those in GloWbE as unlikely to be a function of age, however, allowing for extrapolation from the GloWbE data.

4.3 British and American orthography in Northern and Southern Nigeria

A level of language that is much more salient than prepositions is orthography (cf. section 2.4). Bearing in mind that one must have had at least some education to set up an (English) Twitter account and use it on a regular basis, most young educated Nigerians are well aware that some words have a British English spelling and others an American English one. An example of competing native spelling influences is seen in lemmata ending in British English *-our* versus American English *-or*.

Lemma	British orthography	American orthography
<i>fav[o/ou]r</i>	83.60	84.70
<i>hon[o/ou]r</i>	32.64	29.51
<i>col[o/ou]r</i>	27.01	28.26
<i>lab[o/ou]r</i>	15.73	17.06
<i>neighb[o/ou]r</i>	12.68	12.84
<i>behavi[o/ou]r</i>	9.78	4.38
<i>flav[o/ou]r</i>	5.40	4.15

Table 6: Frequencies (pmw) of the most frequent lemmata with word-final British English *-our* and American English *-or* in NiTE.

The per-million-word frequencies of occurrence of the most frequent lemmata in NiTE (cf. Table 6) are on average multiple times higher than those of the prepositions outlined in the previous section. A closer look at NiTE shows that a general trend of a dominating British English spelling is not true for each individual lemma (cf. Table 6), as expected from the results in the previous section. Particularly the lemmata *favor*, *color*, *labor* and to some degree the lemma *neighbor* show a higher frequency of occurrence in American English orthography than in the British one.

In more general terms, Table 6 is striking because it provides clear evidence against the notion of exclusive use of either an American English variant or a British English one. This clarity cannot be shown as effectively as it could

in Table 2 for prepositions, but the tendency to the American English spelling variants is obvious.

Lemmata	χ^2	$p_{\text{one-tailed}}$ -value
favor : favour	0.09	(0.763)
honor : honour	2.02	(0.156)
color : colour	0.36	(0.547)
labor : labour	0.69	(0.406)
neighbor : neighbor	0.01	(0.912)
behavior : behavior	26.31	<0.001
flavor : flavor	2.10	(0.148)

Table 7: Chi-squared and probability-of-error values for the difference between word-final British English *-our* and American English *-or* in NiTE.

Table 7 shows the chi-squared goodness-of-fit test results for the spelling of the lemmata shown in the previous table. Except for the spelling variants of *labor/labour*, none of the frequency distributions of the two spellings deviate significantly from the expected ones. That is, only British *labour* occurs significantly more often than expected. Although the spellings are merely tendencies without statistical significance, the Nigerian youth seems to consciously orient towards the American English spelling, as both spelling variants occur in the data (cf. section 2.4), and as both spelling variants occur at similar frequencies.

Lemma	British orthography	American orthography
<i>fav[o/ou]r</i>	101.11	70.70
<i>hon[o/ou]r</i>	83.17	51.73
<i>col[o/ou]r</i>	59.30	55.43
<i>neighb[o/ou]r</i>	49.81	36.16
<i>behavi[o/ou]r</i>	36.63	43.59
<i>lab[o/ou]r</i>	22.36	31.28
<i>flav[o/ou]r</i>	8.65	6.38

Table 8: Frequencies of occurrence (pmw) of the most frequent lemmata with word-final British English *-our* and American English *-or* in the Nigerian component of GloWbE.

In case of the rather formal genres in the Nigerian component of GloWbE shown in Table 8, we can observe a similar trend as mentioned above for the informal English used in Twitter (cf. Table 4). Most of the lemmata are spelled more often in a British English manner than in an American English manner, except for the lemma *behavior* and the lemma *labor*. This finding also

corroborates the usage of prepositions shown in section 4 insofar, as the use of British orthography is not exclusive in most of the cases. The tendency to use the American English spelling variant is, however, very slight.

Lemmata	χ^2	$p_{one-tailed}$ -value
favor : favour	229.59	<0.001
honor : honour	312.58	<0.001
color : colour	5.56	0.012
neighbor : neighbour	92.40	<0.001
behavior : behavior	25.76	<0.001
labor : labour	6.35	0.012
flavor : flavor	14.68	<0.001

Table 9: Chi-squared and probability-of-error values for the difference between word-final British English *-our* and American English *-or* in the Nigerian component of GloWbE.

Startlingly, all the lemmata occur significantly more often in their British and American orthography, respectively, than expected (cf. Table 9). This result is in stark contrast to that of NiTE, shown in Table 7, particularly for the lemma *labor*, since it is used significantly more often in the American English spelling in the Nigerian component of GloWbE and significantly more often in the British English spelling in NiTE. The similarity of the spelling behavior in the Nigerian component of GloWbE to that in NiTE is that the lemma *behavior* is used significantly more frequently in the American English spelling. (cf. Table 9). As mentioned earlier, both spelling variants are in use in both corpora, despite the traditional British English spelling prevalence.

Orthography thus provides good support for the increasing influence of American English, as stated by Alo and Mesthrie (2008), because it shows that both variants – the British and the American English one – exist alongside one another. This may reflect an ever-increasing influence of American English through increasing frequencies of occurrence that eventually may surpass the frequencies of British English spelling variants in two to four generations time.

Whether this influence is caused by or “due to American broadcasts (*CNN* and *Voice of America*)” (Alo & Mesthrie 2008: 338) remains highly doubtful. As Foulkes and Docherty (2000), for example, have shown in an L1English setting, television rather serves as a catalyst instead of introducing new features, so that speakers redeploy the linguistic resources already available to them. Their speech may shift towards the linguistic features of imagined socially attractive speakers (cf. Audience Design, Bell 1991). At least, they may change their

attitudes towards existing features through broadcasts (cf. Stuart-Smith 2007) or become more aware of standard forms (cf. Milroy & Milroy 1985). In case of an L2English setting as in Nigeria, it seems that the linguistic forms speakers are aware of (spelling/orthography) show higher American English variants than those speakers are less aware of (prepositional usage) – albeit only in *Twitter English*. This could thus mean that the American English variants have gained value in social attractiveness and that broadcasts like *CNN* have made young and educated Nigerians more aware of orthographical features. Naturally, this interpretation only holds true if the same processes apply for L2English speakers as those mentioned above for L1English speakers.

5 Conclusion

The investigation of tweets from young English speakers in Nigeria has shown that, most importantly, they tend to use the American prepositions more often, which is particularly striking for feature 6 (*on* before street names). These patterns in NiTE might thus be understood as an artefact of the Twitter data or their genre-specificity, as they are only partially corroborated by the web-based English in the Nigerian component of GloWbE. All the prepositional differences found in GloWbE are statistically significant, which does not, however, mean, that the use of a British or an American preposition is exclusive. Both variants are used alongside one another in Nigeria.

On a much more salient level of language – orthography – the findings are a little bit different. In spelling words, the young Nigerians show higher frequencies of occurrences for the majority of American English orthographic realizations in informal Twitter English, which is not, however, statistically significant. The only exception to the lack of statistical significance is the lemma *labour*, which is used significantly more often in its British English spelling. These findings are not corroborated by the Nigerian component of GloWbE, where the dominance of British English spellings is statistically significant. The significantly higher American English spellings are a minority in GloWbE.

These findings may suggest that among the upcoming generation, much of the traditional linguistic differences may be considered to play a continuing, but diminishing role in the future of the country's online discourse. It may be the case that some Nigerian speakers change their spell checker from British English to American English on their mobile phones to use American English spelling, but on a structural level, American English seems to be the newly emerging standard. There is potential to interpret this development as a consequence of the previous conscious orientation towards community-external norms.

As the analysis has shown, Twitter English can be considered specialized discourse, since it shows genre-specific frequencies in the use of American English forms. This specificity means, however, that generalizations to Nigerian English are impossible, and yet, Twitter appears to be a rich and promising source of urban youth speech in African countries such as Nigeria. The tools to collect Twitter data should attempt to collect as much sociolinguistic meta-information about the tweeters as possible. Even if most Nigerians using Twitter will not provide it with their socio-demographic background, the vast number of collectable tweets will help to differentiate tweeters from one another. Additionally, it may even allow for deletion of so many tweets that keeping the ones with the vital sociolinguistic information is still providing sufficient data worth investigating. A greater quantity of tweets would also allow for a fine-grained differentiation of genres within NiTE. Such an approach requires, however, a broader scope than the current paper has, by including a longer data collection period, more prepositions and high-frequency prepositions to investigate, and a more sophisticated means of statistical assessment, such as multivariate regression analyses, decision trees, cluster analysis, etc.

Notes

¹ Please note that the 2008 version is an unaltered reprint of the 2004 version.

² I am using the term register in a narrow sense as defined and used by Trudgill (1974, 1999) and Wardaugh (1986).

References

- Hofmann, M. (in prep.) 'American English and urban Nigerian youth: Investigating the influence in tweets from Lagos to Kano.' In: Oloruntoba-Oju, T. and Schmied, J. (eds) *Indigenous Grammar and the Rhetoric of African Youth Languages* (REAL Studies 11). Göttingen: Cuvillier. 101-119.
- Adedeji, K. (2015) 'RP in Nigeria: Prestige vs. intelligibility.' In: Opeibi, T., Schmied, J., Omoniyi, T. and Adedeji, K. (eds) *Essays on Language in Societal Transformation: A Festschrift in Honour of Segun Awonusi* (REAL Studies 9). Göttingen: Cuvillier. 29-40.
- Albrecht, S. and Schmied, J. (2015) 'TwitCollect'. Retrieved on 28 November 2017.
- Alo, M. A. and Mesthrie, R. (2008) 'Nigerian English: Morphology and syntax.' In: Rajend, M. R. (ed.) *Varieties of English 4: Africa, South and Southeast Asia*. Berlin: Mouton de Gruyter. 278-339.
- Anthony, L. (2018) *AntConc*. (Version 3.5.7) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net/>.
- Bamman, D., Eisenstein, J. and Schnoebelen, T. (2014) 'Gender identity and lexical variation in social media.' *Journal of Sociolinguistics* 18(2), 135-160.
- Bell, A. (1991) 'Audience Accommodation in the Mass Media.' In: Giles, H., Coupland, N. and Coupland, J. (eds) *Contexts of Accommodation: Developments in Applied Sociolinguistics*. Cambridge: Cambridge University Press. 69-102.

- Bernardini, S., Baroni, M. and Evert, S. (2006) 'A WaCky introduction.' In: Baroni, M. and Bernardini, S. (eds) *WaCky!: Working Papers on the Web as Corpus* (Studi interdisciplinari su traduzione, lingue e cultura). Bologna: Gedit Edizioni. 9-40.
- Burghardt, M. (2015) 'Introduction to tools and methods for the analysis of Twitter data.' *10plus1: Living Linguistics* 1, 74-91.
- Davies, M. (2013) *Corpus of Global Web-Based English: 1.9 Billion Words from Speakers in 20 Countries (GloWbE)*. Available online at <https://corpus.byu.edu/glowbe/>.
- Foulkes, P., and Docherty, G. J. (2000) 'Another Chapter in the Story of /r/: 'Labiodental' Variants in British English.' *Journal of Sociolinguistics* 4(1), 30-59.
- Gries, S. T. (2013) *Statistics for Linguistics with R: A Practical Introduction*. 2nd ed. Berlin: De Gruyter Mouton.
- Gut, U. B. (2008) 'Nigerian English: Phonology.' In: Mesthrie, R. (ed.) *Varieties of English 4: Africa, South and Southeast Asia*. Berlin: Mouton de Gruyter. 35-54.
- Halliday, M. A. K. (1989) 'Spoken and written language.' In: Webster (2007) *The Collected Works of M.A.K. Halliday*. Oxford: Oxford University Press. 63-80.
- Heyd, T. and Mair, C. (2014) 'From vernacular to digital ethnolinguistic repertoire: The case of Nigerian Pidgin.' In: Lacoste, V., Jakob, R., Leimgruber, E. and Breyer, T. (eds) *Indexing Authenticity: Perspectives from Linguistics and Anthropology*. Berlin: De Gruyter Mouton. 242-266.
- Hundt, M. and Nesselhauf, N. (eds) (2007) *Corpus Linguistics and the Web*. Amsterdam: Rodopi.
- Hutchinson, A. (2017) 'Top social network demographics 2017 [Infographic].' Online document. Retrieved on 6 November 2017 <<https://www.socialmediatoday.com/social-networks/top-social-network-demographics-2017-infographic>>.
- Jowitt, D. (1991) *Nigerian English Usage: An Introduction*. Lagos: Longman Nigeria.
- Jowitt, D. (1995) 'Nigeria's national language question: Choices and constraints.' In: Bamgbose, A., Banjo, A. and Thomas, A. (eds) *New Englishes: A West African Perspective*. Ibadan, Nigeria: Mosuro. 34-56.
- Kelly, R. (2009) *Twitter Study - August 2009: Twitter Study Reveals Interesting Results about Usage*. San Antonio, Texas.
- Kilgariff, A. (2007) 'Googleology is bad science.' *Computational Linguistics* 33(1), 147-151.
- Lipsman, A. (2009) 'What Ashton vs. CNN foretold about the changing demographics of Twitter.' Online document. Retrieved on 6 November 2017 <http://blog.comscore.com/2009/09/changing_demographics_of_twitter.html>.
- Mair, C. (2013) 'Corpus-approaches to the vernacular web: Post-colonial diasporic forums in West Africa and the Caribbean.' *Covenant Journal of Language Studies* 1(1), 17-30.
- Mesthrie, R. and Bhatt, R. M. (2008) *World Englishes: The Study of New Varieties*. Cambridge: Cambridge University Press.
- Milroy, J. and Milroy, L. (1985) *Authority in Language: Investigating Language Prescription and Standardisation (Language, Education and Society)*. London: Routledge & Kegan Paul.
- Murthy, D., Gross, A. and Pensavalle, A. (2016) 'Urban social media demographics: An exploration of Twitter use in major American cities.' *Journal of Computer-Mediated Communication* 21(1), 33-49.
- Nguyen, D., Gravel, R., Trieschnigg, D. and Meder, T. (2013) 'How old do you think I am?: A study of language and age in Twitter.' In: *Proceedings of the Seventh*

- International AAAI Conference on Weblogs and Social Media, 8-11 July 2013, Cambridge, Massachusetts, USA. Palo Alto, CA: AAAI Press. 439-448.*
- Opeibi, T. (2011) *Discourse Strategies in Election Campaigns in Nigeria*. Saarbrücken: Lampert Academic Publishing.
- Osisanwo, A. (2015) 'A morphological study of the use of pidgin in selected Nigerian electronic media advertisements.' In: Opeibi, T., Schmied, J., Omoniyi, T. and Adedeji, K. (eds) *Essays on Language in Societal Transformation: A Festschrift in Honour of Segun Awonusi* (REAL Studies 9). Göttingen: Cuvillier. 155-172.
- Pomikálek, J. (2014) *jusText*. Brno: Natural Language Processing Centre of Masaryk University.
- R Core Team. (2018) 'R: A language and environment for statistical computing.' Vienna, Austria. Online document. Retrieved from <http://www.R-project.org/> on 22 November 2018.
- Schettler, T. (n.y.) *Twython*. Available online at <http://www.ling.uni-potsdam.de/~scheffler/twitter/>.
- Schmied, J. (1995) 'National standards and the international corpus of English.' In: Bamgbose, A., Banjo, A. and Thomas, A. (eds) *New Englishes: A West African Perspective*. Ibadan: Mosuro. 337-348.
- Schmied, J. (2012) 'Social digital discourse: New challenges for corpus- and sociolinguistics.' *Topics in Linguistics – Approaches to Text and Discourse Analysis* 10, 43-56.
- Schmied, J. (2015) 'Internet English in Nigeria: New data – new discourses?' In: Opeibi, T., Schmied, J., Omoniyi, T. and Adedeji, K. (eds) *Essays on Language in Societal Transformation: A Festschrift in Honour of Segun Awonusi* (REAL Studies 9). Göttingen: Cuvillier. 189-208.
- Schmied, J. and Opeibi, T. (ed.) (2017) *From the Virtual Sphere to Physical Space: Exploring Language Use in Nigerian Democracy* (REAL Studies 13). Göttingen: Cuvillier.
- Schneider, E. W. (2007) *Postcolonial English: Varieties of English around the World*. Cambridge: Cambridge University Press.
- Sousa, A. and Ivanova, A. (2012) 'Constructing digital rhetorical spaces in Twitter: A case-study of @BarackObama.' *Topics in Linguistics – Approaches to Text and Discourse Analysis* 9, 46-55.
- Stuart-Smith, J. (2007) 'The Influence of the Media.' In: Llamas, C., Mullany, L. and Stockwell, P. (eds) *The Routledge Companion to Sociolinguistics*. New York and London: Routledge. 140-148.
- Stuart-Smith, J., Gwilym, P., Claire T. and Barrie, G. (2013) 'Television can also be a Factor in Language Change: Evidence from an Urban Dialect.' *Language* 89(3), 501-536.
- Tagg, C. and Seargeant, P. (2015) 'Facebook and the discursive construction of the social network.' In: Georgakopoulou, A. and Spilioti, T. (eds) *The Routledge Handbook of Language and Digital Communication*. Abingdon: Routledge. 339-353.
- Taiwo, R. (ed.) (2010) *Handbook of Research on Discourse Behavior and Digital Communication: Language Structures and Social Interaction*. Hershey, PA: IGI Global.
- Trudgill, P. (1974) *Sociolinguistics: An Introduction*. New York: Penguin.
- Trudgill, P. (1999) 'Standard English: What it isn't.' In: Bex, T. and Watts, R. J. (eds) *Standard English: The Widening Debate*. London: Routledge. 117-128.
- Wardaugh, R. (1986) *An Introduction to Sociolinguistics*. Oxford: Basil Blackwell.

Matthias Hofmann has been a postdoctoral lecturer and research associate in English Language & Linguistics at Chemnitz University of Technology since 2015. His research interests are in Language & Culture (sociolinguistics, English in Africa) and in Language & Computers (corpus-linguistics, e-learning, internet English). His current research focuses on linguistic analyses of internet data and of national and subnational variation of Englishes in Africa.

Address: Matthias Hofmann, Chemnitz University of Technology, English Language and Linguistics, 09107 Chemnitz, Germany. [e-mail: matthias.hofmann@phil.tu-chemnitz.de]