

## **Corpus-Approaches to the New English Web: Post-Colonial Diasporic Forums in West Africa and the Caribbean**

**Christian Mair**

University of Freiburg, Germany

### **Abstract**

The present contribution reports on research carried out since November 2011 in the framework of the project 'Cyber-Creole' funded by Deutsche Forschungsgemeinschaft (DFG MA 1652/9). The Cyber-Creole project is complemented by 'RomWeb' (DFG PF 699/4), which is headed by Stefan Pfänder of the Department of Romance Languages and Literatures at the University of Freiburg, Germany. The two projects both use data downloaded from web-based forums, employ identical strategies of data collection and mark-up and have developed a shared search and analysis interface, NCAT (= Net Corpora Administration Tool). The two projects also share a similar general theoretical orientation, being motivated by an interest in how the new media impact on the spread of standard and non-standard varieties of European ex-colonial languages in conditions of economic, political, cultural and media globalisation. In this regard, their work is intended to make a substantial contribution to the emerging research paradigm of the sociolinguistics of globalisation (Blommaert 2010, Coupland, ed. 2010, Mair 2013, Mair and Pfänder (forthcoming)). The present brief survey will introduce the theoretical stance the "English" branch of the project takes in the context of contemporary World Englishes research (section 1). Building on this, I will introduce the analytical tool-kit which we have developed to study computer-mediated communication (CMC) in the diasporas at the centre of our attention (section 2) and describe the sociolinguistic profile of the "Cyber-Jamaican" and "Cyber-Nigerian" developed by two such groups (sections 3 and 4). The conclusion (section 5) will summarise the most important insights and outline perspectives for further work.

Kachru and Nelson, eds. 2009, Kortmann and Schneider, eds. 2004).

### **1. Mobilising EWL research: new models for new types of data**

The study of "varieties of English around the world", the "New Englishes" or "World Englishes" emerged at the intersection of dialectology, sociolinguistics and historical linguistics in the early 1980s and has been among the most vibrant sub-fields of English linguistics ever since. Pioneering works include Bailey and Görlach, eds. (1982) or Trudgill (1982), and the research tradition has achieved provisional culmination recently through the publication of several major handbooks (Kachru,

Work in this tradition has generally been carried out under two self-imposed and not unreasonable constraints. First, the focus was on the description of varieties of English, which were typically understood to be based in territorial speech communities and generally described as abstract and decontextualised linguistic systems through listing their phonological, morphosyntactic, and lexical features. Questions of discourse, situated language use, or language ideologies held by speakers and outsiders, on the other hand, tended to recede into the background. Second, EWL studies were regarded as an essentially monolingual

enterprise. Language contact was recognised as a force in the emergence of the New Englishes to the extent that it left structural residues (e.g. substrate influences), but multilingual practices employed by individual agents were usually not at the centre of attention. This is reflected, for example, in the design of major corpora of the New Englishes, such as the International Corpus of English (ICE), in which multilingual practices are not systematically targeted for collection and non-English passages, where they are included, are treated as extra-corpus material for the purposes of the analysis.

Relying on these useful analytical simplifications, the "varieties"-approach served well to document the linguistic legacy of colonialism, charting the typical course of development of new regional standard and non-standard varieties of the language (Schneider 2007). However, it is arguably less well suited to handle important uses of English which have come to the fore in more recent globalisation, for example the role of English as a *lingua franca* in business and the media or in academia – which is well known –, but also, to mention a more off-beat example, among African refugees in continental Europe (see Goglia 2009 for a study of the function of English in multilingual repertoires developed by African asylum seekers in Italy).

In short: in addition to the more traditional fully fledged and territorially based varieties dominantly used in face-to-face interaction, the study of World Englishes increasingly has to reckon with global linguistic flows of deterritorialised vernacular resources, which come to the fore, for example, in digitally mediated communication, where they are available for the fashioning of subcultural

styles or ethnolinguistic repertoires (on the notion of repertoire, as opposed to variety or -lect, see Benor 2010 or Sharma 2011). The more dominant English is globally, the more heterogeneous it becomes internally. The farther the language spreads, the more it is affected by the multilingual settings in which it is being used. "Natural" links between vernaculars and their territories or primary communities of speakers are becoming weaker, as migrations and media encourage the flow of linguistic resources. On such a view, the English language, the undisputed world language of the present era, is *not* an amalgam of separate varieties, each with its own history and each captured by listing its phonological, morphosyntactic, and lexical properties. Rather, English is *a pool of standard and nonstandard resources* of varying and fuzzy regional and temporal reach. Following the lead of sociolinguists such as Eckert (2008) and Blommaert (2005, 2010) or linguistic anthropologists such as Silverstein (2003), we should aim to add a focus on social styles in interaction to our traditional concerns with sounds, words, and constructions.

A world which is marked by criss-crossing global currents of migration of historically unparalleled intensity and an increasingly all-encompassing global *mediascape* (Appadurai 1996) provides a linguistic ecology in which, at least in principle, *all Englishes are everywhere*. However, this does not mean that the world will therefore become a more homogeneous, Anglophone communicative space. To appreciate this, it is sufficient to look at major metropolitan centres in the English-speaking world, such as London, New York, Sydney or Toronto, which have long been or have recently developed into hot spots of urban

multilingualism. Multilingualism will persist, if in a hierarchically layered structure which de Swaan (2002, 2010) has described as the World Language System. As for English itself, the heightened profile of many previously marginal and non-standard varieties in the public sphere and in the media does not betoken across-the-board de-standardisation and the emergence of an egalitarian polyphony of voices. Unequal relationships and restricted access to linguistic resources persist within the community of monolingual and multilingual users of English, as they do in the world at large.

In sum, the aim must be to *mobilise* the study of World Englishes and make it part of a "sociolinguistics of mobile resources and not of immobile languages" (Blommaert 2010: 180). As Blommaert explains:

The shift from language to resources is crucial in all of this. [...] The consequences of that shift, however, are not yet deeply understood. We must, for instance, accept that abandoning a structural notion of language (a linguists' construct, as we know) compels us to replace it by an ethnographic concept such as *voice*, which embodies the experiential and practice dimensions of language and which refers to the way in which people actually deploy their resources in communicative practice. (2010: 180)

The term *super-diversity* (Vertovec 2007) has been coined to mark the transition from patterns of multilingual and multicultural urban life that dominated until the last third of the twentieth century to a more recent situation, encountered particularly in the world's global cities (Sassen 1994), in which

ethnic and linguistic ties have become more diverse and also more unstable. Alongside the global city, the global media, in particular the World Wide Web, has become a second site of super-diversity. The web is becoming more multilingual (Dor 2004; Danet and Herring, eds. 2007), and, as far as English itself is concerned, more multidialectal. In the anonymous, protean and apparently unbounded spaces opened up by the participatory digital media, nonstandard and vernacular linguistic resources assume enormous importance for linguistic self-styling. In the absence of other clues, people become intensely aware of linguistic difference and variation, and it would be surprising indeed if a more extended *passive* knowledge of variation in World English were to remain the only sociolinguistic consequence of this new situation.

## **2. Web-based communities of practice and the emergence of visual and digital ethnolinguistic repertoires**

The three digitally based diasporic communities of practice which are the focus of the present research are the contributors to [www.jamiacans.com](http://www.jamiacans.com), [www.nairaland.com](http://www.nairaland.com) and [www.cameroon-info.net](http://www.cameroon-info.net), three web forums centred around former African and Caribbean colonies of the British (and, in the case of Cameroon, also of the French) colonial empire. In addition, all three forums can be described as diasporic, as one of their main purposes seems to be to provide platforms for communication among Jamaican, Nigerian and Cameroonian residents, globally dispersed migrant communities originating from these three nations and a very

heterogeneous group of additional individuals displaying various kinds of interest or involvement in these three communities. The forums were selected from among a number of alternatives because of the large amount of material available for inspection and subsequent download from the web, because of the diversity of the contributors and the multifarious nature of the topics which are being discussed, which distinguishes them from single-issue forums devoted to specific topics such as, for example, reggae or dancehall music. In order to create a fixed and finite database and a stable corpus-linguistic working environment, the forum data were stored offline in autumn 2008, yielding the following large corpora corresponding to the respective sites.

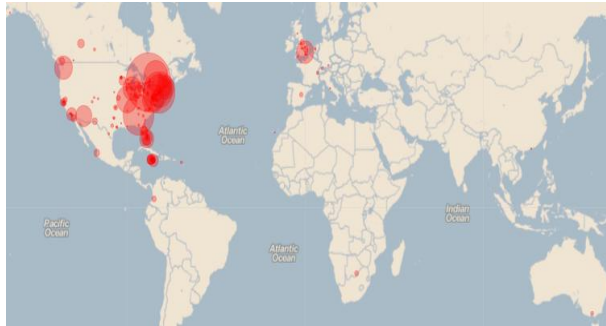
- Corpus of "Cyber-Jamaican" (CCJ) ([← www.jamaicans.com](http://www.jamaicans.com))  
2128 members, 252 015 posts, 16.9 millions tokens, 2000 – 2008
- Corpus of "Cyber-Nigerian" (CCN) ([← www.nairaland.com](http://www.nairaland.com))  
11718 members, 244 048 posts, 17.3 millions tokens, 2005 – 2008
- Corpus of "Cyber-Cameroonian" (CCC) ([← www.cameroon-info.net](http://www.cameroon-info.net))  
3140 members, 179 563 posts, 22.1 millions tokens, 2000 – 2008

These three forums and, by implication, the three corpora which document their linguistic practices represent three different post-colonial sociolinguistic constellations. CCJ, the Jamaican diasporic corpus, is written in English and Jamaican Creole (in this quantitative and hierarchical order), with the two languages frequently blending into each other in ways which are interestingly different from the creole-

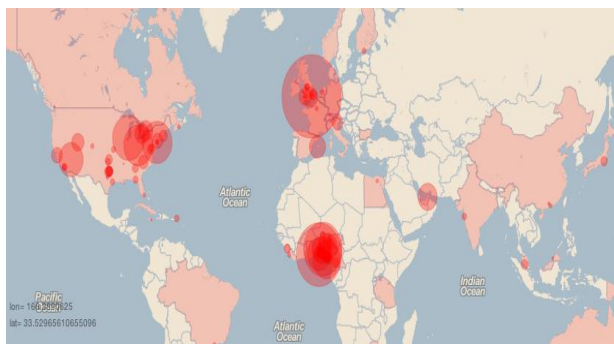
English continuum encountered on the ground in Jamaica itself. The term Cyber-Jamaican can thus be understood fairly literally as a designation of the linguistic repertoire shared by the participants. English is the dominant language also in the Nigerian CCN. Nigeria's and Nigerians' intense multilingualism, however, forbids a similarly literal understanding of the term Cyber-Nigerian. Cyber-Nigerian comprises many sub-repertoires which are not necessarily accessible to all participants. Chief among them is Nigerian Pidgin, but the country's major indigenous languages, such as Yoruba, Igbo and Hausa, also play some role, as do nonstandard forms of English not directly related to Nigeria, such as African American Vernacular English, introduced into the forum by Nigerian emigrants to the US and their US-born descendants. The special feature of the Cameroonian corpus is that while it has less input from indigenous languages except for the local variety of West African Pidgin and locally developed urban slang of the *Camfranglais* type, it features two European ex-colonial languages, namely French and English.

In addition to concordancing, NCAT, the search and analysis interface developed for the project, helps us visualise sociodemographic characteristics of the contributors. Our goal is to "place" our communities in the apparently boundless cyber-space and, ultimately, to write the linguistic geography of this new communicative habitat. A very simple visualisation is based on identifying the geo-location of key contributors to the forums on a map and representing the amount of language produced by them by circles of varying diameter. Figures 1a to one 1c

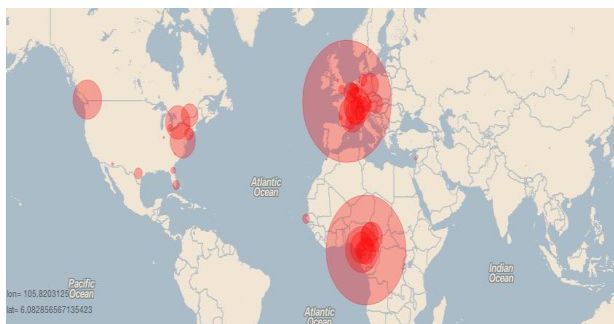
below show that the three forums investigated cast their cyber-spatial anchors in real-world geography quite differently.



**Fig. 1a:** www.jamaicans.com / CCJ: regional spread and intensity of activity of 480 core participants localised at city level



**Fig. 1b:** www.nairaland.com / CCN: regional spread and intensity of activity of 630 core participants localised at city level



**Fig. 1c:** www.cameroon-info.net / CCC: regional spread and intensity of activity of 1000 core participants localised at city level

The forum which represents the most clearly post-colonial constellation is the Cameroonian one. It comprises a sizable community of contributors resident in Cameroon; the biggest single focus of activity, however, is Paris, the dominant colonial metropolis, with London, the secondary metropolis for the minority "Anglophone" regions of the country, making a modest showing only. The post-colonial shadows of empire are much less evident in the constellations represented in 1a and 1b, which show West Africa and the Caribbean in conditions of globalisation. Nigeria, Africa's demographic giant, is an expected major focus of activity in the Nigerian forum, rivalled only by London, the former colonial metropolis. Strong additional focuses of activity, such as the United Arab Emirates in the Gulf, Toronto or several immigrant destinations in the US, could not have been predicted by the colonial historical experience but are the result of more recent currents of migration promoted by globalisation. In contrast to the Nigerian forum, the Jamaican one is characterised by a relatively weaker position of the home island itself and a fading importance of London, the colonial metropolis, in comparison to the more recent immigrant destinations for Jamaicans in the US and Canada.

### 3. Cyber-Jamaican between transcription and evocation: from the oral to the visual vernacular in three stages

As in many other genres of CMC, non-standard spellings and vernacular literary practices are very prominent in our post-colonial forum data, as well, as I will

illustrate with examples from the Corpus of Cyber-Jamaican here. In our work, we do not see spelling primarily as a technique for the visual representation of the sound of a dialect or other non-standard variety, but as much more, namely a partly autonomous semiotic resource germane to the new medium of digital writing and hence as a form of social action (Sebba 2007; Jaffe, Androutsopoulos and Sebba, eds. 2012). Thus, the attempted mimetic representation of "genuine" or "authentic" island Creole, as illustrated in example (1) below, is merely the simplest case in a wide range of spelling practices we are encountering:

- (1) Satdeh time...braps, beef soup done, seh shi neva go a college fi bwile no yam. One mawnin mi a hat up little mackrel an some pepper an shi tell mi seh mi a tink up di house.

*[Saturday(s) – there you have it! – the beef soup is ready, and she says she didn't go to college to cook yams. One morning, I am heating up a little mackerel and some peppers and she tells me that I am stinking up the house]*

This qualifies as a reasonably realistic **transcription** of Jamaican Creole within the confines of the 26 letters of the Roman alphabet and the conventions of English-based "grassroots" spelling of Jamaican Creole. The text would sound natural if read aloud by a competent speaker.

The majority of Creole or Creole-influenced passages in the corpus, however, are **evocations** rather than realistic transcriptions. This means that they aim to style an authentic Jamaican Creole voice which must be convincing first and foremost on the visual plane of the digital medium and not necessarily in live pronunciation.

- (2) so what mi can use **fi** clean *gems, stones, diamonds, etc??* mi have de silver/gold cleaning solution but it nuh seem **fi** do nutten **fi** de **stone dem**..sometimes it look wussa dan when it *went* in. mi read up pon de net an some *sites* say use dish washing liquid..others say that's a big no no. what **unu** use clean **unu** stones? *would it be better* **fimi carry it go a jeweler??** how much dat would cost?

*[so what can I use to clean gems, stones, diamonds, etc.? I have the silver/gold cleaning solution but it does not seem to do anything for the stones ... sometimes it looks worse than when it went in. I read up on the net and some sites say 'use dish washing liquid' ... others say that's a big no no. What do you use to clean your stones? Would it be better for me to take it to the jeweller's?]*

This passage does not make sense in terms of the creole-English continuum of face-to-face interaction in contemporary Jamaica, as it combines a very high frequency of (sometimes very) basilectal features (infinitival marker *fi*, pluralisation with *-dem*, 2<sup>nd</sup> person plural pronoun *unu*, serial verb construction *carry ... go* – here printed in bold) with a high incidence of acrolectal ones (e.g. inflected plurals, *went*, conditional *would* – here printed in italics). It is therefore difficult to imagine it being read out aloud, nor is it likely that the author intended it to be read out aloud when composing it. The text should not be seen as a variety of Jamaican English or Jamaican Creole – a standard by which it would be inauthentic –, but as a use of Jamaican Creole linguistic resources appropriate to the digital medium. At this level, basilectal features of Jamaican Creole are no longer socially stigmatised alternatives to their more prestigious informal-mesolectal or

standard-acrolectal variants. The rules of the sociolinguistic game have changed in a digital community of practice which our sociodemographic investigations have shown to be dominated by professional women: selected elements of the basilect have become visual markers of in-group status in the diaspora and function as contextualisation cues in new ways in the new digital communicative environment.

Jamaican Creole has a phonemic spelling system, the so-called Cassidy-LePage orthography, which is widely practiced among linguists and propagated by the Jamaican Language Unit of the University of the West Indies, but has not been able to replace widely established spellings based on Standard English orthography. The example just discussed largely conforms to these conventionalised non-standard *ad hoc* spellings for Jamaican Creole (referred to as "grassroots" spelling above), which also represent the communal baseline among the forum participants. There are, however, more consciously experimental and idiosyncratic approaches, such as the following, which emphasises the autonomous visual dimensions of Cyber-Jamaican:

- (3) oww manee peeps inn dem caribbean household versus mzungu oousehold. iff itt tekk two ar more caribbean wage earnas wukkinn more owa dan mzungu fi mekk more dan mzungu oousehold widd less peeps wat iss da artikkle seyinn bout da caribbean peeps qualitee aff life

*[How many people are there in those Caribbean households versus white people's households? If it takes two or more Caribbean wage earners working more hours than white people to make more than a white household with fewer people what is the article*

*saying about the Caribbean people's quality of life?]*

As the English gloss makes clear, we are dealing with a relexified version of a fairly typical written English complex sentence here. There is the occasional grammatical creolism (e.g. *fi mekk*, uninflected noun in plural reference in *owa / hours*), but the majority of the grammatical constructions fully conform to the Standard English pattern. The alienation effect produced on the reader is primarily visual. Unusual grapheme combinations such as <ww> in *oww / how* have no conceivable phonetic correlate. The <kk> spellings might be taken to signify shortness of the preceding vowel in *mekk*, but are clearly redundant in this function in *artikkle*. The spelling <ff> in *aff / of*, in suggesting voicelessness, explicitly contravenes the actual pronunciation. With some stretch of the imagination, example (2) could have been characterised as a consciously "anti-formal" usage in the Jamaican context in the sense of Allsopp (1996). Allsopp distinguishes "informal" use of (mostly mesolectal) creolisms, which is the norm in spoken Caribbean Englishes in private and informal contexts even among educated speakers, from a more salient type of anti-formal use (of often basilectal features), which he defines as "[d]eliberately rejecting Formalness; consciously familiar and intimate; part of a wide range from close and friendly through jocular to coarse and vulgar; any Creolized or Creole form or structure surviving or conveniently borrowed to suit context or situation. When such items are used an absence or a wilful closing of social distance is signalled. *Such forms survive profusely in folk-proverbs and sayings, and are widely written with conjectural spellings in attempts at realistic*

*representations of folk-speech in Caribbean literature.*" (Allsopp 1996: lvii)

Example (3), finally, though obviously making use of selected creolisms, does not instantiate any variety of Jamaican English or Jamaican Creole in a straightforward sense. It draws on many linguistic resources available to a member of the Jamaican diaspora resident in the United States: for example the originally US slang abbreviation *peeps* which has more recently become a feature of global "Netspeak" or, somewhat more unexpectedly, the word *mzungu*, originally a Kiswahili term for "European" or "foreigner", but now widely current as a slightly derogatory reference to white people in Southern and Eastern Africa. As the user who produced this text is a high-profile contributor to the forum, some of his idiosyncracies can be shown to have spread among other participants, as well.

#### 4. "For authenticity jus sprinkle it with o ..." – Nigerian Pidgin as a visual resource in the digital diaspora<sup>1</sup>

One prominent feature in the Nairaland data is the high profile of Nigerian Pidgin, both among contributors resident in Nigeria itself and in the international Nigerian diaspora – and notwithstanding the fact that its use is officially discouraged in the forum participation guidelines.<sup>2</sup> Of course, some

---

<sup>1</sup> This section incorporates some joint work with Dr. Theresa Heyd, Freiburg. See Heyd and Mair (forthcoming).

<sup>2</sup> "Nairaland is an English language forum. English happens to be our official language in Nigeria. [I]t also happens to be the language of the web. Please make every effort to use clear English at all times: [...] Avoid pidgin English, but 'put am inside italics' if you need to use pidgin English."

contributors resident in Nigeria and many more from the diaspora have less than full competence in Nigerian Pidgin, for example because they are second- or third-generation emigrants and use of Pidgin was discouraged in the home. This has given rise to a lively grassroots movement for using the forum as a digital classroom for the teaching and learning of Pidgin. The recommendation in the heading of the present section emerged from precisely such a context, the "Pidgin English Thread" (<http://www.nairaland.com/246047/pidgin-english-thread>), which was active with about 700 contributions between March 2009 and July 2012 (and, unlike similar ventures before it, is therefore not included in CCN). Participant *mruknaiboy*, who in his very name indicates his dual affiliation to the UK and Nigeria, indicates that he feels his competence is deficient because he apparently had to learn Pidgin mainly from interaction on the web and invites fellow contributors to the thread to correct him:

(4) nairaland wetin dey, i no sabi blow pidgin like una sake of say dem born me for jand na im make me no sabi am well, i go try sha, i dey learn am small small from una for dis thread, mak una correct me if una see mistake for my pidgin, how una dey?

Rather than immediate corrections, the contribution earns him praise:

(5) Bros eh, de way una nak the pidgin sef suprice me. You no carry last at all. Any way, make we kukuma dey jolly de go as laif allow us. Welcome home, mruknaiboy.

Participant *talina*, who according to her profile is based in Canada, is far less fluent

---

(<http://www.nairaland.com/6/nairaland-forum-participation-guidelines>)



in Pidgin, and also somewhat less conversant with the facts of everyday life in Nigeria, as is shown, for example, by the fact that she has to google the meaning of NEPA, the name of Nigeria's national power provider. Nevertheless her interest in the Pidgin thread is strong enough for her to join the digital conversation with a post consisting of only the stock phrase *wetin dey*. In the ensuing exchanges she soon realises that her Pidgin is probably not up to the task and freely admits her incompetence – in what can only be described as "mock Pidgin" rather than the real thing:

(5) Thats [*wetin dey*] all I can say in pidgin Me no speaky

Since the rules of the game were defined in the very first post of the thread ("This na pidgin english thread. This thread na for we people wey don tire for oyibo talk talk."), *talina* is challenged, if in a facetious way:

(6) Whhhaatt!!! Your yan don shok mi like NEPA. Unbelieveabl. So, na for which language we go fit yan wit you naa

To which she responds by offering to withdraw from the thread. The offer is not taken up, because another participant encourages her to stay on and invites her to learn on the web:

(7) hahaha abeg talina no leave o [...] u try sha! abi mek we teach u small pidgin?

As promised, feedback is given, and a first simple and helpful rule is provided in order to make the digital Pidgin "sound" more natural, or rather – since we are in the digital / visual domain of the forum and no longer in the auditory sphere of vernacular face-to-face interaction – make the Pidgin "look" more idiomatic.

(8) @**talina**, Yessoo but for authenticity jus sprinkle it with o...so u can say *fine o*

Nairaland is thus not only used by participants to showcase their own usage. The forum also serves as a platform to inquire about questions of vocabulary, pronunciation, and Pidgin usage in general. These exchanges are made possible by the community structure of the forum as described in the introduction: it serves as a meeting place for Nigerian locals – who are usually viewed as the authorities on proper Pidgin usage – and other, more heterogeneous participants in the diaspora: first and second generation emigrants and people whose affiliation with Nigeria is even more indirect, for example through a spouse or people in their peer group. As a consequence, the desire to learn Pidgin is often motivated to a certain degree by the participants' Nigerian heritage; but another, frequently expressed motivation is the wish to speak Pidgin in their offline interactions, e.g. with friends who are fluent in Pidgin. In sum, these examples show that there is a grassroots 'market' for learning Pidgin online.

Based on these data, we can investigate how a previously oral vernacular makes the transition into the written domain of computer-mediated communication. How do norms for writing Pidgin evolve? How does the spread of this digitised vernacular proceed? And centrally, how do members of an online community such as Nairaland assess the linguistic competence and authenticity of different Pidgin users? This latter question is of particular importance given the globalised nature of the community: with users hailing both from the Nigerian heartland and the suburbs of New

York City, with residents of Nigeria, recent emigrants and other members with quite heterogeneous ethnic backgrounds, the issue of ethnolinguistic gate keeping becomes central. In digital communication, questions of authenticity, reliability and trustworthiness have played an important role from the earliest days due to the delocalized nature of computer-mediated communication and its relative anonymity (or pseudonymity). As has been noted in recent discussions of such repertoires, charting the metalinguistic perceptions that surround this kind of language use is a powerful analytical tool. Thus, Benor notes that:

it is important to point out that ethnic groups do not exist a priori but are socially-constructed phenomena that come into being through the discourse of members and non-members. When describing an ethnic group's distinctive linguistic repertoire, researchers might begin by identifying ideologies surrounding the group and its boundaries in the discourse of core members, marginal members, and non-members. (2010: 170)

The specific analytical concept which we have developed for the analysis of vernacularised web-forum data is the *digital ethnolinguistic repertoire* (Heyd and Mair forthcoming). Although the default language of forum communication on Nairaland is (Standard) English, the data exhibit high degrees of stylistic variability involving features from several standard and non-standard Englishes, such as Nigerian English, British or American English, Nigerian Pidgin or African American Vernacular English. Nigerian Pidgin, in particular, is involved in various types of

code-switching and code-mixing, as are – to a lesser extent – Nigeria's major indigenous languages, and occasionally even German, Dutch or French (as major languages of some of the destinations of the emigrants). In the analysis of our data, the classical reference descriptions of the relevant languages or "varieties" (e.g. Faraclas 1996 for Nigerian Pidgin) provide useful orientation. However, in our view it is not a productive approach to measure practices on the web against the yard-stick of such reference descriptions. On the one hand, there is an inevitable mimetic deficit whenever vernaculars are reduced to writing, and there is an inevitable truncation of speaker competence when vernaculars are used by second- and third-generation Nigerian emigrants who have only limited input in the diaspora. On the other hand, in some respects the digital vernaculars are actually richer than the ones spoken on the ground, because writers may exploit visual resources (e.g. expressive spellings) or explore new modes of contact and combination among varieties and languages for which the new medium provides opportunities.

Rich and varied though it presents itself, "Cyber-Nigerian" is thus not a "variety", or an ethnic dialect of a more general "Netspeak" (itself a notion which hardly survives closer scrutiny). Rather, it should be seen as similar to the ethnic "style repertoire" described for Asians in Britain by Sharma (2011) or to the "ethnolinguistic repertoire" described for orthodox Jews in the United States by Benor (2010). In her introduction, Benor defines her notion of the ethnolinguistic repertoire by demarcating it against an older understanding of "ethnic dialect":

Ethnolinguistic repertoire [...] is defined as a fluid set of linguistic resources that members of an ethnic group may use variably as they index their ethnic identities. In this paper, I describe the ethnolinguistic repertoire approach and show how it solves five theoretical problems with the notion of 'ethnic language variety' as a bounded entity.

- Intra-group variation.
- Intra-speaker variation.
- Out-group use.
- Delineating the ethnic group.
- Delineating 'ethnolect'. (Benor 2010: 160)

As the analyses in sections 3 and 4 have made clear, the five problems mentioned address our most important concerns in the analysis of the forum data, too. We assume that extending the repertoire approach to the analysis of CMC is easy and appropriate. Trivially, *digital ethnolinguistic repertoires* are special because they lack the phonetic dimension. But as we have seen, this truncation is partly compensated for by the visual semiotic potential of expressive orthography or CMC-specific iconography. There is also a massive truncation of real-life ethnolinguistic repertoires in CMC at the macro-level. With around 500 languages in use in face-to-face interaction, Nigeria is an intensely multilingual society. However, of these 500 languages, only English, Pidgin, Yoruba, Igbo and Hausa regularly make it into the Nairaland forum. Other indigenous languages seem to be excluded for lack of a suitable script or because of insufficient demographic weight. On the side of gains, on the other hand, linguistic resources are freed up for new uses in CMC, if only because language is an even more important

means of self-authenticating and indexing ethnicity in the fluid and ill-defined space of the digital forum than in face-to-face interaction on the ground.

## 5. Conclusion

Research on the language of CMC is indebted to several traditions. Pioneering early studies aimed to characterise the language of the internet as a whole, for example as Netspeak (cf., e.g. Crystal 2001 [2<sup>nd</sup> ed. 2006]). This has been and will continue to be useful up to a point, but clearly fails to address the results of the one internet trend which is probably most interesting to study for linguists at the moment, namely the increasingly multilingual, multi-dialectal and multi-modal nature of the World-Wide Web (e.g. Danet and Herring, eds. 2007). Another influential tradition has been the discourse-analytical one, which – among other things – has been preoccupied with CMC as a hybrid form of discourse combining features of orality and literacy (e.g. Baron 2008). In sociolinguistics, linguistic variation on the web attracted little attention as long as the focus of research was on the spontaneous use of the vernacular in face-to-face interaction. This has certainly changed, however, in more recent qualitative and interactional approaches, which are more suited to the investigation of vernacular performances and conscious styling (cf., e.g., Androutsopoulos, ed. 2006; Squires 2010).

With research that is situated at the intersection of corpus linguistics, World Englishes studies and the discourse analysis of CMC, the present paper has aimed at integrating these traditions, making two important points. First, an analysis of non-

standard spelling (using the Jamaican material as a study example) has shown that the task facing the investigator is not to scour the web for faithfully rendered transcriptions of spoken non-standard language (Jamaican Creole, in our case), but to study the ways in which non-standard linguistic resources of various and occasionally unexpected origin are combined by diasporic communities of practice in order to fashion their own ethnolinguistic repertoires suitable to the digital medium. In this sense, Cyber-Jamaican, Cyber-Nigerian and Cyber-Cameroonian are not cut off from the sociolinguistic orders prevailing on the ground, but the digital medium certainly makes possible partial autonomy – most obviously in the substitution of the visual expressiveness of spelling for accent and prosody, but also in the generally different assignation of sociolinguistic prestige and stigma to the "same" linguistic feature online and off-line, as was seen in the discussion of the examples in section 3 above and as is demonstrated more extensively and systematically in Mair (2011), Mair and Lacoste (2012), and Moll (forthcoming). Second, the present paper is a plea for the compilation of web-based CMC corpora. Searching the web direct will of course get us to the passages we have analysed for as long as they remain archived by the forum moderators, and interesting data will continue to accumulate under our very eyes. However, we still need to "freeze" the data into a corpus through download because this is the only way to ensure that the results of the studies remain replicable and different quantitative and qualitative-interactional explorations of the material undertaken at different times actually build up into a coherent over-all

analysis. Also, readily available strategies for searching the web directly are not usually sophisticated enough to fully satisfy the (socio)linguist. As was shown, offline corpora and NCAT, a customised search and visualisation tool, were necessary to draw the maps of cyber-space represented in Figures 1a to 1c above. If the offline CMC corpora derived from the Web are large enough for the task at hand, as they are in the Cyber-Creole and RomWeb projects, these advantages outweigh their one disadvantage, which is the fact they cannot document the most recent developments in CMC practices emerging post-download.

## References

- Allsopp, Richard (1996): *Dictionary of Caribbean English Usage*. Oxford. OUP
- Androutsopoulos, Jannis ed. (2006): Special issue on "The Sociolinguistics of Computer-Mediated Communication." *Journal of Sociolinguistics*, 10(5)
- Appadurai, Arjun (1996): *Modernity at Large: Cultural Dimensions of Globalisation*. Minneapolis MN: University of Minnesota Press
- Baron, Naomi S. (2008): *Always On: Language in an Online and Mobile World*. Oxford: OUP
- Benor, Sarah Bunin (2010): "Ethnolinguistic Repertoire: Shifting the Analytic Focus in Language and Ethnicity," *Journal of Sociolinguistics*, 14, 159-183
- Blommaert, Jan (2010): *The Sociolinguistics of Globalization*. Cambridge: CUP
- Blommaert, Jan (2005): *Discourse*. Cambridge: CUP
- Bailey, Richard; Görlach, Manfred eds. (1982): *English as a World Language*. Ann Arbor MI: University of Michigan Press

- Coupland, Nikolas, ed. (2010): *The Handbook of Language and Globalization*. Malden MA: Blackwell
- Crystal, David (2001 [2<sup>nd</sup> ed., 2006]): *Language and the Internet*. Cambridge: CUP
- Danet, Brenda; Herring, Susan eds. (2007): *The Multilingual Internet: Language, Culture, and Communication Online*. Oxford: OUP
- de Swaan, Abram (2002): *The World Language System: A Political Sociology and Political Economy of Language*. Cambridge: Polity
- de Swaan, Abram (2010): "Language Systems," in: Coupland, Nikolas (ed.): *The Handbook of Language and Globalization*. Malden MA: Blackwell. 56-76
- Dor, Daniel (2004): "From Englishization to Imposed Multilingualism: Globalization, the Internet, and the Political Economy of the Linguistic Code," *Public Culture*, 16: 97-118
- Eckert, Penelope (2008): "Variation and the Indexical Field," *Journal of Sociolinguistics*, 12, 453-476
- Faraclas, Nicholas (1966). *Nigerian Pidgin*. London: Routledge.
- Goglia, Francesco (2009): "Communicative Strategies in the Italian of Igbo-Nigerian Immigrants in Italy: a Contact-linguistic Approach," *Sprachtypologie und Universalienforschung / Language Typology and Universals*, 62, 224-240
- Heyd, Theresa, and Christian Mair (forthcoming, 2013). "From Vernacular to Digital Ethnolinguistic Repertoire: The case of Nigerian Pidgin," in: Lacoste, Véronique; Leimgruber, Jakob; Breyer, Thiemo (eds.): *Indexing Authenticity: Sociolinguistic Perspectives*. Berlin/New York: de Gruyter.
- Jaffe, Alexandra; Androutsopoulos, Jannis; Sebba, Mark eds. (2012): *Orthography as Social Action: Scripts, Spelling, Identity and Power*. Berlin: Mouton de Gruyter
- Kachru, Braj B.; Kachru, Yamuna; Nelson, Cecil L. eds. (2009): *The Handbook of World Englishes*. Malden MA: Blackwell
- Kortmann, Bernd; Schneider, Edgar (with Kate Burridge, Rajend Mesthrie, Clive Upton) eds. (2004): *A Handbook of Varieties of English*. 2 vols. Berlin: Mouton de Gruyter
- Mair, Christian (2011): "Corpora and the New Englishes: Using the 'Corpus of Cyber-Jamaican' (CCJ) to Explore Research Perspectives for the Future," in: Meunier, Fanny; De Cock, Sylvie; Gilquin, Gaëtanelle; Paquot, Magalie (eds.): *A Taste for Corpora: In Honour of Sylviane Granger*. Amsterdam: Benjamins, 209-236
- Mair, Christian; Lacoste, Véronique (2012): "A Vernacular on the Move: Towards a Sociolinguistics of Mobility for Jamaican Creole," *Cahiers de Linguistique*, 38(2), 87-110
- Mair, Christian; Pfänder, Stefan (forthcoming): "Using Vernacular Resources to Create Digital Spaces: Towards a Sociolinguistics of Diasporic Web Forums," in: Auer, Peter; Hilpert, Martin; Stukenbrock, Anja; Szmrecsanyi, Benedikt (eds.): *Space in Language and Linguistics: Geographical, Interactional, and Cognitive Perspectives* [working title]. Berlin/New York: de Gruyter
- Mair, Christian (2013): "The World System of Englishes: Accounting for the Transnational Importance of Mobile and Mediated Vernaculars," *English World-Wide*, 34: XX-XX
- Moll, Andrea (2012): *Jamaican Creole Goes Web: Sociolinguistic Styling and*

- Authenticity in a Digital Yaad*. PhD,  
University of Freiburg
- Sassen, Saskia (1994 [4<sup>th</sup> ed., 2012]): *Cities  
in a World Economy*. Thousand Oaks  
CA: Sage / Pine Forge Press
- Schneider, Edgar (2007): *Postcolonial  
English: Varieties around the World*.  
Cambridge: Cambridge University Press
- Sebba, Mark (2007): *Spelling and Society:  
The Culture and Politics of Orthography  
around the World*. Cambridge: CUP
- Sharma, Devyani (2011): "Style Repertoire  
and Social Shift in British Asian  
English," *Journal of Sociolinguistics*, 15,  
464-492
- Silverstein, Michael (2003): "Indexical  
Order and the Dialectics of  
Sociolinguistic Life," *Language and  
Communication*, 23, 193-229
- Squires, Lauren (2010): "Enregistering  
Internet Language," *Language in  
Society*, 39, 457-492
- Trudgill, Peter; Hannah, Jean (1982 [5<sup>th</sup> ed.  
2008]): *International English: A Guide  
to Varieties of Standard English*.  
London: Arnold
- Vertovec, Steven (2007): "Super-diversity  
and its Implications," *Ethnic and Racial  
Studies*, 30, 1024-1054

## **About the Author**

Prof. Dr. Dr. h.c. Christian Mair is a Professor of English Linguistics in the Department of English, University of Freiburg Germany.

Email. [christian.mair@anglistik.uni-freiburg.de](mailto:christian.mair@anglistik.uni-freiburg.de)