CCN ("Corpus of Cyber-Nigerian"): Concordancing, annotation and visualisation in a very large web-derived CMC corpus

Christian Mair
Englisches Seminar, Universität Freiburg

DFG MA 1652/9-1: 'Cyber-Creole'
Roadmap

1. Introduction: scenes from contemporary life – mobile speakers, mobile languages
2. The research agenda: from "English as a World Language" (EWL) to the "Sociolinguistics of Globalisation" / The World System of (Standard and Non-Standard) Englishes
3. [www.nairaland.com](http://www.nairaland.com) and CCN: from the community to the corpus (and back)
4. Conclusion and outlook
1. Introduction: scenes from contemporary life – mobile speakers, mobile languages
(i) the Italian from England and the Nigerians from the US meeting in cyberspace …
(Topic: "How to convince an Ibo girl that I am truly in love with her," www.XXX.com, 2008)

XXX ["sono di CITY IN SOUTHERN ITALY ma vivo in UK"]: please please please i beg you to help me!!!
I am an Italian white man who has met this beautiful Ibo girl. [...] this girl i am in love with, she is married too to a English man, but he doesn't treat her the way she should be. we are very closed friends and [...] any help from you Ibo people will be really appreciated. i thank you all in advance for all your advises.
Houston TX and New York City joining the conversation …

YYY: Oka what now?
Are you sure you are not on the rebound? Cause you sound sort of desperate. Also the chick is married and you dare disrespect her marriage by involving herself in it? C'mon now fella, i know you smarter than that. My diagnosis, is for you to lay off the chick, and really think to yourself that you ACTUALLY LOVE her, not infatuated by her.
For your future reference, it is Igbo not ibo.

ZZZ: if you married her you'D be headed for divorce soon … maybe you'll be seeking help from Wollof folks this time.
(ii) The Nigerian from Treviso saving the day ...

XXX: nonostante tu dici che il tuo Italiano non sia buono, ti esprimi benissimo e sai scrivere bene. complimenti. [...] 

AAA [= "Gospel from CITY IN SOUTHERN ITALY"?]: Grazie per il tuo complimento. una domanda solo quanto anni sei e quanto anni e lei?

XXX: io 33 e lei 26
2. The research agenda: from "English as a World Language" (EWL) to the "Sociolinguistics of Globalisation" / The World System of Standard and Non-standard Englishes
From "varieties of English" to Englishes in a multilingual global language ecology …

- varieties → styles → ideologies
- structural features → communicative resources
- (CMC) spelling → (digital) literacy practice
- (monolingual) native-speakers → multilingual and / or truncated repertoires
- local / vernacular community → deterritorialised community of practice
- face-to-face interaction → mediated communication
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 and primary communities are becoming weaker, as migrations and media encourage the transnational and global flow of linguistic resources.

Even in a world in which "all Englishes are everywhere", there is no egalitarian polyphony of voices. Unequal relationships and restricted access to linguistic resources persist.

- the World Language System as "a surprisingly efficient, strongly ordered, hierarchical network, which ties together – directly or indirectly – the 6.5 billion inhabitants of the earth at the global level" (de Swaan 2010: 56)
English – contact language for the world ...

- **hyper-central** language: English, the hub of the world language system
- **super-central** languages: French, Hindi, Mandarin, Spanish, Turkish, Arabic, ...
- **central** languages: Dutch, Finnish, Cambodian, ...
- **peripheral** languages: 6,000+
The World System of Standard and Non-standard Englishes (Mair 2013)

... the English Language Complex as ... a surprisingly efficient, strongly ordered, hierarchical constellation of varieties, styles and registers which ties together – directly or indirectly – the 1-billion-plus regular users of English at the global level ...
• the hub / hyper-central variety:
  Standard American English

• super-central varieties:
  - standard: BrE, AusE, IndE, NigE, …
  - non-standard: AAVE, Jam CreoleE, popular LondonE, …
  - further domain-specific ELF uses:
    science, business, international law

• central varieties:
  - standard: IrE, NZE, JamEng, …
  - non-standard: US "Southern," …

• peripheral varieties: all traditional rurally based non-standard dialects, plus a large number of ex-colonial varieties including pidgins and creoles
3. [www.nairaland.com](http://www.nairaland.com) and CCN: from the community to the corpus (and back)
Data:

- Corpus of "Cyber-Jamaican" (CCJ)  
  www.jamaicans.com  
  2,128 members; 252,015 posts  
  16.9 million tokens  
  2000 – 2008

- Corpus of "Cyber-Cameroonian" (CCC)  
  www.cameroon-info.net  
  3,140 members; 179,563 posts  
  22.1 million tokens  
  2000 – 2008

- Corpus of "Cyber-Nigerian" (CCN)  
  www.nairaland.com  
  11,718 members; 244,048 posts  
  17.3 million tokens  
  2005 – 2008
Search item a: Nairaland is a globally dispersed digital diaspora.
A look at the underlying raw data …

<table>
<thead>
<tr>
<th>docs / country</th>
<th>docs / city</th>
<th>docs / month</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;none&quot; - 70539</td>
<td>&quot;none&quot; - 158917</td>
<td>…</td>
</tr>
<tr>
<td>Nigeria - 63811</td>
<td>London - 19646</td>
<td>04.2007 - 11797</td>
</tr>
<tr>
<td>US - 54570</td>
<td>Lagos - 17128</td>
<td>05.2007 - 13925</td>
</tr>
<tr>
<td>UK - 33411</td>
<td>Chicago - 5969</td>
<td>06.2007 - 11429</td>
</tr>
<tr>
<td>Canada - 4548</td>
<td>New York - 5328</td>
<td>07.2007 - 10159</td>
</tr>
<tr>
<td>Spain - 2515</td>
<td>Pt Harcourt - 4592</td>
<td>08.2007 - 7651</td>
</tr>
<tr>
<td>Germany - 1881</td>
<td>Abuja - 3590</td>
<td>09.2007 - 10077</td>
</tr>
<tr>
<td>Malaysia - 1397</td>
<td>Columbus - 3370</td>
<td>10.2007 - 7999</td>
</tr>
<tr>
<td>Ghana - 1360</td>
<td>Palma - 2491</td>
<td>11.2007 - 7147</td>
</tr>
<tr>
<td>Italy - 1206</td>
<td>Liverpool - 2238</td>
<td>12.2007 - 12632</td>
</tr>
<tr>
<td>Australia - 1112</td>
<td>Las Vegas - 2058</td>
<td>01.2008 - 10937</td>
</tr>
<tr>
<td>UAE - 960</td>
<td>Cardiff - 1629</td>
<td>02.2008 - 8284</td>
</tr>
<tr>
<td>Fiji - 747</td>
<td>Toronto - 1574</td>
<td>03.2008 - 6905</td>
</tr>
<tr>
<td>Sierra L. - 741</td>
<td>Washington - 880</td>
<td>04.2008 - 7456</td>
</tr>
<tr>
<td>Netherl. - 618</td>
<td>Houston - 773</td>
<td>05.2008 - 7402</td>
</tr>
<tr>
<td>Ireland - 476</td>
<td>Suva - 747</td>
<td>06.2008 - 11052</td>
</tr>
<tr>
<td>China - 475</td>
<td>Brisbane - 744</td>
<td>07.2008 - 10833</td>
</tr>
<tr>
<td>Belgium - 455</td>
<td>Brisbane - 744</td>
<td>08.2008 - 7857</td>
</tr>
<tr>
<td>Norway - 440</td>
<td>Ibadan - 700</td>
<td>10.2008 - 365</td>
</tr>
<tr>
<td>...</td>
<td>Boston - 635</td>
<td>12.2008 - 365</td>
</tr>
</tbody>
</table>
Search item a (# posts by country): Nairaland is a globally dispersed *and gendered* digital diaspora.
Search item *abi*: Nigerian Pidgin is globally dispersed …
Search item **stuffs**: … more dispersed in fact than Nigerian English.
Search item *Mum and Dad*: Empire still casts its shadow.
Search item *Mom and Dad* shows selective assimilation of Nigerian immigrants to US norms.
AAVE on the move: Search item [j*] ass in CCN & GloWBE
(1) if you are over 25 and your dad talks to you in a rude manner... Dude, it's high time you "win your own bread" just like seun said. Get your ass out of the house. (CNN, hot-angel [5754])

(2) GOOD FOR EMEKA, IF I WERE HIM NA POLICE I GO TAKE ARREST YOUR DUMB ASS ! NEXT TIME YOUR STUPID ASS WILL NOT BET ON UNCERTAINTY, ESPECIALLY ON A TEAM THAT BARELY MADE IT TO THE SECOND ROUNDS. (CCN, Sweet T [5409])

(3) My views on divorce are: I am willing to divorce any stupid-ass man that messes with my heart. (CNN, hot-angel [5754])
4. Conclusion & outlook: discourse analysis of CMC

"[…] networked writing questions the adequacy of the feature-based approach and spoken language bias that have dominated conceptions of language change in sociolinguistics. […] a change of scale in the volume and publicness of vernacular writing; a diversification of old and new vernacular patterns; an extension of written language repertoires, and a concomitant pluralisation of written language norms." (Androutsopoulos 2011: 153)

"In domains of unregimented writing, stylistic appropriateness is opened up to localised negotiation, for example with regard to spelling and punctuation or the representation of regional dialects." (Androutsopoulos 2011: 155)

→ What is it that we "hear" when we "read" the digitally mediated vernacular voices?
Conclusion & outlook: World Englishes research

weakness of the postcolonial nation state $\leftrightarrow$ weakness of the emerging national standard?

inversion of prestige of Standard New Englishes and the corresponding pidgins and creoles in the diaspora and the global mediascape?

Crisis of the Outer Circle?
References


