Multilingual resources and authentication in the computer-mediated communication of U.S.-Nigerians

The role of African American Vernacular English in the repertoire

Mirka Honkanen, M.A.
mirka.honkanen@anglistik.uni-freiburg.de
University of Freiburg, Germany
Today’s talk

(1) Introducing my corpus of Nigerian digital communication

(2) The Net Corpora Administration Tool (NCAT)

(3) My project: describing the digital linguistic repertoires of Nigerian immigrants in the U.S.

(4) First findings: U.S.-Nigerians’ usage of African American Vernacular English (AAVE)
The CyberCreole/RomWeb project

• My PhD is associated with a larger research project at Uni Freiburg (2011–2014)
• Investigating “contact varieties and the effects of digital globalized communication”
• Principal investigators: Christian Mair & Stefan Pfänder
• Corpus collection and maintenance: Daniel Alcón
My dataset

• Downloaded from a Nigerian online discussion forum nairaland.com

• Meeting point for Nigerians all over the world
• Discussions on a wide variety of topics
**Int'l Friendly: Nigeria Vs Cameroon 3.0 (Live) - Sports - Nairaland**

The Super Eagles will look forward to consolidate on their 0-2 defeat against the Democratic Republic of Congo when they face the Indomitable Lions of Cameroon in an international friendly at the 11,000-capacity Edmond Machertens Stadion, Brussels, Belgium. On Saturday, 11th of October, 2015.

But before then, let's take a look at the five cracking games involving Nigeria and Cameroon...

**Re: Int'l Friendly: Nigeria Vs Cameroon 3.0 (Live)**

CAMEROON vs NIGERIA (1984 AFCON FINAL)

The Indomitable Lions of Cameroon won their first African Cup of Nations title at the 14th edition of the tournament at the expense of the Green
Corpus facts:
airaland corpus

2005–2014
19,169,867 posts
843.1 million tokens
302,714 members
Comparison with another large CMC corpus—Corpus of Global Web-based English

**GloWbE corpus (Davies 2013)**
- **42.6** million words from Nigeria
- Data from websites located in Nigeria
- 60% informal blogs, 40% more formal online texts

**My Nigerian corpus**
- **843.1** million words from Nigerians
  - **20x** the size of GloWbE’s Nigerian data
- Data from Nigerians all around the world
  - Language in globalization
- Very informal and interactional discussion forum data

[Map of the world with highlighted regions]
Downside of informality

• ‘Messiness’ of data, typical for online language:
  • Multilingualism
  • Nonstandard orthography
  • Emoticons
  • Truncated sentences
  • Quoting, etc.

➢ NO automatic syntactic parsing or tagging
Tagging the corpus

• Information about gender taken over from user profiles
• Manual tagging for geolocation on the most frequent users
  • Searching for phrases such as “here in” or “live in”
  • At the moment, 1,869 frequent users tagged for location at a country-level, 877 at a city-level
• Work in progress!
Search string: a|the|is|be|i|are|na|dey|to

- Not balanced in terms of location
- For visualizing the spread of a feature
- For comparing competing (e.g. orthographic) variants

- Global community
- Blue circles for country-level, red circles for city-level location
- **Size** of each circle roughly represents the # of posts from each location
Geolocation map – North America
“Immigrant population from Nigeria” by the Migration Policy Institute

- Our online data reliably reflect patterns found “on the ground”
- Bringing the same level of sophistication and credibility to linguistic studies

Source: MPI analysis of 2010–12 ACS, pooled (MPI 2014: 7)
Net Corpora Administration Tool (NCAT)

- Geolocational maps produced by a web-based tool NCAT
- Developed at Uni Freiburg
- Retrieval and analysis platform
- Specifically for working on discussion forum data
- Searching, visualization, reports of data
• Searching for strings and substrings
• Searching for specific user(s)
• Searching for specific time periods
• Retains the hierarchical structure (whole NL > subfora > threads > posts)
• Context available
Displaying results: raw view

The raw view imitates the layout of the forum.
Displaying results: text view

The **text view** shows plain text.

- Its actually *ebonics* Just a type of SLangBut ya english is spoken differently all around the world Even in England they have their own slang.

- I don't think you should allow it to affect your English writing ya we have ours in UK but street data where it ends, it's just gon cost u alot if u write *ebonics* abt aucanics in ur papers.

- Well all you have to implement and be aware of is what we ebonic speakers call "Code Switching". That is you speak or use a varying degree of *ebonics* depending on social context. I'm currently on Industrial Placement and I work with a Government Agency and don't use *ebonics* when I'm in work related mode. If you happen to associate with different people of different backgrounds code switching will become sub-conscious to you before you know it. Being in a foreign country definitely helps because you are mostly going to work study and work with people from other racial and social background than yours. I think the subject of conversation may also initiate code switching as I've found even when I'm with my homies. As I said once it's a sub-conscious thing you barely have control over it. Your brain just flip modes and you begin to use the appropriate language form. If you've read my post on other subjects I've replied to you can see there's a chance of mode and pattern here.

- Dear Continental Africans et al. Please do not be alarmed, this letter is another African American pep talk. African Americans put these out every other week or so. We like to scare each other into striving higher and higher. Yawn. The person actually slips into *Ebonics* LOL Smiles.
The **concordance view** shows only the immediate context.
Flexible tagging

• Each researcher can create his/her own tags (‘categories/labels’) which can be added to relevant posts
• Searching for/within tagged posts
• Great feature for more qualitative research questions
• E.g. When interacting with the data, I tag interesting posts for ‘authenticity’ (another key interest)
  ➢ Easy to find later
Developing the NCAT

• Ongoing project, some debugging to do
• Possible further features:
  • New search filters (e.g. gender, quotes, emoticons…)
  • Different data exportation options
  • Language recognition
My project – basics

• Aim: describing the **digital multilingual repertoires** of Nigerians in the United States

• Study of **language in globalization**: data from intersection of migration and computer-mediated communication (CMC)

• Jan Blommaert’s (2010) “**sociolinguistics of globalization**”: focus on mobile language resources and individuals
On repertoires

• “Ethnolinguistic repertoire” (Benor 2008, 2010) = “a fluid set of linguistic resources that members of an ethnic group may use variably as they index their ethnic identities” (Benor 2010: 160)

• Heyd and Mair (2014: 251) apply to CMC settings:
  • “Digital ethnolinguistic repertoires” lack the phonetic dimension
    ➢Partly compensate with nonstandard spellings and iconography

• Extending concept further to multilingual situations
Relevant languages and varieties – Nigeria vs. diaspora

• Nigeria
  • Multilingual, multiethnic
  • Former British colony: English as official language
  • Other languages used in informal and home contexts:
    • 400–500 indigenous languages (Yoruba, Hausa, Igbo...)
    • Nigerian Pidgin
      • Contact language developed during colonial period
      • Appears on the forum frequently
      ...despite previously stigmatized status and lack of standardized orthography

Indigenous mother tongue

Nigerian Pidgin

Nigerian English + other English varieties

Typical educated Nigerian in Nigeria
• Nigerians in the U.S. diaspora
  • Ca. 380,000 people
  • Mostly recently immigrated
  • Relatively well educated
    (Migration Policy Institute 2014)

• Linguistic behavior different from their compatriots’ in Nigeria?
  • Restricted access to Pidgin and indigenous languages
    ➢ “Truncated repertoires” (Blommaert, Collins & Slembrouck 2005)
  • Easier access to standard and nonstandard varieties of English spoken in North America
    • Standard American English
    • Colloquial American English
    • African American Vernacular English (AAVE)
    ...

Role of AAVE in the repertoire?

• **Method** for investigation:

Choosing 50 (mainly) frequent users
Going through a **large part** of posts **qualitatively**
Confirming findings with the NCAT **quantitatively**

• Problem: Many AAVE features are in fact omissions
  ➢ Cannot be searched for in an untagged corpus

• Individuals’ repertoires ➢ Digital ethnolinguistic repertoire of U.S.-Nigerians

**Looking for...**
• Nonstandard syntax
• Nonstandard lexis
• Metalinguistic comments
• Users’ own descriptions of language skills

**Omissions in AAVE**
• No copulas
• No genitive suffix
• Little verbal inflection
<table>
<thead>
<tr>
<th>LOCATION</th>
<th>AGE</th>
<th>TRIBE</th>
<th>SEX</th>
<th>BORN IN</th>
<th>LOCATION</th>
<th>AGE</th>
<th>TRIBE</th>
<th>SEX</th>
<th>BORN IN</th>
</tr>
</thead>
<tbody>
<tr>
<td>New York</td>
<td>17</td>
<td>Hausa/AA</td>
<td>M</td>
<td>USA</td>
<td>New York</td>
<td>23–32</td>
<td>Yoruba/Igbo</td>
<td>M</td>
<td>NGR</td>
</tr>
<tr>
<td>Boston</td>
<td>15–19</td>
<td>Igbo</td>
<td>F</td>
<td>NGR</td>
<td>Tampa</td>
<td>23–25</td>
<td>Igbo</td>
<td>F</td>
<td>NGR</td>
</tr>
<tr>
<td>Boston</td>
<td>Adult</td>
<td>Mixed</td>
<td>F</td>
<td>NGR</td>
<td>Chicago</td>
<td>30–45</td>
<td>Yoruba</td>
<td>M</td>
<td>NGR</td>
</tr>
<tr>
<td>LA State</td>
<td>20–25</td>
<td>Igbo</td>
<td>F</td>
<td>USA</td>
<td>Houston</td>
<td>20–25</td>
<td>Igbo</td>
<td>F</td>
<td>USA</td>
</tr>
<tr>
<td>New Orleans</td>
<td>20</td>
<td>Yoruba</td>
<td>F</td>
<td>NGR</td>
<td>TX State</td>
<td>18–25</td>
<td>?</td>
<td>M</td>
<td>BEL</td>
</tr>
<tr>
<td>Indianapolis</td>
<td>25–30</td>
<td>Yoruba</td>
<td>M</td>
<td>NGR</td>
<td>GA State</td>
<td>15–16</td>
<td>Igbo</td>
<td>F</td>
<td>USA</td>
</tr>
<tr>
<td>TX State</td>
<td>15–20</td>
<td>Yoruba</td>
<td>F</td>
<td>NGR</td>
<td>MD/TX</td>
<td>25–30</td>
<td>Yoruba</td>
<td>M</td>
<td>NGR</td>
</tr>
<tr>
<td>?</td>
<td>15–20</td>
<td>Yoruba</td>
<td>F</td>
<td>NGR</td>
<td>CA State</td>
<td>?</td>
<td>Yoruba?</td>
<td>M</td>
<td>NGR?</td>
</tr>
<tr>
<td>TX State</td>
<td>?</td>
<td>?</td>
<td>M</td>
<td>NGR</td>
<td>CA State</td>
<td>17–18</td>
<td>Yoruba?</td>
<td>M</td>
<td>NGR</td>
</tr>
</tbody>
</table>
Preliminary findings: AAVE in the repertoire

I have identified **five user types**, differing in:

- Amount of AAVE influence
- Grammatical or only lexical AAVE features?
- Distinctive grammatical features, or features that appear in many other varieties as well?
User type 1: Non-user

• Seven users (~1/3) do not employ any AAVE features
• Many negative attitudes and stereotypes about African Americans and AAVE circulating in the Nairaland community

“butchered English”  “uneducated akata stuck in the ghetto”
“no family values”  “welfare”
“Ebonics speaking thugs, baby mamas and baby daddies”
“synonymous with crime, drugs, miseducation and dysfunctional family households”  “Improper lazy English (ebonics)”
User type 2: Occasional users

• Rely mainly on lexical and some non-distinctive grammatical features

“i speak english fluently, ofcourse, but i also speak ibonics with some of my friend.” (2006, f)

“I talk like that sometimes, u know when u trying to pull the ‘shaniqua’ look.” (2006, f)

➢ AAVE being used consciously for identity purposes
Competent users (~1/3)

- Extensive usage of complex grammatical and lexical features
- User type 3: Consistent expert users
- User type 4: Occasional expert users

Note: None of the three consistent AAVE users contributed many posts

➢ Not a repertoire for an easy integration into the community?
User type 5: Borrowers

• Only employ AAVE lexis
• Parts of it have apparently already crossed over to American English as well

• E.g. “ain’t as a negated form of be”
  • Pervasive feature of all AAVE varieties
  • Exists (“not pervasive but not extremely rare”) in colloquial American English (Kortmann & Lunkenheimer [eds.] 2013)
• Has definitely spread outside of North America
• Has clearly been taken up also by Nigerians who do not use any (other) AAVE features

78,903 posts from 9,944 users, 40,135 of which tagged for geolocation (50.9%)
“ain’t […] before a main verb”

- Pervasive feature in urban AAVE (Wolfram 2013)
- Extremely rare in colloquial American English (Simon 2013)

“ain’t know”|“aint know”:
111 documents from 73 users

Spillover of AAVE to the African continent
Example: auxiliary verb finna

• AAVE feature from the interface of grammar and lexis
• Short for “fixin’ to”
• Indicates that something is happening “in the immediate future” (Green 2002: 70)

• Benefits of a large dataset:
  earlier version of the corpus (17.3 M words) finds 7 relevant hits; current corpus (843.1 M words) offers 277 instances
• Different spellings finna, fina, feena, fixna
finna|fina|feena|fixna in the corpus

277 documents by 129 members
Last qualitative example...

- Young Nigeria-born male living in California
- 4,918 posts in 2009
- Interesting usage of AAVE
- Competently writes entire posts in Standard English...

---

Politics / Re: Fashola Launches Surveillance Helicopters
by [Author] 7:02am On Dec 10, 2009

Yeah, I don't grasp the purpose of this helicopter unless someone can explain it to me in bits and pieces. I wouldn't deny it being for a good cause, I just can't imagine how it would be effective in curbing chaos if it breaks out. We already have a hard time commuting to-and-fro emergencies by land, let alone flying helicopters through the air. If havoc bobs its ugly head at Yaba market would be Helicopter be able to find a suitable landing zone? If not, then how can it take action?

Yes, I completely agree that nothingness does imply a lack of existence, as I said before. However, a lack of nothingness did not just come about by way of "nothing", so something must have caused that sudden disappearance of nothingness. This is where the intersection is felt. The Prime Mover itself is that intersection. It would be befuddling to think that the Prime Mover itself also existed while there was nothing, as this would completely wallop the idea of a "nothing". In short, there was a something all along.

All of this aside, my own argument for God's own "divine existence" is that the existence of God is perfectly independent of time. Time, as we know, is the yard stick for measuring the length of one's "normal" existence. Lol
...in Nigerian Pidgin...

**Fashion / Re: A Close Up At Miss Nigeria**

by [username] 2:55am On Dec 12, 2009

No mind me abeg. I jus dey make sure say no b ghost i dey see. i dey kampe. i for holler u but i no sabi

---

**Wetin we do ooo madam, sey na lie i dey lie? you fit correct me if i'm wrong. the maga run go marry madam wey go chop hin money, wey no sabi cook or give am rough sex.**

---

**How You Take Make Geisha?**

by [username] 10:54pm On Jun 14, 2009

I greet una sabi sabi people wey dey for neraland. abeg una, how you take make gaysha stew?
...and in Yoruba.

Romance / Re: Lets Toast In Our Native Language! by unnamed 2:10am On Dec 10, 2009

E wo baby yen bo se n ju sibi, ju soun. gbe sibi, gbe soun. oya gba sibe. gbe mi.

Re: Wishing A Happy Birthday To Two Lovely Ladies: Sisi Jinx And Karmamod :) I

ile la n wo ki a to so omo ni oruko.

well i had a dream sha

ileri ayo yio se

lojo kan

olorun mimo ti so
• Occasionally employs some AAVE grammar and lexis
• AAVE features usually cluster together

- Deletion of *have*
- *ain’t* as negation of *have*
- Negative concord
- Possessive *they*
- *ain’t* as negation of *be*
- Habitual *be*
Grammatical features
• Deletion of *be* before an adjective phrase
• Regularized *was* for past *be*
• Negative concord
• Deletion of *be* before a noun phrase
(Green 2002)

Lexical features
• Pseudo (nonreflexive) pronoun *ass*
• *nigga*
• *baby mama* (‘baby’s mama’)
• *bus one’s nuts* (‘have an orgasm’)
(Smitherman 2002)
User A: “ain’t nobody feena be jealous of pretty boy”

User B: “what is feena?”

User A: “fixing to, about to, getting ready to, going to, gonna. i don’t talk like that in person though, omg for real. i’m feena go to the gym marra of fact”
Selective use of clusters of AAVE features
Awareness of AAVE
AAVE plays a part in his playful online repertoire
Comment on linguistic assimilation:

AAVE usage need not amount to an ‘inauthentic’ identity
Coming up…

• Analysis of c. 30 more individual repertoires
  ➢ Extending to group-level ethnolinguistic repertoire
• Nigerian Pidgin in the digital repertoire
• Comparing NigE vs. AmE vs. BrE (quantitative?)
• Descriptions of negotiation of authenticity (qualitative!)

...
Thank you for your attention!

Kiitos mielenkiinnostasi!

Questions? Comments? Suggestions?
References


References


Davies, Mark. 2013 *Corpus of Global Web-Based English: 1.9 billion words from speakers in 20 countries*. Available online at [http://corpus.byu.edu/glowbe/](http://corpus.byu.edu/glowbe/).


