Web forum corpora on globalized varieties: Notes from the user perspective

Theresa Heyd
Albert-Ludwigs-Universität Freiburg
February 14, 2013
Technische Universität Dortmund
Sociolinguistic context of the project

Changing research paradigms:
- Third wave sociolinguistics:
Increased attention to indexicality and social meaning, to sociocultural and constructivist factors in variation, to construction of style, stance and identity... (e.g. Eckert 2008, Bucholtz and Hall 2005, Rampton 1995)
- Sociolinguistics of globalization:
From locally anchored indigenous varieties to ethnolinguistic repertoires/mobile resources (e.g. Blommaert 2010, Coupland 2010, Benor 2010)
- CMC studies:
From a niche discipline with a highly specific agenda to a fully integrated and substantial part of sociolinguistics (e.g. Squires 2010, Danet and Herring 2007)
Why CMC?

“The boundary between private and public life-domains is less clear than it was previously, and mass-mediated language is often based on informality and intimacy as well as on formality and distance. (...) The media are increasingly inside us and us in them.” (Coupland 2007: 28)

→ CMC as an integral part of the “mediascape” (Appadurai 1990) that globalized varieties operate in
→ online communities as a repository for, and a catalyst of, digitally mediated vernaculars
CyberCreole and RomWeb:  
Some assumptions

• Linguistic diversity is gradually increasing on the Internet (Danet & Herring 2007):
  
  On a global scale: from Anglophone to multilingual settings  
  For individual languages: from standard to nonstandard varieties

• Superdiversity and the digital sphere: “migration movements from the 1990s onwards have coincided with the development of the Internet and mobile phones, and these have affected the cultural life of diaspora communities of all kinds.” (Blommaert and Rampton 2011)

• The sociotechnical conditions of CMC are conducive to the emergence of
  
digital
  deterritorialized  
  diasporic

           vernaculars
Evidence of

- heightened **metalinguistic awareness** (Herring 1996)
- an (early-stage) process of **commodification** (Johnstone 2009)
- **ethnolinguistic repertoires**: “a fluid set of linguistic resources that members of an ethnic group may use variably as they index their ethnic identities.” (Benor 2010: 160)
- emerging **superdiversity**: “the diversification of diversity”; “a level and kind of complexity surpassing anything previously experienced in a particular society” (Vertovec 2007)
Some areas of investigation

Linguistic factors:
• Orthographic variation
• Lexical features
• Semantic fields
• Discourse structure
• Narrative episodes
• …

Social factors:
• Gender
• Geolocation
• Race/ethnicity
• Social networks
• …
Publications and work in progress


Mapping the community
Mapping the community: gender vs. geolocation

Interaction between the top 200 participants, coded by gender

Gender of top 1000 Nairaland users by geolocation

→ A community of practice that is dominated by male interaction
→ But: strong interaction between geolocation and gender
Race and ethnicity
Race and ethnicity

Racial terms and semantic shift...

...and narratives of belonging:
me, my papa na oyibo man i no go lie..ma mama na Nigeria woman..i was born n bred in Nigeria..ever so proud to b nigerian..nigerian for life but i wont deny ma father's land as well.
Semantic fields:
the metacommunicative lexicon

<table>
<thead>
<tr>
<th>form</th>
<th>yab</th>
<th>kowatiate</th>
<th>bosh</th>
</tr>
</thead>
<tbody>
<tr>
<td>flow</td>
<td>no fit shout</td>
<td>kowa</td>
<td>opata</td>
</tr>
<tr>
<td>yarn</td>
<td>yabbing</td>
<td>blow</td>
<td>fabu</td>
</tr>
<tr>
<td>misyarn</td>
<td>amebo</td>
<td>relate</td>
<td>ote</td>
</tr>
<tr>
<td>murder</td>
<td>broken</td>
<td>make I hear</td>
<td>knack</td>
</tr>
<tr>
<td>slaughter</td>
<td>tori</td>
<td>word</td>
<td></td>
</tr>
<tr>
<td>hola</td>
<td>origo</td>
<td>na yam</td>
<td></td>
</tr>
<tr>
<td>phonetics</td>
<td>warri</td>
<td>get mouth</td>
<td></td>
</tr>
<tr>
<td><strong>grammar</strong></td>
<td>gist</td>
<td>expo</td>
<td></td>
</tr>
<tr>
<td>tok</td>
<td>expatiate</td>
<td>tabon</td>
<td></td>
</tr>
</tbody>
</table>
Semantic fields: the metacommunicative lexicon

E.g. *grammar*: “unnecessarily convoluted, incomprehensible language; mumbo jumbo”:
I hate to dey yarn *grammer*, y una no dey like yarn broken?
Abeg make I go find for my dictionary. All this *grammar* today, are mind blowing.

<table>
<thead>
<tr>
<th></th>
<th>all meanings</th>
<th>literal meaning</th>
<th>NP meaning</th>
<th>other meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>grammar</td>
<td>336</td>
<td>114</td>
<td>195</td>
<td>27</td>
</tr>
<tr>
<td>grammer</td>
<td>110</td>
<td>35</td>
<td>71</td>
<td>4</td>
</tr>
<tr>
<td>gramma</td>
<td>21</td>
<td>6</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>grama</td>
<td>6</td>
<td>0</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>all spellings</td>
<td>473</td>
<td>155</td>
<td>286</td>
<td>32</td>
</tr>
</tbody>
</table>

Contrastive: metalinguistic analysis in the other corpora: Camfranglais:
moi paticulierement je *whitise* avec les blancs sinon ils ne me comprennent pas ,mais avec les refres et resses ,je *speak* comme au pays
je crois que si tu pif mon style de *tchate* c'est tout simplement parce que tu ya ce que je *tell*, et ceci parce que mon ma facon de *topo* t'es aussi familiere.
Issues and conceptual questions

...regarding:

• Methodology: implications of corpus work, quantitative/qualitative approaches, complementary methods...

• Genre: choice, delineation, demographics...

• Locating the users: space vs. place and the nature of deterritorialized discourse...
Methodology

The corpus comes alive? Possibilities of doing (quasi-)ethnographic work with CMC corpus data; ethical and methodological limitations?

From...

...to...

...and
Genre

• Making abstractions about the data: what are we analyzing? A variety, a genre, a community of practice?

• Genre demographics: what do we/should we/can we know about the communicants?

• Web forums: a web 1.0 relic, or a persistent CMC genre? How to align with social media material/multimodal resources?
Worldwide Trends
- Anden

winner winner winner winner winner! Eagles U don win oh winner! #AFCON2013final #NaJia

#NaJia is now trending in South Africa trendsmap.com/za

HILLBROW IS CRAZY TONIGHT #AFCON2013 #AFCON2013FINAL #AFCON2013Winner #BossoKeMang #TeamStephenKeshi #AfconChamps #NAIJA #AfconChamps

RT @Super_Spotter: Nigeria !!!!!!! pic.twitter.com/VhX4oUV6p

#NaJia

© 2013 Twitter. Über uns Hilfen Bedingungen Privatsphäre Blog Status Apps Ressourcen Jobs Interessenten Unternehmen Medien Entwickler Adressbuch
Geolocation as an extralinguistic factor that
- is (relatively) easily available for extraction, analysis and visualization
- is relevant for the sociolinguistic exploration of diasporic communities and their language use

Problems and limitations:
- Accuracy
- Snapshot effect: unable to capture the high mobility/super-diversity of users
- Explanatory power: how does geolocational analysis mesh with more complex social factors (race, ethnicity, citizenship, identity)?
- Deterritorialization: how can geolocational analysis be reconciled with notions such as (physical) place vs. (digital) space, global/local/glocal, mobility/ubiquitous computing?