CyberCreole/RomWeb: An Overview

Theresa Heyd
(Daniel Alcón)
Albert-Ludwigs-Universität Freiburg
February 14, 2013
Technische Universität Dortmund
RomWeb  
CyberCreole

Contact varieties and the effects of
digital globalized communication
Funding period: 2011 - 2014

Stefan Pfänder
DFG Pf699/4
Oliver Ehmer
Philipp Dankel

Christian Mair
DFG Ma 1652/9
Theresa Heyd
Andrea Moll

Daniel Alcón and others...

Andean Spanish
(Peru, Argentina, Bolivia, Ecuador)
West African French varieties
(Congo, Cameroon)

Jamaican Creole
West African English varieties
(Nigerian Pidgin, Cameroonian Pidgin)
Shared research goals

Cross-linguistic perspective on
• Diversification and pluricentric standardization processes of the world languages English, Spanish and French in post-colonial settings
• Transformation of locally anchored vernaculars under the effect of globalization, diasporic usage and computer-mediated communication
Specific research goals

RomWeb:
• Selection, stylization and emerging conventions in the choice of linguistic structures
• Spatio-temporal dynamics of the online communities
• Cross-linguistic modeling

CyberCreole:
• Orthographic, morphosyntactic and lexical conventions
• Authenticity and authentication
• Effects of crossing, enregisterment and commodification
The Data

The project is based on (a) web forums organized in (b) large-scale corpora.

(a) Why web forums?
- Easy availability of large-scale datasets; genre characteristics such as conversational structure of threads, (rudimentary) social network features, topic subsections

(b) Why a corpus approach?
- Creating a persistent dataset that is amenable to systematic analysis (moving target problem); most work in the field so far geared toward qualitative analysis/ethnographic datasets or small-scale corpora
The Data

www.jamaicans.com
2128 members
252015 posts
16,9 million tokens
2000 - 2008

www.cameroon-info.net
3140 members
179553 posts
22,1 million tokens
2000 – 2008

congopage.com/forums
1002 members
47474 posts
8,9 million tokens
2002 - 2012

www.forosenperu.com
871 members
171665 posts
4,4 million tokens
2007 - 2009

www.forosperu.net
8438 members
374258 posts
19,9 million tokens
2005 – 2008

www.futbol-argento.com.ar
1017 members
159315 posts
7,8 million tokens
2007 - 2008

hinchadasdebolivia.mforos.com
526 members
27530 posts
2,1 million tokens
2006 - 2008

foro.univision.com/ecuador
674 members
16099 posts
2 million tokens
2005 - 2012

skyscraperlife.com/latin
3059 members
404408 posts
25 million tokens
2006 - 2012

English
French
Spanish
Corpus compilation

Crawler

Data

Indexer

Index

User Interface
User interface

First generation: Linguistic Search Engine, Matthias Koch and Prof. Georg Lausen, Databases and Information Systems, Freiburg University
Second generation: Net Corpora Administration Tool (NCAT), Daniel Alcón, Christian Mair, Stefan Pfänder

• Web-based
• Shared → cross-linguistic perspectives; facilitates collaborative work
• Various search parameters: string/substring; members; date range; raw vs. ‘clear’ data...
• Various levels of discourse complexity display (concordance vs. conversational context)
• Metadata: gender, geographical location...
• Annotation/editing for corpus users
• Embedded tools for spatiotemporal visualization
NCAT

http://cyber-creole.phil2.uni-freiburg.de/ncat/index.php
KING GUZU

Re: Why is the man the head of the family?
« #20 on May 31, 2003, 06:29 PM »

It's very funny that people have tried to dignify there difference to this in diverse ways but it smacks of ignorance to even think on the contrary. Right from time immemorial, God Almighty said "let us make man in our image and after our likeness" and after this WONDERFUL creation called man God found out that man was lonely and decided to make a help for him in the form of a woe-man since her creation neither was she wise nor God has known peace and you now think that she should head the family? Just the little opportunity she had to take an important in the 'wiseness' decided to consort with Satan and agreed that the 'apple' was fit for man to eat and see where we are today. Abeg make we manage the 'shift' they have put us in its enough? who knows what they will do next?

legs

Managing Your In-laws
« on June 20, 2005, 12:12 PM »

It's a common myth that dealing with your in-laws is a nightmare, especially when you are a woman. So Nairaland users help me out because some of need to know - personal or borrowed experiences will be appreciated, as well as tips on how to handle difficult situations.

And please no one should divert this topic to the issue of virginity. Abeg.
**Search**

<table>
<thead>
<tr>
<th>String/Substring</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mrug Ads</td>
<td>546</td>
</tr>
<tr>
<td>Plombs Sautès</td>
<td>1546</td>
</tr>
<tr>
<td>Un Autre Tunisien</td>
<td>2546</td>
</tr>
<tr>
<td>chosen vessel</td>
<td>3546</td>
</tr>
<tr>
<td>coolshade1</td>
<td>4546</td>
</tr>
<tr>
<td>Busta</td>
<td>5460</td>
</tr>
<tr>
<td>Zandra1</td>
<td>8461</td>
</tr>
<tr>
<td>spottl</td>
<td>5462</td>
</tr>
<tr>
<td>chigo</td>
<td>5463</td>
</tr>
<tr>
<td>jkpretty</td>
<td>5404</td>
</tr>
<tr>
<td>titlook</td>
<td>5465</td>
</tr>
<tr>
<td>Pain</td>
<td>5466</td>
</tr>
<tr>
<td>bulului</td>
<td>5467</td>
</tr>
<tr>
<td>Chuckdoo</td>
<td>5468</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Member</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zandra1</td>
<td>546</td>
</tr>
</tbody>
</table>

**Re: Igbo music**

- #23 on: April 20, 2007, 09:31 AM

> check radiopalwine.com.

**Re: Can You Pass The Test?**

- #19 on: April 26, 2007, 03:23 AM

> For real jaguar1, what mohawkchic said is right. Now we can separate the men from the boys. Keep up with them good values of ours cause they always amount to good. I know they are men that think like You but at times they are rare to find. Your baby is really lucky. Hope You two remain happy with each other. (I don't see a word?) my hat for You.

**Re: What Rights Has Your Wife Over Your Mobile Handset Or Vice Versa?**
### Manager
A place to edit the members' metadata and to add/edit "locations" (country, region, city, street).

### Members
Select a forum:

<table>
<thead>
<tr>
<th>Select a forum</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

### Locations

<table>
<thead>
<tr>
<th>Country</th>
<th>Code</th>
<th>Link Wikipedia?</th>
<th>Pop.</th>
<th>Area</th>
<th>Nr. assoc. members</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antigua and Barbuda</td>
<td>ATG</td>
<td><img src="wikipedia" alt="link" /></td>
<td>44</td>
<td>83039</td>
<td>44</td>
</tr>
<tr>
<td>Algeria</td>
<td>DZA</td>
<td><img src="wikipedia" alt="link" /></td>
<td>238174</td>
<td>32854150</td>
<td>238174</td>
</tr>
<tr>
<td>Azerbaijan</td>
<td>AZE</td>
<td><img src="wikipedia" alt="link" /></td>
<td>8260</td>
<td>8352021</td>
<td>8260</td>
</tr>
<tr>
<td>Albania</td>
<td>ALB</td>
<td><img src="wikipedia" alt="link" /></td>
<td>2740</td>
<td>3153731</td>
<td>2740</td>
</tr>
<tr>
<td>Armenia</td>
<td>ARM</td>
<td><img src="wikipedia" alt="link" /></td>
<td>2820</td>
<td>3017661</td>
<td>2820</td>
</tr>
</tbody>
</table>
Plans for public/open access?

• Middle-term: remote guest access to existing corpora by request

• Long term: public availability of software/web interface for purpose-built corpora

→ Overcoming technical and legal (privacy/copyright) issues