Le Diasema

Lexical Diachronic semantic maps

Representing and explaining meaning extension

A short introduction to the project
Le Diasema

• Duration: December 2016 – December 2018

• Main research question: how semantic maps make significant predictions about language change at the lexical level?

• Funding schemes
Le Diasema

Specific objectives

- To incorporate the diachronic dimension into semantic maps of content words

- To extend the method so as to also include information about the cognitive and cultural factors behind the development of the various meanings

- To create an online platform for automatically plotting diachronic semantic maps based on polysemy data from the languages of the world
Le Diasema

Filling a gap

- Adding a diachronic dimension to semantic maps of content words

“[T]he best synchronic semantic map is a diachronic one”
(van der Auwera 2008: 43)
Le Diasema

Filling a gap

- Adding a diachronic dimension to semantic maps of content words

“[T]he best synchronic semantic map is a diachronic one”
(van der Auwera 2008: 43)
**Le Diasema**

**Data sources**

- **Synchronic (polysemy) data**
  - Online databases containing large language samples
    - **CLiCs**: an online database of synchronic lexical associations ([http://clics.lingpy.org](http://clics.lingpy.org); List et al., 2014)
    - **WordNet** ([http://wordnet.princeton.edu](http://wordnet.princeton.edu); Princeton University, 2010) and **Open Multilingual Wordnet** ([http://compling.hss.ntu.edu.sg/omw/index.html](http://compling.hss.ntu.edu.sg/omw/index.html)): large lexical databases of words grouped into sets of cognitive synonyms each expressing a distinct concept
    - **Concepticon** ([http://concepticon.clld.org/](http://concepticon.clld.org/); List et al., 2016)
  - etc.
Le Diasema

Data sources

- **Diachronic data**
  - Ancient Greek (8th – 1st c. BC)
    - Dictionaries, grammars
  - Ancient Egyptian (26th c. BC – 10th c. AD)
    - Thesaurus Linguae Aegyptiae ([http://aaew.bbaw.de/tla/](http://aaew.bbaw.de/tla/))
    - The Ramses corpus ([http://ramses.ulg.ac.be](http://ramses.ulg.ac.be)),
    - Lexical resources (Coptic etymological dictionaries)
  - The Catalogue of Semantic Shifts in the Languages of the World
    - (Zalizniak, 2006; Zalizniak et al., 2012; [http://semshifts.iling-ran.ru/](http://semshifts.iling-ran.ru/))
Le Diasema

Building diachronic semantic maps of content words

- Manually… not an option
Le Diasema

Building diachronic semantic maps of content words

- **Manually… not an option**
  - From max. 60 grammatical functions to $60,000^+ \text{ lexical meanings}$

A semantic map of functions associated with allative markers (based on Rice & Kabata, 2007: 54 allative grams in 44 genetically and areally diverse languages)
Le Diasema
Building diachronic semantic maps of content words

- Manually... not an option
  - From max. 60 grammatical functions to 60.000+ lexical meanings

- Automatically, which implies the following steps
Le Diasema

Building diachronic semantic maps of content words

- **Manually… not an option**
  - From max. 60 grammatical functions to 60,000+ lexical meanings

- **Automatically, which implies the following steps**
  - Converting synchronic polysemy data into a lexical matrix

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>banana tree</td>
<td>tree</td>
<td>bul_std:darvó</td>
</tr>
<tr>
<td>banana tree</td>
<td>wood</td>
<td>bul_std:darvó</td>
</tr>
<tr>
<td>club</td>
<td>firewood</td>
<td>am_std:mamil://pue_std:ipuk</td>
</tr>
</tbody>
</table>

Polysemy data from CLiCs ([http://clics.lingpy.org/download.php](http://clics.lingpy.org/download.php))
Le Diasema

Building diachronic semantic maps of content words

- Manually... not an option
  - From max. 60 grammatical functions to 60,000+ lexical meanings

- Automatically, which implies the following steps
  - Converting synchronic polysemy data into a lexical matrix

```
Tmap = [Tsenses]
for t in Tclon:
    split_langWord = t[2].split('://')
    for couple in split_langWord:
        langWord = couple.split(':')
        line = [langWord[0], langWord[1]]
        for i in range (2,len(Tsenses)):
            line.append('0')
        line[Tsenses.index(t[0])] = '1'
        line[Tsenses.index(t[1])] = '1'
        Tmap.append(line)
```

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>bul_std</td>
<td>darvd</td>
<td>1</td>
<td></td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>cat_std</td>
<td>pal</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>cof_std</td>
<td>a/co</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>ąkv_Northern</td>
<td>чули</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>ąkv_Southern</td>
<td>чули</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>ąva_Batlukh</td>
<td>тил</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>ąva_Karakh</td>
<td>тил</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Python script  α

Lexical matrix
Le Diasema

Building diachronic semantic maps of content words

- **Manually**... not an option
  - From max. 60 grammatical functions to 60,000+ lexical meanings

- **Automatically**, which implies the following steps
  - Converting synchronic polysemy data into a lexical matrix
  - Inferring a semantic map from a lexical matrix (Regier et al. 2013)

```python
# CREATE INITIAL GRAPH
# graph G: add each term's nodes, no edges in graph yet.
G = nx.Graph()  # create empty graph (undirected)
PossE = []  # list of possible edges, filled below
for t in T:
    # add all nodes in t, if not already in graph
    for n in t:
        if not G.has_node(n):
            G.add_node(n)
    # add to PossE a link between each pair of nodes in t
    # adding a link between every node in G is needless and slower
    for pair in allpairs(t):
        u = pair[0]
        v = pair[1]
        if not (((u, v) in PossE) or ((v, u) in PossE)):
            PossE.append((u, v))
```

Python script
Building diachronic semantic maps of content words

A weighted semantic map of the meanings associated with [TREE] and [WOOD] based on data from CLiCs (incl. modularity analysis)
Le Diasema
Building diachronic semantic maps of content words

- Manually... not an option
  - From max. 60 grammatical functions to 60,000+ lexical meanings

- Automatically, which implies the following steps
  - Converting synchronic polysemy data into a lexical matrix
  - Inferring a semantic map from a lexical matrix (Regier et al. 2013)
  - Inferring oriented edges based on diachronic data

```python
H = G.to_directed()  # convert the graph 'G' into a directed graph
# all the possibilities as regards the rel
# (i.e., both A -> B and B -> A) for all the
# not only A -> R
nx.set_edge_attributes(H, 'type', 'undirected')  # set the default

for u, v, e in H.edges(data=True):
    for t in T_Full:
        if t.count(u) == 1 and t.count(v) == 0:
            # if the metadata
            # while the
            LongWord = t[0:2]  # store temporarily
            Did = t[2]  # store temporarily
            for c in T_Full:
                if c[0:2] == LongWord and c[2] > 0:
                    H[u][v]['type'] = 'directed'

Python script γ
Le Diasema

Building diachronic semantic maps of content words

- **Manually... not an option**
  - From max. 60 grammatical functions to 60,000+ lexical meanings

- **Automatically, which implies the following steps**
  - Converting synchronic polysemy data into a lexical matrix
  - Inferring a semantic map from a lexical matrix (Regier et al. 2013)
  - Inferring oriented edges based on diachronic data

- **Labelling the types of polysemy**, so as to identify shared cognitive motivations and to assess the potential impact of cultural factors on the evolution of various lexical domains
  - The role of Metaphor
  - The role of Metonymy
  - Areal and cultural reasons accounting for different types polysemy
Le Diasema

Talks, workshop and Website

- Diasema talks

- Workshop
  - ‘Semantic maps: where do we stand and where are we going?’, Liège 2018.06.27-29

- Website: [http://web.philo.ulg.ac.be/lediasema/](http://web.philo.ulg.ac.be/lediasema/)
Thanks!

athanasios.georgakopoulos@ulg.ac.be
s.polis@ulg.ac.be