Thanasis Georgakopoulos (Post-doc in ULg)
Stéphane Polis (F.R.S.-FNRS / ULg)

Liège – 2017.03.10

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

Université de Liège

HR EXCELLENCE IN RESEARCH

BeIPD-COFUND

euraxess

Le Diasema

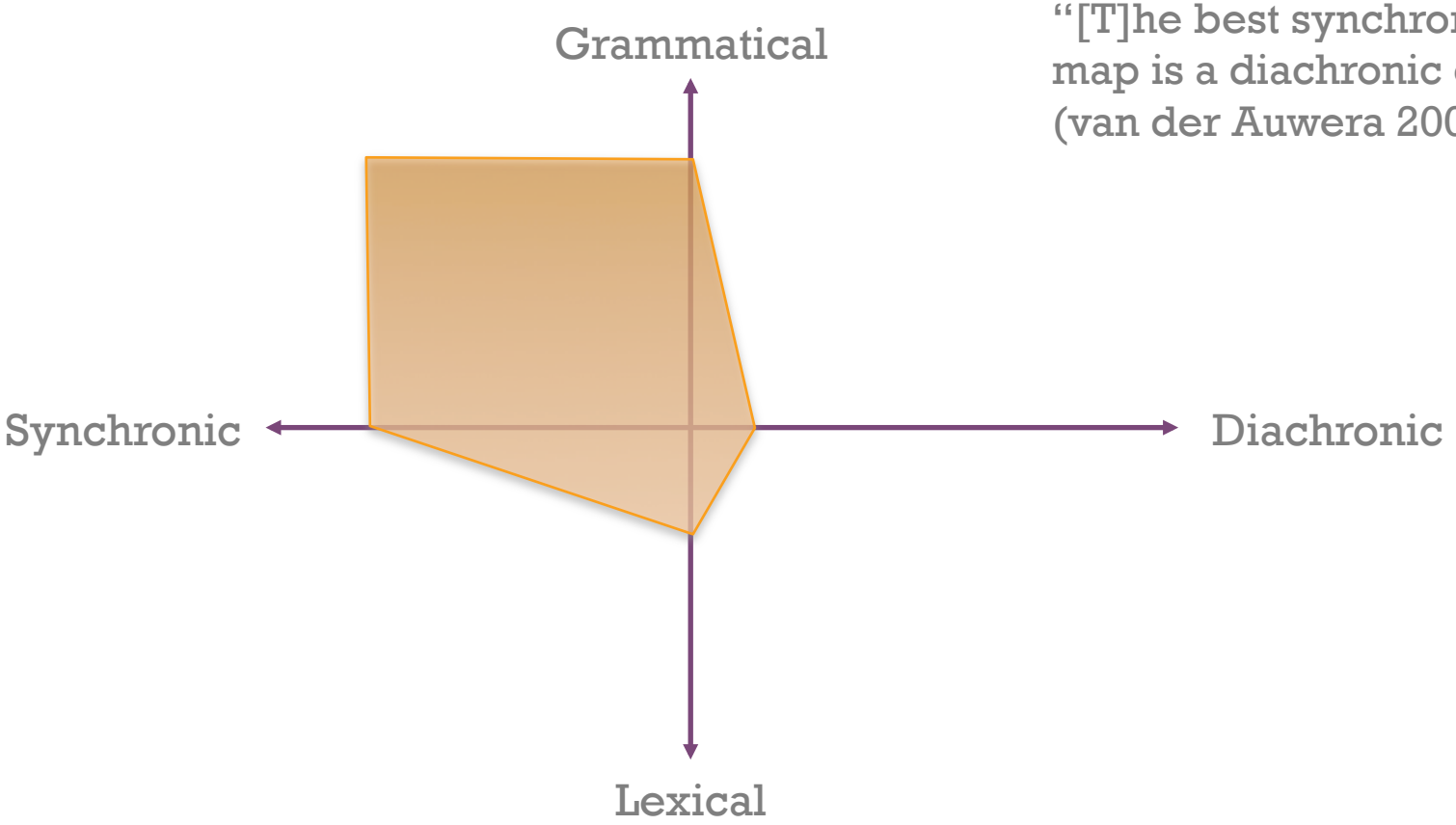
Specific objectives

- 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



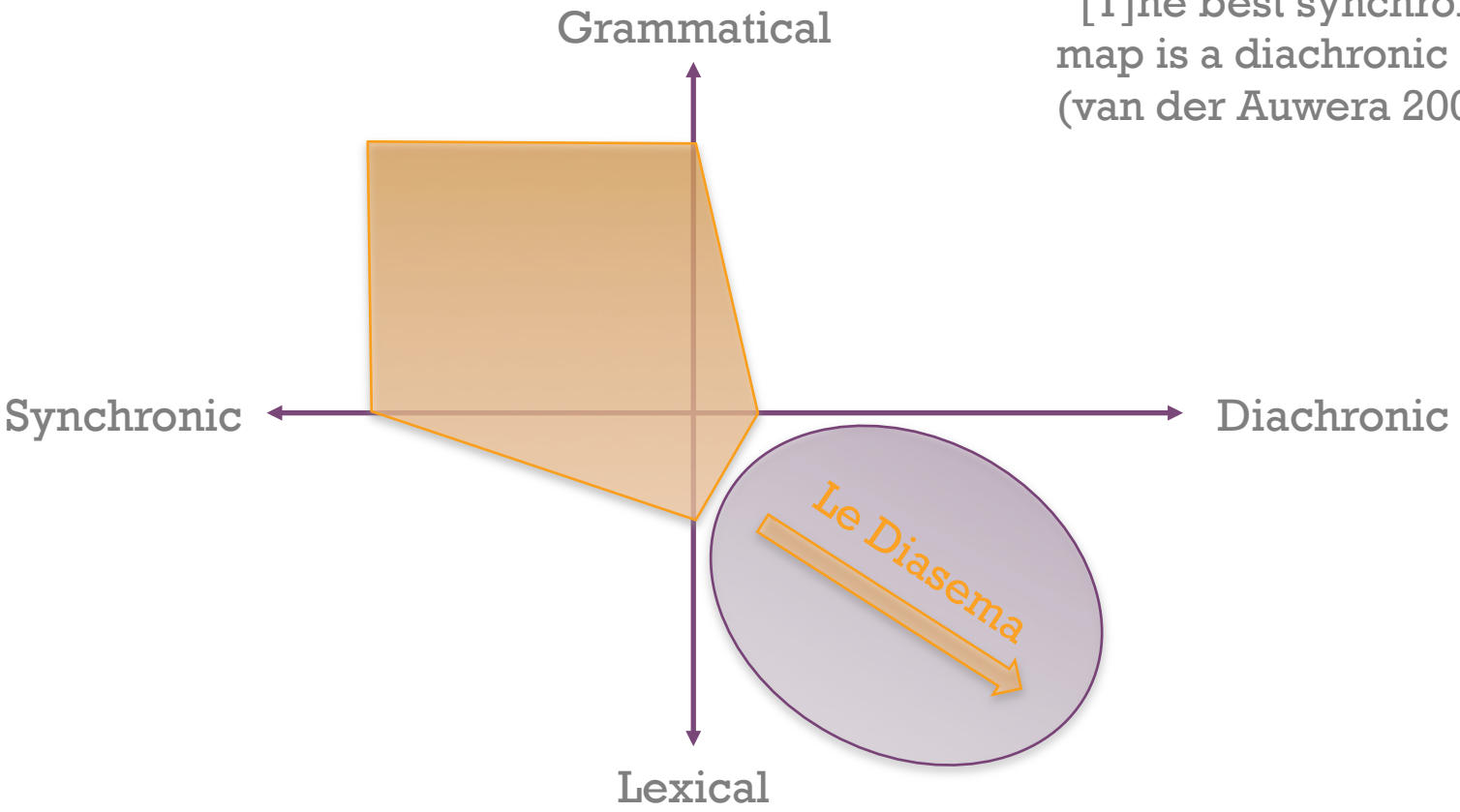
"[T]he best synchronic semantic map is a diachronic one"
(van der Auwera 2008: 43)

Le Diasema

Filling a gap

- Adding a diachronic dimension to semantic maps of content words

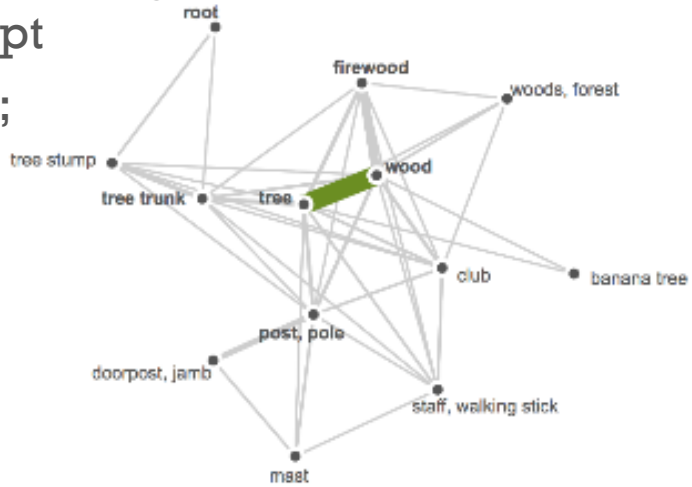
"[T]he best synchronic semantic map is a diachronic one" (van der Auwera 2008: 43)



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; List et al., 2014)
    - **WordNet** (http://wordnet.princeton.edu; Princeton University, 2010) and **Open Multilingual Wordnet** (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/; List et al., 2016)
    - etc.



Le Diasema

Data sources

- **Diachronic data**
  - Ancient Greek (8th – 1st c. BC)
    - Perseus digital library (http://www.perseus.tufts.edu/hopper/)
    - Dictionaries, grammars
  - Ancient Egyptian (26th c. BC – 10th c. AD)
    - Thesaurus Linguae Aegyptiae (http://aaew.bbaw.de/tla/)
    - The Ramses corpus (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/)

Le Diasema

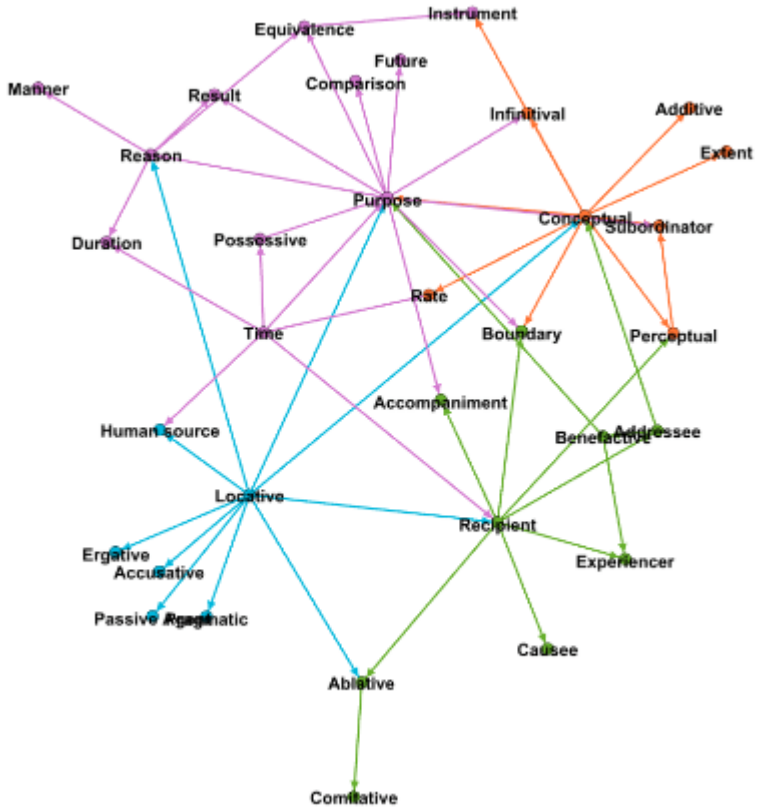
Building diachronic semantic maps of content words

- **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+ 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

	A	B	C
1	banana tree	tree	bul_std:dərvó
2	banana tree	wood	bul_std:dərvó
3	club	firewood	arn_std:mamilʏ//pue_std:ɨpuk
4	club	post, pole	bul_std:prə̀t//cat_std:pal//cof_std:ci²de//haw_std:lā?au//ito_std:abite//qvi_std:kaspi
5	club	staff, walking stick	akv_Northern:чlулu//akv_Southern:чlулu//ava_Batlukh:тlил//ava_Karakh:ил//ava_Ka
6	club	tree	arn_std:mamilʏ//haw_std:lā?au//ito_std:abite//ket_std:ōks//pbh_std:iye//pue_std:ɨpuk/

Polysemy data from CLiCs (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 Tclean:
        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)
```

Python script α

	A	B	C	D	E	F
1			banana tree	tree	wood	club
2	bul_std	dərvó	1	1	1	0
3	cat_std	pal	0	0	0	1
4	cof_std	ci²de	0	0	0	1
5	akv_Northern	чIули	0	0	0	1
6	akv_Southern	чIули	0	0	0	1
7	ava_Batlukh	тIил	0	0	0	1
8	ava_Karakh	ил	0	0	0	1
9	ava_Karakh	тIил	0	0	0	1

Lexical matrix

Le Diasema

Building diachronic semantic maps of content words

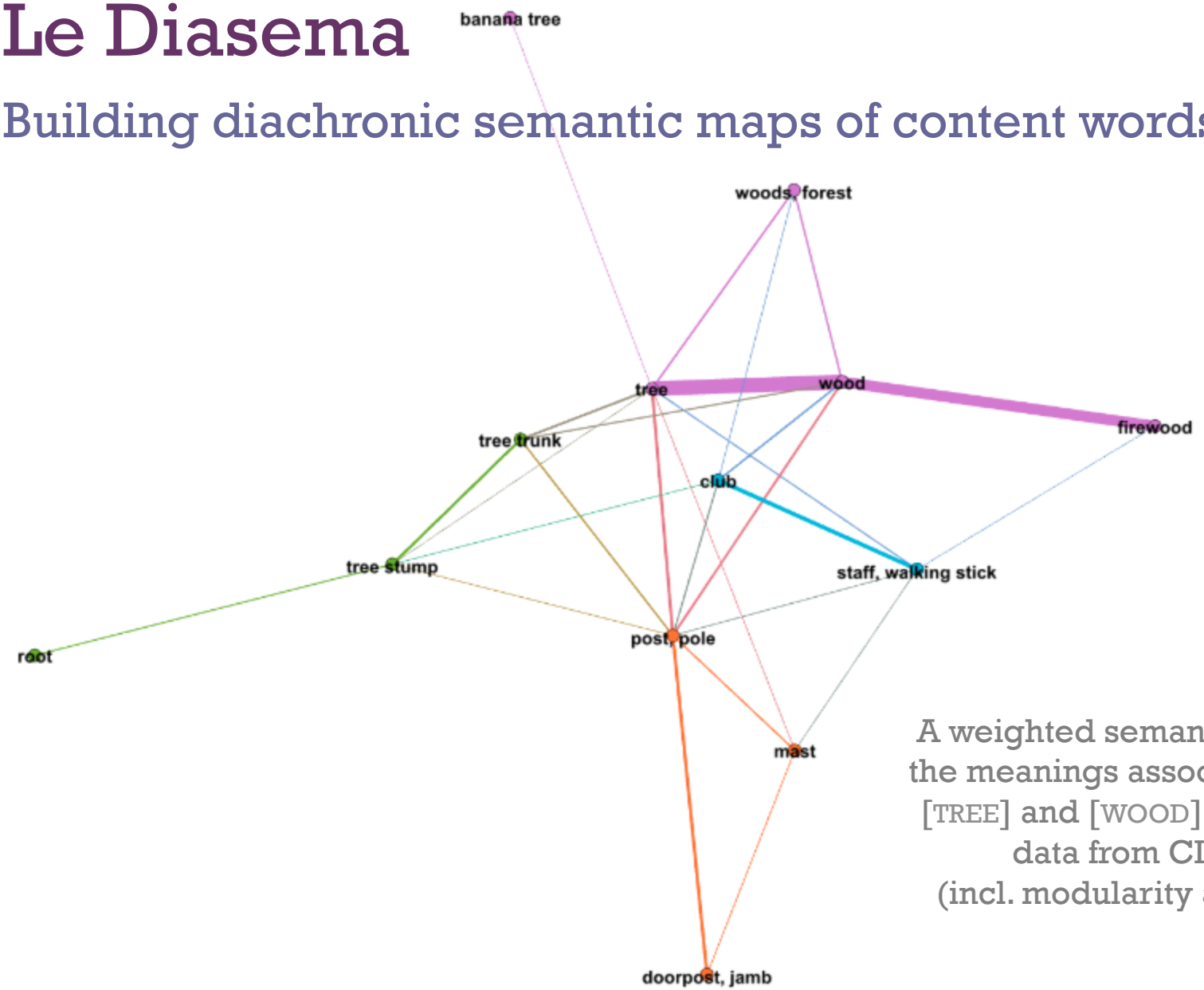
- **Manually… not an option**
  - From max. 60 grammatical functions to 60.000+ lexical meanings
- **Automatically, which implies the following steps**
  - Converting synchronic polysemy data into a lexical matrix
  - Inferring a semantic map from a lexical matrix (Regier et al. 2013)

```
# 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 *β*

Le Diasema

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

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

for u,v,e in H.edges(data=True):        # loop over all the edges i
        for t in T_Full:                # look at the metadata and
                if t.count(u) == 1 and t.count(v) == 0: # if the me
                                                        # while the
                        LangWord = t[0:2]       # store temporarily
                        Dia = t[2]              # store temporarily
                        for c in T_Full:        # then look at the
                                if c[0:2] == LangWord and c[2] > Di


                                        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/

Thanks!

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