Krzysztof Nowak
Lexicon Mediae et Infimae Latinitatis Polonorum

ENeL WG 4 Task Group on

(Meta)Lexicography

EUROPEAN NETWORK OF e-LEXICOGRAPHY
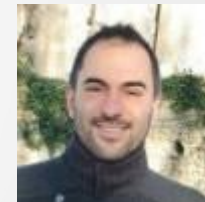
IJP

Institute of Polish Language,
Polish Academy of Sciences

PAN

Sources

- 3rd Task "Lexicography, Metalexicography" as defined in the minutes of the Leiden meeting

- WG4 objectives as elaborated in the ENeL MoU

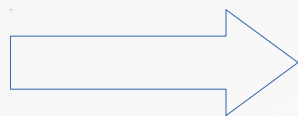- suggestions and research interests of 3 current members of the Task Group

Bruno Bon (FR) IRHT CNRS	*Novum Glossarium Mediae Latinitatis*	historical semantics, wiki interface, corpora, statistics	
Nathalie Mederake (DE) GAW	*Deutsche Wörterbuch (Neubearbeitung)*	lexical loans, historical dictionaries, innovative interfaces	
Krzysztof Nowak (PL) IPL PAS	*Dictionary of Polish Medieval Latin*	electronic dictionaries, corpora, language change	

Goals

**lexicographic data:
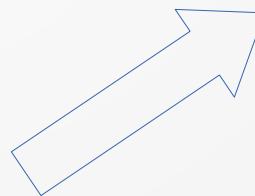dealing with heterogenity**

- structural
- syntactic
- semantic

e-lexicography: making sense of data

- making data **accessible**
- making data **meaningful**
- making data **reusable** in new research contexts

I. Dictionary content access

II. Research-driven integration

III. Emerging tools and examples to follow

IV. Promising interfaces

Task Group
Metalexicography

I. Dictionary content access

- user input (already WG 1 topic)

  - varying scripts

  - error handling

  - search suggestions

- choosing basis for integration

  - etyma?

  - roots?

  - other?

- choosing formalisation

Possible links
→ WG 1 (R. Lew et al.)

Possible links
→ Alina et al.
root dictionary
→ Aleksandra Jarosova
etymology

Possible links
→ Thierry
Semantic Web modelling
→ L. Romary: lexical
content representation

I. Dictionary content access

- testing non-lemmatic dictionary access

  - thesauri

  - concept-, ontology-based

  - browsing

    - linguistic evidence

    - maps

    - patterns (e.g. regular polysemy, conceptualisation patterns etc.)

- building knowledge retrieval paths and attractive narratives

Possible links
→ M. Zock
conceptual access

→ N. Tadmor
tracking change
in conceptualisation

II. Research-driven integration: data quality issues

- mapping EDP design choices to precise research questions

- e-lexicography as a tool of research and source of research questions

- changing picture

  - new research objects?

  - consequences of data scarcity

    and incompleteness

  - linguistic reality

- data reuse: towards theory-agnostic lexicography?

Possible links
→ Alina, Naomi, Jadwiga, Chris shortcomings and ways of improving current lexicography practice

II. Research-driven integration: data quantity issues

- handling large data amounts:

  tools of exploratory analysis

  - visualisation
  - summaries
  - statistical insights

- tracking patterns

  - lexical borrowing
  - lexical lacunae
  - changing orthography

Possible links
→ R. Theron
visualisation for lexicography

Possible links

→ Chris, ALE, Jadwiga
visualising lexical variation

III. Emerging tools and examples to follow

- standards
  - representing lexicographic data
- formats
  - storing
  - retrieving
  - enriching

- tools
  - database solutions
  - query systems
  - visualisation libraries

IV.   Promising interfaces

dictionary-corpus interface

- enriching dictionary with corpus data

Links

→ WG 3

→ <u>Dan</u>
linking dictionaries with digital libraries

dictionary-encyclopaedia

- towards dictionary hybrids

- enriching dictionary with world-knowledge

- enriching dictionary with images

Links
→ <u>R. Navigli</u>
retrieval and reuse of BabelNet data

Methods / Deliverables

case studies

mashups

working prototypes

showcases

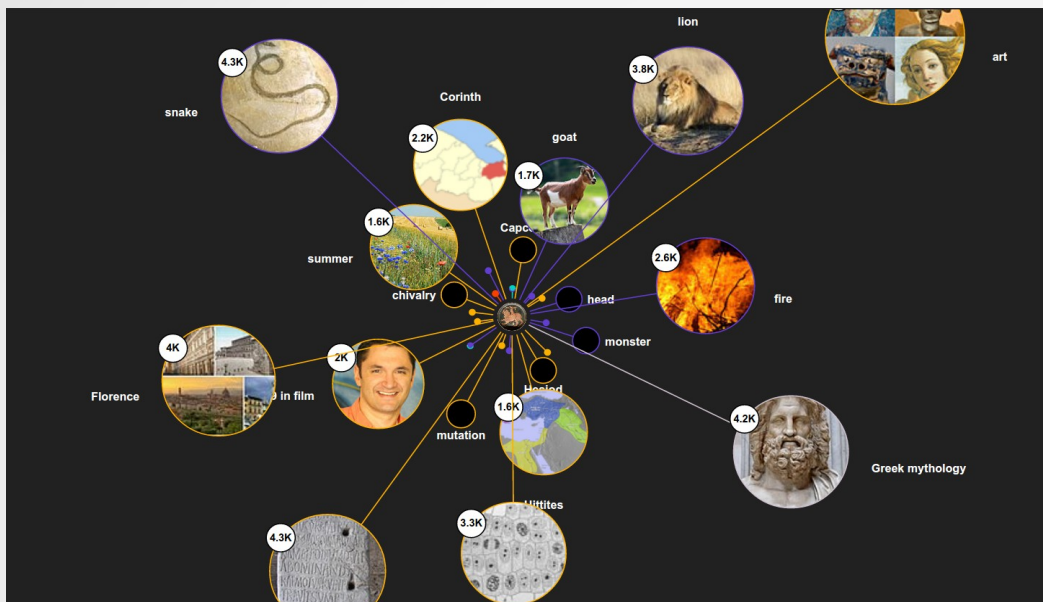discussion

guidelines

best practices

2015

- STSM: NM → KN

  (tracking loanwords with wiki-interface)

- Prototype of a Dictionary-Corpus-Encyclopedia Mashup (BB & KN)

- working session during the ENeL Sussex meeting

Next

- linking more people to the Task

- e-lexicography session during the International Congress of Medieval Latin (2017)?

- workshops

  – alternative forms of dictionary content access?

  – dictionary-encyclopaedia interface?

More

Visualisation of the BabelNet entry for CHIMAERA

- **chimaera** symbolises an awkward combination of incompatible parts
- **BabelNet** caption represents dictionary-encyclopaedia interface and visualisation of lexicographic data

- task description in full may be found at the WG 4 Vienna Meeting page

- contact

krzysztofn@ijp-pan.krakow.pl