



The lexical reflection of
cultural dynamics.
Large-scale corpus-driven
neologism analysis

Trojina, Uni. Ljublijana, INL, KU Leuven
(submitted project proposal)



KU Leuven
RU Quantitative Lexicology and Variational
Linguistics

Project type

Joint bilateral programme:

- FWO (lead): Flemish Science Foundation
- ARRS: Slovene Science Foundation

Stress on basic (non-applied) research

-> But strong methodological part with
high relevance for lexicography

4 years, approx. 130K per year

2 FT researcher (1 senior, 1ESR)

NOTIFICATION OF APPROVAL: November 2014



Project partners

Partners (officially...):

- Trojina institute (Slovenia)
 - Iztok Kosem
- University of Ljublijana (Slovenia)
 - Simon Krek
- QLVL, University of Leuven (Belgium)
 - Dirk Speelman
 - Kris Heylen
- INL(in NL, but bi-national B/NL)
 - Martin Everaert



Project partners

Partners (officially...):

- Trojina institute (Slovenia)
 - Iztok Kosem
- University of Ljublijana (Slovenia)
 - Simon Krek
- QLVL, University of Leuven (Belgium)
 - Dirk Speelman
 - Kris Heylen
 - Dirk Geeraerts
- INL(in NL, but bi-national B/NL)
 - Martin Everaert
 - Carole Tiberius,...



Concept-based Neology

- Starting point: Neologisms reflect cultural dynamics (~ culturomics)
- New *concepts* introduced in language community
- New concepts can take different linguistic shapes:
 - New forms:
 - New formations (morphology)
 - New multiword units
 - Loan words
 - Semantic extensions of existing words
- COMBINE INTO ONE STUDY



Comparative Neology

- Diachronic fluctuations in the rate of neologism creation?
 - in different lexical fields?
 - Within and between the languages?
 - Reflections of different cultural dynamics?
- Which type is chosen? (new creation, loan, semantic extension) factors:
 - language (Dutch vs Slovene)
 - concept characteristics
 - Year
 - genre (newspaper articles, internet blogs,
- What determines survival? (cf FUDGE)

Methodology Neology Detection

AUTOMATIC PREPROCESSING

- Single words
 - building on INL's Blacklab environment
- Multiword units
 - building on Trojina's experience with collocation extraction
- Semantic extensions
 - Building on QLVL's experience with automatic Word Sense Induction

MANUAL CLASSIFICATION

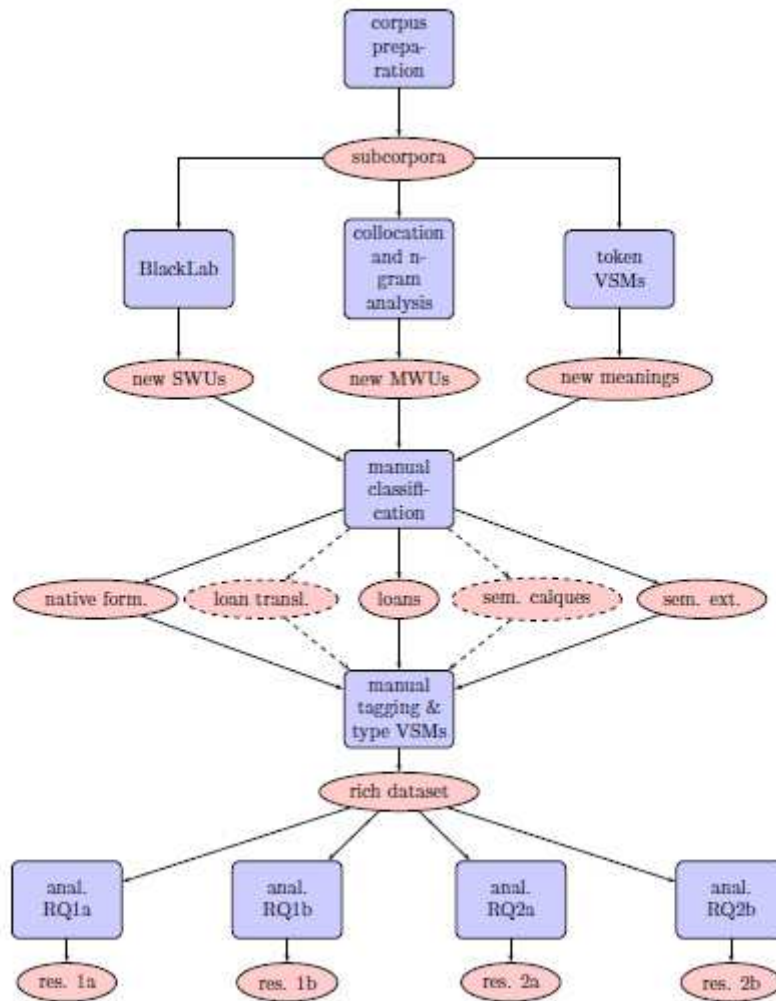
- Loan translations, semantic calques

STATISTICAL DATA ANALYSIS

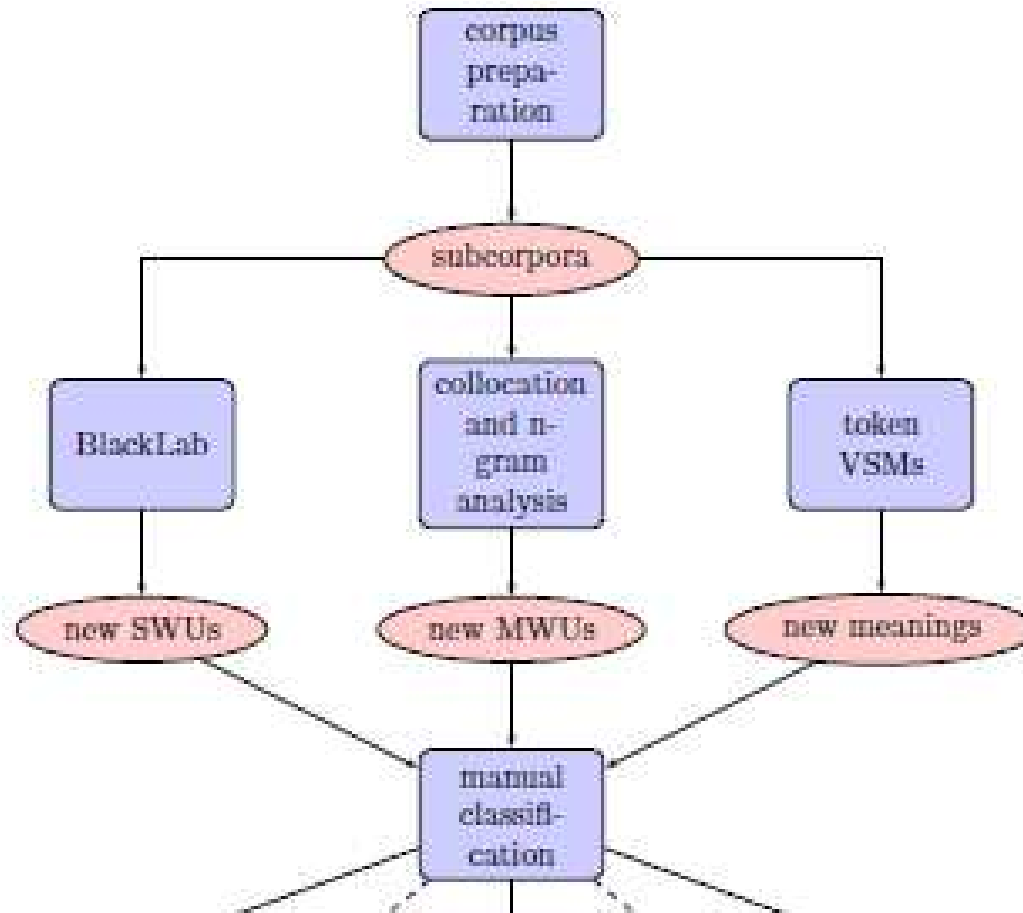
- Relation between variables



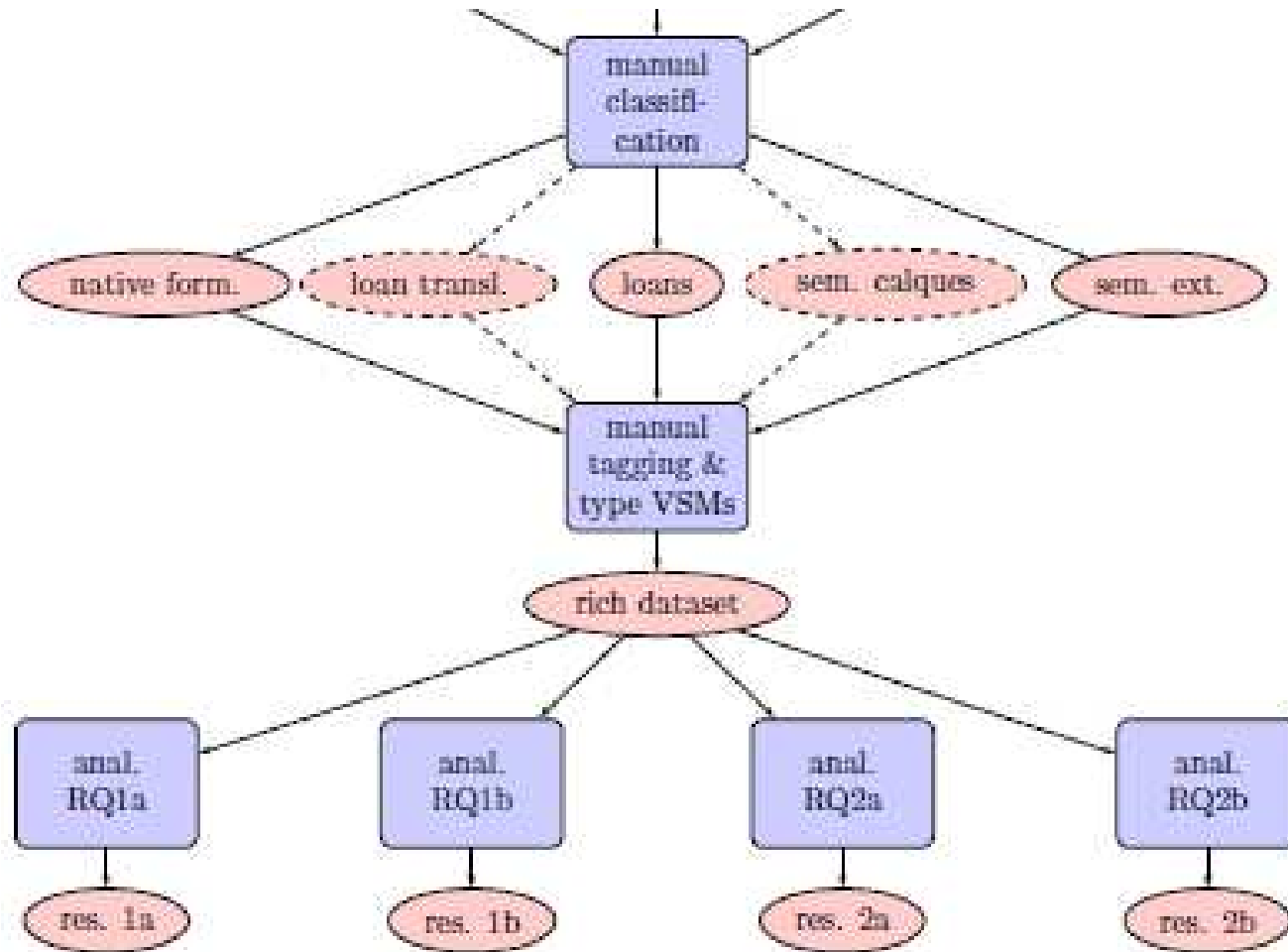
Project work flow



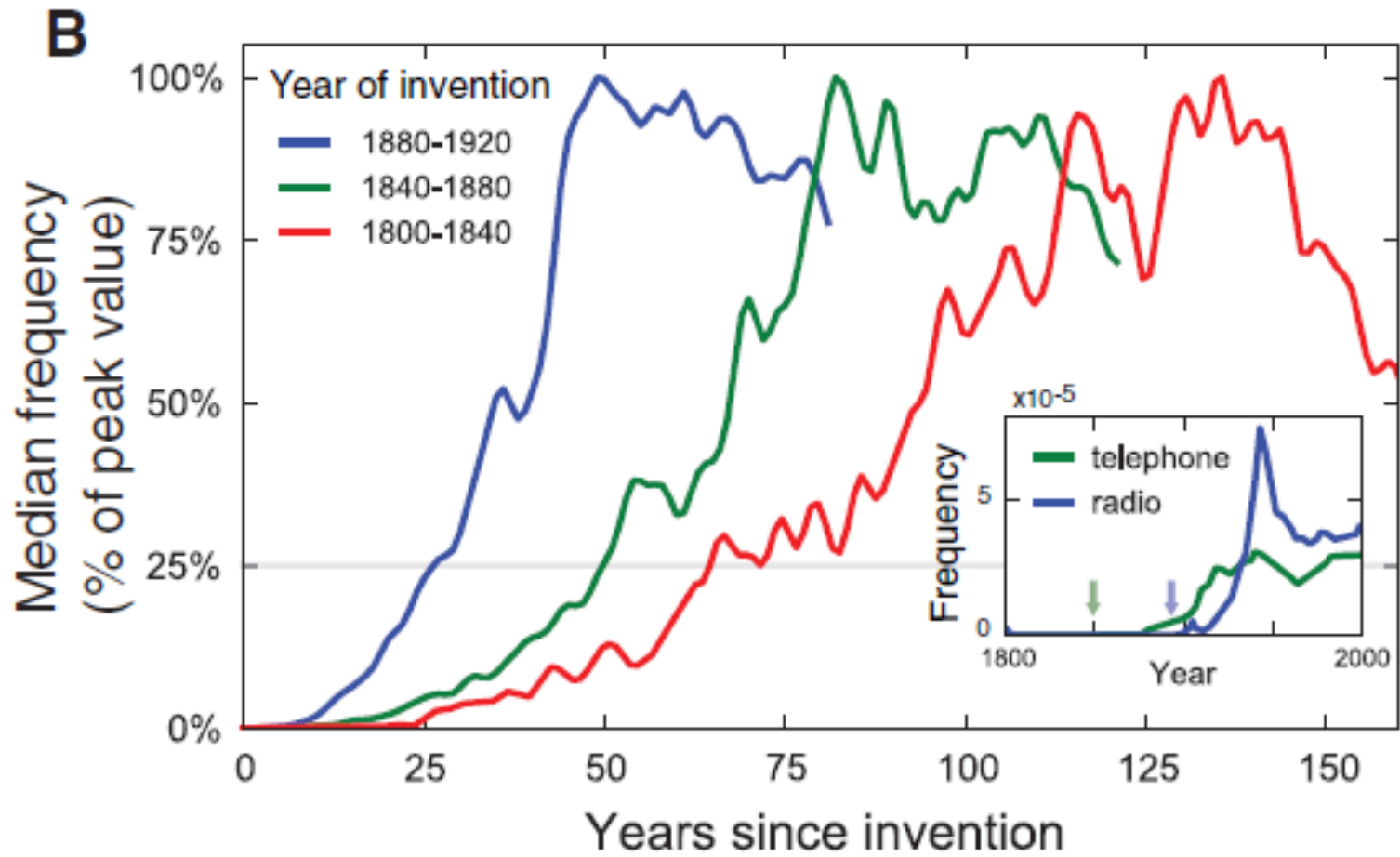
Project work flow



Project work flow



Culturomics



for further information:

Kris.heylen@kuleuven.be

<http://wwling.arts.kuleuven.be/qlvl>

QM