



You get what you annotate: a pedagogically annotated corpus of coursebooks for Swedish as a Second Language

Elena Volodina, Ildikó Pilán, Stian Rødven Eide, Hannes Heidarsson

University of Gothenburg, Sweden

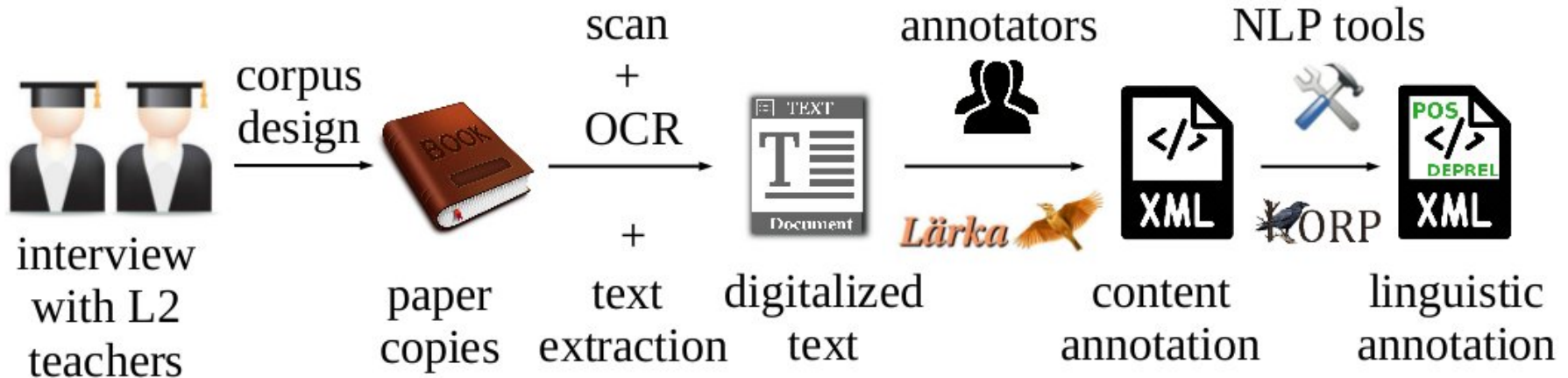


COCTAILL time

- **COCTAILL - Corpus of CEFR Textbooks as Input for Learner Level modeling**
- January 2013-October 2014
- CLT-funded (University of Gothenburg)



Brewing COCTAILL





COCTAILL content “ingredients”

ELEMENTS

Coursebook

└ Extras

└ Lesson

└ Text

└ Genre

└ Subheadings

└ Activity instruction

└ Task

└ List

└ Language Example

ATTRIBUTES

title, author etc.

CEFR level, (title)

topic (e.g. arts), (title)

genre family (e.g. facts)

macrogenre (e.g. dialogue)

skills/competences,

activity type/format

skills/competences,

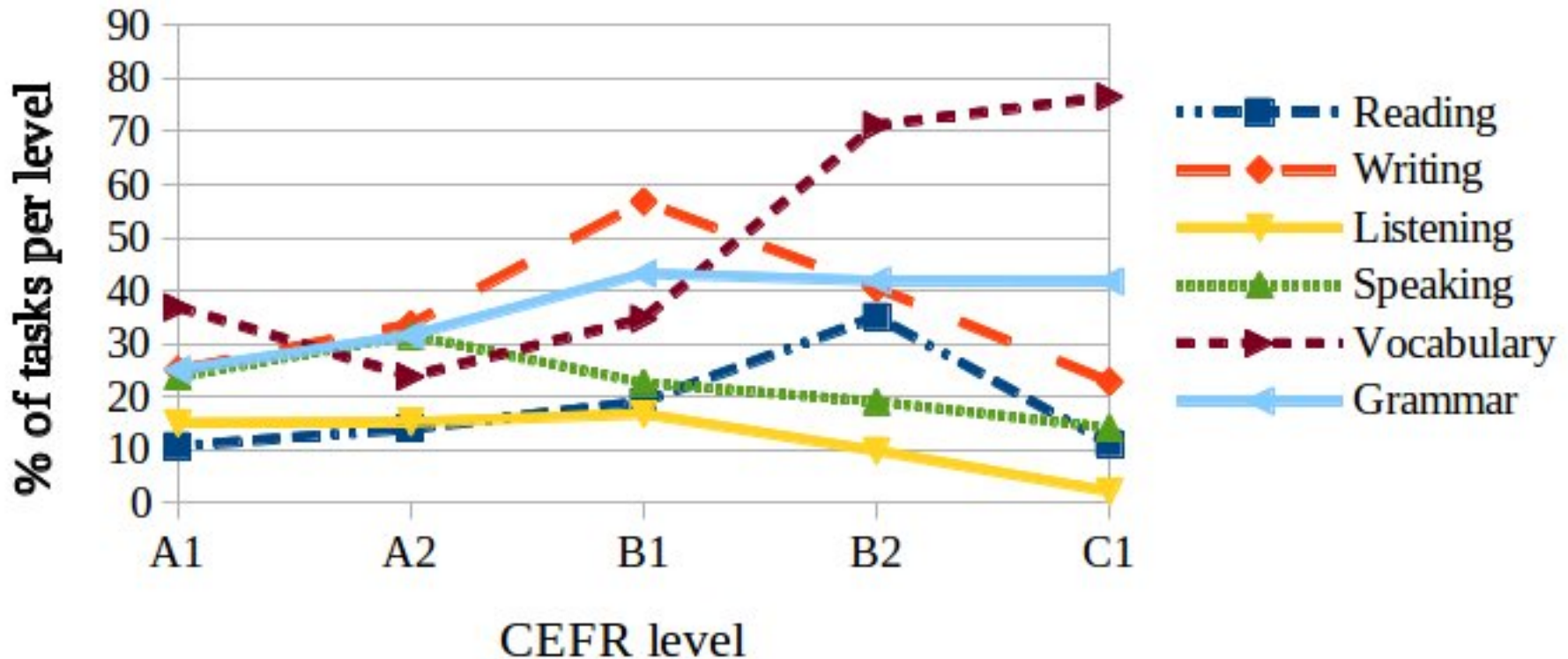
linguistic unit

Common attributes:

id, ref

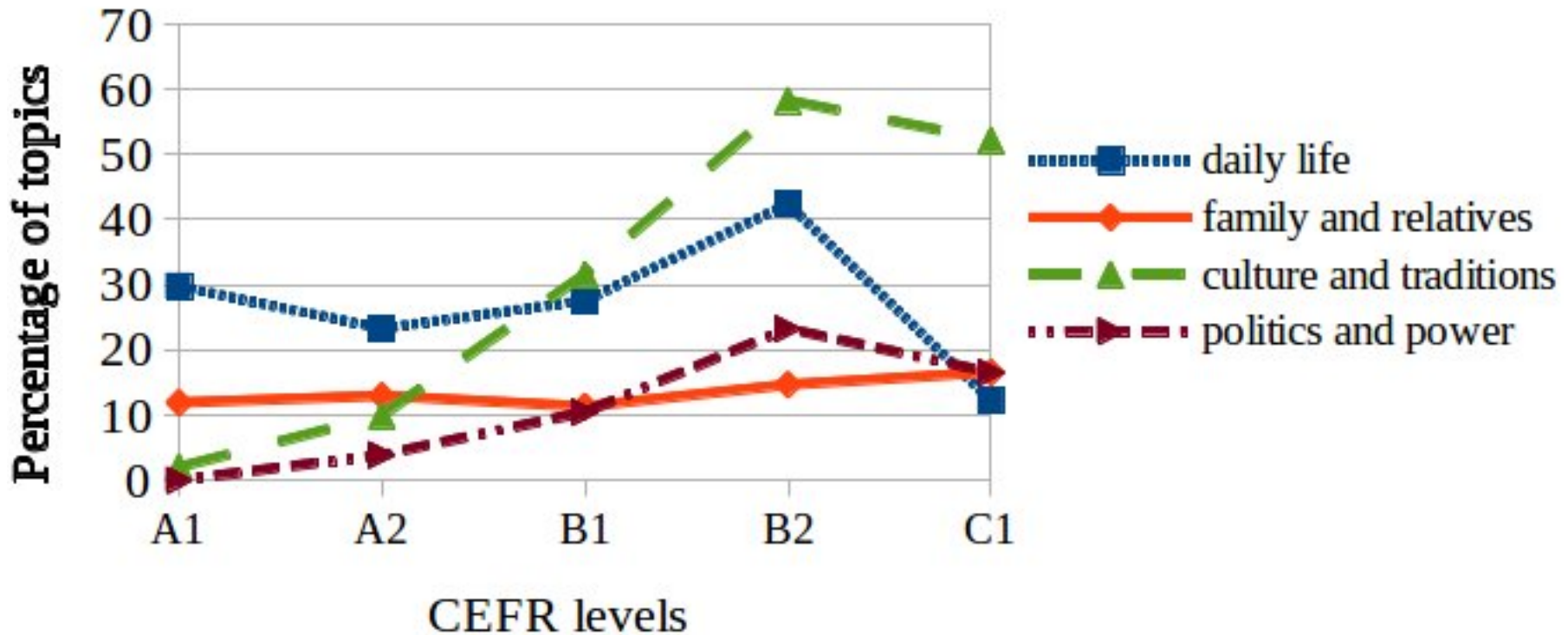


COCTAILL quantitative explorations: target skills across levels





COCTAILL quantitative explorations: topics across levels





COCTAILL impact

- A possibility to study coursebooks in their entirety, correlating proficiency levels, L2 input, pedagogical and textual variables
- SLA-aware ICALL-relevant studies, such as CEFR profiling, vocabulary and grammar profiling, studies on sentence and text readability, automatic genre id, etc.
- A crystallized form of what should be taught, at which level and in which format -> for better material generation



Taste our COCTAILL at the poster stand:



e.g. browse the corpus,
test the annotation tool,
zoom into inter-annotator agreement results, etc.