

Syntax meets Digital Humanities?

DELPH-IN Summit 2025

Olga Zamaraeva, July 10, 2025

Digital Humanities

- A big field on its own
- More and more papers at *ACL conferences
 - **main** (my impression)
 - **+100%** size in the NLP4DH SIG
- Related to **linguistics**
 - Could we **connect** to them?

Digital Humanities Research

...historically

- Building **resources**
 - starting from OCR...
- Performing **search**
 - corpus linguistics
 - character networks, etc., based on regex, word counts...
- LLMs: Historical/literary simulations, story/poetry writing, translations, ...
 - Testing literary theories about narratives, archetypes...
 - Measuring biases

Example: Translating subtitles

(of complex, highly creative speech/text)

- Subtitles are chunks of text
 - They are better translated if first combined into a whole text
 - ...but then they need to be split back, meaningfully
- **Constituency parsing:** Helpful for meaningful chunking
 - possibly will help translation itself, not just splitting the result
 - “**formal equivalence**” in translation involves **syntax**

Other connection points?

- Distributions of grammar types?
- Evaluation?..
- Wordnets: comparative
- Authorship signatures with ERG parses
 - maybe not a solved problem yet
- Detecting text re-use
 - with DMRS
 - especially for low-resource scenarios

Other connection points?

- Database of motifs (Brezkin)
 - DMRS for tracking structure
- Organize a workshop to bring together linguists and researchers in digital humanities
- Discourse connectives
- Parsing historical/older texts to help build historical databases
 - (maybe not so easy to adapt the grammar)

Other connection points?

- Colocations of semantic relations
- Rhetorical device of parallel expression
- Sources of ambiguity
- Adapting DELPH-IN tools to do e.g. appraisal theory annotation?