Automatic Disambiguation in the Corpora of Greek and Yiddish

Elizaveta Kuzmenko  eakuzmenko_2@edu.hse.ru
Elmira Mustakimova  egmustakimova_2@edu.hse.ru
NRU HSE

Moscow, 30/05/2015
Why

Disambiguation
Better corpora \(\Rightarrow\) better search results \(\Rightarrow\) better research!

Automatic
By hands \(\Rightarrow\) Takes TIME

Greek and Yiddish
No gold standard
Desperate need of disambiguation
How
Three methods

01
Brill Algorithm
Transformation-based error-driven learning

02
N-gram model
Learning on non-ambiguous bigrams and trigrams extracted from corpora

N-gram based user interface
Requires a linguist, but a lot faster and more efficient
What results

- Disambiguation for under-resourced languages
- Resolution of ambiguities in the two corpora
- Effective order of several methods
Thank you for your attention!