COMPLEXITY AND INDECISION

A PROOF-OF-CONCEPT EXPLORATION OF LEXICAL COMPLEXITY AND LEXICAL SEMANTIC CHANGE

David Alfter
Gothenburg Research Infrastructure in Digital Humanities (GRIDH)
Department of Literature, History of Ideas and Religion
University of Gothenburg, Sweden
david.alfter@gu.se

UNIVERSITY OF GOTHENBURG

INTRODUCTION

Lexical Complexity Prediction (LCP)

Lexical Semantic Change Detection (LSCD)

Investigate the potential link between lexical complexity and lexical semantic change. Both phenomena are influenced by polysemy. Does this mean that we can use LCP for LSCD?

METHODOLOGY

LCP

Example sentence with TARGET word

Target complexity

WiC

Example sentence 1 with TARGET word
Example sentence 2 with TARGET word

Target complexity 1
Target complexity 2

Train a LCP classifier on LCP data, then predict complexity for words in the WiC task.
Take the difference in complexity as measure for semantic change.

RESULTS

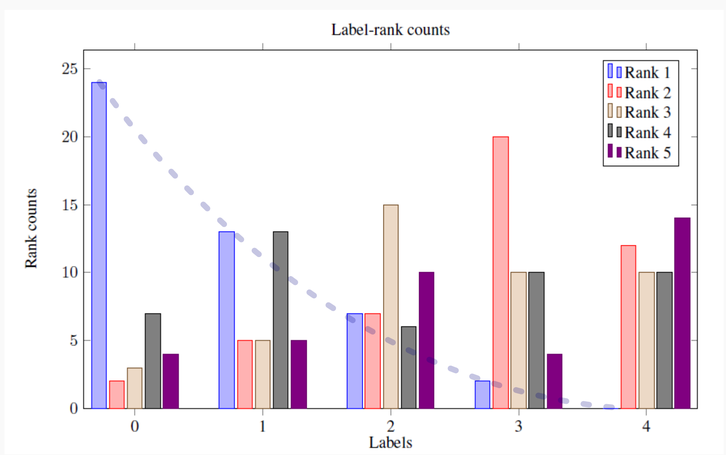
LCP for LSCD
- **0.444 LCP model**
- **0.422 2020 best model**
- **0.757 current SoTA**

The complexity difference model would have won the 2020 LSCD shared task. The current state of the art is much more advanced.

Lexical complexity differences seem to indicate human indecision in the manual annotation of WiC data.

Manual WiC annotation
- **1: different sense**
- **2: distantly related**
- **3: closely related**
- **4: identical**
- **0: cannot decide**

Rank labels by complexity difference



LIMITATIONS

NN in LCP data
Mostly NN in LSCD data
English data

Necessary training data is a bottle neck; especially graded LCP data (multiple instances across different complexity levels)

Due to the presence of mostly/only nouns in the LCP and LSCD data, the results may not generalize to other PoS. There might also be a genre influence.

FUTURE WORK

Investigate underlying factors

Extend to other languages
- **German, Swedish, Latin**

Hybrid approaches
- **LCP as component in LSCD**
- **LCP as component in annotation**