

(three) Ideas for master projects

Malvina Nissim
m.nissim@rug.nl
room 1311.421

IK meeting, 9 December 2016

Project 1: Complexity in Language Learning

Complexity in Language Learning

with Mircea Lungu (software engineer, JBI, FMNS, RuG)

Background and state of things:

- current systems for the assessment of a texts complexity and for text simplification are normally based on general models of complexity;
- but: what is simple for someone might not be simple for someone else
- so that a general, undistinguished model of simple vs complex text might not be representative of actually anyone.

Complexity in Language Learning

with Mircea Lungu (software engineer, JBI, FMNS, RuG)

Background and state of things:

- current systems for the assessment of a texts complexity and for text simplification are normally based on general models of complexity;
- but: what is simple for someone might not be simple for someone else
- so that a general, undistinguished model of simple vs complex text might not be representative of actually anyone.

Project: adaptive user-specific complexity assessment for L2 learners

- existing features to assess complexity can be directly ported at the single user's level
- complexity models will be trained on specific users' data (we already know how to acquire gold training labels in this sense)
- done in the context of an already existing and running application

Project 2: To Normalise or Not to Normalise? (POS Tagging on Twitter)

POS Tagging on Twitter

with Barbara and Rob

State of things:

- POS tagging on (English) newswire: 97-98%
- POS tagging on (English) tweets: < 90%

Easy to understand why, but what can be done?

Project: compare methods (structure of project partially already in place)

POS Tagging on Twitter

with Barbara and Rob

State of things:

- POS tagging on (English) newswire: 97-98%
- POS tagging on (English) tweets: < 90%

Easy to understand why, but what can be done?

Project: compare methods (structure of project partially already in place)

- normalise Twitter language to more “standard” language (but what is standard?)
- build “native” taggers for Twitter
 - enough **gold** data for training?
 - exploring semi-supervised techniques
 - come up with theoretical stance

Project 3: Tracing metaphors through diachronic representations

Tracing metaphors in time

with Hessel

Background and state of things

- quite a lot of work on metaphors
- but little on the **dynamics** of metaphors
- pilot work on Italian (small set of words): change in embeddings over time correlates with the rise of new figurative meanings.
- pilot work on Dutch: same (collected data, trained consecutive models, pre-selected a few words to work with. CLIN abstract)

Tracing metaphors in time

with Hessel

Background and state of things

- quite a lot of work on metaphors
- but little on the **dynamics** of metaphors
- pilot work on Italian (small set of words): change in embeddings over time correlates with the rise of new figurative meanings.
- pilot work on Dutch: same (collected data, trained consecutive models, pre-selected a few words to work with. CLIN abstract)

Project: turn top-down pilot into bottom-up approach for automatic detection of shifts

- finalise pilot work with set of words
- evaluation strategies
- develop predictive measures to automatically detect shifts in word meaning in a completely unsupervised fashion from raw text