



University of Stuttgart
Germany

German in Flux:

Detecting Metaphoric Change via Word Entropy

August 4, 2017

Dominik Schlechtweg, Stefanie Eckmann, Enrico Santus,
Sabine Schulte im Walde, Daniel Hole

`dominik.schlechtweg@gmx.de`, `stefanie.eckmann@campus.lmu.de`,
`esantus@mit.edu`, `schulte@ims.uni-stuttgart.de`, `holedan@gmail.com`

Introduction

- ▶ Our aim:
 - ▶ **overall**: build a computational model detecting semantic change
 - ▶ **in this paper**: distinguish metaphoric change from semantic stability
- ▶ How we do it:
 - ▶ exploit the idea of **semantic generality** from hypernym detection
 - ▶ apply **entropy** to **distributional semantic model**
 - ▶ sample language German
 - ▶ introduce the first resource for evaluation of models of metaphoric change

Shortcomings of Related Work

- ▶ Previous work includes mainly:
 - (i) spatial displacement models
 - (ii) word sense induction models
- ▶ quantify the degree of **overall change** rather than being able to qualify different **types**
- ▶ do not examine metaphoric change

Metaphoric Change

- ▶ frequent and important type of semantic change
- ▶ source and target concept are related by similarity or a reduced comparison (cf. Koch, 2016, p. 47)

earlier: ... *muß ich mich vmbweltzen / vnd kan keinen schlaff in meine augen bringen*

'... I have to turn around and cannot bring sleep into my eyes.'

later: *Kinadon wollte den Staat umwälzen ...*

'Kinadon wanted to revolutionize the state ...'

- (i) creates **polysemy**
 - (ii) often results in more abstract or **general** meanings
- assumption: (i) and (ii) imply extension and dispersion in the range of linguistic contexts

Corpus

- ▶ *Deutsches Textarchiv (erweitert)* (DTA)
- ▶ **large**: provides more than 2447 lemmatized and POS-tagged texts (with more than 140M tokens)
- ▶ **covers long time period**: late 15th to the early 20th century
- ▶ **balanced**: includes literary and scientific texts as well as functional writings

Word Entropy

- ▶ corresponds to **entropy of word vector**
- ▶ is assumed to reflect **semantic generality** in hypernym detection
- ▶ is given by

$$H(C) = - \sum_{i=1}^n P(c_i | w) \log_2 P(c_i | w)$$

where $P(c_i | w)$ is the occurrence probability of context word c_i given target word w

- ▶ measures the **unpredictability** of w 's co-occurrences

Evaluation

- ▶ **no standard test set** of semantic or metaphoric change
- ▶ we create a small but first test set via annotation (**28 items**)
- ▶ annotators judged **560 context pairs** for a metaphorical relation

Workflow:

- (i) preselect 14 changing words
- (ii) add 14 stable distractors
- (iii) identify a date of change
- (iv) extract 20 contexts for each target from before and after date of change
- (v) for each word combine contexts between time periods randomly
- (vi) annotation of context pairs

Annotation

- ▶ steps to identify metaphoric relation of C1 to C2:
 1. Does any of these hold?:
 - ▶ C1 is less concrete than C2
 - ▶ C1 is less human-oriented than C2
 - ▶ C1 is not related to bodily action in contrast to C2
 - ▶ C1 is less precise than C2
 2. if yes: does C1 contrast with C2 but can be understood in comparison with it?
- ▶ **agreement**: κ (Fleiss' Kappa) between .40 and .46
- ▶ result is **gold ranking** of targets for strength of metaphoric change

Annotation Results

target	POS	type	date	meaning	score
Donnerwetter	N	met	1805	thunderstorm > thunderstorm, blowup	0.78
...					
Unhöflichkeit	N	sta	1605	discourtesy	0.1
...					

Table 1 : Sample of test set items ordered by their annotated degree of metaphoric change.

Results

	1700-1800	1800-1900	all
entropy	.64***	.10	.39*
frequency	.29	-.07	.26

Table 2 : Correlation (ρ) between predicted and gold ranks. Significance is determined with a t-test.

Result Analysis

▶ **ausstechen**

1605: *Von einem Bawren / welcher einem Kalbskopff die Augen außstach.*

'About a Farmer / who cut out the eyes of a calf's head.'

1869: *Sie wollen ihre Aufgabe nicht nur lösen, sondern auch elegant, d. h. rasch lösen, um Nebenbuhler auszustechen.*

'They not only wanted to solve their task, but also elegantly, i.e., solve it fast, in order to excel rivals.'

- ▶ gold rank: 12/28, entropy: 13, frequency: 17

▶ **Donnerwetter**

1631: *Die Lufft ist heiß / vnd gibt viel Blitzen vnd Donnerwetter ...*
'The air is hot / and there are many lightnings and thunderstorms ...'

1893: *Potz Donnerwetter!*

'Man alive!'

- ▶ gold rank: 1/28, entropy: 27, frequency: 15

Conclusions

- ▶ you *can* annotate semantic change in a corpus (so do it)
- ▶ entropy correlates strongly and significantly with degree of metaphoric change
- ▶ frequency correlates moderately, but non-significantly on small data set
- ▶ annotation and model are **generalizable** to different types of semantic change

<https://github.com/Garrafao/MetaphoricChange>