On the difficulty of making concreteness concrete

Hans van Halteren, Daphne Theijssen, Nelleke Oostdijk and Lou Boves
Centre for Language Studies, Radboud University Nijmegen

Introduction

- Semantic features often used in annotation/modelling without a clear or even any definition
- We take the concreteness of nouns as a case study
  - Found with many different interpretations: defined as 'tangibility' or 'specificity', established for types or tokens
  - Still, often deemed so obvious in literature that conclusions are given without a real explanation

Research question 1

- In what way do the actual annotations of the concreteness of nouns change when we use various approaches?

Method 1

- Using 4 approaches, we compare the annotations of nouns in a corpus
- We calculate correlation coefficients and compare annotations per WordNet noun class
- Data set of 68,484 noun tokens (8,476 types) in SEMCOR

Research question 2

- In what way do the conclusions in a syntactic study change when using various approaches to annotate the concreteness of nouns?

Method 2

- Using 6 approaches, we compare the conclusions in a study of the dative alternation (*give him a book* vs. *give a book to him*)
- We build 6 logistic regression models and compare the effects found
- Data set of 619 'themes' (*a book*) in ICE-GB

Approaches for establishing concreteness

Approach	Description	Question	Concrete examples	Abstract examples
MRC	look up noun lemma in MRC phycholinguistic database	1+2	*book*, *flower*	*harm*, *chance*
BOOTS	bootstrapping; iteratively find syntactic patterns of concrete/abstract nouns to find new concrete/abstract nouns (using BNC), seeding with examples in Garretson (2003)	1+2	*time*, *flower*	*right*, *harm*
WN-HIER	count nodes from noun sense to root in WordNet	1+2	*flower*, *one*	*indication*, *message*
WN-PHYS	check if noun sense is part of 'physical entity' in WordNet	1+2	*book*, *flower*	*harm*, *chance*
MANUAL	manually assign value; largely based on Garretson (2003)	2	*book*, *flower*	*harm*, *chance*
BOOTS-OBJ	bootstrapping; same as BOOTS except that we only look at syntactic patterns of nouns that are direct objects	2	*room*, *good*	*harm*, *indication*

Results 1

Table: Spearman correlations between the approaches for the 68,484 instances in SEMCOR (or the 44,395 also present in MRC, marked with *)

	MRC*	BOOTS	WN-HIER	WN-PHYS
MRC*	1.00*	0.65*	0.29*	0.60*
BOOTS	0.65*	1.00	0.13	0.45
WN-HIER	0.29*	0.13	1.00	0.17
WN-PHYS	0.60*	0.45	0.17	1.00

- Highest correlations between MRC and BOOTS
- WN-HIER differs most; defines concreteness as specificity, not tangibility

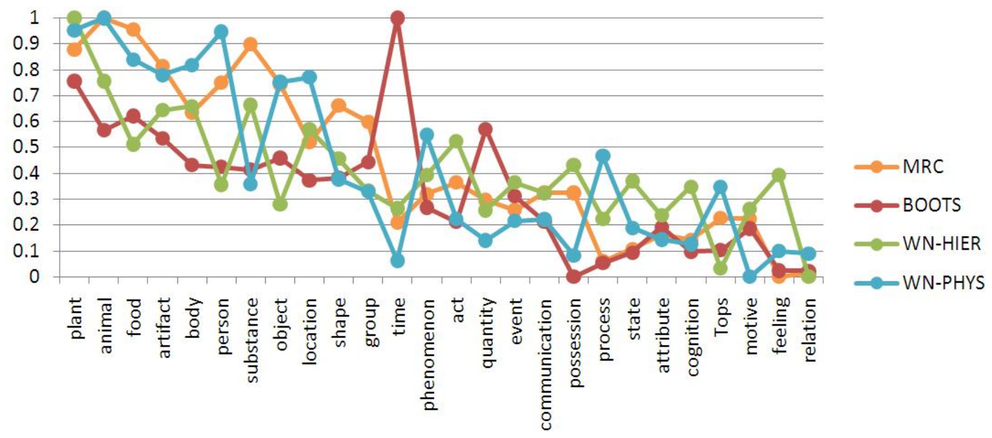

Figure: Average normalised concreteness score per WordNet noun class, for each approach

- BOOTS: 'time' (e.g. *minute*) and 'quantity' (e.g. *inch*) concrete
- WN-HIER: 'feeling' (e.g *trouble*) concrete

Results 2

- 11 features: 1 is the Medium, 10 describe theme and recipient (*give the theme to the recipient*): Concreteness, Animacy, Definiteness, Discourse givenness, Length difference, Number, Person, Pronominality
- For MRC, WN-HIER and WN-PHYS, data is missing. We follow the standard approach: removing instances with a missing value

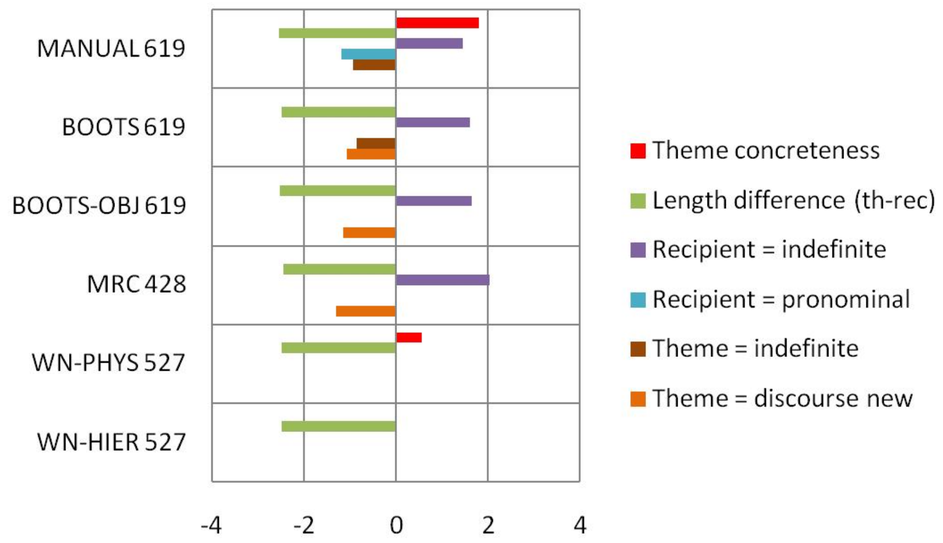

Figure: Regression coefficients for the significant features, for each model (positive coefficients favouring the variant with *to*, negative ones the variant without *to*)

- Concreteness is only significant for MANUAL and WN-PHYS
- Missing values for MRC, WN-PHYS and WN-HIER make other significant effects disappear

Conclusion

1. Different approaches lead to locally very different annotations
2. Different approaches lead to very different conclusions for the feature in question, and can influence conclusions for other features

- Coverage is a major concern, especially since significance measurements are often hindered by data set size and composition
- Bootstrapping process is risky: the bootstrapped distinction can start shifting because of strong newly induced seeds