EBERHARD KARLS UNIVERSITÄT TÜBINGEN

SFB 833
BEDEUTUNGSKONSTITUTION
PROJECTS A3 & A1

A Corpus-Based Model of Semantic Plausibility for German Bracketing Paradoxes

Corina Dima · Jianqiang Ma · Sebastian Bücking · Frauke Buscher · Johanna Herdtfelder · Julia Lukassek · Anna Prysłopska · Erhard Hinrichs · Daniël De Kok · Claudia Maienborn

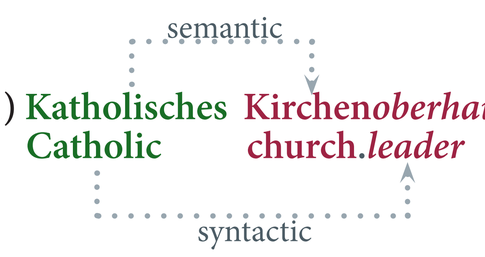
SFB 833, Deutsches Seminar and Seminar für Sprachwissenschaft, University of Tübingen, Germany

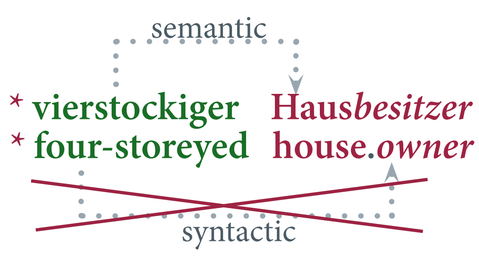
Topic

We investigate German constructions consisting of an adjective (A) and a two-part nominal compound (N_1N_2), focusing on two aspects:

(i) What are the prerequisites for semantically possible A-N_1N_2 constructions?

(ii) Which semantic factors determine the availability of **bracketing paradox (BP)** readings (i.e. where the adjective modifies the first noun) in such constructions?





Theoretical Model for Bracketing Paradoxes

The ungrammaticality of Ex. (2) suggests a prerequisite for any A-N_1N_2 construction:

H_1: if A-N_1N_2 is *semantically possible*, then A-N_2 is *semantically possible*.

The preference for the bracketing paradox reading is modeled by:

H_2: for examples where H1 holds, the higher the *semantic plausibility* of A-N_1 *relative* to A-N_2, the more likely it is that A-N_1N_2 is a *bracketing paradox*.

Frequency-Based Semantic Plausibility Model for Bracketing Paradoxes

Given a corpus, compute for an A-N_1N_2 construction:

- $freq_{A\text{-}N1}$, frequency (number of corpus occurences) of A-N_1
- $freq_{A\text{-}N2}$, frequency of A-N_2
- $freq_{A\text{-}N1N2}$, frequency of A-N_1N_2
- $rf_{A\text{-}N1N2} = freq_{A\text{-}N1}/freq_{A\text{-}N2}$, the relative frequency of A-N_1 and A-N_2

- a construction is considered to be *semantically possible* if its frequency is > 0.
- the *semantic plausibility score* of a construction is defined to be its frequency count
- the *relative semantic plausibility score* of an A-N_1N_2 construction is the relative frequency of the two adjective-noun pairs

Annotation

Five annotators annotated the 77 semantically possible A-N_1N_2 constructions used for testing H_1 with regard to the questions:

Q1: Is the A-N_1N_2 construction grammatical as a whole? yes/no

Q2: Which reading is preferred? BP, non-BP, equal preference

The majority of the annotators had to agree for an item to be categorized.

Results & Discussion

Grammaticality (Q1):

- 73 of the 77 constructions were considered grammatical, 2 ungrammatical and 2 could not be categorized (moderate agreement, Fleiss' $\kappa = 0.45$).

Preference (Q2):

- the 73 constructions judged as grammatical were further annotated for the preferred reading.
- the annotators categorized 62 constructions: 46 items were judged to have a non-BP reading; 16 items were judged to have a BP reading. 11 items could not be categorized (moderate agreement, Fleiss' $\kappa = 0.58$).
- no items were annotated as having equal preference between the iconic and anti-iconic readings.
- several of the uncategorized examples are prototypes for bracketing paradoxes according to the theoretical literature (e.g. *katholisches Kirchenoberhaupt*)

scan4paper

Testing H_1

- data: the *semantically possible* constructions from our dataset (77 out of 198 examples with frequency > 0)
- verify for these constructions the H_1 hypothesis, which states that if A-N_1N_2 is semantically possible, A-N_2 should also be semantically possible; use the corpus frequency of the A-N_2 construction as an indicator

Results

For 90.9% of the data (70 examples), the model correctly predicts A-N_2 to be semantically possible, based on the frequency information.

Discussion

Ex. (2), *vierstockiger Hausbesitzer*, appears 10 times in the corpus in meta-discussions concerning its semantic impossibility, and is therefore considered semantically possible by the model. However, the construction *vierstockiger Besitzer* never occurs, therefore pointing to a logical discrepancy.

Testing H_2

- data: 62 constructions catogorized by the annotators as BP or non-BP
- classify each construction as BP or non-BP using a logistic regression classifier and the normalized *relative semantic plausibility score* as an indicator. Results for 10-fold cross-validation.

Results

Data set	F1 score	Accuracy (%)
62 (16BP, 46 non-BP)	0.90	95.71

Discussion

The relative semantic plausibility score is a very good predictor for the preferred interpretation of a particular construction, despite the imbalanced number of instances for each class.

Conclusions

- the corpus-based frequency model confirmed hypotheses H_1 and H_2
- for all A-N_1N_2 constructions, A-N_2 must be semantically possible, irrespective of whether A can modify N_1 or not (H_1)
- the higher the relative semantic plausibility score of a construction, the more likely it is that the construction is a bracketing paradox (H_2)
- *katholisches Kirchenoberhaupt*, the textbook example for bracketing paradoxes, received, surprisingly, mixed ratings from the annotators; this suggests that the distinction might not be a binary one, but one that involves graded judgments
- future extension: rating study with graded judgments regarding the BP-ness of A-N_1N_2 constructions
- limitations of the frequency-based model: (i) in *intelligenter Tierarzt (intelligent animal.doctor)*, the pair *intelligentes Tier* is more frequent than *intelligenter Arzt* (a doctor is generally implied to be intelligent), leading to an erroneous BP interpertation (*doctor for intelligent animals*); (ii) predictions cannot be made in the absence of frequency information, resulting in only a partial analysis of the initial dataset
- future extension: use distributional semantics models, which integrate information about the semantics of the constitiuents, to improve the generalization capability of the model