



Date: 2026-05-13

---

## notes on papers

snc

- t1

### (4) annotations: Frankowsky\_2022.pdf

paper: Frankowsky (2022)

page

note

comment

2

Frankowsky

#cite\_Frankowsky\_2022.pdf\_\_

11

the date of attestation for each formation was identified.

17

Data concerning the date of attestation were not available for all 2,337 cases of ProtICCs. The year of the utterance could be identified in 772 cases (33.0%).

25

Schäfer, Roland & Bildhauer, Felix. 2012. Building Large Corpora from the Web Using a New Efficient Tool Chain. LREC: 486–493.

**(1) annotations: Wilkinson et al. 2016.pdf**

paper: @

page

note

comment

1

Findability, Accessibility, Interoperability, and Reusability

**(18) annotations: Zinsmeister 2013.pdf**

paper: Zinsmeister (2013)

page

note

comment

1

Heike Zinsmeister

#cite\_Zinsmeister 2013.pdf\_\_

2

Our corpus-based study is grounded in the approach of Distributional Semantics, which assumes that the meaning of a word can be modeled by collecting information about its distribution in a corpus

2

we distinguish between three different semantic compound types. The first type is traditionally referred to as a determinative (or endocentric) compound (cf. Olsen 2000). Its meaning is compositionally derived from the meanings of its parts: The meaning of the head noun denotes the type of referent, and the meaning of the modifier noun modifies this in some semantic relation to the head noun, such that it further specifies or modifies the meaning of the head noun.

2

The distributional information is collected by examination of co-occurring words, structures, or relations. The sum of this contextual information can be viewed as a kind of mold, a container that shapes the meaning of the word in a characteristic fashion.

3

The final semantic type is traditionally called the copulative (or dvandva) compound

3

includes all noun-noun compounds that are lexicalized such that their meanings are opaque with respect to the meanings of their parts.(..)(..)They are semantically

3

is the least frequent type of compound in German. In Fanselow's (1981b) definition, the meaning of a copulative compound is created by applying the basic relation 'AND' to the meanings of its parts

3

The second semantic type is traditionally referred to as the possessive (also bahuvrihi or exocentric) compound.

4

the distributions of the target nouns are vectors.

5

In such a vector space, each context word corresponds to a dimension, and the distribution of a target word is a vector in this multidimensional space. The similarity between two words can then be measured in terms of the distance between their vectors in the space

5

The similarity of two distributions is interpreted as the distance between their vectors. A common way to measure this distance is to calculate the cosine of the angle between the two vectors (e.g., Erk 2012: 637). If two vectors point in the same direction, meaning that the angle between them is  $0^\circ$ , then the cosine will be 1. Otherwise, the cosine will be less than 1; it will be -1 if the two vectors point in exactly opposite directions

5

The basic elements in the standard model are the co-occurrence frequencies of context words. Context is defined as a window on the token string.

6

The Kullback-Leibler divergence

6

skew divergence as a variant of the KullbackLeibler divergence. Instead of calculating the divergence using the observed probability distributions, this function smoothes the values of  $q$ , meaning that  $q$  will be positive for all dimensions for which  $p$  has a positive value

6

Lee (1999, 2001) has introduced skew divergence

7

extracted all words that were annotated as common nouns in the corpus and analyzed them further using SMOR (Schmid et al. 2004

7

skewed distribution of head family size

what?

8

skew divergence distributions D. These scores range from a minimum of 0.093, observed for the divergence of a head from its compound (cf. leftmost boxplot  $D(c(..)(..)h)$ ), to a maximum of 2.293, observed for the divergence of a compound from its modifier

Frankowsky, Maximilian. 2022. *Extravagant Expressions Denoting Quite Normal Entities: Identical Constituent Compounds in German*. John Benjamins Publishing Company. <https://www.degruyterbrill.com/document/doi/10.1075/slcs.223.07fra/html>.

Zinsmeister, Heike. 2013. "Corpus-Based Modeling of the Semantic Transparency of Noun-Noun Compounds." In *A Festschrift for Susan Olsen*, edited by Holden Härtl. Akademie Verlag. <https://doi.org/doi:10.1524/9783050063799.303>.