Association Norms of German Noun Compounds

Sabine Schulte im Walde
Institut für Maschinelle Sprachverarbeitung
Universität Stuttgart
schulte@ims.uni-stuttgart.de

Susanne Borgwaldt
Institut für Germanistik
TU Braunschweig
s.borgwaldt@tu-bs.de

Ronny Jauch
Institut für Maschinelle Sprachverarbeitung
Universität Stuttgart
jauchry@ims.uni-stuttgart.de

Overview: Noun Compounds, Compositionality, and Associations

German Noun Compounds

- Compounds: combinations of two or more simplex words; examples for German:
  Ahornblatt **`maple leaf´,** ***Nähmaschine*** **`sewing machine´,** ***Rotkohl*** **`red cabbage´**
- Basis: 450 concrete, depictable noun compounds from von der Heide and Borgwaldt (2009)

Compositionality

- Interest: degrees of compositionality of German noun compounds
  - ➢ cognitive and computational research on
    - lexical-semantic properties of compounds and their constituents
    - semantic relatedness between the compounds and their constituents
  - ➢ psycholinguistic research on compounds in the mental lexicon:
    - are compounds stored as single units, or are they stored decomposed?
    - are the meanings of the constituents activated during processing?

Association Norms

- Motivation: association norms have a long tradition in psycholinguistic research to investigate semantic memory
- Implicit notion: associates reflect meaning components of words
- Assumption: associations are useful for research on degrees of compound compositionality, to identify salient properties of the compound components
- Result: new lexical-semantic resource – associations of compounds and their constituents
- General procedure to collect associations and obtain association norms:
  - present *target stimuli* to participants in an experiment
  - participants provide *associate responses*, i.e., words spontaneously called to mind
  - quantification over target-association pairs: *association norms*

Obtaining Association Norms from a Web Experiment

- Web experiment: 442 compound nouns and their constituents → 996 target stimuli;
  example stimuli: *Ahornblatt, Ahorn, Blatt; Nähmaschine, nähen, Maschine; Rotkohl, rot, Kohl;*
  random division into 12 separate experimental lists of 83 nouns each
- Data: 268 participants, between 14 and 28 for each data set;
  maximum number of responses: 3 associations; average: 2.6 associations;
  collection: 46,989/29,221 association tokens/types in 17,906 trials;
  861 trials without response (mostly for cranberry morphemes, e.g., *Him* in *Himbeere)*
- Create association norms: for each stimulus (i.e., the compounds and their constituents), quantify over all responses in the experiment
- Two modes: perform quantification
  (i) considering only the first response in each trial, and
  (ii) considering all responses, disregarding the order of the associates (presented below);
  background: association chain effects (McEvoy and Nelson, 1982), e.g., *tree – leaf – to float*

Part-of-Speech Tag Distribution of Responses

	N	ADJ	V	OTHER
TOTAL FREQ	33,322	6,835	5,264	1,827
TOTAL PROB	71%	14%	11%	4%
Dose `can´	92%	3%	5%	0%
Notenschlüssel `clef´	96%	2%	0%	3%
Faden `thread´	43%	10%	44%	3%
Türklinke `door handle´	54%	3%	38%	5%
Zitrone `lemon´	20%	74%	3%	3%
Wollschal `woollen scarf´	37%	49%	13%	1%

Association Frequencies for Example Noun Compounds and their Constituents (mode (ii): all associations)

Ahornblatt **`maple leaf´**			***Ahorn*** **`maple´**			***Blatt*** **`leaf´**		
Kanada	`Canada´	8	*Baum*	`tree´	14	*Papier*	`paper´	8
Baum	`tree´	7	*Sirup*	`syrup´	11	*schreiben*	`write´	4
Herbst	`autumn´	7	*Kanada*	`Canada´	9	*grün*	`green´	4
Sirup	`syrup´	4	*Blatt*	`leaf´	6	*Herbst*	`autumn´	2
rot	`red´	3	*Blätter*	`leaves´	4	*Blume*	`flower´	2
Wald	`forest´	2	*rot*	`red´	2	*Käfer*	`bug´	1
Form	`shape´	2	*Ahornsirup*	`maple syrup´	2	*fallen*	`fall´	1
bunt	`colourful´	2	*grün*	`green´	2	*Blattadern*	`leaf veins´	1
zackig	`jagged´	2	*Herbst*	`autumn´	1	*weiß*	`white´	1
Ahornsirup	`maple syrup´	2	*Kindheit*	`childhood´	1			

Fliegenpilz **`fly agaric´**			***Fliege*** **`fly/bow tie´**			***Pilz*** **`mushroom´**		
giftig	`poisonous´	12	*nervig*	`annoying´	4	*Wald*	`forest´	10
rot	`red´	7	*summen*	`buzz´	2	*Fliegenpilz*	`fly agaric´	8
Wald	`forest´	5	*lästig*	`annoying´	2	*sammeln*	`collect´	4
Gift	`poison´	2	*Insekt*	`bug´	2	*giftig*	`poisonous´	4
Hut	`cap´	1	*Tier*	`animal´	2	*Schimmel*	`mould´	2
Glück	`fortune´	1	*Fliegenklatsche*	`fly flap´	2	*Suche*	`search´	2
Kinderbuch	`children's book´	1	*Krawatte*	`tie´	2	*Hut*	`cap´	1
Pflanze	`plant´	1	*Sommer*	`summer´	2	*Pilzpfanne*	`mushroom pan´	1
Muster	`pattern´	1	*Anzug*	`suit´	1	*essbar*	`eatable´	1
weiß	`white´	1	*fangen*	`catch´	1	*Suppe*	`soup´	1

Case Study: Compositionality of German Noun Compounds

- Goal: explore whether associations to compounds provide insight into salient properties that in turn could be useful for computational models of compound compositionality
- Method: simple association overlap measure to predict the degree of compositionality of the experiment compound nouns, i.e.,
  use proportion of shared associations of a compound and a constituent in comparison to the total number of associations of the compound
- Example: *Ahornblatt* receives a total of 39 associations, out of which it shares 31 with the first constituent *Ahorn* and 14 with the second constituent *Blatt*.
  Thus, the predicted degrees of compositionality are 31/39 = 0.79 for *Ahornblatt–Ahorn*, and 14/39 = 0.36 for *Ahornblatt–Blatt*.
- The predicted degrees of compositionality (system scale: 0 to 1) are compared against the mean compositionality judgements (human scale: 1/opaque to 7/strong compositionality) by von der Heide and Borgwaldt, using the Spearman rank-order correlation coefficient.
- Surprisingly successful correlation: $r_s = 0.5228$, $p < .000001$.

Compositionality Scores for Modifiers and Heads (System vs. Human)

Compound	Modifier	Head	Modifier Scores		Head Scores	
			System	Human	System	Human
Ahornblatt `maple leaf´`	*Ahorn* `maple´`	*Blatt* `leaf´`	.69	5.63	.35	5.70
Badeanzug `bathing costume´`	*baden* `bath´`	*Anzug* `suit´`	.68	6.13	.00	3.03
Feuerwerk `fireworks´`	*Feuer* `fire´`	*Werk* `opus´`	.02	4.20	.02	2.80
Fliegenpilz `fly agaric´`	*Fliege* `fly/bow tie´`	*Pilz* `mushroom´`	.00	1.93	.47	6.55
Nähmaschine `sewing machine´`	*nähen* `sew´`	*Maschine* `machine´`	.16	6.03	.00	4.93
Rotkohl `red cabbage´`	*rot* `red´`	*Kohl* `cabbage´`	.03	2.70	.30	5.83
Schlittenhund `sledge dog´`	*Schlitten* `sledge´`	*Hund* `dog´`	.16	5.70	.13	5.10