Linguistic Constructions Related to the Distinction between Emotion-Denoting and Affect-Denoting German Nouns

Roland Schäfer
German Grammar Group, FU Berlin

SLE 2012, Stockholm (version of August 30, 2012)

Outline

We are here. . .

What are we talking about?

(1) Die Tibeter leben in ständiger Angst vor ihren
the tibetans live in permanent fear of their
Unterdrückern und Besatzern.
suppressors and occupying forces
The tibetans live in permanent fear of their suppressors and the occupying forces.

(2) Aber er konnte nichts hören, und Angst befiel ihn.
but he could nothig hear, and fear beset him
He couldn't hear anything and fear beset him.

We are here. . .

Affective Lexicon

- Affective Lexicon: Lexemes which **descriptively** characterize emotions.
- Emotion words **name** emotions (*love*, *hate*),
- . . . they **don't express** emotions (or only marginally).
- As opposed to: emotive words like *hopefully*, interjections, etc.
- (Hermanns, 1995, 1996; Jäger and Plum, 1988)
- This talk is **only about nouns**!

Linguistic realizations of emotions

"Emotions, of course, are not linguistic things. However, the most convenient nonphenomenological access we have to them is through language. Thus one reasonable way to separate emotions from nonemotions is to consider what are the referents of putative emotion words." (Ortony et al., 1987, 342)

"If, in the evolution of languages, certain types of distinctions between different types of emotional processes have been considered important enough for communication to generate different words or expressions, social and behavioral scientists should consider these distinctions worthy of study." (Scherer, 2005, 707)

Emotions vs. affective states

"Emotions are always about something, and this 'aboutness' is a useful way to distinguish emotions from other affectice states such as moods." (Clore and Ortony, 2000)

- The experiencer may not be aware of the aboutness object.
- The object is not necessarily syntactically realized.

This (exploratory) study

- The idea is based on Clore and Ortony (2000).
- Can we classify emotion-denoting **nouns** by their co-occurrences with object-denoting expressions?
- Even if the object is not necessarily syntactically realized, **certain nouns might prefer such a realization** more than others.
- Initially, the idea was to look at potential **near-synonyms**, but see below.

Collocations and meaning

- **Collocation** is about the frequency of co-occurrence of two words
- The degree to which (content) words collocate is obviously (but not in a trivial way) related to the meaning of the words.
- An overview of collocations and context-based meaning extraction can be found in (Manning and Schütze, 1999, 151–189, 294–308).
- Some methods try to reduce meaning of a word to the set of its contexts (informally speaking).
- One such theory is Latent Semantic Analysis (Landauer and Dumais, 1997; Landauer et al., 1998).

Constructions/Collostructions

- Construction Grammar (Goldberg, 1995, 2006) defines constructions (form-meaning pairs) as the basic linguistic unit.
- No restrictions as to the level of abstraction in a construction: "In other words, a construction is any linguistic expression, no matter how concrete or abstract, that is directly associated with a particular meaning or function, . . ." (Stefanowitsch and Gries, 2003, 212)
- Collostructional Analysis (Stefanowitsch and Gries, 2003; Gries and Stefanowitsch, 2004) provides a means of calculating the affinity of words to constructions.

How far do we want to go with psycholinguistic claims based on corpora?

A caveat. . .

- It is a matter of debate whether/to which extent information extracted from corpora can be interpreted in terms of cognitive linguistics.
- I subscribe to a careful view of Corpus Linguistics as. . .

> *". . . psycholinguistically informed, (cognitively-inspired) usage-based linguistics."* (Gries, 2010, p.334)

Freie Universität Berlin

⋆ Experimental methods and complementation from corpora

However, it appears to the outsider that a mixed approach (experimental/corpus based) could be of mutual benefit.

"Valid conclusions from scaling studies require that the items in the stimulus set do indeed belong to the domain being studied, and are appropriately sampled." (Ortony et al., 1987, 343)

"The difficulty with simply asking subjects to rate whether a particular term is a good example of an emotion is that the linguistic context in which subjects implicitly consider the terms is usually uncontrolled and unknown." (Ortony et al., 1987, 343)

Freie Universität Berlin

We are here. . .

Freie Universität Berlin

Goal of the study

- The meaning of an emotion word (or any word) influences the affinity to certain constructions.
- We define a very abstract **Aboutness Object Construction (AOC)**: An emotion-denoting **noun** occurs with some linguistic realization of a cognitive object associated with the emotion.
- The respective words could thus be classified as denoting more or less an **emotion** or rather another **affective state**.
- For near-synonyms, the attraction to or repulsion from this construction could be a distinguishing factor.

Freie Universität Berlin

⋆ Previous study using the DWDS Kernkorpus

- Two years ago, occurrences of 4 pairs of potential near-synonyms were annotated for whether they were realized in the AOC or not.
- Source: the small balanced DWDS Kernkorpus of the 20th century (Geyken, 2006).
- It turned out that the samples were not annotated consistently by some assistants.

DECOW2012 web corpus

- For the real study, new samples were taken from the 9.1 billion token web corpus DECOW2012 (Schäfer and Bildhauer, 2012).
- The corpus data was downloaded in 2011.
- A crippled form (just sentences, no documents) dan be downloaded: `http://hpsg.fu-berlin.de/cow/`
- The following slides show the composition as evaluated post-hoc according to the COWCat scheme (derived from Sharoff, 2006).

Authorship, Mode, Audience

	DECOW2012		ESCOW2012	
Type	Percentage	CI ±%	Percentage	CI ±%
Authorship				
Single, female	**6.0**	2.8	**5.0**	2.5
Single, male	**11.5**	3.7	**16.5**	4.3
Multiple	**36.0**	5.6	**16.5**	4.3
Corporate	**21.0**	4.7	**20.5**	4.7
Unknown	**25.5**	5.0	**41.5**	5.7
Mode				
Written	**71.0**	5.0	**86.0**	4.0
Spoken	**1.0**	3.0	**2.5**	1.8
Quasi-Spontaneous	**22.5**	4.9	**3.5**	2.1
Blogmix	**4.5**	2.4	**8.0**	3.2
Audience				
General	**75.5**	5.0	**94.0**	2.8
Informed	**17.0**	4.4	**2.5**	1.8
Professional	**7.5**	3.0	**3.5**	2.1

categorization results for DECOW2012/ESCOW2012; N=200; confidence level=90%

⋆ Aim, Domain

	DECOW2012		ESCOW2012	
Type	Percentage	CI ±%	Percentage	CI ±%
Aim				
Recommendation	**12.5**	3.8	**7.0**	3.0
Instruction	**4.5**	2.4	**6.0**	2.8
Information	**36.0**	5.5	**41.5**	5.7
Discussion	**47.0**	5.8	**44.5**	5.8
Fiction	**0.0**	0.0	**1.0**	1.2
Domain				
Science	**2.5**	1.8	**5.0**	2.5
Technology	**14.0**	4.0	**6.5**	2.9
Medical	**4.5**	2.4	**4.0**	2.3
Pol., Soc., Hist.	**21.5**	4.8	**21.0**	4.7
Business, Law	**10.0**	3.5	**12.5**	3.8
Arts	**8.5**	3.2	**8.5**	3.2
Beliefs	**5.0**	2.5	**3.0**	2.0
Life, Leisure	**34.0**	5.5	**39.5**	5.7

categorization results for DECOW2012/ESCOW2012; N=200; confidence level=90%

Sample size and distributions

Lemma	Transl.	AOC	¬AOC
Abscheu	disgust	114	81
Angst	fear	88	107
Begierde	desire	118	78
Eifersucht	jealousy	169	25
Ekel	disgust	130	70
Furcht	fear	91	101
Hochmut	pride	188	7
Hoffnung	hope	97	98
Lust	lust	47	144
Neid	envy	165	20
Sehnsucht	longing	96	93
Stolz	pride	135	43
Verlangen	desire	78	117
Wut	anger	171	27
Zorn	wrath	163	18
Zuversicht	confidence	164	29

What are object encoding expressions?

- Operationalization was problematic only in some cases.
- Some lemmas have a clear preference for PPs or even a specific preposition.
- Others tend to embed complement clauses.
- Consequently: different preferences for the syntactic constituent to denote objects, states, event.
- For today's purposes, these complications are ignored.

Examples I

(3) Meine Stimme trieft nur so von Abscheu.
My voice oozes just so with disgust
My voice is just full of digust.

(4) Und es wird sich da zeigen, ob du vor ihnen
and it will itself there show, whether you from them
Abscheu bekommen wirst.
disgust get will
And then we'll see whether you will be disgusted by them.

Examples II

(5) Seine Zuversicht, diese Marke zu erreichen, gründet sich
his confidence, this mark to reach, grounds itself
auch auf die Quote von mehr als 30% verbandsfreier
also on the proportion of more than 30% union-free
Händler.
traders.
His confidence to be able to reach this mark is based on the fact that 30% of the traders are not organized in a union.

(6) Wünsche euch gutes Wetter und viel Zuversicht.
wish yall good weather and much confidence.
I wish you all good weather and much confidence.

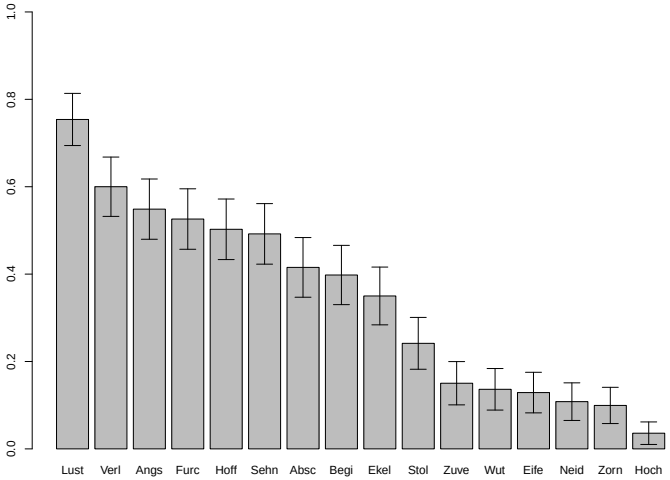
Examples III

(7) Der Ekel lähmte sie.
the disgust paralyzed them
They were paralyzed in disgust.

(8) Anfangs hatte ich auch Ekel vor dem Tieren.
at the beginning had I also disgust for the animals
At the beginning, I was disgusted by the animals, too.

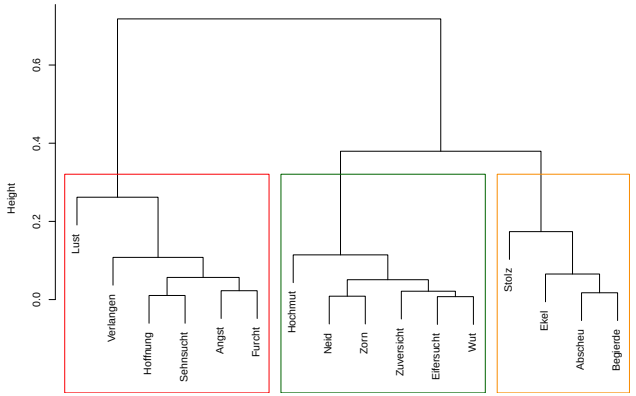
(9) [...] und das macht Lorelai wohl Angst.
[...] and that makes Lorelai maybe fear
This seems to be what causes Lorelai's fear.

Proportions



proportions of each lemma in AOC; $N = 200$ per lemma (3,200 total); confidence intervals at 95% level.

Failure of the original "near-synonyms" hypothesis?



HACA on proportion of cases with object; `hclust()`, distance: Manhattan, method: complete; **harmonic means:** 0.559, 0.087, 0.336

Not quite. . .

Significantly different (potential) near-synonyms:

Lemmas	p_{Fisher}	odds ratio
Begierde/**Lust**	< 0.001	4.615
Hochmut/**Stolz**	< 0.001	8.511
Hoffnung/Zuversicht	< 0.001	0.176
Lust/Verlangen	< 0.01	0.49
Sehnsucht/**Verlangen**	< 0.05	1.542

Collostructional Analysis

- CS would allow us to quantify the association strength in a way that takes into account the overall frequencies of the lemmas and constructions in the corpus.
- In CS, **p values of Fisher Exact Tests** are used to measure **association/repulsion** strength.

	construction	¬ construction
lemma	...	...
¬ lemma	...	...

- **The smaller p, the stronger the association.**

Why not Collostructional Analysis

- Usually, the "fourth field" (¬lemma & ¬ construction) is a matter of guesswork (cf. also Schmid, 2010).
- In this case, the **third field (¬ lemma & construction)** is even more problematic.
- Since the construction is so abstract, we cannot hope to quantify over the total number of occurrences of all the diverse sub-types of the construction (different PPs, NP-internal complement clauses, causatives constructions, etc.) in which none of the lemmas occurs.
- At least, the required effort would be extreme.

We are here. . .

Achievements

- exploratory study of a (quite abstract) aboutness object construction
- overall: clear but graded differences attraction/repulsion strengths observable
- some near-synonym pairs: very similar w. r. t. AOC
- some others: attraction to AOC clearly a distinguishing factor

Possible further work

- other parts of speech
- full differentiation of specific constructions (diverse PPs, complement clauses, causatives, etc.)
- other psychological factors to consider, e. g., valency (positive/negative)

References I

Clore, Gerald and Ortony, Andrew. 2000. Cognition in Emotion: Always, Sometimes, or Never? In Richard D. Lane and Lynn Nadel (eds.), *Cognitive Neuroscience of Emotion*, pages 24–61, New York: OUP.

Geyken, Alexander. 2006. The DWDS Corpus: A Reference Corpus for the German Language of the 20th Century. In Christiane D. Fellbaum (ed.), *Collocations and Idioms: Linguistic, Lexicographic, and Computational Aspects*, London: Continuum Press.

Goldberg, Adele. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: OUP.

Goldberg, Adele E. 1995. *Constructions. A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Gries, Stefan Th. and Stefanowitsch, Anatol. 2004. Extending collostructional analysis. A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics* 9(1), 97–129.

Gries, Stefan Thomas. 2010. Corpus Linguistics and Theoretical Linguistics – A Love-Hate Relationship? Not Necessarily. *International Journal of Corpus Linguistics* 15(3), 327 – 343.

Hermanns, F. 1995. Kognition, Emotion, Intention. Dimensionen lexikalischer Semantik. In H.E.Wiegand (ed.), *Wörterbücher in der Diskussion. Vorträge aus dem Heidelberger Lexikographischem Kolloquium*, pages 256—-278, Tübingen: Niemeyer.

Hermanns, F. 1996. Emotion im Wörterbuch. Zur Lexikologie von affektiver Lexik. In G. Harras (ed.), *Die Ordnung der Wörter. Kognitive und lexikalische Strukturen*, pages 138–178, Berlin: De Gruyter.

Jäger, L. and Plum, S. 1988. Historisches Wörterbuch des deutschen Gefühlswortschatzes. Theoretische und methodische Probleme. In L. Jäger (ed.), *Zur historischen Semantik des deutschen Gefühlswortschatzes*, pages 5–55, Aachen: Alano.

References II

Landauer, T. K. and Dumais, S. T. 1997. A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review* 104, 211–40.

Landauer, T. K., Foltz, P. W. and Laham, D. 1998. Introduction to Latent Semantic Analysis. *Discourse Processes* 25, 259–84.

Manning, Christopher D. and Schütze, Hinrich. 1999. *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.

Ortony, A., Clore, G.L. and Foss, M.A. 1987. The referential structure of the affective lexicon. *Cognitive Science* 11, 341–364.

Schäfer, Roland and Bildhauer, Felix. 2012. Building Large Corpora from the Web Using a New Efficient Tool Chain. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Ugur Dogan, Bente Maegaard, Joseph Mariani, Jan Odijk and Stelios Piperidis (eds.), *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, pages 486–493, Istanbul: ELRA.

Scherer, K.R. 2005. What are emotions? And how can they be measured? *Social Science Information* 44(4), 693–727.

Schmid, Hans-Jörg. 2010. Does frequency in text instantiate entrenchment in the cognitive system? In Dylan Glynn and Kerstin Fischer (eds.), *Quantitative methods in cognitive semantics: Corpus-driven approaches*, pages 101–133, Berlin etc.: De Gruyter.

Sharoff, Serge. 2006. Creating General-Purpose Corpora Using Automated Search Engine Queries. In Marco Baroni and Silvia Bernardini (eds.), *WaCky! Working papers on the Web as Corpus*, Bologna: GEDIT.

Stefanowitsch, Anatol and Gries, Stefan Th. 2003. Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics* 8(2), 209–243.