

Building an OLIF-based lexical database for representing constructions

Luder Marc
University of Zurich, Switzerland
m.luder@psychologie.uzh.ch

This paper demonstrates the implementation of a small monolingual lexical database for German (currently 7000 entries) for the purpose of manual and automated lexical queries. The lexicon is part of a web based text analysis application which serves to systematically analyze patient's narratives from psychotherapy sessions. The narratives, small stories from everyday life, are conceived as dramaturgically constructed and performed linguistic productions (Boothe, 2004). The specific function of the lexicon in this context is to provide the means for the lexical coding of the story vocabulary, which is an important step in the narrative analysis¹. The interdisciplinary project is located in psychology, but involves likewise linguistics, corpus linguistics, lexicography, interactional linguistics and conversation analysis.

The lexical database is implemented in the OLIF format (Open Lexicon Interchange Format)², the XML structure is mapped to a MySQL-database. OLIF allows for the entry of single words as well as multi word units. We customize slightly the annotation of lexical entries to support a Construction Grammar (Croft, 2001; Goldberg, 1995) approach with fully lexicalized constructions, providing as much as possible syntactic, semantic and pragmatic information (Luder, Clematide, & Distl, 2008). Constructions are single or multi word units, including idioms, metaphors and other phraseological and collocational patterns of different types (Granger & Paquot, 2008; Villavicencio, Bond, Korhonen, & McCarthy, 2005), according to the language in the transcripts of spoken language (German with regional dialect characteristics). The goal of the project is to establish a prototypical lexicon for constructions as pairings of form and meaning with a rich set of linguistic features for disambiguating the query results. Aside from the attributes provided by the OLIF structure, we use data from different dictionaries and lexical resources in order to enrich the lexicographic description of the entries, such as "Der Kleine Wahrig" (Wahrig, 2007), Duden 11 (Scholze-Stubenrecht & Wermke, 2008), VALBU (Schumacher, Kubczak, Schmidt, & de Ruiter, 2004), Dornseiff (Dornseiff, 2004) and FrameNET (Fillmore, Johnson, & Petruck, 2003). The lexicon entries include a conceptual category of our narrative analysis coding system. In this regard the main emphasis lies on the actions told and performed by the narrator, i.e. on the coding of verbs and verbal constructions. For this reason we are especially interested in semantic and pragmatic verb classifications of German, as described and proposed by various authors (Čulo, Erk, Padó, & Schulte Im Walde, 2008; Fellbaum, Geyken, Herold, Koerner, & Neumann, 2006; Fellbaum, 2007; Hanks, Urbschat, & Gehweiler, 2006).

The paper presents the structure of lexical multi-word entries and the potential of the analysis in matching the form-meaning pairs of the lexicon against the expressions in the transcribed discourse and disambiguating them (e.g. "mit einem blauen Auge

¹ Erzählanalyse JAKOB: <http://www.jakob.uzh.ch>

² <http://www.olif.net>

davonkommen”, “to get off with a slap on the wrist”; “ein Zeug machen”, “to make a fuss”). Currently the matching works for lexicalized expressions. We are planning to extend the application by implementing more abstract, only partially lexicalized constructions (e.g. “haben zu +[Infinitiv]”, “to have to +[infinitive]” or “typisch +[human-animate]”, “typically +[human-animate]”).

An intention of our project is to integrate theoretical perspectives of conversation analysis and interactional linguistics with perspectives of corpus linguistics. Speakers use for their utterances a big inventory of prefabricated patterns, from entirely fixed units to more loose phraseological constructs (Moon, 2008). “Human beings store in their brains not just words in isolation, but also sets of stereotypical syntagmatic patterns associated with each word.” (Hanks, 2004, p. 245). The data background of the lexicon is therefore the language use in our corpus of therapy conversations. A typical way to create a new lexicon entry is as follows: After looking for salient examples of a construction in the therapy transcripts, we look for the constructions in the whole data corpus. Selected quotations are then analysed and interpreted using techniques of conversation analysis (Goodwin & Heritage, 1990). We then search for further quotation examples in alternative corpora (e.g. internet documents) to approve or refuse our hypotheses about the meaning of the construction.

A special issue is the constitution of meaning (Deppermann & Spranz-Fogasy, 2006): Interactional linguists advocate the constitution of meaning as emergent, accordingly meaning is created by speakers locally and in the process of discourse (Günthner, 2007). An investigation of the use of constructions over time, e.g. the analysis of the construction “ein Zeug machen” (“to make a fuss”) over the period of a long-term psychotherapy (300 sessions) shows evidence that the specific meaning of this idiom is more and more fixed over time and that the therapist is adopting the construction from the client (Luder, 2009). This could be an argument for Hoey’s (2005) theory of lexical priming: words and collocation patterns get their individual meaning by frequent use in specific contexts, and interaction with others results in shared meaning constitution. The handling of the emergence phenomenon and the question of how to integrate it in a lexicon represents an interesting problem for further studies.

References

- Boothe, B. (2004). *Der Patient als Erzähler in der Psychotherapie* (2nd ed.). Giessen: Psychosozial-Verlag.
- Croft, W. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Čulo, O., Erk, K., Padó, S., & Schulte Im Walde, S. (2008). Comparing and combining semantic verb classifications. *Language Resources and Evaluation*, 42, 265–291.
- Deppermann, A., & Spranz-Fogasy, T. (Eds.) (2006). *Stauffenburg Linguistik: Vol. 27. Be-deuten: Wie Bedeutung im Gespräch entsteht* (2nd ed.). Tübingen: Stauffenburg-Verlag.
- Dornseiff, F. (2004). *Der deutsche Wortschatz nach Sachgruppen* (8th ed.). Berlin: de Gruyter.

- Fellbaum, C. (2007). Argument selection and alternations in VP idioms. In C. Fellbaum (Ed.), *Idioms and collocations. Corpus-based linguistic and lexicographic studies* (pp. 188–202). London: Continuum.
- Fellbaum, C., Geyken, A., Herold, A., Koerner, F., & Neumann, G. (2006). Corpus-based Studies of German Idioms and Light Verbs. *International Journal of Lexicography*, 19(4), 349–360.
- Fillmore, C. J., Johnson, C. R., & Petruck, M. R. L. (2003). Background to Framenet. *International Journal of Lexicography*, 16(3), 235–250.
- Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: Univ. of Chicago Press.
- Goodwin, C., & Heritage, J. (1990). Conversation Analysis. *Annual Review of Anthropology*, 19, 283–307.
- Granger, S., & Paquot, M. (2008). Disentangling the phraseological web. In S. Granger & F. Meunier (Eds.), *Phraseology. An interdisciplinary perspective* (pp. 27–50). Amsterdam: John Benjamins.
- Günthner, S. (2007). Zur Emergenz grammatischer Funktionen im Diskurs: wo-Konstruktionen in Alltagsinteraktionen. In H. Hausendorf (Ed.), *Studien zur deutschen Sprache: Vol. 37. Gespräch als Prozess. Linguistische Aspekte der Zeitlichkeit verbaler Interaktion* (pp. 125–155). Tübingen: Gunter Narr.
- Hanks, P. (2004). The Syntagmatics of Metaphor and Idiom. *International Journal of Lexicography*, 17(3), 245–274.
- Hanks, P., Urbach, A., & Gehweiler, E. (2006). German Light Verb Constructions in Corpora and Dictionaries. *Int J Lexicography*, 19(4), 439–457.
- Hoey, M. (2005). *Lexical priming: A new theory of words and language*. London: Routledge.
- Luder, M. (in press). Konstruktionen in der Erzählanalyse JAKOB: Sonderheft Erzählen, Träumen und Erinnern. *Erträge klinischer Erzählforschung. Psychoanalyse. Texte zur Sozialforschung*, 23(2), 226–237.
- Luder, M., Clematide, S., & Distl, B. (2008). Ein elektronisches Lexikon im OLIF-Format für die Erzählanalyse. In E. Bernal & J. DeCesaris (Eds.), *Proceedings of the XIII. Euralex International Congress* (pp. 729–735). Barcelona.
- Moon, R. (2008). Sinclair, Phraseology, and Lexicography. *International Journal of Lexicography*, 21(3), 243–254.
- Scholze-Stubenrecht, W., & Wermke, M. (2008). *Duden 11 - Redewendungen: Wörterbuch der deutschen Idiomatik* (3rd ed.). *Der Duden in zwölf Bänden: Vol. 3*. Mannheim: Dudenverlag.
- Schumacher, H., Kubczak, J., Schmidt, R., & Ruiter, V. de (2004). *VALBU - Valenzwörterbuch deutscher Verben*. Tübingen: Narr.
- Villavicencio, A., Bond, F., Korhonen, A., & McCarthy, D. (2005). Introduction to the special issue on multiword expressions: Having a crack at a hard nut: Special issue on Multiword Expression. *Computer Speech & Language*, 19(4), 365–377.
- Wahrig, G. (2007). *Der kleine Wahrig: Wörterbuch der deutschen Sprache*. München: Bertelsmann.