

**Tschichold, Cornelia**

***Multiword Units and Computational Lexicography***

*Wednesday 10:00, B19*

In this presentation I will attempt to show the challenges that multi- word lexemes pose for lexical databases. Such lexemes have been called "a pain in the neck" (Sag et al. 2002) for Natural Language 33

FLaRN 6

Processing, the processes which underlie computational linguistics. All multi-word lexemes that are not completely fixed, such as *by and large*, pose a problem for electronic dictionaries and similar databases and any applications reliant on them. Different types of multi-word lexemes pose different kinds of challenges, depending largely on the variability they exhibit in running text. For V+NP expressions of the type *to spill the beans*, the verb can be inflected in the usual way, there may be modification by an adverb (*He always spills the beans.*) and the whole expression can undergo passivization (*The beans were spilled at the meeting.*). Such modifications need to be coded into the database in order for the resulting NLP system to be able to recognize the expression in a text.

Reference

Sag, I. A., Baldwin, T., Bond, F., Copestake, A., & Flickinger, D. (2002). Multiword Expressions: A Pain in the Neck for NLP. In *Computational Linguistics and Intelligent Text Processing* (pp. 1- 15). Springer Berlin Heidelberg.