

Mehrworteinheiten

Einführung in das Seminar

Timm Lichte

HHU Düsseldorf

1. Sitzung, 10.10.2017



Kanonische Beispiele für Mehrworteinheiten

- (1) kingdom come ('Jenseits'), by and large
- (2) jemanden im Stich lassen
- (3) ins Gras beißen
- (4) spill the beans
- (5) ein Bad nehmen (Funktionalverbgefüge)
- (6) Gib's auf. (Verb-Partikel-Konstruktionen)
- (7) bis der Arzt kommt
Ich glaube, mich tritt ein Pferd.^[2]
- (8) John loaded the hay onto the wagon.
John loaded the wagon with the hay.
- (9) Telefonhäuschen, frische Luft
#Telefonhüttchen, #neue Luft (false friends)

Complex expressions which have surprising properties not predictable from their component words. (Baldwin & Kim 2010)

Was sind Mehrworteinheiten?

Complex expressions which have surprising properties not predictable from their component words. (Baldwin & Kim 2010)

Viele (aber möglicherweise nicht deckungsgleiche) Bezeichnungen: feste Redewendungen, Kollokationen, Idiome, Konstruktionen, tote Metaphern, Phraseologismen, ...

⇒ systematische Abrenzung?

Was sind “komplexe Ausdrücke”? Nur mit lexikaler Instanziierung?
Oder auch abstrakte, unlexikalisierte Konstruktionen?

Was sind “überraschende Eigenschaften”?

Voraussetzung: Es gibt gewisse “erwartbare” Eigenschaften (Semantik, Morphosyntax, Pragmatik, ...).

⇒ Verstoß gegen das Kompositionalitätsprinzip

⇒ Verstoß gegen das Flexibilitätsprinzip

Was sind **keine** Mehrworteinheiten?

Wenn die “Überraschung” nur von einem Wort hervorgerufen wird.

Unikale Elemente?

(siehe codii-Sammlung)

(10) jdm. einen **Bärendienst** erweisen

Was sind **keine** Mehrworteinheiten?

Wenn die “Überraschung” nur von einem Wort hervorgerufen wird.

Unikale Elemente? (siehe codii-Sammlung)

(10) jdm. einen **Bärendienst** erweisen

Monolexikale oder nicht-lexikalisierte Metonymien:

(11) Warum Frauen seit **Köln** verunsicherter denn je sind.

Was sind **keine** Mehrworteinheiten?

Wenn die “Überraschung” nur von einem Wort hervorgerufen wird.

Unikale Elemente? (siehe codii-Sammlung)

(10) jdm. einen **Bärendienst** erweisen

Monolexikale oder nicht-lexikalisierte Metonymien:

(11) Warum Frauen seit **Köln** verunsicherter denn je sind.

Monolexikale oder nicht-lexikalisierte Metaphern:

(12) Wir werden sie **jagen**. Wir werden Frau Merkel oder wen auch immer **jagen**. Und wir werden uns unser Land und unser Volk **zurückholen**.

Warum sind Mehrworteinheiten interessant?

“Pain the neck for NLP” (Sag et al. 2002)

- identification (types and tokens)
- lexicographic description
- grammar theory
- parsing
- machine translation

Aber: auch “reizvoll” für andere Felder (Semantik, Psycholinguistik, Philosophie, Kognitionswissenschaft)

Letztlich geht es um die Frage: Wie halten wir’s mit “Irregularitäten” bei der Komposition von Worten?

- 1 Verständnis der Empirie (Definition, Deskription)
- 2 Verständnis der Rolle in Theorie und Praxis
 - Lexikographie, Syntax und Semantik
 - **Computerlinguistik**
 - Psycholinguistik

⇒ holistische Perspektive mit CL-Schwerpunkt
- 3 Umgang mit Texten als Leser und Produzent
("Lektüreseminar")

- 1 Empirie und Begriffe
- 2 Grammatikmodelle (GG, HPSG, LFG, TAG)
- 3 Lexikographie (Formate)
- 4 Identifikation (qua Korpora, DS)
- 5 Parsing
- 6 Maschinelle Übersetzung

- Leistungsnachweis:
 - AP: Referat + Hausarbeit (benotet)
 - BN: Referat
- Zielgruppe: Fortgeschrittene (Sprachwissenschaft, Computerlinguistik)
- Website: http://timm-lichte.de/?page_id=378

- [1] Baldwin, Timothy & Su Nam Kim. 2010. Multiword expressions. In Nitin Indurkha & Fred J. Damera (eds.), *Handbook of natural language processing*, 2nd. CRC Press.
- [2] Richter, Frank & Manfred Sailer. 2009. Phraseological clauses in constructional HPSG. In Stefan Müller (ed.), *Proceedings of the 16th international conference on Head-Driven Phrase Structure Grammar, University of Göttingen, Germany*, 297–317. Stanford,CA: CSLI Publications.
- [3] Sag, Ivan A, Timothy Baldwin, Francis Bond, Ann Copestake & Dan Flickinger. 2002. Multiword expressions: A pain in the neck for NLP. In Alexander Gelbukh (ed.), *Computational linguistics and intelligent text processing* (Lecture Notes in Computer Science 2276), 1–15. Berlin: Springer.