

Building Lexicons out of a Database for Idioms

Ricarda Dormeyer and Ingrid Fischer

Speaker: Eleni Metheniti

08.12.2016

Introduction

- ▶ Keil, Martina (1997). Wort für Wort: Repräsentation und Verarbeitung verbaler Phraseologismen (Phraseo-Lex).
- ▶ Database for representation of German verbal idioms:
 - ▶ Computational dictionary for humans:
 - ▶ Lexica generator for NLP systems.
- ▶ Humans: Design formal representations of idioms
- ▶ Machines: Extract contents and generate phenomena
- ▶ Mapping: Keil, M.(1998). The lexical entry of phrasal idioms.
- ▶ Idioms: Diversity, arbitrariness, varied syntactic behavior

Paper Overview

- ▶ Idiom characteristics and features in *Phraseo-Lex*
- ▶ Idiom representation
- ▶ Mapping algorithm to convert entries to NLP framework

Syntax and Semantics of Verbal Idioms

- ▶ *Phraseo-Lex*: extensive description of German idioms with syntactic, semantic and pragmatic criteria

Lemma and Base Lexemes

- ▶ Lemma - Dictionary entry with base lexemes
- ▶ See ex.1 and ex.2, section 2.1

Syntactic Features

- ▶ Idiom's internal structure - phrase structure tree
- ▶ Valency: subject's position
- ▶ Syntactic transformations: possible/impossible/undecidable
- ▶ List of syntactic anomalies (e.g. "*sich bei jemandem lieb Kind machen*") and unique components

Semantic Features

- ▶ Three categories:
 - ▶ Fixed expressions
 - ▶ Free combinations
 - ▶ Partially compositional idioms
- ▶ Idiom definition: Literal, non-idiomatic paraphrase
- ▶ Syntactic transformations: possible/impossible/undecidable
- ▶ List of syntactic anomalies (e.g. "*sich bei jemandem lieb Kind machen*") and unique components

The Goal System

- ▶ A prototypical natural language processing system consisting of:
 - ▶ A chart parser
 - ▶ Syntactic component using a unification grammar formalism
 - ▶ a separate semantic component which combines Discourse Representation Theory with lambda-calculus
- ▶ Separate representation to each of the idiom's composing elements.
- ▶ Head of idiom: Verb
- ▶ Grammar rules to express relations of features
- ▶ *val2* feature in the verbal entry, head feature, *det sem* feature, noun feature

lamda-DRT

- ▶ Abstract over a discourse referent (predicative DRS) or over a complete DRS (partial DRS)
- ▶ Functional composition
- ▶ *see Fig. 3*
- ▶ Semantically meaningful (non-idioms, appear in context)
vs. meaningless (w/ empty semantic representation) idiom parts

The Mapping process

- ▶ *Phraseo-Lex* allows limited range of phenomena
- ▶ Generates lexical entries for all possible forms of lexeme
- ▶ **Idioms are built from the normal lexical material of a language**
- ▶ Unique components are assigned inflection by the determiner.
- ▶ Unfamiliar items are dealt as unique components
- ▶ Valency in verbs is encoded, can be changed
- ▶ Functions for internal/external valency

The Mapping process (cont.)

- ▶ Verb valency is represented as lambda-x, NP internal valency as lambda-e
- ▶ If valency = 0, *empty-det* DRS and *empty-noun* DRS assigned
- ▶ Noun-DRSs can be constructed too

Phraseo-Lex Screenshots

Syntax

Phraseo-Lex

Lemma: (jmd.) jmdm. einen Baeren aufbinden

syntaktischer Typ: VPL1

Phrasen

- VP
- NP_nenn
- NP_akk
- NP_dat
- PP_mit
- PP_in
- PP_auf
- PP_zur
- PP_ueber
- PP_bei
- PP_zu
- PP_vor

Valenz

Stabilität

Variabilität

Liste

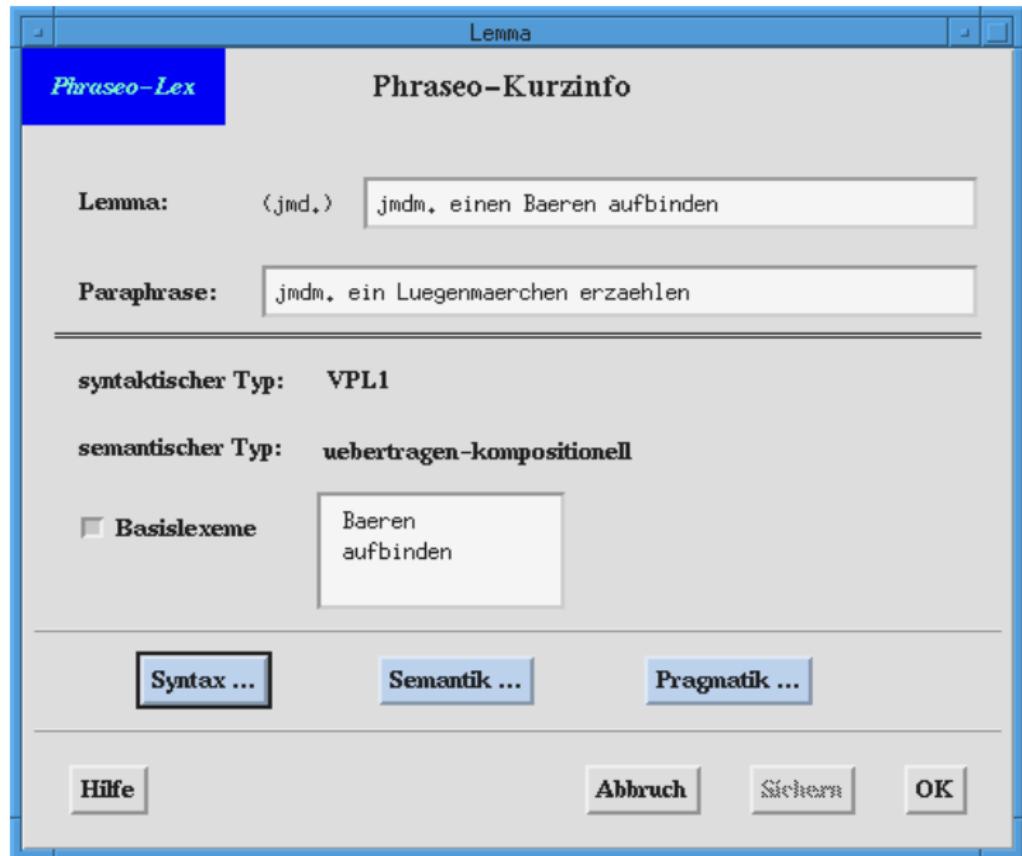
Hilfe

Abbruch

OK

```
graph TD; VPL1[VPL1  
[+idiom]] --> NP_dat[NP_dat]; VPL1 --> NP_akk[NP_akk  
[+mod, +idiom]]; VPL1 --> Verb[Verb  
[+inf, +idiom]]; NP_akk --> Artikel[Artikel  
[+idiom]]; NP_akk --> Nomen[Nomen  
[+idiom]]; Artikel --> ein[ein]; Nomen --> Baeren[Baeren]; Verb --> aufbinden[aufbinden]
```

Phraseo-Lex Screenshots



Phraseo-Lex Screenshots

Semantik

Phraseo-Lex

Semantik

Lemma: (jmd.) jmdm. einen Baeren aufbinden

Paraphrase: jmdm. ein Luegenmaerchen erzaehlen

semantischer Typ: uebertragen-kompositionell

Semantische Repräsentation

	Rollen	Referentenpotential
NP_nom: jmd.	-----> Agens	jmd.
NP_dat:	-----> Empfaenger	jmdm.
NP_akk: einen Baeren	-----> Patiens	ein Luegenmaerchen

Klassifikation

Wertigkeit

Modifikation

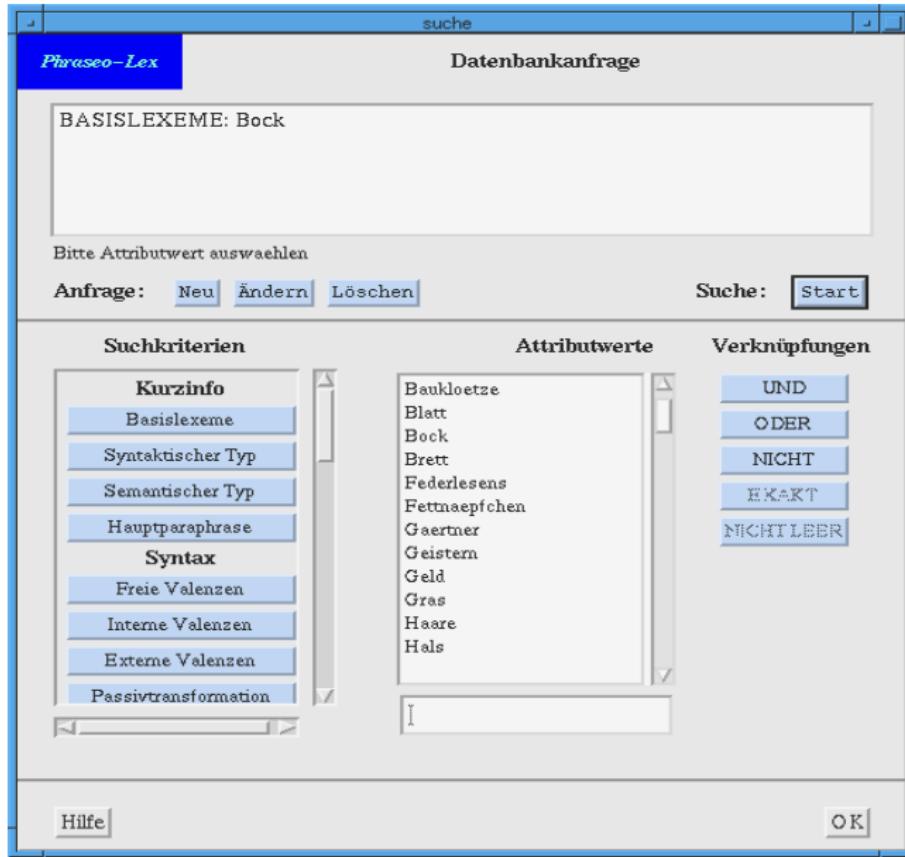
Paraphrasierungen

Synonyme & Antonyme

aufbinden(<_,_,_>) erzaehlen(X,Y,Z); Luegenmaerchen(Z)P

Hilfe Abbruch OK

Phraseo-Lex Screenshots



Bibliography

- ▶ Ricarda Dormeyer, Ingrid Fischer(1998). Building Lexicons out of a Database for Idioms. IMMDII, University of Erlangen
- ▶ Ricarda Dormeyer, Ingrid Fischer, Martina Keil (1998). A Database for Verbal Idioms. IMMDII, University of Erlangen
- ▶ Keil, Martina (1997). Wort fur Wort — Repräsentation und Verarbeitung verbaler Phraseologismen (Phraseo-Lex). Tübingen, Max Niemeyer Verlag.