

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 lamda-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

- ▶ 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 form 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 λx , NP internal valency as λe
- ▶ If valency = 0, *empty-det* DRS and *empty-noun* DRS assigned
- ▶ Noun-DRSs can be constructed too

Phraseo-Lex Screenshots

The screenshot shows the 'Syntax' window of the 'Phraseo-Lex' application. The window title is 'Syntax'. The 'Lemma' field contains '<jmd.> jmdm. einen Baeren aufbinden'. The 'syntaktischer Typ' is 'VPL1'.

On the left, a 'Phrasen' list contains various grammatical categories: VP, NP_acc, NP_akk, NP_dat, PP_mit, PP_in, PP_auf, PP_zu, PP_ueber, PP_bei, PP_zu, and PP_vor. The 'NP_akk' category is currently selected.

The central area displays a hierarchical syntax tree for the phrase 'jmdm. einen Baeren aufbinden'. The root node is 'VPL1' with feature '[+idiom]'. It branches into three nodes: 'NP_dat', 'NP_akk' (with feature '[+mod,+idiom]'), and 'Verb' (with feature '[+inf,+idiom]'). The 'NP_akk' node further branches into 'Artikel' (with feature '[+idiom]') and 'Nomen' (with feature '[+idiom]'). The 'Artikel' node leads to the word 'einen', and the 'Nomen' node leads to the word 'Baeren'. The 'Verb' node leads to the word 'aufbinden'.

On the right side, there are three buttons: 'Valenz', 'Stabilität', and 'Variabilität'. At the bottom right is a 'Liste' button.

At the bottom left is a 'Hilfe' button. At the bottom right are 'Abbruch' and 'OK' buttons.

Phraseo-Lex Screenshots

The screenshot shows a window titled 'Lemma' with a blue header bar. Inside, there is a blue tab labeled 'Phraseo-Lex'. The main content area is titled 'Phraseo-Kurzinfo'. It contains the following fields and controls:

- Lemma:** A text box containing '(jmd.) jmdm. einen Baeren aufbinden'.
- Paraphrase:** A text box containing 'jmdm. ein Luegenmaerchen erzaehlen'.
- syntaktischer Typ:** A label followed by the value 'VPL1'.
- semantischer Typ:** A label followed by the value 'uebertragen-kompositionell'.
- Basislexeme:** A checkbox (currently unchecked) followed by a text box containing 'Baeren aufbinden'.
- Navigation buttons:** Three buttons labeled 'Syntax ...', 'Semantik ...', and 'Pragmatik ...' are arranged horizontally.
- Footer buttons:** Four buttons labeled 'Hilfe', 'Abbruch', 'Sichern', and 'OK' are arranged horizontally at the bottom.

Phraseo-Lex Screenshots

Semantik

Phraseo-Lex

Semantik

Lemma: (jmd.) jmdm. einen Baeren aufbinden

Paraphrase: jmdm. ein Luegenmaerchen erzaehlen

semantischer Typ: uebertragen-kompositionell

Semantische Repraesentation

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

aufbinden(.,.,.)

erzaehlen(X,Y,Z); Luegenmaerchen(Z)P

Hilfe

Abbruch

OK

Klassifikation

Wertigkeit

Modifikation

Paraphrasierungen

Synonyme & Antonyme

Phraseo-Lex Screenshots

suche

Phraseo-Lex **Datenbankanfrage**

BASISLEXEME: Bock

Bitte Attributwert auswählen

Anfrage: Suche:

Suchkriterien	Attributwerte	Verknüpfungen
Kurzinfo Basislexeme Syntaktischer Typ Semantischer Typ Hauptparaphrase	Baukloetze Blatt Bock Brett Federlesens Fettnaepfchen Gaertner Geistern Geld Gras Haare Hals	UND ODER NICHT EXAKT NICHTLEER
Syntax Freie Valenzen Interne Valenzen Externe Valenzen Passivtransformation		

- ▶ 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.