

Die Sammlung unikalener Wörter des Deutschen

Aufbauprinzipien und erste Auswertungsergebnisse

Manfred Sailer und Beata Trawiński

SFB 441

Universität Tübingen

{mf, trawinsk}@sfs.uni-tuebingen.de.

Überblick

- Allgemeiner Hintergrund
- Struktur der Sammlung
- Kurzdemo
- Andere Sammlungen
- Erste Auswertungsergebnisse
- Ausblick





Allgemeiner Hintergrund

Distributionsidiosynkrasien (DI)

- Projekt A5 des SFB 441, Tübingen



Distributionsidiosynkrasien (DI)

- Projekt A5 des SFB 441, Tübingen
- Das Auftretensmuster eines Wortes kann nicht vollständig durch syntaktischen Kategorie, Bedeutung, Selektion und allgem. Prinzipien erfasst werden.



Distributionsidiosynkrasien (DI)

- Projekt A5 des SFB 441, Tübingen
- Das Auftretensmuster eines Wortes kann nicht vollständig durch syntaktischen Kategorie, Bedeutung, Selektion und allgem. Prinzipien erfasst werden.
- (Distributions-)Kontinuum:

frei	Pron	PE	unikal
<i>Buch</i>	<i>sich</i>	<i>beileibe</i>	<i>Tacheles</i>



Distributionsidiosynkrasien (DI)

- Projekt A5 des SFB 441, Tübingen
- Das Auftretensmuster eines Wortes kann nicht vollständig durch syntaktischen Kategorie, Bedeutung, Selektion und allgem. Prinzipien erfasst werden.

- (Distributions-)Kontinuum:

frei	Pron	PE	unikal
<i>Buch</i>	<i>sich</i>	<i>beileibe</i>	<i>Tacheles</i>

- Orthogonal zum Konstruktionskontinuum:

freie Konstr.	Phraseolschablone	voll lex. Konstr.
<i>Bücher lesen</i>	<i>Der und heiraten?!</i>	<i>frei Haus</i>
<i>wissen, wo Barthel den Most holt</i>		



DIs in der Linguistik

- Keine übergreifende Erfassung der Distributionsdaten.
- Keine systematische Analyse von DIs.



DIs in der Linguistik

- Gute Dokumentation der Unikalia in der phraseolog. Literatur.
- Keine Behandlung aus DI-Perspektive
- Außer van der Wouden: kein Zusammenhang zwischen Unikalia und Polaritätselementen



CoDII: Collection of Distributionally Idiosyncratic Items

- Systematische Sammlung von Elementen mit DI
- Angelegt für verschiedene Sprachen und verschiedene Typen von Elementen
- Vorhanden: SuWD
- weitere Teilsammlungen geplant (englische Unikalia, deutsche PEs)





Sammlung unikalener Wörter des Deutschen (SuWD/CoDII-BW.de)

Auswahl der Elemente

Alle potenziellen unikalen Wörter:

- ausgehend von Dobrovol'skij 1988
- andere phraseologische Veröffentlichungen
- okkasionelle Beobachtungen
(*jdm. etw. **madig** machen*)
- Momentan: ca. 450 Einträge



Block 1: Identifikation

- Unikales Element
- mit Wendung und Übersetzung
- Verweise auf Beispiele



Block 2: Klassifikation

- Klassifikationen (Dobrovol'skij 1988, 1989, Dobr.& Piirainen 1994, Nunberg u.a. 1994)
- ursprüngliche Angaben und evtl. Übertragung auf andere Elemente.
- weitere Klassifikationen möglich



Block 3: Syntax

- syntaktische Struktur
- syntaktische Flexibilität (Passiv, Pronominalisierung, ...)
- Verweise aus Beispiele



Block 4: Recherche

- Beispielqueries für öffentlich zugängliche Korpora (IDS, DWDS, google)
- Verweise auf Beispiele



Technische Realisierung

- XML-Datei
- Über XSL-Transformationen in XHTML-Datei umgewandelt
- Zugänglich auf:

<http://www.sfb441.uni-tuebingen.de/a5/codii/index.xhtml>



SuWD: Kurzdemonstration



Andere Sammlungen

Kollokationen im Wörterbuch

- beschränkt auf VPs
- betrachtet Idiome
- geht von Idiomsammlungen aus
- basiert auf eigenem Korpus (dwds)
- breites Zielpublikum



Usuelle Wortverbindungen

- geht von hochfrequenten Wörtern aus
- untersucht statistisch signifikante Kookkurrenzpartner
- Datenbasis wird rein quantitativ erhoben
- basiert auf eigenem Korpus (IDS)
- Ziel: Erfassung von typischen “Gebrauchsclustern”



Abgrenzung

- Kandidaten werden kompetenzorientiert ermittelt
- nicht beschränkt auf bestimmte Wortart oder Frequenz
- kein eigenes Korpus, sondern Verweise und Beispielqueries
- linguistisches Zielpublikum





Erste Auswertungsergebnisse

Hypothesen zu Unikalia

Soehn & Sailer 2003: Eine Wendung mit unikalem Element kann nicht die Form haben:

- **Modifikator Kopf**
(zu strikt: *der **große** Zampano*)
- [_{XP} **Komplement** [... Y ...] X]
(bei Idiomen: *die Katze aus dem Sack lassen*)



Hypothesen zu Unikalia

Ein unikales Element (**UE**) erfordert die Präsenz eines bestimmten Lexems (**LEX**), wobei gilt:

- UE selegiert LEX: *Angst einjagen*
- LEX selegiert UE: *Tacheles reden*
- LEX₁ selegiert LEX₂ und LEX₂ selegiert UE:
[zu₂ Potte] kommen₁
- aber nicht 3 Schritte:
[zu₃ UE] kommen₂ lassen₁
(bei Idiomen: *[fest [im Sattel]] sitzen*)





Ausblick: Datenbank und weitere Teilsammlungen

Technische Verfeinerung

Umwandlung in eine Datenbank

- dynamischer Zugriff
- verfeinerte Suchmöglichkeiten
- geplant für 2005



Inhaltlicher Ausbau

- Vervollständigung der bisherigen Einträge (bis Ende 2004)
- Integration von weiterer linguistischer Information (Abhängigkeitsstruktur, logische Form)
- CoDII-BW.en: englische Unikalia (2005)
- CoDII-PI.de: deutsche Polaritätselemente (2005–2008)

