

# Korpuslinguistische Verfahren zur Optimierung lexikalisch-semantischer Beschreibungen

Jörg Asmussen

Det Danske Sprog- og Litteraturselskab, DSL

14. März 2006

DSL

---

Überblick

Wörterbücher

Sprachtechnologie

Rolle des Korpus

---

Digitale Wörterbücher

---

Semantik im DDO

---

Genera proxima im DDO

---

Differentiae im DDO

---

Ausblick

---

# **Gesellschaft für dänische Sprache und Literatur (DSL): Wörterbücher, Korpora und mehr**

DSL

Überblick

Wörterbücher

Sprachtechnologie

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Ausblick

## Gesellschaft für dänische Sprache und Literatur:

- Außeruniversitäre Forschungseinrichtung, seit 1911
- Unterliegt heute dem dänischen Kulturministerium
- Teils öffentlich, teils durch private Stiftungen finanziert
- Ca. 45 Mitarbeiter

DSL

Überblick

Wörterbücher

Sprachtechnologie

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Ausblick

## Gesellschaft für dänische Sprache und Literatur:

- Außeruniversitäre Forschungseinrichtung, seit 1911
- Unterliegt heute dem dänischen Kulturministerium
- Teils öffentlich, teils durch private Stiftungen finanziert
- Ca. 45 Mitarbeiter

## Aufgaben:

- Texteditionen
- Bibliographien
- Sprachgeschichtliche Darstellungen
- Wörterbücher und Sprachtechnologie

DSL

Überblick

Wörterbücher

Sprachtechnologie

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Ausblick

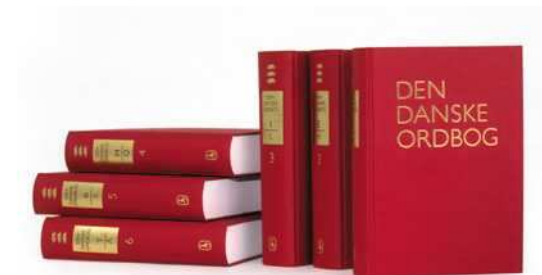
## Ordbog over det danske Sprog

neuzeitliches Dänisch, 1700–1950, 28+5 Bände, auch retrodigitalisiert im Web:  
[ods.ordnet.dk](http://ods.ordnet.dk)



## Den Danske Ordbog

Dänisch der Gegenwart, 1950–2000, 6 Bände, digitale Version in Vorbereitung



DSL

Überblick

Wörterbücher

Sprachtechnologie

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Ausblick

## Den Danske Ordbogs Korpus (DDOC):

- 40 Millionen Textwörter umfassendes Referenzkorpus
- 43.000 Texte einer Vielzahl von unterschiedlichen Textsorten aus dem Zeitraum 1983-1992
- Detaillierte bibliographische sowie kommunikativ und soziolinguistisch orientierte Annotation auf Textebene

## Korpus 2000:

- Im Web zugänglich gemachtes Referenzkorpus:  
[www.korpus2000.dk](http://www.korpus2000.dk)
- Jeweils 28 Millionen Textwörter aus den Zeiträumen 1983-1992 (Korpus 90) sowie 1998-2002 (Korpus 2000)
- Morphosyntaktisch annotiert

**Weitere Spezialkorpora** sowie diverse lexikalische Ressourcen

DSL

---

Rolle des Korpus

---

Korpus als Quelle

Lexikogr. Beschreibung

Weitere Möglichkeiten?

Digitale Wörterbücher

---

Semantik im DDO

---

Genera proxima im DDO

---

Differentiae im DDO

---

Ausblick

---

# Die Rolle des Korpus in der Lexikographie

# Das Korpus als Quelle in der Lexikographie

DSL

Rolle des Korpus

Korpus als Quelle

Lexikogr. Beschreibung

Weitere Möglichkeiten?

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Ausblick

**Prämisse:** ein Korpus ist eine Stichprobe des tatsächlichen Sprachgebrauchs

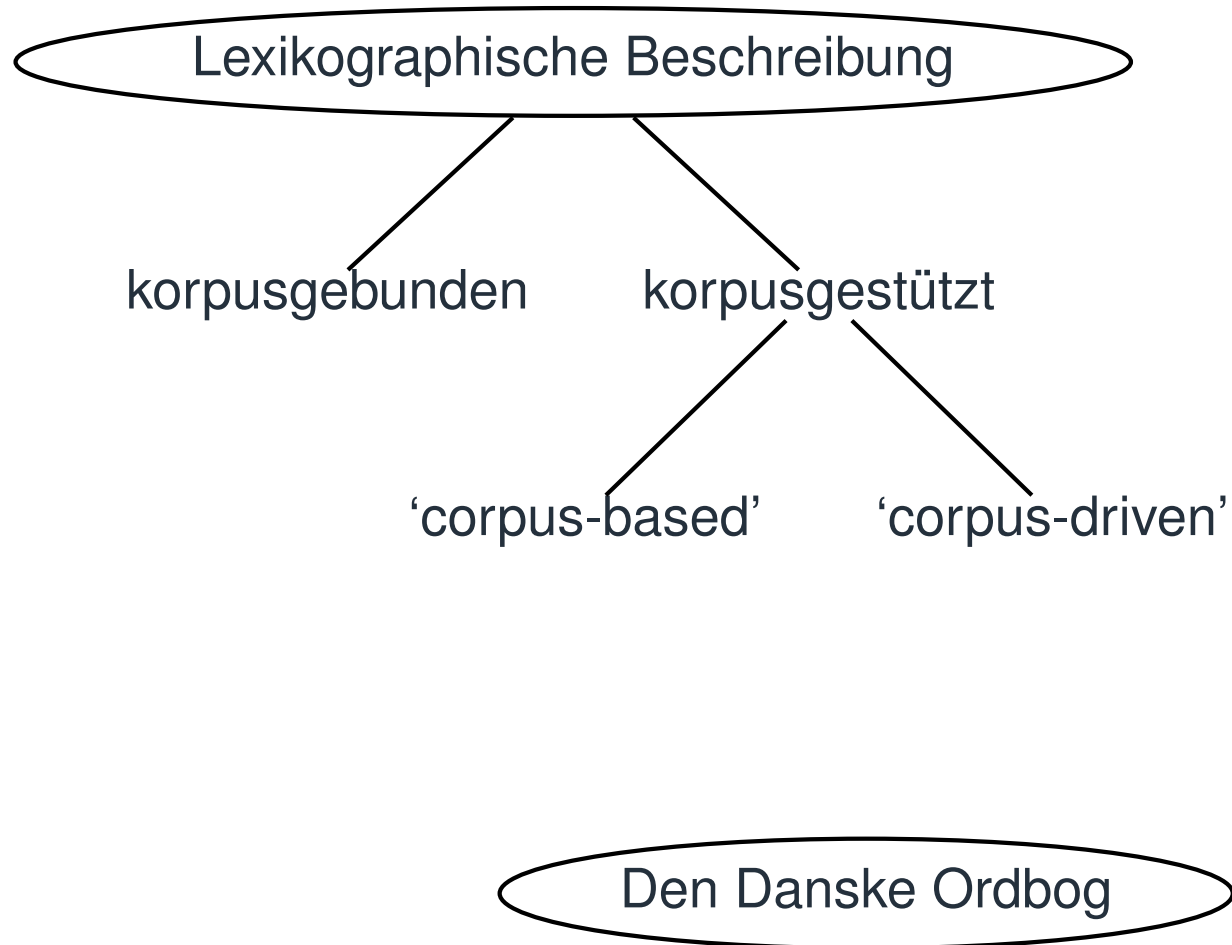
**Konsequenz:** es kann als Quelle für die lexikographische Beschreibung dienen

- Stichwortlisten
- Lesarten und ihre Häufigkeit
- Kollokationen
- Illustrative Textbelege
- Orthographie
- Flexion

**Beispiel:** Den Danske Ordbog



# Arten der lexikographischen Beschreibung



DSL

Rolle des Korpus

Korpus als Quelle

Lexikogr. Beschreibung

Weitere Möglichkeiten?

Digitale Wörterbücher

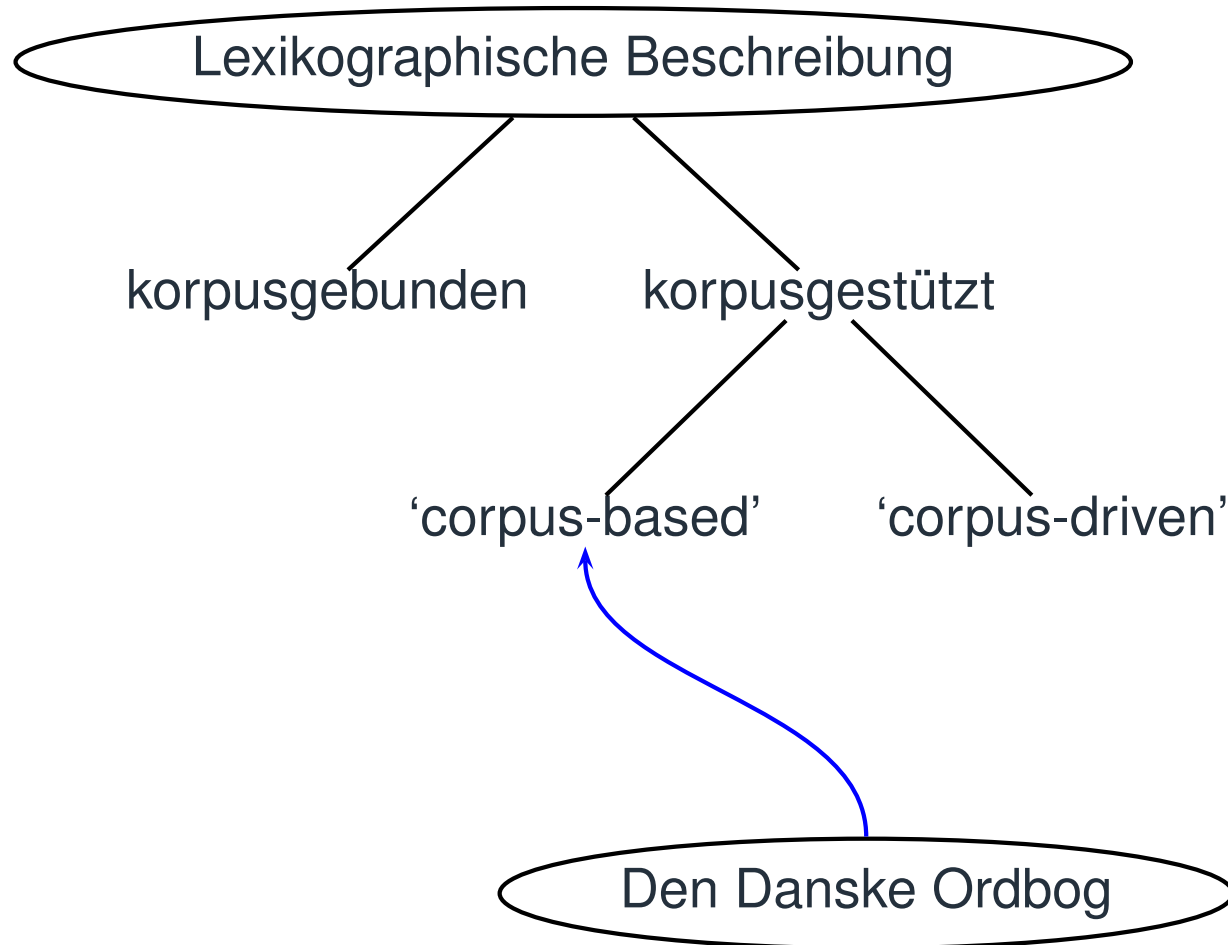
Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Ausblick

# Arten der lexikographischen Beschreibung



DSL

Rolle des Korpus

Korpus als Quelle

Lexikogr. Beschreibung

Weitere Möglichkeiten?

Digitale Wörterbücher

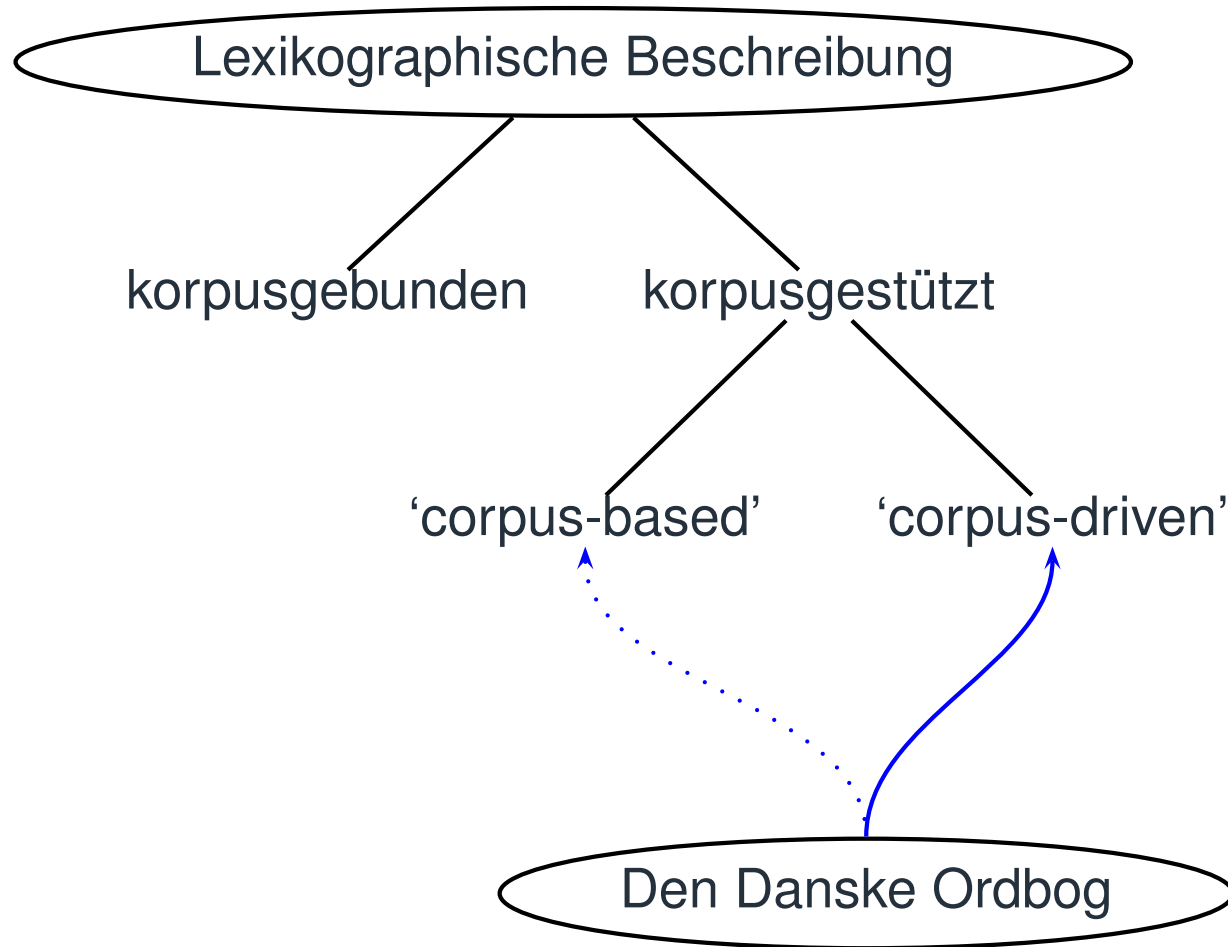
Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Ausblick

# Arten der lexikographischen Beschreibung



DSL

Rolle des Korpus

Korpus als Quelle

Lexikogr. Beschreibung

Weitere Möglichkeiten?

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Ausblick

# Weitere Einsatzmöglichkeiten von Korpora?

DSL

Rolle des Korpus

Korpus als Quelle

Lexikogr. Beschreibung

Weitere Möglichkeiten?

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Ausblick

**Quellkorpus als Grundlage** für die lexikalisch-semantische  
Beschreibung:

- Vokabular und Aufbau von Bedeutungserklärungen
- Hyperonymierelationen

# Weitere Einsatzmöglichkeiten von Korpora?

DSL

Rolle des Korpus

Korpus als Quelle

Lexikogr. Beschreibung

Weitere Möglichkeiten?

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Ausblick

**Quellkorpus als Grundlage** für die lexikalisch-semantische Beschreibung:

- Vokabular und Aufbau von Bedeutungserklärungen
- Hyperonymierelationen

**Korpuslinguistische Verfahren** angewandt auf das Wörterbuch:

- Optimierung von Hyperonymieangaben
- Optimierung von Differentia-Angaben
- Optimierung des Vokabulars für Bedeutungserklärungen
- Optimierung des Aufbaus von Bedeutungserklärungen

**Ziel solcher Optimierungen:** Eine generalisierte, multifunktionale lexikalisch-semantische Ressource

DSL

Rolle des Korpus

Digitale Wörterbücher

Hintergrund: *ordnet.dk*

Eigenschaften

Onomasiolog. Zugriff

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Ausblick

# Wörterbücher im digitalen Kontext

**Ziel des Projektes:** Wörterbücher und Korpora in einem web-basierten System zu vereinigen

DSL

Rolle des Korpus

Digitale Wörterbücher

Hintergrund: *ordnet.dk*

Eigenschaften

Onomasiolog. Zugriff

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Ausblick

DSL

Rolle des Korpus

Digitale Wörterbücher

Hintergrund: *ordnet.dk*

Eigenschaften

Onomasiolog. Zugriff

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Ausblick

**Ziel des Projektes:** Wörterbücher und Korpora in einem web-basierten System zu vereinigen

## Teilbereiche:

- Ordbog over det danske Sprog in einer Web-Version
- Weiterentwicklung des DDO zu einem digitalen Wörterbuch
- Unterprojekt: *DanNet* – ein dänisches Wordnet
- Weiterentwicklung und Integration der Korpus-Ressourcen



DSL

Rolle des Korpus

Digitale Wörterbücher

Hintergrund: *ordnet.dk*

Eigenschaften

Onomasiolog. Zugriff

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Ausblick

**Ziel des Projektes:** Wörterbücher und Korpora in einem web-basierten System zu vereinigen

## Teilbereiche:

- Ordbog over det danske Sprog in einer Web-Version
- Weiterentwicklung des DDO zu einem digitalen Wörterbuch
- Unterprojekt: *DanNet* – ein dänisches Wordnet
- Weiterentwicklung und Integration der Korpus-Ressourcen

## Entwicklungsarbeit:

- Konzept für digitale Wörterbücher: Generalisierung, Multifunktionalität
- Korpuslinguistische Methoden

# Prinzipielle Eigenschaften digitaler „Wörterbücher“

DSL

Rolle des Korpus

Digitale Wörterbücher

Hintergrund: *ordnet.dk*

Eigenschaften

Onomasiolog. Zugriff

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Ausblick

**Prämisse:** Ein „elektronisches Wörterbuch“ ist kein Buch, sondern eine strukturierte Sammlung lexikalischer Daten

**Struktur:** Strukturelle Bindungen durch das Buchmedium werden aufgehoben:

**Platz:** mehr Stichwörter, mehr Beispiele

**Organisation:** Entkoppelung von Artikelstruktur und Präsentation

**Funktion:** Funktionale Bindungen durch das Buchmedium werden aufgehoben

**Multifunktionalität:** „herkömmliches“ Wörterbuch, sprachtechnologische Ressource

**Zugriff:** onomasiologischer Zugriff wird durch ‘Umkehrung’ des semasiologischen ermöglicht

DSL

Rolle des Korpus

Digitale Wörterbücher

Hintergrund: *ordnet.dk*

Eigenschaften

Onomasiolog. Zugriff

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Ausblick

## Anforderungen an die semantische Beschreibung

- Konsistente Hyperonymieangaben
- Festes Inventar von übrigen Angaben
- Feste Angaben-Abfolge
- Vererbung von Merkmalen

DSL

Rolle des Korpus

Digitale Wörterbücher

Hintergrund: *ordnet.dk*

Eigenschaften

Onomasiolog. Zugriff

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Ausblick

## Anforderungen an die semantische Beschreibung

- Konsistente Hyperonymieangaben
- Festes Inventar von übrigen Angaben
- Feste Angaben-Abfolge
- Vererbung von Merkmalen

## Fragen

1. Welche Anpassungen sind notwendig, damit die bestehenden Bedeutungsparaphrasen im DDO einen onomasiologischen Zugriff erlauben?
2. Welche Hilfsmittel ermöglichen einen Überblick über den Zustand der Bedeutungsparaphrasen im DDO?

DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Zum Begriff

Verfahren

Beispiele

Korpusling. Optimierg.

Genera proxima im DDO

Differentiae im DDO

Ausblick

# Semantische Beschreibung im DDO

## Semantische Beschreibung (= Bedeutungsbeschreibung) Bedeutungserklärung: Angabe von Bedeutungsparaphrasen (und von Synonymen)

```
<☐ Semem >  
  <☐ Restspec >  
    <☐ Sysfag > zoo  
  <☐ Denbet > lille gnaver med tilspidset snude, forholdsvis store  
ører og en ofte lang, nøgen hale  
    <☐ Genprox > gnaver  
  <☐ Onym >  
    <☐ Syn >  
      <☐ KomOnym > familien  
      <☐ txt > Muridae  
  <☐ Encykl > omfatter bl. a. husmus, skovmus, rotter og dværgmus  
  <☐ Rel >  
    <☐ Adled >  
      <☐ txt > lille ~  
  <☐ Rel >  
    <☐ Typsam >  
      <☐ txt > ~ og rotter  
  <☐ Typsam >  
    <☐ txt > forsøg med ~
```

DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Zum Begriff

Verfahren

Beispiele

Korpusling. Optimierg.

Genera proxima im DDO

Differentiae im DDO

Ausblick

## Genus proximum:

- Kein prädefiniertes ontologisches System
- Nicht korpusgestützt
- Pragmatisch orientierter Zugang, sprich: Introspektion
- Explizite Angabe des sprachlichen Zeichens, welches das genus proximum bezeichnet (*Genprox*-Element)

## Differentia specifica:

- Kein vorgegebenes semantisches Beschreibungsmodell (Merkmale, Relationen)
- Kein vorgegebenes Vokabular
- Keine vorgegebene Syntax

# Bedeutungsparaphrasen – Beispiele

DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Zum Begriff

Verfahren

Beispiele

Korpusling. Optimierg.

Genera proxima im DDO

Differentiae im DDO

Ausblick

1. **grammofon**  $\approx$  **apparat** til at **afspille grammofonplader** med  
wörtlich: „Apparat um zu abspielen Schallplatten mit“  
▷ Gerät zum Abspielen von Schallplatten  
DUW: Plattenspieler  $\approx$  Gerät zum Abspielen von Schallplatten
2. **cd-afspiller**  $\approx$  **apparat** til at **afspille cd'er** med  
wörtlich: „Apparat um zu abspielen CDs mit“  
▷ Gerät zum Abspielen von CDs  
DUW: CD-Player  $\approx$  Abspielgerät für Compact Discs

**Genus proximum**

**Differentia specifica (als Angabe des Verwendungszwecks)**



# Bedeutungsparaphrasen – Beispiele

DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Zum Begriff

Verfahren

Beispiele

Korpusling. Optimierg.

Genera proxima im DDO

Differentiae im DDO

Ausblick

1. **grammofon**  $\approx$  **apparat** til at **afspille grammofonplader** med  
wörtlich: „Apparat um zu abspielen Schallplatten mit“  
▷ Gerät zum Abspielen von Schallplatten  
DUW: Plattenspieler  $\approx$  **Gerät** zum Abspielen von Schallplatten
2. **cd-afspiller**  $\approx$  **apparat** til at **afspille cd'er** med  
wörtlich: „Apparat um zu abspielen CDs mit“  
▷ Gerät zum Abspielen von CDs  
DUW: CD-Player  $\approx$  **Abspielgerät** für Compact Discs

**Genus proximum**

**Differentia specifica (als Angabe des Verwendungszwecks)**

**Hyperonymie-Inkonsistenz im DUW**

# Korpuslinguistische Optimierung

DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Zum Begriff

Verfahren

Beispiele

Korpusling. Optimierg.

Genera proxima im DDO

Differentiae im DDO

Ausblick

**Frequenzprofile der GP-Bezeichner :** Generalisierung niederfrequenter GP-Bezeichner  $\Rightarrow$  drastische Reduktion der Anzahl der GP-Bezeichner

**Ermittlung einander ähnlicher BP:**

Aufdecken hierarchischer Inkonsistenzen

**Frequenzprofile des BP-Vokabulars:** Generalisierung niederfrequenter Wortformen  $\Rightarrow$  Reduktion (Vereinfachung?) des BP-Vokabulars

**Vergleiche von  $n$ -Grammen ( $G^2$ ):** Aufdecken signifikanter syntagmatischer Muster in den BP

**KWIC-Konkordanzen:** Auffinden syntagmatischer Muster und kontextueller Verwendung bestimmter Wortformen in den BP

**„Feature Detection“ (Mutual Information):** Aufdecken augenfälliger Merkmalsbezeichner in den BP

# Korpuslinguistische Optimierung

DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Zum Begriff

Verfahren

Beispiele

Korpusling. Optimierg.

Genera proxima im DDO

Differentiae im DDO

Ausblick

**Frequenzprofile der GP-Bezeichner:** Generalisierung niederfrequenter GP-Bezeichner  $\Rightarrow$  drastische Reduktion der Anzahl der GP-Bezeichner

**Ermittlung einander ähnlicher BP:**

**Aufdecken hierarchischer Inkonsistenzen**

**Frequenzprofile des BP-Vokabulars:** Generalisierung niederfrequenter Wortformen  $\Rightarrow$  Reduktion (Vereinfachung?) des BP-Vokabulars

**Vergleiche von  $n$ -Grammen ( $G^2$ ):** Aufdecken signifikanter syntagmatischer Muster in den BP

**KWIC-Konkordanzen:** Auffinden syntagmatischer Muster und kontextueller Verwendung bestimmter Wortformen in den BP

**„Feature Detection“ (Mutual Information):** Aufdecken augenfälliger Merkmalsbezeichner in den BP

DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Hierarchie-Zuordnung

Hierarchie-Korrekturen

Niederfrequente GP

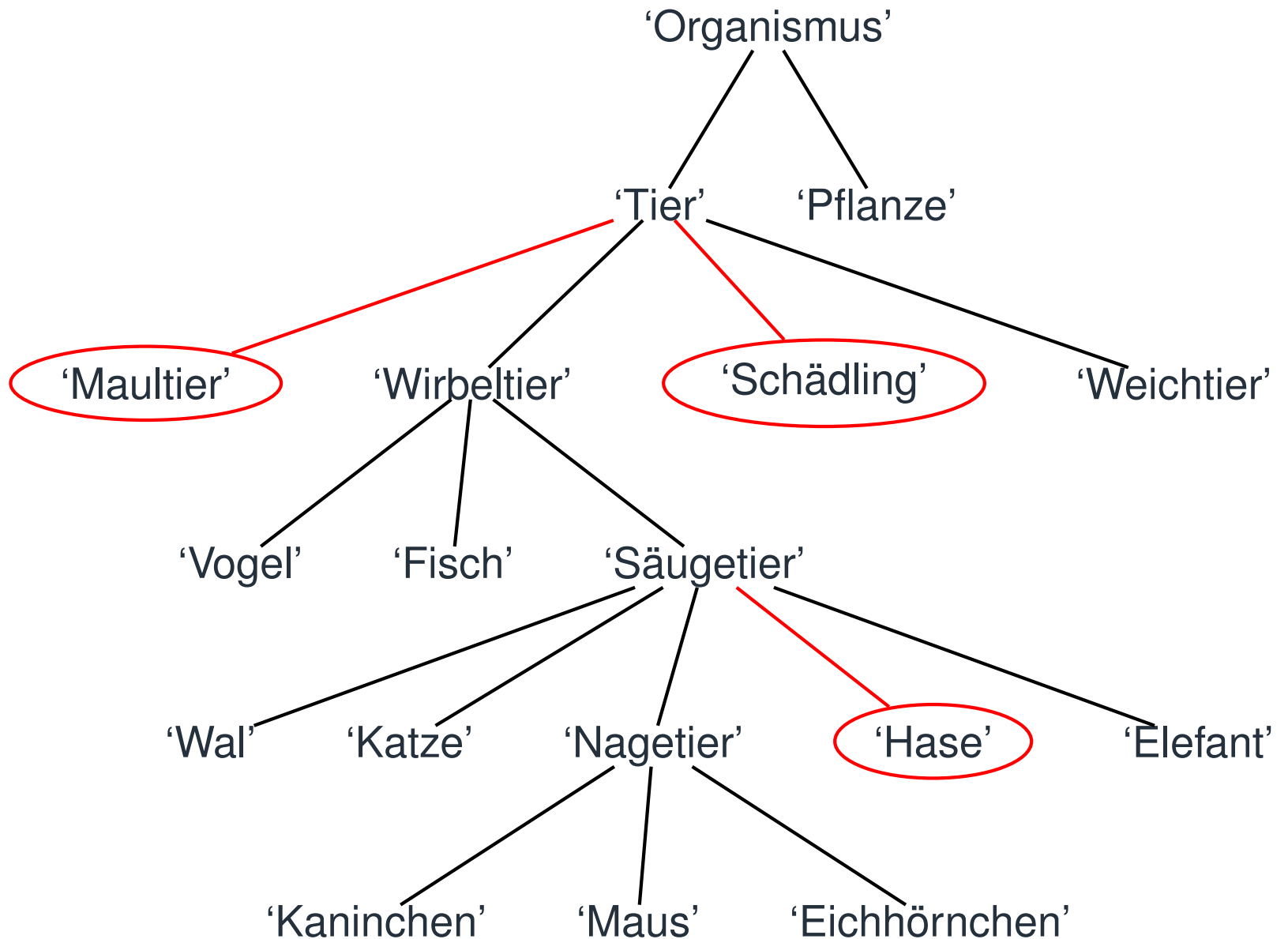
Differentiae im DDO

Ausblick

# Genera proxima im DDO und ihre mögliche Optimierung

1. **kanin**  $\approx$  **gnaver** der ligner en lille hare, men har forholdsvis kortere ører og bagben og graver gange i jorden  
▷ **Kaninchen**  $\approx$  **Nagetier**, das einem kleinen Hasen ähnelt, aber verhältnismäßig kürzere Ohren und Hinterbeine hat und Gänge in der Erde gräbt
2. **hare**  $\approx$  gråbrunt **pattedyr** med hvid bug, to store fortænder i over- og undermunden, lange ører og kraftige bagben  
▷ **Hase**  $\approx$  graubraunes **Säugetier** mit weißem Bauchfell, zwei großen Schneidezähnen im Ober- und Unterkiefer, langen Ohren und kräftigen Hinterbeinen
3. **gnaver**  $\approx$  **pattedyr** med to store fortænder i over- og undermunden der er særligt egnede til at gnave med  
▷ **Nagetier**  $\approx$  **Säugetier** mit zwei großen Schneidezähnen im Ober- und Unterkiefer, die besonders zum Nagen geeignet sind

# Hierarchie-Korrekturen



DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Hierarchie-Zuordnung

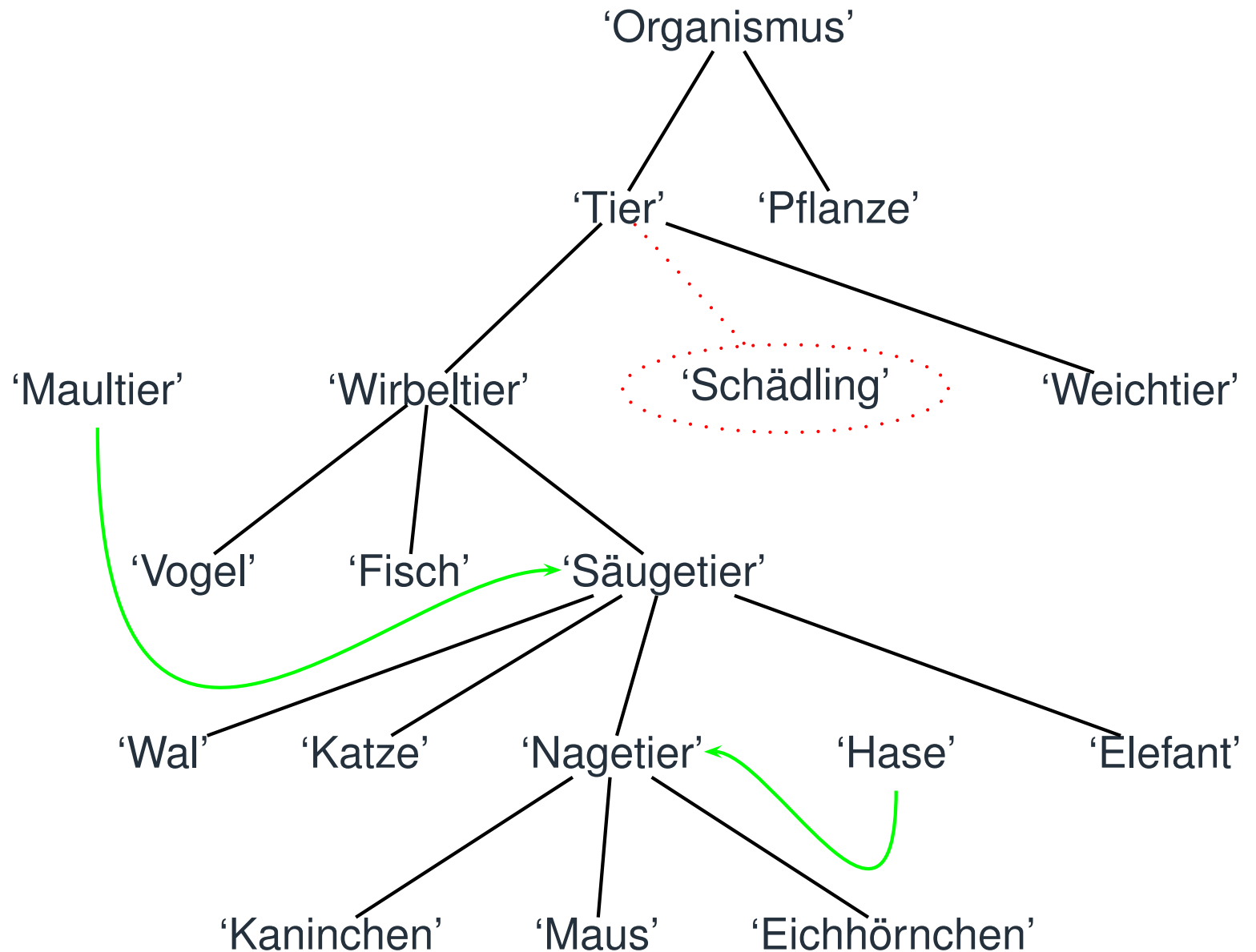
Hierarchie-Korrekturen

Niederfrequente GP

Differentiae im DDO

Ausblick

# Hierarchie-Korrekturen



- DSL
- Rolle des Korpus
- Digitale Wörterbücher
- Semantik im DDO
- Genera proxima im DDO
- Hierarchie-Zuordnung
- Hierarchie-Korrekturen
- Niederfrequente GP
- Differentiae im DDO
- Ausblick

# Niederfrequente genera proxima

Etwa die Hälfte aller genera proxima wird nur einmal verwendet  
6776 von 13748 (49%) – Beispiele (nicht im DDOC):

GP	Dt.
pergament <b>håndskrift</b>	Pergament <b>handschrift</b>
jernt <b>ang</b>	Eisen <b>zange</b>
teater <b>figur</b>	Theater <b>figur</b>
steger <b>redskab</b>	Brat <b>werkzeug</b>
mund <b>del</b>	Mund <b>teil</b>
konkurrence <b>princip</b>	Konkurrenz <b>prinzip</b>
minde <b>bygning</b>	Denkmal <b>gebäude</b>
farve <b>opløsning</b>	Farb <b>lösung</b>
kviksølv <b>forbindelse</b>	Quecksilber <b>verbindung</b>
symaskine <b>del</b>	Nähmaschine <b>teil</b>
lakrids <b>kugle</b>	Lakritz <b>kugel</b>
okse <b>fedt</b>	Rinder <b>fett</b>

DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Hierarchie-Zuordnung

Hierarchie-Korrekturen

Niederfrequente GP

Differentiae im DDO

Ausblick



DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Manuelle Ermittlung

Häufige GP

Statistische Ermittlung

Erste Ergebnisse

Weitere Ergebnisse

Vereinheitlichung

Weitere Anwendungen

Ausblick

# Differentia-Angaben im DDO und ihre mögliche Optimierung

# Manuelle Ermittlung typischer Differentia-Angaben

DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Manuelle Ermittlung

Häufige GP

Statistische Ermittlung

Erste Ergebnisse

Weitere Ergebnisse

Vereinheitlichung

Weitere Anwendungen

Ausblick

## Hyponyme zu 'Nagetier'

rotte

markmus

studsmus

mus

marsvin

murmeldyr

egern

bæverrotte

bæver

chinchilla

kanin

mosegris

hamster

sumpbæver

rødmus

lemming

hulepindsvin

hasselmus

syvsover

ørkenrotte

præriehund

jordegern

# Manuelle Ermittlung typischer Differentia-Angaben

DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Manuelle Ermittlung

Häufige GP

Statistische Ermittlung

Erste Ergebnisse

Weitere Ergebnisse

Vereinheitlichung

Weitere Anwendungen

Ausblick

## Hyponyme zu 'Nagetier'

rotte

markmus

studsmus

mus

marsvin

murmeldyr

egern

bæverrotte

bæver

chinchilla

kanin

mosegris

hamster

sumpbæver

rødmus

lemming

hulepindsvin

hasselmus

syvsover

ørkenrotte

præriehund

jordegern

## Merkmale in den Bedeutungsparaphrasen

'klein' oder 'mausähnlich'

'Haustier'

'nicht-endemisch'

'schwanzlos'

# Manuelle Ermittlung typischer Differentia-Angaben

- DSL
- Rolle des Korpus
- Digitale Wörterbücher
- Semantik im DDO
- Genera proxima im DDO
- Differentiae im DDO
- Manuelle Ermittlung
- Häufige GP
- Statistische Ermittlung
- Erste Ergebnisse
- Weitere Ergebnisse
- Vereinheitlichung
- Weitere Anwendungen
- Ausblick

## Hyponyme zu 'Nagetier'

rotte  
markmus  
studsmus  
mus  
marsvin  
murmeldyr  
egern  
bæverrotte  
bæver  
chinchilla  
kanin  
mosegris  
hamster  
sumpbæver  
rødmus  
lemming  
hulepindsvin  
hasselmus  
syvsover  
ørkenrotte  
præriehund  
jordegern

## Merkmale in den Bedeutungsparaphrasen

'klein' oder 'mausähnlich'

'Haustier'

'nicht-endemisch'

'schwanzlos'

# Manuelle Ermittlung typischer Differentia-Angaben

- DSL
- Rolle des Korpus
- Digitale Wörterbücher
- Semantik im DDO
- Genera proxima im DDO
- Differentiae im DDO
- Manuelle Ermittlung
- Häufige GP
- Statistische Ermittlung
- Erste Ergebnisse
- Weitere Ergebnisse
- Vereinheitlichung
- Weitere Anwendungen
- Ausblick

## Hyponyme zu 'Nagetier'

- rotte
- markmus
- studsmus
- mus
- marsvin
- murmeldyr
- egern
- bæverrotte
- bæver
- chinchilla
- kanin
- mosegris
- hamster
- sumpbæver
- rødmus
- lemming
- hulepindsvin
- hasselmus
- syvsover
- ørkenrotte
- præriehund
- jordegern

## Merkmale in den Bedeutungsparaphrasen

'klein' oder 'mausähnlich'

'Haustier'

'nicht-endemisch'

'schwanzlos'

# Manuelle Ermittlung typischer Differentia-Angaben

- DSL
- Rolle des Korpus
- Digitale Wörterbücher
- Semantik im DDO
- Genera proxima im DDO
- Differentiae im DDO
- Manuelle Ermittlung
- Häufige GP
- Statistische Ermittlung
- Erste Ergebnisse
- Weitere Ergebnisse
- Vereinheitlichung
- Weitere Anwendungen
- Ausblick

## Hyponyme zu 'Nagetier'

- rotte
- markmus
- studsmus
- mus
- marsvin
- murmeldyr
- egern
- bæverrotte
- bæver
- chinchilla
- kanin
- mosegris
- hamster
- sumpbæver
- rødmus
- lemming
- hulepindsvin
- hasselmus
- syvsover
- ørkenrotte
- præriehund
- jordegern

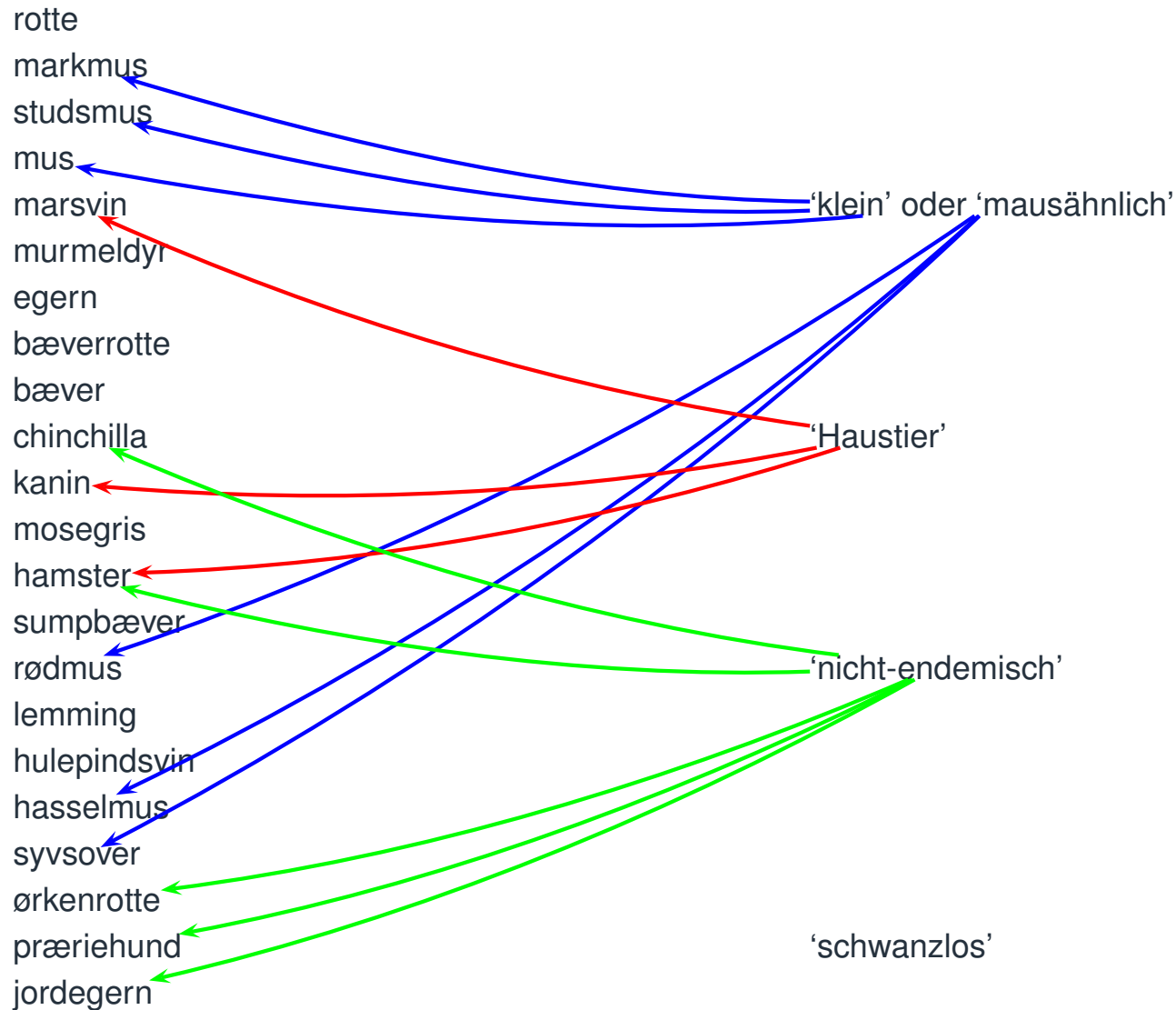
## Merkmale in den Bedeutungsparaphrasen

'klein' oder 'mausähnlich'

'Haustier'

'nicht-endemisch'

'schwanzlos'



# Manuelle Ermittlung typischer Differentia-Angaben

- DSL

---

- Rolle des Korpus

---

- Digitale Wörterbücher

---

- Semantik im DDO

---

- Genera proxima im DDO

---

- Differentiae im DDO

---

- Manuelle Ermittlung
- Häufige GP
- Statistische Ermittlung
- Erste Ergebnisse
- Weitere Ergebnisse
- Vereinheitlichung
- Weitere Anwendungen

---

- Ausblick

## Hyponyme zu 'Nagetier'

rotte  
 markmus  
 studsmus  
 mus  
 marsvin  
 murmeldyr  
 eger  
 bæverrotte  
 bæver  
 chinchilla  
 kanin  
 mosegris  
 hamster  
 sumpbæver  
 rødmus  
 lemming  
 hulepindsvin  
 hasselmus  
 syvsover  
 ørkenrotte  
 præriehund  
 jordeger

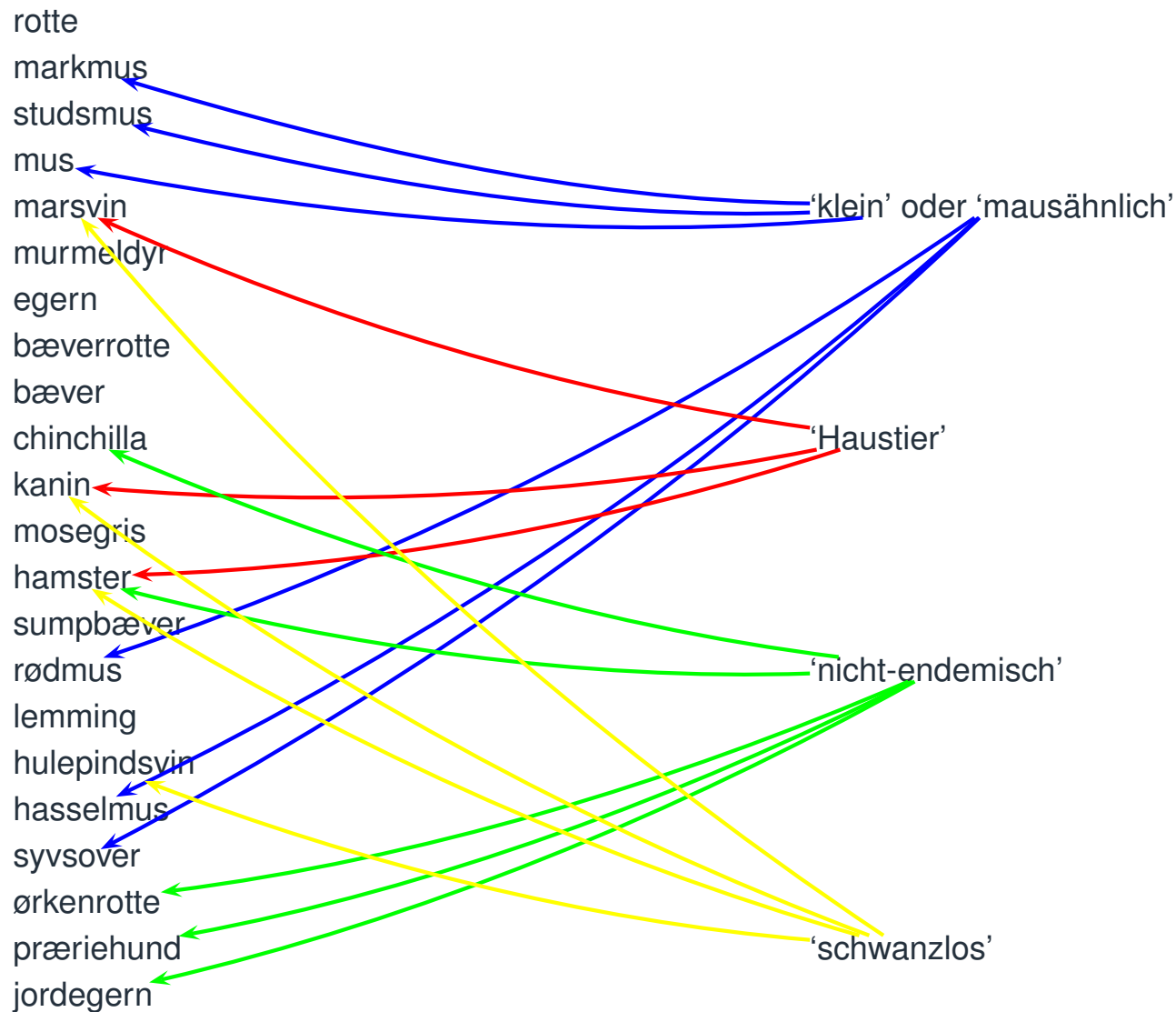
## Merkmale in den Bedeutungsparaphrasen

'klein' oder 'mausähnlich'

'Haustier'

'nicht-endemisch'

'schwanzlos'



# Häufige genera proxima

Genus proximum	Dt. Entsprechung	Häufigkeit
person	Person	4382
del	Teil	799
sted	Ort/Stelle	651
område	Gebiet	422
gruppe	Gruppe	421
plante	Pflanze	389
måde	Art/Weise	345
mængde	Menge	326
stof	Stoff	325
kvinde	Frau	318
farve	Farbe	314
genstand	Gegenstand	291

DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Manuelle Ermittlung

Häufige GP

Statistische Ermittlung

Erste Ergebnisse

Weitere Ergebnisse

Vereinheitlichung

Weitere Anwendungen

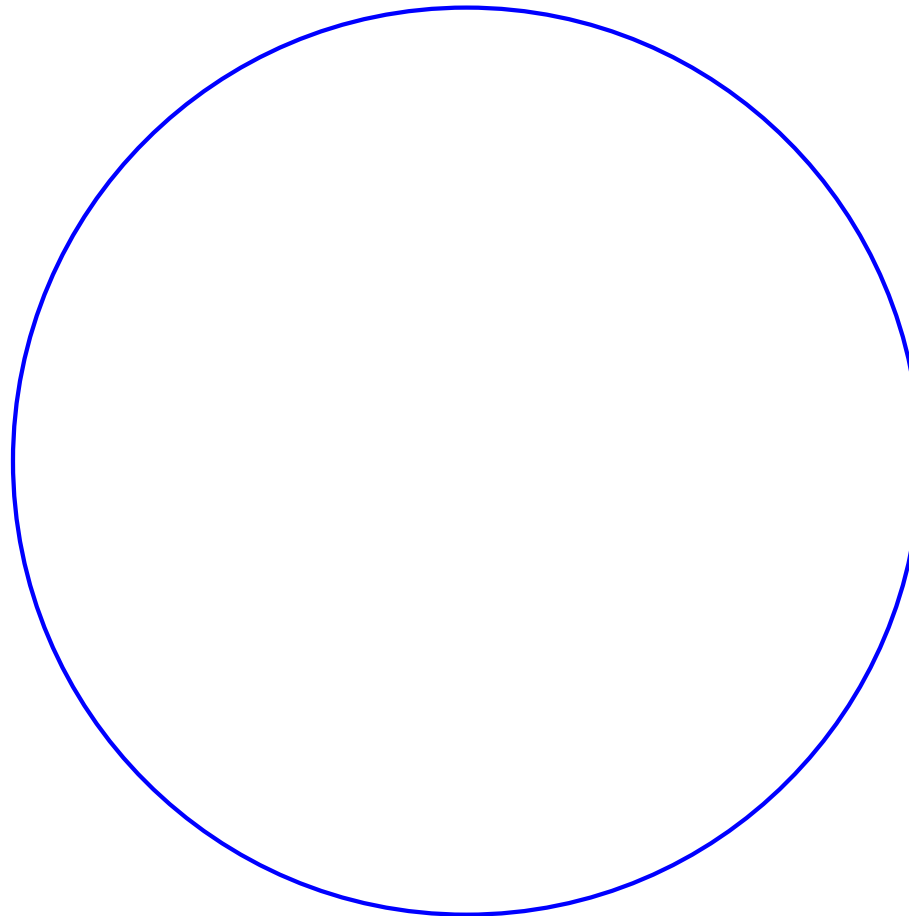
Ausblick



# Statistische Ermittlung von Differentia-Angaben

## Methode: „Feature Detection“ (Prinzip: Mutual Information)

Korpus aus sämtlichen DDO-BP



DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Manuelle Ermittlung

Häufige GP

Statistische Ermittlung

Erste Ergebnisse

Weitere Ergebnisse

Vereinheitlichung

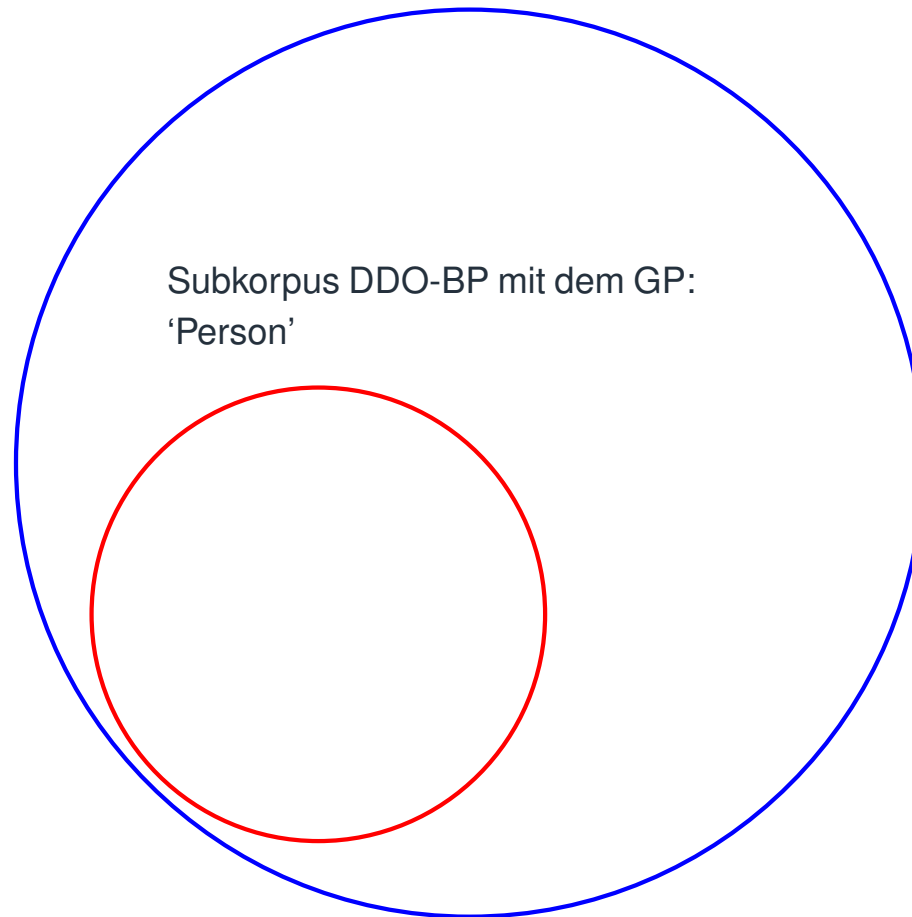
Weitere Anwendungen

Ausblick

# Statistische Ermittlung von Differentia-Angaben

## Methode: „Feature Detection“ (Prinzip: Mutual Information)

Korpus aus sämtlichen DDO-BP



DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Manuelle Ermittlung

Häufige GP

Statistische Ermittlung

Erste Ergebnisse

Weitere Ergebnisse

Vereinheitlichung

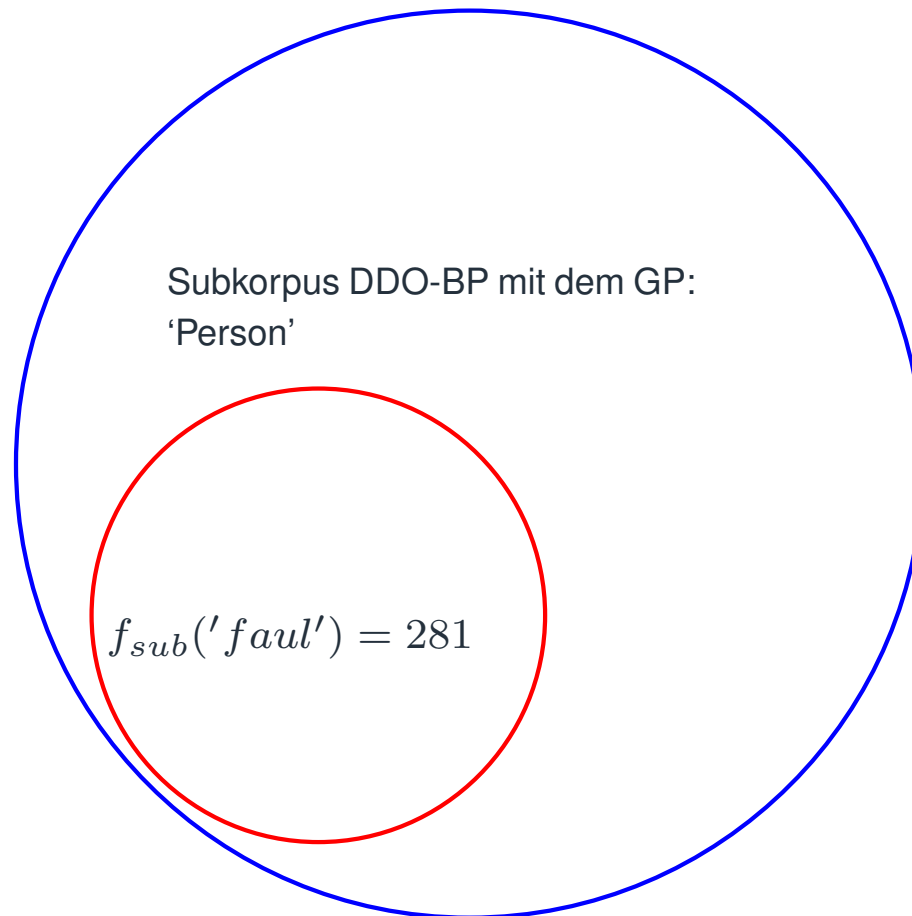
Weitere Anwendungen

Ausblick

# Statistische Ermittlung von Differentia-Angaben

## Methode: „Feature Detection“ (Prinzip: Mutual Information)

Korpus aus sämtlichen DDO-BP



DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Manuelle Ermittlung

Häufige GP

Statistische Ermittlung

Erste Ergebnisse

Weitere Ergebnisse

Vereinheitlichung

Weitere Anwendungen

Ausblick

# Statistische Ermittlung von Differentia-Angaben

## Methode: „Feature Detection” (Prinzip: Mutual Information)

Korpus aus sämtlichen DDO-BP

$$f_{total}('faul') = 19$$

Subkorpus DDO-BP mit dem GP:  
'Person'

$$f_{sub}('faul') = 281$$

DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Manuelle Ermittlung

Häufige GP

Statistische Ermittlung

Erste Ergebnisse

Weitere Ergebnisse

Vereinheitlichung

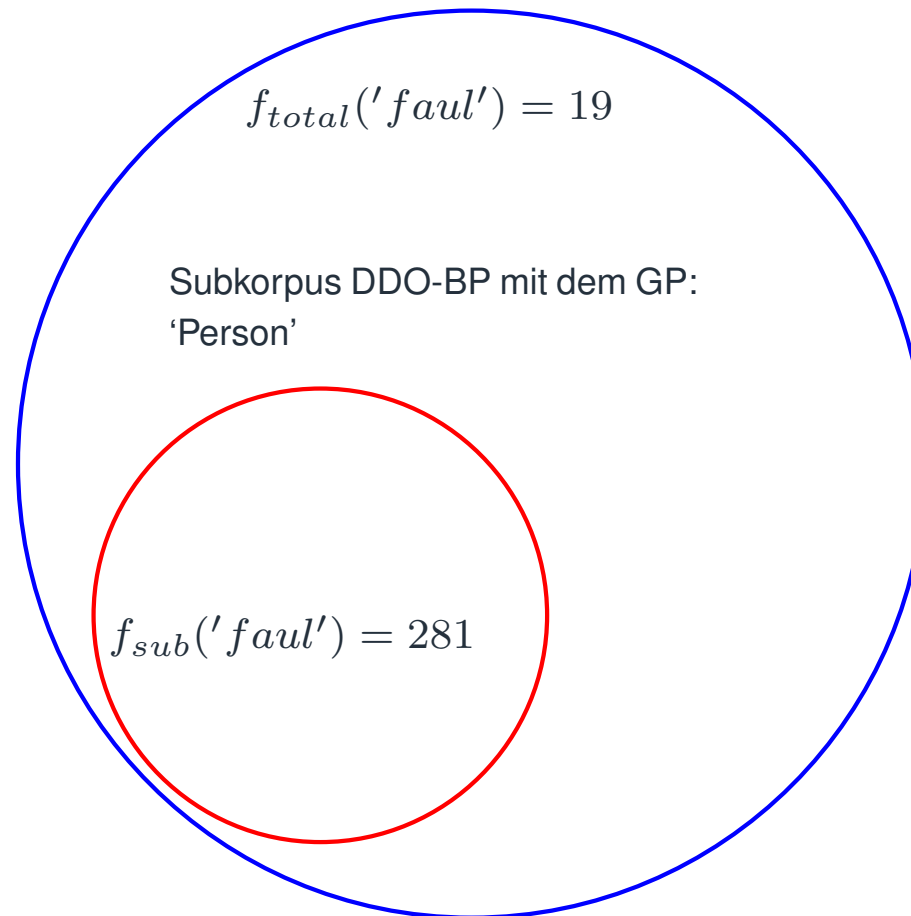
Weitere Anwendungen

Ausblick

# Statistische Ermittlung von Differentia-Angaben

## Methode: „Feature Detection“ (Prinzip: Mutual Information)

Korpus aus sämtlichen DDO-BP



$$\begin{aligned} \text{Score} &= \\ \frac{f_{sub}('faul')}{f_{total}('faul')} &= \frac{281}{19} \\ &\approx 15 \end{aligned}$$

DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Manuelle Ermittlung

Häufige GP

Statistische Ermittlung

Erste Ergebnisse

Weitere Ergebnisse

Vereinheitlichung

Weitere Anwendungen

Ausblick

DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Manuelle Ermittlung

Häufige GP

Statistische Ermittlung

Erste Ergebnisse

Weitere Ergebnisse

Vereinheitlichung

Weitere Anwendungen

Ausblick

## Merkmale mit den höchsten Scores (GP 'gnaver', 'Nagetier')

- bæverlignende 'bieberähnlich'
- fodsåler 'Fußsohlen'
- hårbeklædt 'haarbedeckt'
- nataktiv 'nachtaktiv'
- nutria 'Nutria'
- prærieområder 'Präriegebiete'
- Sibiriens 'Sibiriens'
- tundra 'Tundra'
- ungskov 'Jungwald'

DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Manuelle Ermittlung

Häufige GP

Statistische Ermittlung

Erste Ergebnisse

Weitere Ergebnisse

Vereinheitlichung

Weitere Anwendungen

Ausblick

## Merkmale mit den höchsten Scores (GP 'gnaver', 'Nagetier')

- bæverlignende 'bieberähnlich'
- fodsåler 'Fußsohlen'
- hårbeklædt 'haarbedeckt' → *studsmus*
- nataktiv 'nachtaktiv' → *hulepindsvin*
- nutria 'Nutria'
- prærieområder 'Präriegebiete'
- Sibiriens 'Sibiriens'
- tundra 'Tundra'
- ungskov 'Jungwald'

DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Manuelle Ermittlung

Häufige GP

Statistische Ermittlung

Erste Ergebnisse

Weitere Ergebnisse

Vereinheitlichung

Weitere Anwendungen

Ausblick

## Merkmale mit den höchsten Scores (GP 'gnaver', 'Nagetier')

- bæverlignende 'bieberähnlich'
- fodsåler 'Fußsohlen'
- hårbeklædt 'haarbedeckt' → *studsmus*
- nataktiv 'nachtaktiv' → *hulepindsvin*
- nutria 'Nutria'
- prærieområder 'Präriegebiete'
- Sibiriens 'Sibiriens'
- tundra 'Tundra'
- ungskov 'Jungwald'

Niederfrequente Wortformen bekommen einen (zu?) hohen Score



DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Manuelle Ermittlung

Häufige GP

Statistische Ermittlung

Erste Ergebnisse

Weitere Ergebnisse

Vereinheitlichung

Weitere Anwendungen

Ausblick

## Merkmale mit den höchsten Scores (GP 'gnaver', 'Nagetier')

- bæverlignende 'bieberähnlich'
- fodsåler 'Fußsohlen'
- hårbeklædt 'haarbedeckt' → *studsmus*
- nataktiv 'nachtaktiv' → *hulepindsvin*
- nutria 'Nutria'
- prærieområder 'Präriegebiete'
- Sibiriens 'Sibiriens'
- tundra 'Tundra'
- ungskov 'Jungwald'

Niederfrequente Wortformen bekommen einen (zu?) hohen Score

Nicht im DDOC vorkommende Wortformen

DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Manuelle Ermittlung

Häufige GP

Statistische Ermittlung

Erste Ergebnisse

Weitere Ergebnisse

Vereinheitlichung

Weitere Anwendungen

Ausblick

## Merkmale mit den höchsten Scores (GP 'gnaver', 'Nagetier')

- bæverlignende 'bieberähnlich'  $\implies$  som ligner en bæver
- fodsåler 'Fußsohlen'
- hårbeklædt 'haarbedeckt'  $\implies$  dækket af hår
- nataktiv 'nachtaktiv'  $\implies$  aktiv om natten
- nutria 'Nutria'
- prærieområder 'Präriegebiete'  $\implies$  prærier
- Sibiriens 'Sibiriens'
- tundra 'Tundra'
- ungskov 'Jungwald'  $\implies$  ung skov

Niederfrequente Wortformen bekommen einen (zu?) hohen Score

Nicht im DDOC vorkommende Wortformen

# Weitere Ergebnisse (nach Filterung)

## Merkmale mit den höchsten Scores (GP 'kvinde', 'Frau')

'arrig'	'boshaft'	310
'prostituere'	'prostituiert'	241
'stridbar'	'streitbar'	233
'gift'	'verheiratet'	94
'husholdning'	'Haushalt'	93
'pige'	'Mädchen'	91
'tiltrækkende'	'anziehend'	78
'parforhold'	'Zweierbeziehung'	67
'smuk'	'hübsch'	53

DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Manuelle Ermittlung

Häufige GP

Statistische Ermittlung

Erste Ergebnisse

Weitere Ergebnisse

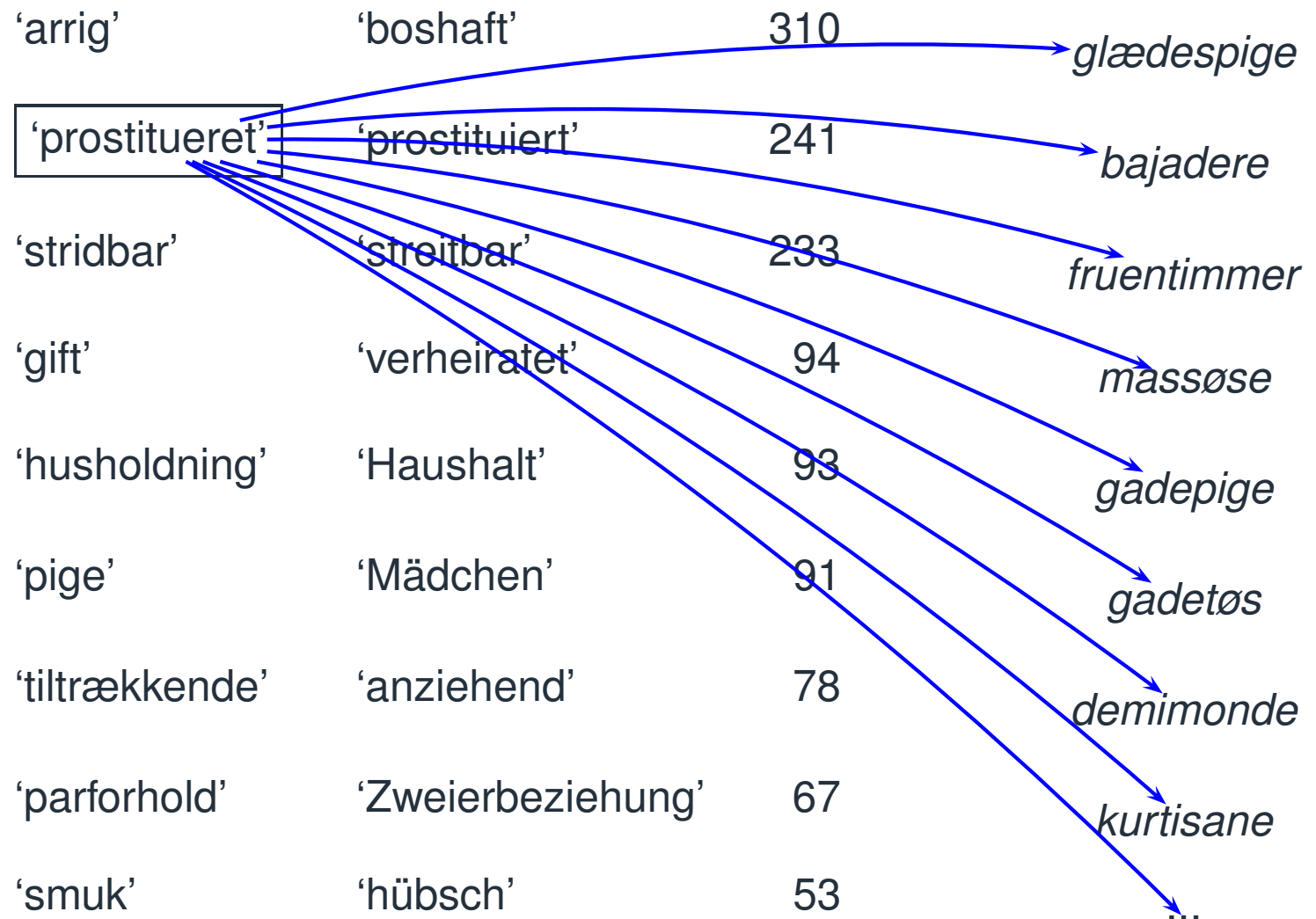
Vereinheitlichung

Weitere Anwendungen

Ausblick

# Weitere Ergebnisse (nach Filterung)

## Merkmale mit den höchsten Scores (GP 'kvinde', 'Frau')



DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Manuelle Ermittlung

Häufige GP

Statistische Ermittlung

Erste Ergebnisse

Weitere Ergebnisse

Vereinheitlichung

Weitere Anwendungen

Ausblick

# Vereinheitlichung von Bedeutungsparaphrasen

DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Manuelle Ermittlung

Häufige GP

Statistische Ermittlung

Erste Ergebnisse

Weitere Ergebnisse

Vereinheitlichung

Weitere Anwendungen

Ausblick

## Wiederkehrendes Muster in den Bedeutungsparaphrasen

- Genus proximum: 'kvinde'
- Merkmal: 'prostituertet'

BP-Muster prostituertet kvinde som ...

## Wörter, die 'kvinde' bezeichnen – ohne signifikante Merkmale

*... talskvinde finke suffragette siv sylfide nonne bolsje sirene spåkone  
skøjteprinsesse alt langhåret korpige blondine sexkilling dukke alenemor  
dulle danserinde ludder squaw nymfoman medusa gimpe rejekælling  
madamme dannekvinde bordelmutter sag damemenneske moderdyr lady  
lebbe datter nattergal groupie trolldkvinde ophavskvinde sanglærke frille  
garderobedame kvabso kvindemenneske skøge rugemor dame ...*

# Vereinheitlichung von Bedeutungsparaphrasen

DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Manuelle Ermittlung

Häufige GP

Statistische Ermittlung

Erste Ergebnisse

Weitere Ergebnisse

Vereinheitlichung

Weitere Anwendungen

Ausblick

## Wiederkehrendes Muster in den Bedeutungsparaphrasen

- Genus proximum: 'kvinde'
- Merkmal: 'prostitueret'

BP-Muster prostitueret kvinde som ...

## Wörter, die 'kvinde' bezeichnen – ohne signifikante Merkmale

... *talskvinde finke suffragette siv sylfide nonne bolsje sirene spåkone skøjteprinsesse alt langhåret korpige blondine sexkilling dukke alenemor dulle danserinde **ludder** squaw nymfoman medusa gimpe rejekælling madamme dannekvinde bordelmutter sag damemenneske moderdyr lady lebbe datter nattergal groupie troldkvinde ophavskvinde sanglærke frille garderobedame kvabso kvindemenneske **skøge** rugemor dame ...*

BP kvinde som tilbyder ... mod betaling

BP kvinde som ernærer sig ved prostitution

# Vereinheitlichung von Bedeutungsparaphrasen

DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Manuelle Ermittlung

Häufige GP

Statistische Ermittlung

Erste Ergebnisse

Weitere Ergebnisse

Vereinheitlichung

Weitere Anwendungen

Ausblick

## Wiederkehrendes Muster in den Bedeutungsparaphrasen

- Genus proximum: 'kvinde'
- Merkmal: 'prostituertes'

BP-Muster prostituertes kvinde som ...

## Wörter, die 'kvinde' bezeichnen – ohne signifikante Merkmale

.. *talskvinde finke suffragette siv sylfide nonne bolsje sirene spåkone skøjteprinsesse alt langhåret korpige blondine sexkilling dukke alenemor dulle danserinde **ludder** squaw nymfoman medusa gimpe rejekælling madamme dannekvinde bordelmutter sag damemenneske moderdyr lady lebbe datter nattergal groupie troldekvinde ophavskvinde sanglærke frille garderobedame kvabso kvindemenneske **skøge** rugemor dame ...*

BP kvinde som tilbyder ... mod betaling

BP kvinde som ernærer sig ved prostitution

DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Manuelle Ermittlung

Häufige GP

Statistische Ermittlung

Erste Ergebnisse

Weitere Ergebnisse

Vereinheitlichung

Weitere Anwendungen

Ausblick

## Primitiver onomasiologischer Zugriff auf das DDO

- ‘alter Mann’
- ‘langsam bewegen’
- ‘sehr groß’
- ‘hübsche Frau’



DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Manuelle Ermittlung

Häufige GP

Statistische Ermittlung

Erste Ergebnisse

Weitere Ergebnisse

Vereinheitlichung

Weitere Anwendungen

Ausblick

## Primitiver onomasiologischer Zugriff auf das DDO

- ↔ 'alter Mann'
- 'langsam bewegen'
- 'sehr groß'
- 'hübsche Frau'

Træk: gammel  
gubbe sb.  
knark sb.  
støder sb.  
stabejs sb.  
jubelolding sb.  
gammelfar sb.  
patriark sb.  
julemand sb.

## Primitiver onomasiologischer Zugriff auf das DDO

- ‘alter Mann’
- ↔ ‘langsam bewegen’
- ‘sehr groß’
- ‘hübsche Frau’

Træk: **langsomt**  
krybe vb.  
drysse vb. bet. 3  
kravle vb.  
dorsk adj.  
snegle vb. snegle sig  
varpe vb. bet. 3  
trille vb. bet. 4  
orme vb. orme sig  
trække vb. bet. 12  
snige vb. snige sig  
slæbe vb. slæbe sig  
snirke vb.  
svæve vb. bet. 3  
sive vb. bet. 3  
luffe vb.  
glide vb. glide op  
fremtrængende adj.  
liste vb.  
drive vb. bet. 5  
knirke vb.  
snile vb. bet. 4

DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Manuelle Ermittlung

Häufige GP

Statistische Ermittlung

Erste Ergebnisse

Weitere Ergebnisse

Vereinheitlichung

Weitere Anwendungen

Ausblick

## Primitiver onomasiologischer Zugriff auf das DDO

- 'alter Mann'
- 'langsam bewegen'
- ↪ 'sehr groß'
- 'hübsche Frau'

Træk: meget  
mægtig adj.  
gevaldig adj.  
gigant- førsteled  
mammut- førsteled  
himmelvid adj.  
monstrøs adj.  
uhyre adj.  
størst adj. bet. 2  
vældig adj.  
milevid adj.  
kæmpe adj.  
drabelig adj.  
kodyl adj.  
kolossal adj.  
tordnende adj. bet. 2  
omfangsrig adj.  
uhyrlig adj.  
umanérlig adj.  
formidabel adj.  
himmelhøj adj. bet. 2  
gigantisk adj.

DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Manuelle Ermittlung

Häufige GP

Statistische Ermittlung

Erste Ergebnisse

Weitere Ergebnisse

Vereinheitlichung

Weitere Anwendungen

Ausblick

## Primitiver onomasiologischer Zugriff auf das DDO

- ‘alter Mann’
- ‘langsam bewegen’
- ‘sehr groß’
- ↪ ‘hübsche Frau’

Træk: **smuk**  
**dejlighed sb.** bet. 2  
**gratie sb.**  
**skønjomfru sb.**  
**yndighed sb.** bet. 2  
**nymfe sb.**  
**lamseben sb.**  
**larve sb.** laber larve  
**gudinde sb.**  
**badenymfe sb.**  
**cirkusprinsesse sb.**  
**huri sb.**

DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Manuelle Ermittlung

Häufige GP

Statistische Ermittlung

Erste Ergebnisse

Weitere Ergebnisse

Vereinheitlichung

Weitere Anwendungen

Ausblick

DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Ausblick

Fragen und Antworten

Zukunftsvision

# Ausblick

## Korpuslinguistik und Erkenntnisfortschritt?

- Ja!
  - Als Quelle für die lexikographische Arbeit
  - Als Werkzeug bei der lexikographischen Arbeit:
    - konsistenterer Aufbau von Bedeutungsparaphrasen

## Sprachgebrauch des Korpus = Sprachgebrauch in den BP?

- Hinsichtlich verwendeter Hyperonyme
- Hinsichtlich des Vokabulars und der Syntax
- Problem der Metasprache in der Lexikographie:
  - Bewusster Aufbau einer (formalen) BP-Metasprache?

DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Ausblick

Fragen und Antworten

Zukunftsvision

## Das Ziel

- Eine multifunktionale lexikalische Ressource, die
  - in unterschiedlichen Buchversionen erscheinen kann
  - als digitales Nachschlagewerk verwendet werden kann
  - auch für NLP-Zwecke verwendet werden kann

DSL

Rolle des Korpus

Digitale Wörterbücher

Semantik im DDO

Genera proxima im DDO

Differentiae im DDO

Ausblick

**Vielen Dank für Ihre  
Aufmerksamkeit!**