

# CCDB

Eine korpuslinguistische Denk-  
und Experimentierplattform

Holger Keibel

19. März 2008

TUFS, Global COE International Workshop  
„Neues aus der korpusling. Forschung am IDS Mannheim“

---

---

# Übersicht

- Einleitung
- CCDB-Daten: Kookkurrenzprofile
  - Kollokationen höherer Ordnung
  - syntagmatische Muster
- CCDB: gegenwärtige Methoden
  - Ähnlichkeit von Kookkurrenzprofilen
  - Analyse der Gebrauchsaspekte eines Wortes anhand ähnlicher Kookkurrenzprofile
    - hierarchisches Clustering
    - SOM
  - Kontrastierung von Quasisynonymen

---

# Allgemeiner Hintergrund

- Auftrag des IDS:  
„... die deutsche Sprache in ihrem gegenwärtigen Gebrauch zu erforschen und zu dokumentieren ...“
- empirische Fundierung erforderlich
  - korpuslinguistischer Ansatz
    - sehr große Sprachkorpora
    - elaborierte Korpusanalysemethoden

---

# CCDB

- „eine korpuslinguistische Denk- und Experimentierplattform für die Erforschung und theoretische Begründung von systemisch-strukturellen Eigenschaften von Kohäsionsrelationen zwischen den Konstituenten des Sprachgebrauchs“
- Cyril Belica (2001-2007)

<http://corpora.ids-mannheim.de/ccdb/>

---

# Korpus

Datengrundlage:

ein 2,2 Milliarden Wörter umfassendes Teilkorpus des  
DEUTSCHEN REFERENZKORPUS (DEREKO)

– größtes Korpusarchiv für geschriebene deutsche  
Gegenwartssprache.

<http://www.ids-mannheim.de/kl/projekte/korpora/>

---

# Kollokationen: mögliche Konzepte

## – *N-Gramm*

- zusammenhängende Folge von  $N$  Wörtern, die überzufällig häufig in genau dieser Abfolge auftreten

- Beispiele:

breite Mehrheit, weiße Weste, immer wieder, ab und zu, viel Geld gekostet

- nur begrenzt hilfreich:

Viele Verwendungsmuster sind diskontinuierlich.

– z.B. für komplexe Verben wie *loslassen*:

es **lässt** ihn nicht mehr **los**

– aber auch für praktisch jedes andere Wort

er **fragte** seine Schwester, **ob** ...

---

# Kollokationen: mögliche Konzepte 2

– *positionelles N-Gramm*

- Menge von  $N$  Wörtern, die überzufällig häufig mit bestimmten festen relativen Wortpositionen auftreten
- die folgenden Beispiele stellen also unterschiedliche positionelle N-Gramme dar:
  - Beispiel 1: fragte \_\_\_\_\_ ob  
etwa in: er fragte sie, ob ...
  - Beispiel 2: fragte \_\_\_\_\_ \_\_\_\_\_ ob  
etwa in: er fragte seine Schwester, ob ...
- zu starres Konzept für unsere Zwecke:  
Beispiele 1 und 2 sollten eher als dasselbe Verwendungsmuster betrachtet werden.

---

# Kollokation höherer Ordnung

- komplexe Verbindung von  $N$  Wörtern, die überzufällig häufig gemeinsam vorkommen
- möglicherweise diskontinuierlich
- variable relative Wortpositionen  
→ insbesondere: Wortreihenfolge kann variieren
- **Beispiel:** Kookkurrenz der Wörter **weshalb**, **Grund**, **der** :
  - (1) Das ist **der Grund**, **weshalb** man immer ...
  - (2) Das ist **der** einzige **Grund**, **weshalb** man ...
  - (3) Doch **der** Lehrer wusste keinen besseren **Grund**, **weshalb** ...
  - (4) Sie fragte, **weshalb der Grund** so wichtig sei.
  - (5) Nun fragt sich **der** Student natürlich, **weshalb** man überhaupt einen **Grund** braucht, um ...



---

# Inwiefern „höherer Ordnung“?

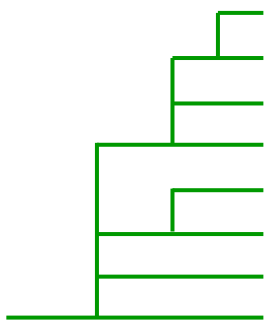
Der Algorithmus, mit dem Kollokationen höherer Ordnung aufgespürt werden, ist eine iterativ angewendete Erweiterung des Standardalgorithmus zum Aufspüren von einfachen Kollokationen (N-Grammen).

## Beispiel:

|                       |                       |
|-----------------------|-----------------------|
| gegebenes Bezugswort: | weshalb               |
| nach Schleife 1:      | weshalb Grund         |
| nach Schleife 2:      | weshalb Grund der     |
| nach Schleife 3:      | weshalb Grund der ... |

Beispiel:  
*machen*

|               |                                    |      |  |
|---------------|------------------------------------|------|--|
| [-2 -1 11238] | <b>aufmerksam</b> werden           | 31   | 96% darauf <b>aufmerksam</b> gemacht [...] werden  |
| [-2 -1 11238] | aufmerksam wollte                  | 17   | 52% <b>aufmerksam</b> machen wollte  |
| [-2 -1 11238] | aufmerksam Öffentlichkeit          | 12   | 66% die <b>Öffentlichkeit</b> [darauf auf ...] <b>aufmerksam</b> [zu] <b>machen</b> daß de |
| [-2 -1 11238] | aufmerksam                         | 1617 | 45% <b>aufmerksam</b> [zu] <b>machen</b>   |
| [-1 5 6970]   | <b>Spaß</b> richtig viel           | 1    | 100% <b>macht</b> ... richtig viel <b>Spaß</b>   |
| [-1 5 6970]   | Spaß richtig                       | 53   | 50% <b>macht</b> [...] richtig [...] <b>Spaß</b>   |
| [-1 5 6970]   | <b>Spaß</b> viel                   | 172  | 37% Es <b>macht</b> [...] sehr] viel [...] <b>Spaß</b>                                     |
| [-1 5 6970]   | Spaß einfach                       | 63   | 50% Es <b>macht</b> [...] einfach [...] <b>Spaß</b>  |
| [-1 5 6970]   | Spaß                               | 1459 | 40% <b>macht</b> [...] <b>Spaß</b>   |
| [-2 4 6186]   | <b>deutlich</b> habe               | 59   | 76% <b>habe</b> [...] <b>deutlich</b> gemacht daß dass ...                                 |
| [-2 4 6186]   | deutlich                           | 1965 | 23% <b>deutlich</b> [...] gemacht dass daß ...   |
| [-2 -1 6051]  | <b>geltend</b> werden können       | 26   | 69% <b>geltend</b> gemacht werden [...] können   |
| [-2 -1 6051]  | geltend werden                     | 106  | 90% <b>geltend</b> [...] gemacht [...] werden  |
| [-2 -1 6051]  | <b>geltend</b> können              | 60   | 45% <b>geltend</b> machen [...] können   |
| [-2 -1 6051]  | geltend hatten                     | 12   | 83% <b>hatten</b> [in ...] <b>geltend</b> gemacht daß ...                                  |
| [-2 -1 6051]  | geltend                            | 810  | 44% <b>geltend</b> [zu] <b>machen</b>  |
| [-2 -1 4188]  | <b>rückgängig</b> werden kann wird | 1    | 100% <b>wird</b> ... <b>rückgängig</b> gemacht werden kann                                 |
| [-2 -1 4188]  | <b>rückgängig</b> werden kann      | 15   | 73% <b>rückgängig</b> gemacht werden kann  |
| [-2 -1 4188]  | <b>rückgängig</b> werden wird      | 2    | 50% <b>wird</b> ... <b>rückgängig</b> gemacht werden                                       |
| [-2 -1 4188]  | <b>rückgängig</b> werden           | 86   | 94% <b>rückgängig</b> gemacht [...] werden   |
| [-2 -1 4188]  | <b>rückgängig</b> kann wird        | 2    | 50% <b>wird</b> ... <b>rückgängig</b> gemacht ... kann                                     |
| [-2 -1 4188]  | <b>rückgängig</b> kann             | 26   | 42% <b>rückgängig</b> gemacht werden kann  |
| [-2 -1 4188]  | <b>rückgängig</b> wird             | 22   | 50% die ... <b>rückgängig</b> gemacht [...] wird   |
| [-2 -1 4188]  | <b>rückgängig</b>                  | 445  | 56% <b>rückgängig</b> [zu] <b>machen</b>   |
| [-1 -1 3829]  | <b>Fehler</b> haben worden         | 1    | 100% <b>haben</b> ... <b>Fehler</b> gemacht ... worden                                     |
| [-1 -1 3829]  | Fehler haben habe                  | 1    | 100% <b>haben</b> ... <b>Fehler</b> gemacht habe   |
| [-1 -1 3829]  | Fehler haben                       | 109  | 45% Wir <b>haben</b> [...] einen ...] <b>Fehler</b> gemacht                                |
| [-1 -1 3829]  | <b>Fehler</b> worden               | 43   | 95% <b>Fehler</b> [...] gemacht [...] worden   |
| [-1 -1 3829]  | Fehler habe                        | 83   | 61% Ich <b>habe</b> [einen] <b>Fehler</b> gemacht  |
| [-1 -1 3829]  | Fehler                             | 965  | 48% einen <b>Fehler</b> [...] gemacht  |
| [-2 -1 3721]  | <b>Gebrauch</b> wird Angebot       | 1    | 100% <b>Angebot</b> wird ... <b>Gebrauch</b> gemacht                                       |
| [-2 -1 3721]  | Gebrauch wird                      | 25   | 48% <b>wird</b> ... <b>Gebrauch</b> gemacht  |



---

# Kohäsion / Kohäsivität

- Kollokation (höherer Ordnung)  
= überzufällig häufiges gemeinsames Vorkommen
- Kohäson  
= Stärke dieses überzufällig häufigen gemeinsamen Vorkommens  
→ intuitiv: wie viel häufiger als durch puren Zufall?
- formale Kohäsionsmaße:
  - log-likelihood ratio (LLR)
  - mutual information (MI)

Beispiel:  
*machen*

|         |       |                                    |      |      |  |
|---------|-------|------------------------------------|------|------|--|
| [-2 -1] | 11238 | <b>aufmerksam</b> werden           | 31   | 96%  | darauf <b>aufmerksam</b> gemacht [...] werden  |
| [-2 -1] | 11238 | aufmerksam wollte                  | 17   | 52%  | <b>aufmerksam</b> machen wollte  |
| [-2 -1] | 11238 | aufmerksam Öffentlichkeit          | 12   | 66%  | die <b>Öffentlichkeit</b> [darauf auf ...] <b>aufmerksam</b> [zu] <b>machen</b> daß de |
| [-2 -1] | 1123  | <b>aufmerksam</b>                  | 1617 | 45%  | <b>aufmerksam</b> [zu] <b>machen</b>   |
| [-1 5]  | 6970  | <b>Spaß</b> richtig viel           | 1    | 100% | <b>macht</b> ... richtig viel <b>Spaß</b>  |
| [-1 5]  | 6970  | Spaß richtig                       | 53   | 50%  | <b>macht</b> [...] richtig [...] <b>Spaß</b>   |
| [-1 5]  | 6970  | Spaß viel                          | 172  | 37%  | Es <b>macht</b> [...] sehr] viel [...] <b>Spaß</b>                                     |
| [-1 5]  | 6970  | Spaß einfach                       | 63   | 50%  | Es <b>macht</b> [...] einfach [...] <b>Spaß</b>  |
| [-1 5]  | 6970  | Spaß                               | 1459 | 40%  | <b>macht</b> [...] <b>Spaß</b>   |
| [-2 4]  | 6186  | <b>deutlich</b> habe               | 59   | 76%  | <b>habe</b> [...] <b>deutlich</b> gemacht daß dass ...                                 |
| [-2 4]  | 6186  | deutlich                           | 1965 | 23%  | <b>deutlich</b> [...] gemacht dass daß ...   |
| [-2 -1] | 6051  | <b>geltend</b> werden können       | 26   | 69%  | <b>geltend</b> gemacht werden [...] können   |
| [-2 -1] | 605   | geltend werden                     | 106  | 90%  | <b>geltend</b> [...] gemacht [...] werden  |
| [-2 -1] | 605   | geltend können                     | 60   | 45%  | <b>geltend</b> machen [...] können   |
| [-2 -1] | 605   | geltend hatten                     | 12   | 83%  | <b>hatten</b> [in ...] <b>geltend</b> gemacht daß ...                                  |
| [-2 -1] | 605   | <b>geltend</b>                     | 810  | 44%  | <b>geltend</b> [zu] <b>machen</b>  |
| [-2 -1] | 4188  | <b>rückgängig</b> werden kann wird | 1    | 100% | <b>wird</b> ... <b>rückgängig</b> gemacht werden kann                                  |
| [-2 -1] | 4188  | rückgängig werden kann             | 15   | 73%  | <b>rückgängig</b> gemacht werden kann  |
| [-2 -1] | 4188  | rückgängig werden wird             | 2    | 50%  | <b>wird</b> ... <b>rückgängig</b> gemacht werden                                       |
| [-2 -1] | 4188  | rückgängig werden                  | 86   | 94%  | <b>rückgängig</b> gemacht [...] werden   |
| [-2 -1] | 4188  | rückgängig kann wird               | 2    | 50%  | <b>wird</b> ... <b>rückgängig</b> gemacht ... kann                                     |
| [-2 -1] | 4188  | rückgängig kann                    | 26   | 42%  | <b>rückgängig</b> gemacht werden kann  |
| [-2 -1] | 4188  | rückgängig wird                    | 22   | 50%  | die ... <b>rückgängig</b> gemacht [...] wird   |
| [-2 -1] | 4188  | rückgängig                         | 445  | 56%  | <b>rückgängig</b> [zu] <b>machen</b>   |
| [-1 -1] | 3829  | <b>Fehler</b> haben worden         | 1    | 100% | <b>haben</b> ... <b>Fehler</b> gemacht ... worden                                      |
| [-1 -1] | 3829  | Fehler haben habe                  | 1    | 100% | <b>haben</b> ... <b>Fehler</b> gemacht habe  |
| [-1 -1] | 3829  | Fehler haben                       | 109  | 45%  | Wir <b>haben</b> [...] einen ...] <b>Fehler</b> gemacht                                |
| [-1 -1] | 3829  | Fehler worden                      | 43   | 95%  | <b>Fehler</b> [...] gemacht [...] worden   |
| [-1 -1] | 3829  | Fehler habe                        | 83   | 61%  | Ich <b>habe</b> [einen] <b>Fehler</b> gemacht  |
| [-1 -1] | 3829  | Fehler                             | 965  | 48%  | einen <b>Fehler</b> [...] gemacht  |
| [-2 -1] | 3721  | <b>Gebrauch</b> wird Angebot       | 1    | 100% | <b>Angebot</b> wird ... <b>Gebrauch</b> gemacht  |
| [-2 -1] | 3721  | Gebrauch wird                      | 25   | 48%  | <b>wird</b> ... <b>Gebrauch</b> gemacht  |

---

# Syntagmatische Muster

- Kollokation höherer Ordnung:  
Reihenfolge der Wörter (Kollokatoren) kann variieren
- *syntagmatisches Muster*  
= eine dieser Reihenfolgen plus *wild-card*-Symbole,  
die anzeigen, wo andere Wörter zwischen den  
Kollokatoren beobachtet wurden
- zur verbesserten Lesbarkeit:  
Syntagmatische Muster werden üblicherweise  
zusammen mit Wörtern präsentiert, die besonders oft  
in einer dieser *wild-card*-Positionen beobachtet  
wurden.

---

# Syntagmatische Muster: Beispiele

- *sich zu vergewissern daß|ob|dass ... auch wirklich*
- *darauf aufmerksam [zu] machen [...] dass*
- *Das ist [...] insofern [...] bemerkenswert als ...*
- *alle [...] Hände [...] voll [zu] tun*
- *besser [...] als [...] erwartet*
- *ungefähr auf|das|dem Niveau von ...*
- *die|alle Fakten [...] auf den Tisch [zu] legen*
- *genug [...] um zu wissen daß|dass ...*

Beispiel:  
*machen*

|         |       |                             |      |      |   |
|---------|-------|-----------------------------|------|------|---|
| [-2 -1] | 11238 | aufmerksam worden           | 31   | 96%  | darauf aufmerksam gemacht [...] worden                            |
| [-2 -1] | 11238 | aufmerksam wollte           | 17   | 52%  | aufmerksam machen wollte  |
| [-2 -1] | 11238 | aufmerksam Öffentlichkeit   | 12   | 66%  | die Öffentlichkeit [darauf auf ...] aufmerksam [zu] machen daß de |
| [-2 -1] | 11238 | aufmerksam                  | 1617 | 45%  | aufmerksam [zu] machen  |
| [-1 5]  | 6970  | Spaß richtig viel           | 1    | 100% | macht ... richtig viel Spaß                                       |
| [-1 5]  | 6970  | Spaß richtig                | 53   | 50%  | macht [...] richtig [...] Spaß                                    |
| [-1 5]  | 6970  | Spaß viel                   | 172  | 37%  | Es macht [...] sehr] viel [...] Spaß                              |
| [-1 5]  | 6970  | Spaß einfach                | 63   | 50%  | Es macht [...] einfach [...] Spaß                                 |
| [-1 5]  | 6970  | Spaß                        | 1459 | 40%  | macht [...] Spaß  |
| [-2 4]  | 6186  | deutlich habe               | 59   | 76%  | habe [...] deutlich gemacht daß dass ...                          |
| [-2 4]  | 6186  | deutlich                    | 1965 | 23%  | deutlich [...] gemacht dass daß ...                               |
| [-2 -1] | 6051  | geltend werden können       | 26   | 69%  | geltend gemacht werden [...] können                               |
| [-2 -1] | 6051  | geltend werden              | 106  | 90%  | geltend [...] gemacht [...] werden                                |
| [-2 -1] | 6051  | geltend können              | 60   | 45%  | geltend machen [...] können                                       |
| [-2 -1] | 6051  | geltend hatten              | 12   | 83%  | hatten [in ...] geltend gemacht daß ...                           |
| [-2 -1] | 6051  | geltend                     | 810  | 44%  | geltend [zu] machen   |
| [-2 -1] | 4188  | rückgängig werden kann wird | 1    | 100% | wird ... rückgängig gemacht werden kann                           |
| [-2 -1] | 4188  | rückgängig werden kann      | 15   | 73%  | rückgängig gemacht werden kann                                    |
| [-2 -1] | 4188  | rückgängig werden wird      | 2    | 50%  | wird ... rückgängig gemacht werden                                |
| [-2 -1] | 4188  | rückgängig werden           | 86   | 94%  | rückgängig gemacht [...] werden                                   |
| [-2 -1] | 4188  | rückgängig kann wird        | 2    | 50%  | wird ... rückgängig gemacht ... kann                              |
| [-2 -1] | 4188  | rückgängig kann             | 26   | 42%  | rückgängig gemacht werden kann                                    |
| [-2 -1] | 4188  | rückgängig wird             | 22   | 50%  | die ... rückgängig gemacht [...] wird                             |
| [-2 -1] | 4188  | rückgängig                  | 445  | 56%  | rückgängig [zu] machen  |
| [-1 -1] | 3829  | Fehler haben worden         | 1    | 100% | haben ... Fehler gemacht ... worden                               |
| [-1 -1] | 3829  | Fehler haben habe           | 1    | 100% | haben ... Fehler gemacht habe                                     |
| [-1 -1] | 3829  | Fehler haben                | 109  | 45%  | Wir haben [...] einen ...] Fehler gemacht                         |
| [-1 -1] | 3829  | Fehler worden               | 43   | 95%  | Fehler [...] gemacht [...] worden                                 |
| [-1 -1] | 3829  | Fehler habe                 | 83   | 61%  | Ich habe [einen] Fehler gemacht                                   |
| [-1 -1] | 3829  | Fehler                      | 965  | 48%  | einen Fehler [...] gemacht  |
| [-2 -1] | 3721  | Gebrauch wird Angebot       | 1    | 100% | Angebot wird ... Gebrauch gemacht                                 |
| [-2 -1] | 3721  | Gebrauch wird               | 25   | 48%  | wird ... Gebrauch gemacht   |

---

# Kookkurrenzprofile in der CCDB

- gegebenes Bezugswort
- das Kookkurrenzprofil dieses Wortes beinhaltet:
  - die kohäsivsten Kollokationen höherer Ordnung, in denen es vorkommt
  - zu jeder dieser Kollokationen: das jeweils häufigste syntagmatische Muster
- Kookkurrenzprofil  $\approx$  Gebrauchsprofil
- Kookkurrenzprofile für ca. 220.000 Bezugswörter
- derzeit: zwei getrennte Datensätze/Ansichten (*mit bzw. ohne Synsemantika*)
- Grundlage: deutsche Schriftsprache (2,2 Milliarden Wörter umfassendes Teilkorpus von DEREKO)



---

# Einsatzmöglichkeiten der CCDB

- CCDB: Kookkurrenzprofile als
  - linguistische Ergebnisse,  
direkt interpretierbar  
(→ Lexikografie, DaF, linguistische Theoriebildung, ...)
  - Datengrundlage für weitere Analysemethoden  
zur Erforschung der deutschen Sprache:  
CCDB als Denk- und Experimentierplattform,  
um neue Analysemethoden
    - zu erforschen
    - zu entwickeln
    - zu evaluieren

---

# derzeitige Analysemethoden

- grundlegendes Ziel:  
entdecke und erforsche Strukturen  
im Gebrauch von Wörtern
  - Ähnlichkeit von Kookkurrenzprofilen
  - Analyse der Gebrauchsaspekte eines Wortes  
anhand ähnlicher Kookkurrenzprofile
  - Kontrastierung von Quasisynonymen

---

# Ähnlichkeit von Kookkurrenzprofilen

- Kookkurrenzprofil eines Bezugsworts  
≈ Profil der Gebrauchsmuster/Gebrauchsaspekte dieses Wortes
- Wörter, die sich in ihrem Gebrauch ähneln  
↔ ähnliche Kookkurrenzprofile
- Vergleich von Kookkurrenzprofilen (als Ganzes!)
- komplexe, nichttriviale Herausforderung:  
Entwicklung einer Ähnlichkeitsmetrik, die quantifiziert,  
wie ähnlich sich zwei Kookkurrenzprofile sind  
(viele verschiedene Aspekte von Ähnlichkeit)
- Testen verschiedener Ansätze

---

# Ähnlichkeit von Kookkurrenzprofilen

- für jeden Datensatz (mit/ohne Synsemantika):  
eine eigene Ähnlichkeitsmetrik
- primäre Anwendung:  
suche zu einem gegebenen Bezugswort  
Wörter mit ähnlichen Gebrauchseigenschaften  
(also mit ähnlichen Kookkurrenzprofilen)  
→ CCDB-Methode „**show related collocation profiles**“
- aufgespürte linguistische Beziehungen:  
semantisch, paradigmatisch, syntagmatisch,  
Terminologie, idiomatisch, ...

---

# Ähnlichkeit von Kookkurrenzprofilen

- weitere Anwendungen:  
Aufdecken von komplexeren Strukturen im „Ähnlichkeitsraum“, der von den Kookkurrenzprofilen aufgespannt wird  
→ im Raum, in dem sich alle Bezugswörter befinden und in dem ähnliche Bezugswörter nahe beieinander liegen
- Ähnlichkeit von Kookkurrenzprofilen ist die Grundlage aller weiteren Analysemethoden

---

# Analyse der Gebrauchsaspekte eines Wortes anhand ähnlicher Kookkurrenzprofile

- gegebenes Bezugswort
- Ziel: Exploration der Gebrauchsaspekte dieses Wortes
- Grundidee:
  - verwende die Menge der Wörter, die diesem Wort am ähnlichsten sind
  - visualisiere die (komplexe) Ähnlichkeitsstruktur zwischen diesen Wörtern in vereinfachender Weise
- zwei komplementäre Ansätze:
  - hierarchische Cluster-Analyse
    - CCDB-Methode „**semantic proximity model**“
  - selbstorganisierende Karten (SOMs)
    - CCDB-Methode „**topographic profile SOM**“

---

# „semantic proximity model“

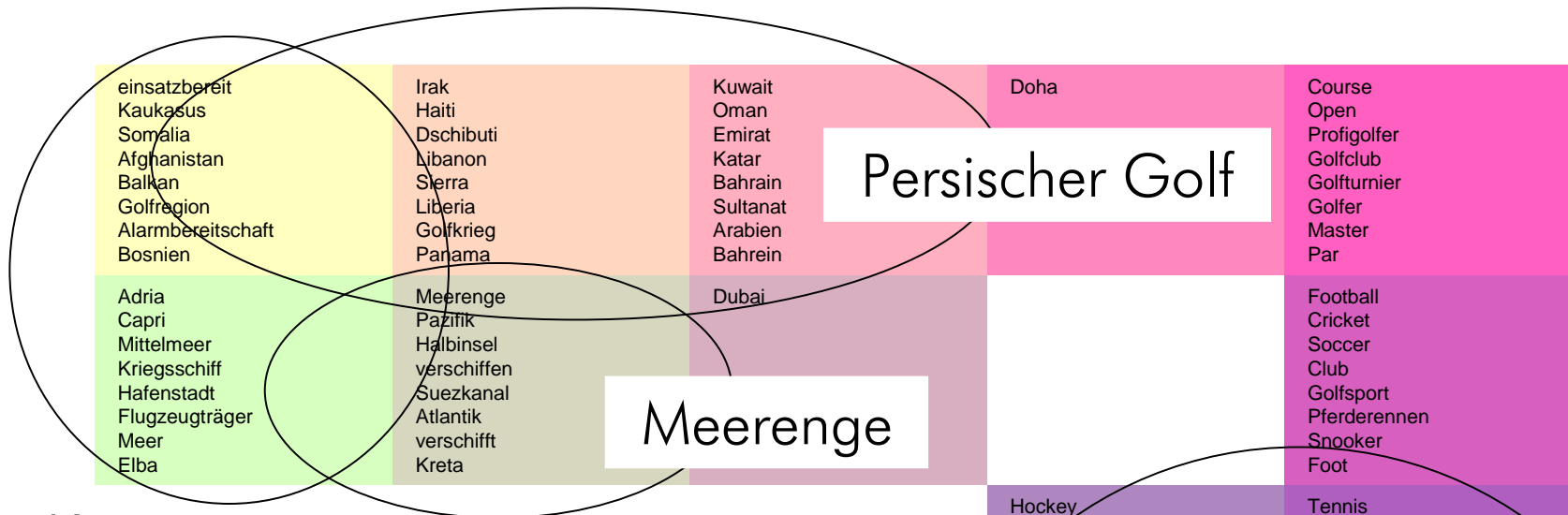
- hierarchische Cluster-Analyse:  
untersuche Ähnlichkeitsbeziehungen zwischen den Wörtern und stelle sie als einen hierarchischen Baum dar
- je ähnlicher zwei Wörter, desto „früher“ werden sie in dem Baum zusammen gefasst
- v.a. dann sinnvoll, wenn man für das gegebene Bezugswort davon ausgeht, dass das Spektrum seiner Gebrauchsaspekte/Bedeutungaspekte eine hierarchische Struktur besitzt

---

# „topographic profile SOM“

- selbstorganisierende Karten  
(*self-organizing maps, SOMs*):  
ordnen Wörter so auf einem zweidimensionalen Gitternetz an,  
dass ähnliche Wörter im Gitter nahe beieinander liegen,  
und weniger ähnliche Wörter weiter entfernt
- Algorithmus:
  - zunächst zufällige Verteilung der Wörter
  - in vielen Analysezyklen:
    - ähnliche Wörter ziehen sich an (→ bewegen sich aufeinander zu)
    - unähnliche Wörter stoßen sich ab (→ bewegen sich voneinander weg)
  - Mehrfache Durchführung derselben Analyse führt i.a. zu unterschiedlichen Ergebnissen.





Kriegsregion

Meerenge

Persischer Golf

Sportart

Fahrzeugmarke

---

# Kontrastierung von Quasisynonymen

- gegebenes Bezugswort und ein Vergleichswort
- Ziel: Visualisierung und Exploration der gemeinsamen und der kontrastierenden Gebrauchsaspekte der beiden Wörter
- Grundidee:
  - verwende die Menge der Wörter, die dem Bezugswort am ähnlichsten sind, und die Menge der Wörter, die dem Vergleichswort am ähnlichsten sind
  - visualisiere die (komplexe) Ähnlichkeitsstruktur zwischen diesen Wörtern durch eine SOM
  - färbe Gitterfelder, je nachdem welchem der beiden Wörter (Bezugswort, Vergleichswort) sie mehr ähneln
- CCDB-Methode „**contrast near-synonyms**“

---

# Relevanz

- linguistische Theoriebildung
- Fremdsprachendidaktik (DaF)
- Lexikografie
- ...

---

**Vielen Dank!**

`korpuslinguistik@ids-mannheim.de`

---