

Using Collocations to Uncover the »Meaning Potential« of Lexical Items

Rainer Perkuhn

New Approaches in Corpus Linguistics
Tokyo, 18.3.2008

My Objective

- Don't tell them what you know about linguistics, don't be »linguologists«!
- Tell them what you know about language!

Peter Eisenberg's answer to the question:
»What can linguistics do for teachers?«

IDS in a Nutshell

The Institute for the German Language (IDS) in Mannheim is the central institution for the study and documentation of the contemporary usage and recent history of the German language. Together with 83 other non-university research institutes and service facilities, it belongs to the Leibniz Association, one of the four major research organisations in Germany.

- 2007: 73 researchers 38 non-scientific employees
 > 50 research assistants (students)
 23 mid- and long-term projects + 3 partnerships
-

Structure of the IDS

Department "Grammar"

Department "Pragmatics"

Department "Lexis"

– Programme Research Areas

- "Lexicology und Lexicography"
- "Corpus Linguistics"

Administration and Technical Support

Public Relations and Library

Structure of the IDS

- Programme Research Area “Corpus Linguistics”
(C. Belica, H. Keibel, M. Kupietz, R. Perkuhn)
 - Archive of General Reference Corpora of Contemporary Written German
 - Methods of Corpus Analysis and Corpus Mining
-

Motivation

- Mission of the Institute for the German Language:
 - “... to investigate and document the German language in its current use ...”
-

Corpora

- empirical base for linguistic research
 - technically: full-text database in computer-readable form
 - and a gold mine for latent linguistic structures and phenomena, uncovered by sophisticated analysis methods
-

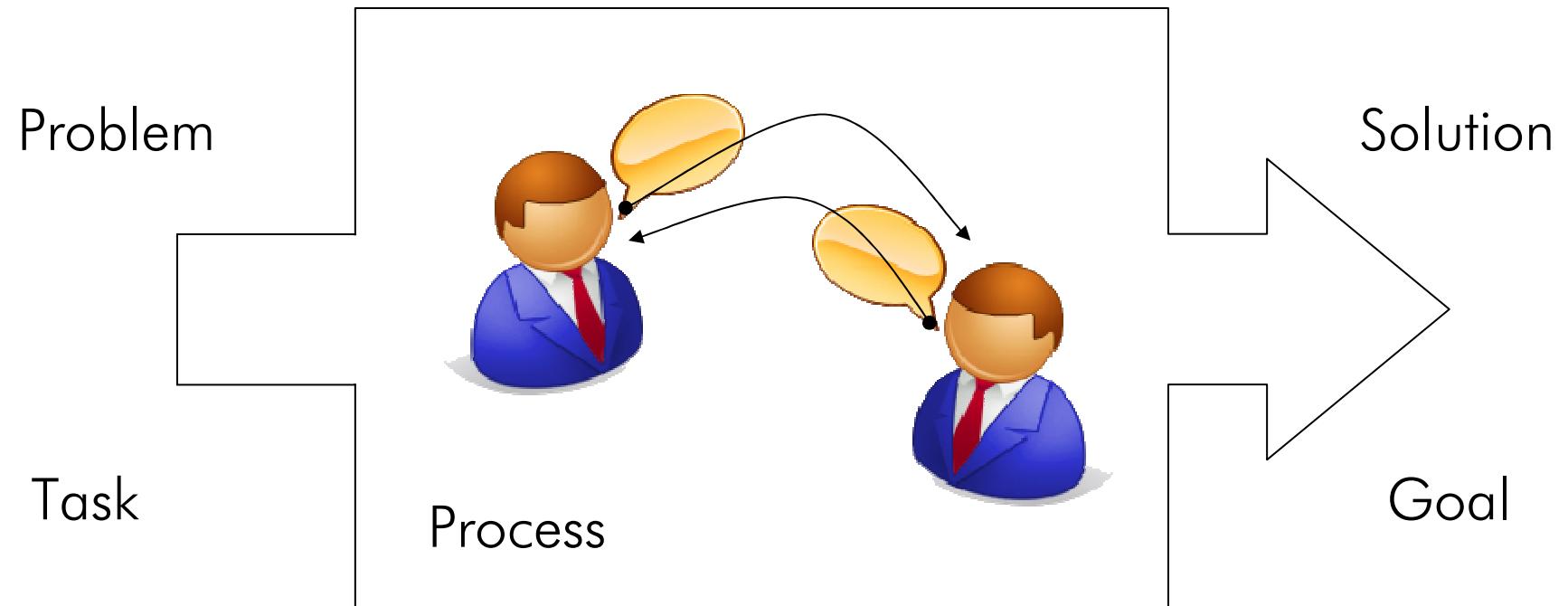
Corpora at the IDS

- long tradition since 1964
 - since 1992 dedicated task, working group, today research programme area
 - task wrt. to rules of IDS, not bound to persons
 - considered a permanent task (even 40 years later on ;-)
 - DEREKO (I) as pilot project 1999-2002
 - nowadays the name of the overall archive of texts of the corpora of the IDS
-

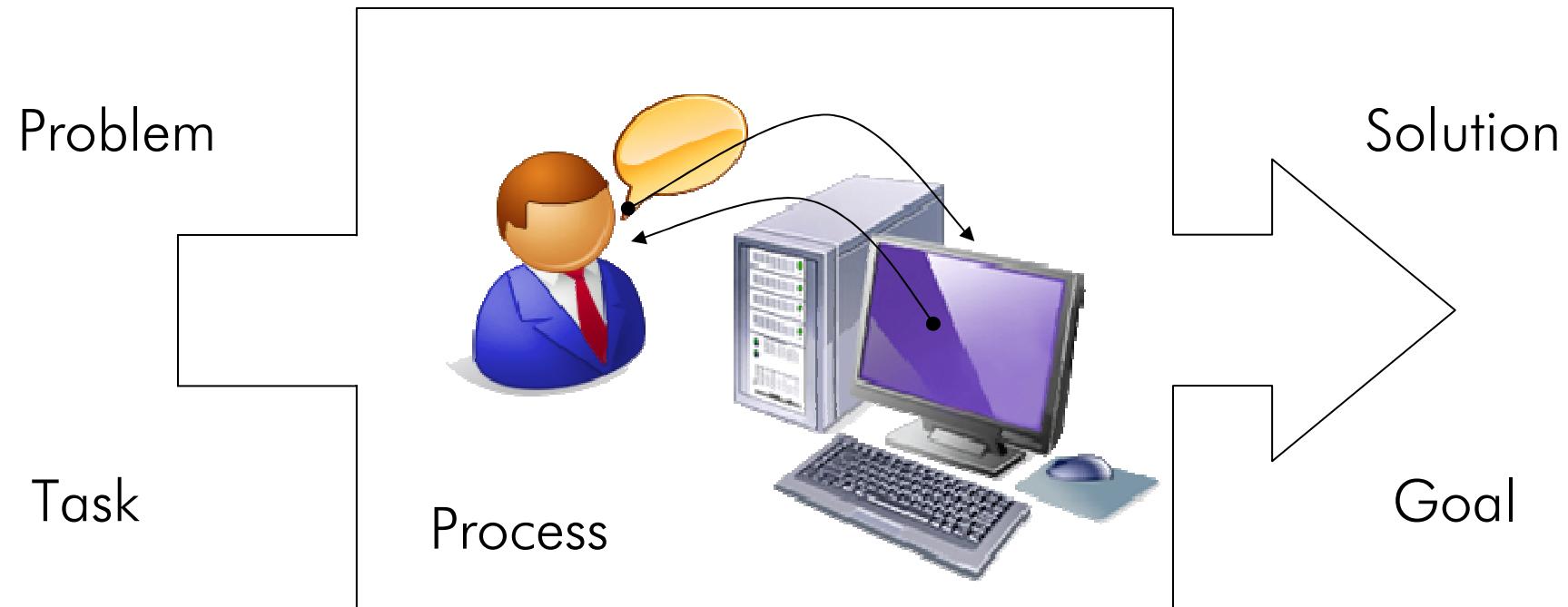
Computational Linguistics vs. Corpus Linguistics

- language as obstacle in solving a problem, computer should help in accomplishing a task
 - language as subject of investigation, computer should help to understand linguistic phenomena
-

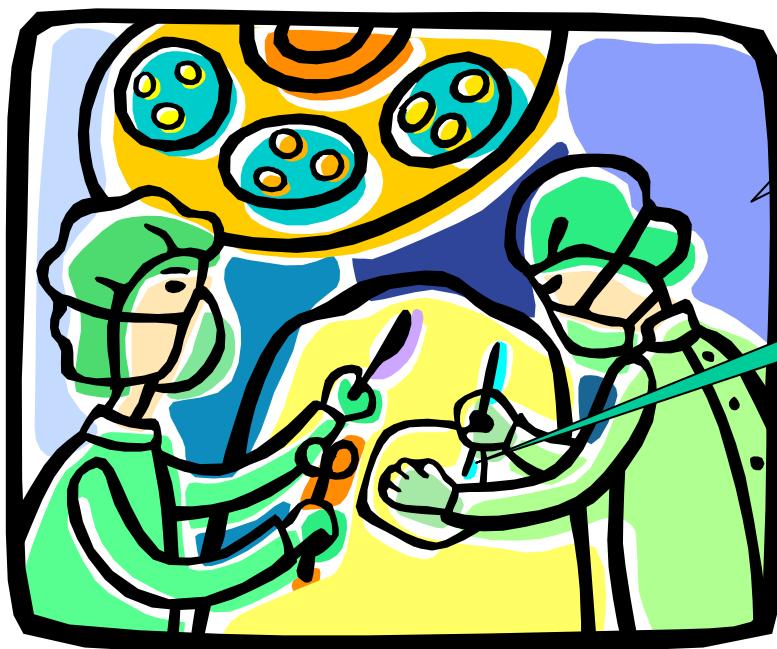
View Computational Linguistics



View Computational Linguistics



View Corpus Linguistics



Scalpel, swab,

Language

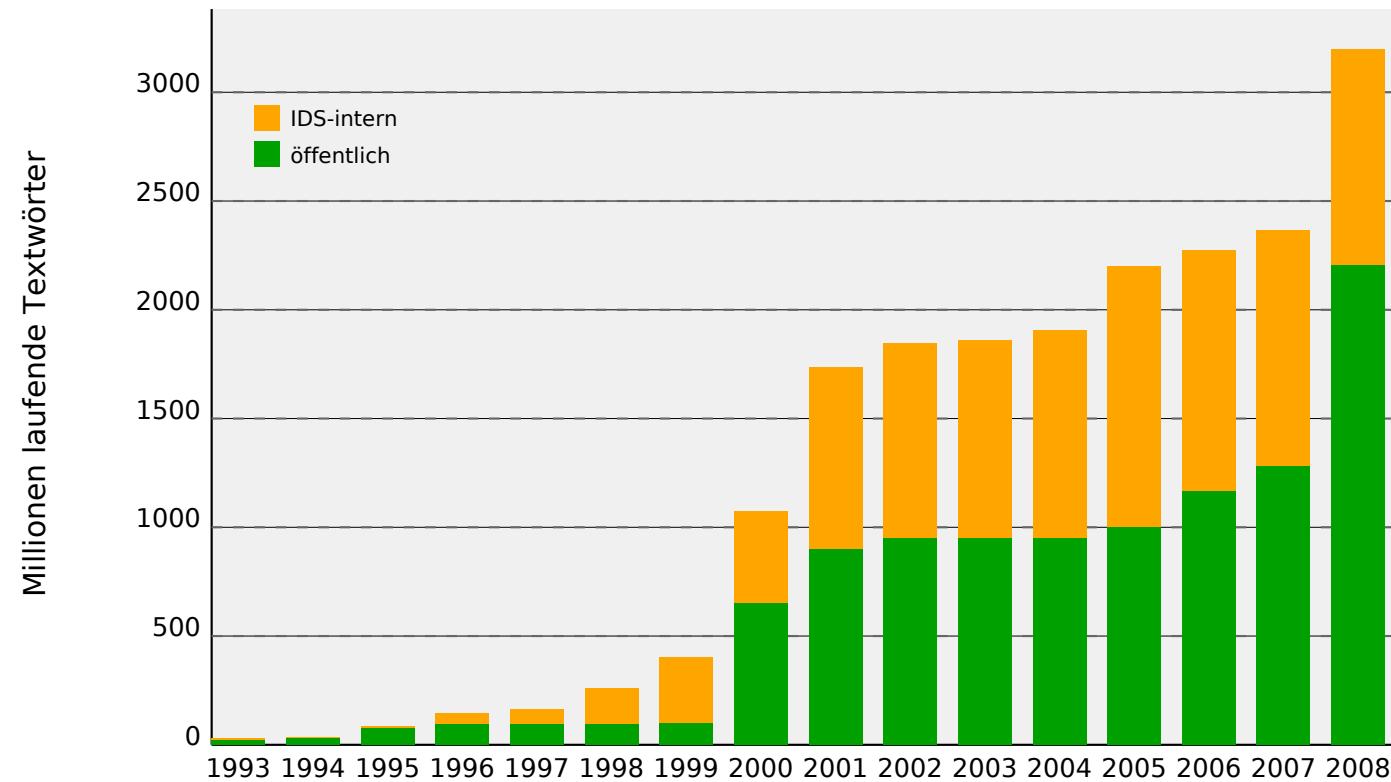
(as anatomy learnt
about the inner
structure of our body)

Huge Corpora: DEREKO

- more than 3 billion words
 - German written texts since 1956
 - belletristic, (popular) scientific texts, texts from newspapers and other text types
 - continuously extended in a controlled way
 - main maximes of extension:
size/scale and variability, legality
-

Huge Corpora: DEREKO

Entwicklung des Gesamtumfangs der Korpora geschriebener Sprache

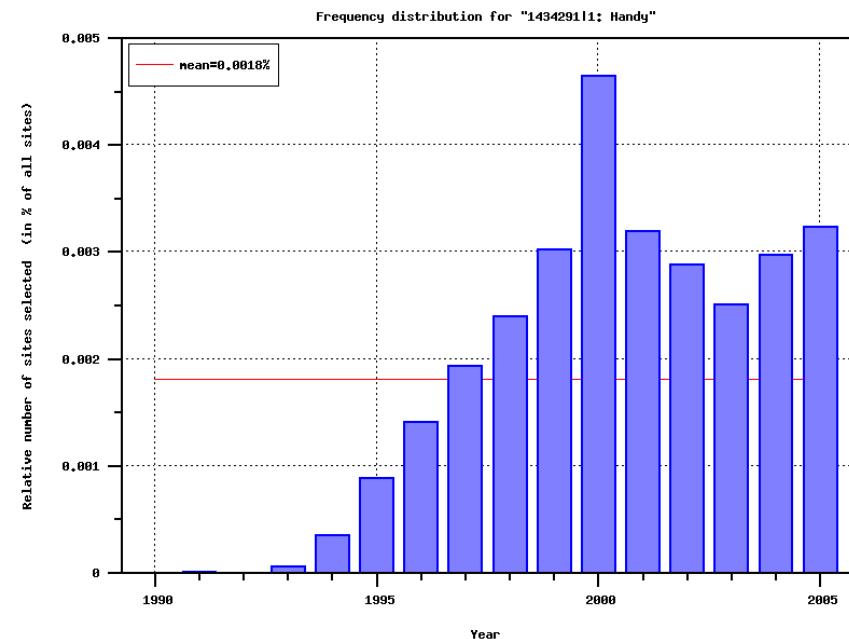


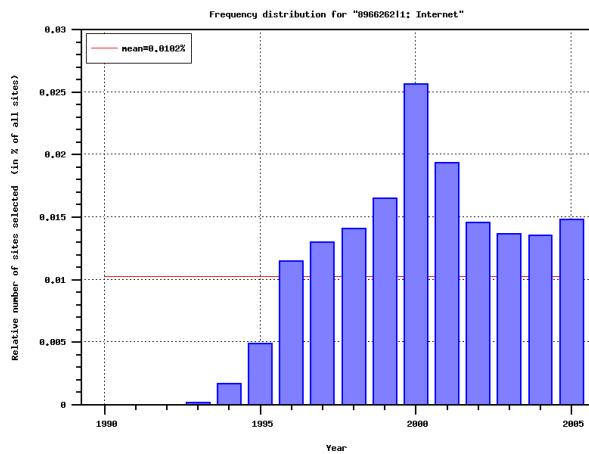
Magnitude of Corpus Size

- results from permanent acquisition with the aim of broad coverage
 - enables (intended) application of statistical data analysis methods (cf. data mining)
 - computational
 - correlational
 - intertwining of data and methods is part of our paradigm
-

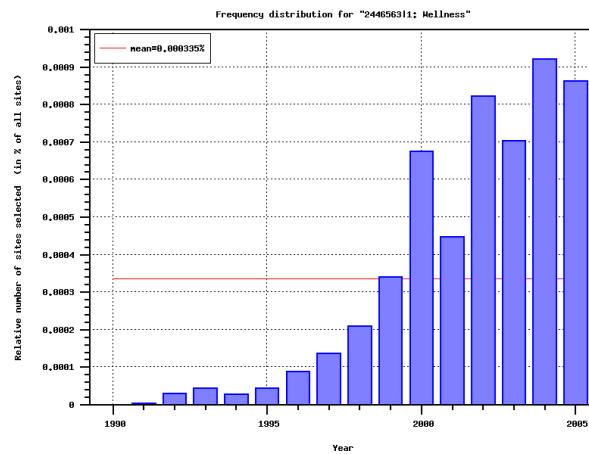
Correlation Analysis

- e.g. dimension time
- occurrences of words in certain time ranges (e.g. interesting for detection of neologisms)

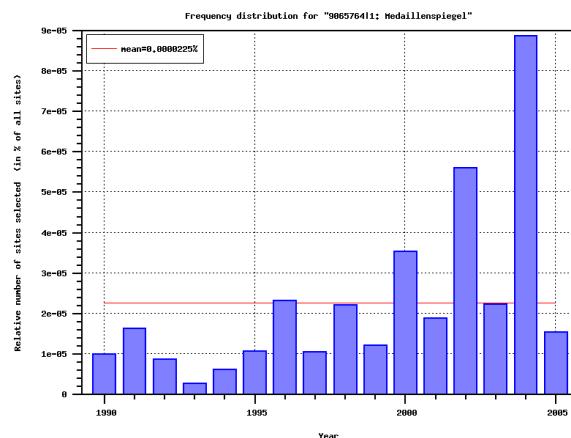




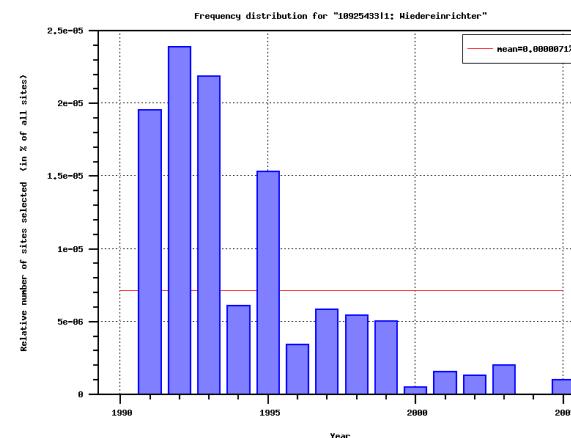
Internet



Wellness



Medaillenspiegel



Wiedereinrichter

What about ‘meaning’?

- up to now no word about ‘meaning’
 - intuitive notion
 - no extensional/referential semantics,
no frame or prototype semantics
 - let’s try to forget any of the known
approaches
-

What about 'meaning'?

- a word has neither one 'meaning' nor many 'meanings'
 - a word in a context has 'a meaning'
(as narrow as possible to let communication work)
 - new contexts might contribute to meaning potential
 - a word without context has a 'meaning potential'
 - that's the way children learn their mother language and even adults learn by and by a more and more clearer 'notion' of formerly unknown terms
-

What about ‘meaning’?

- “Meaning is in the use of a word.”
(Wittgenstein)
 - “You shall know a word by the company it
keeps.” (Firth)
 - many dimensions / nuances of ‘meaning’: connotations, style, rhetorical structure
-

Assumption

- fundamental mechanismn of language
(surely: of its acquisition,
presumably also: of its cognitive processing)
 - relies heavily on tacit knowledge
 - is (partially) functional equivalent to a statistical
assessment of the context behaviour of words in
'language input'
 - cf. mathematics when crossing a road
-

$$s_{auto} = v_{auto} \cdot t_{auto}$$

$$s_{ich} = v_{ich} \cdot t_{ich}$$

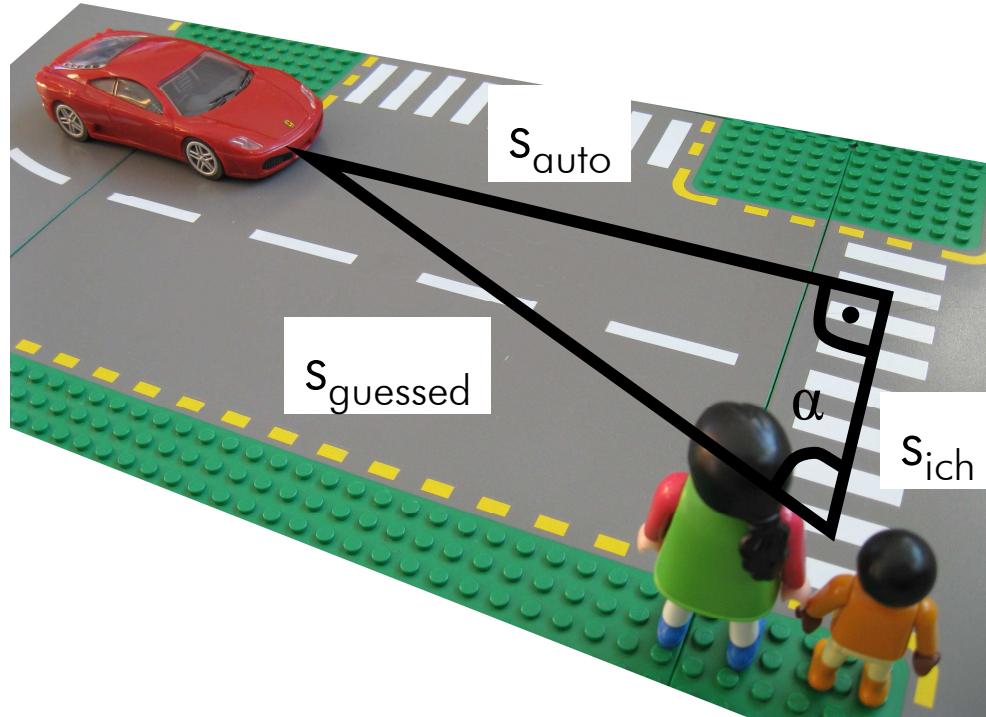
$$\tan(\alpha) = \frac{s_{auto}}{s_{ich}}$$

$$\tan(\alpha) = \frac{v_{auto} \cdot t_{auto}}{v_{ich} \cdot t_{ich}}$$

$$\tan(\alpha) \cdot \frac{v_{ich}}{v_{auto}} = \frac{t_{auto}}{t_{ich}}$$

$$t_{ich} < t_{auto} \Leftrightarrow 1 < \frac{t_{auto}}{t_{ich}}$$

$$1 < \tan(\alpha) \cdot \frac{v_{ich}}{v_{auto}}$$



Free of (Linguistic) Premises

- as few linguistic premises as possible
 - minimal assumption (J. Sinclair)
 - no a-priori-model of language
 - persistent distinction between
 - recorded observation
 - text-external annotation (regional, diachron, ...)
 - postponed intellectual interpretation
 - operationalization of outcomes not necessarily intended
-

Intuitive Notion of 'Meaning (Potential)'

- two lexical items have the same meaning (potential) if they can be interchanged in every context they are used (??)
 - but: trivial contexts, contexts that do not contribute to meaning (potential)
 - dummy contexts (that occur too often)
 - occasional contexts (that occur too irregularly)
 - notion of non-trivial 'typical context' necessary
 - meaning differences/relations correspond to different kinds of overlapping of common (typical) contexts
-

The Idea

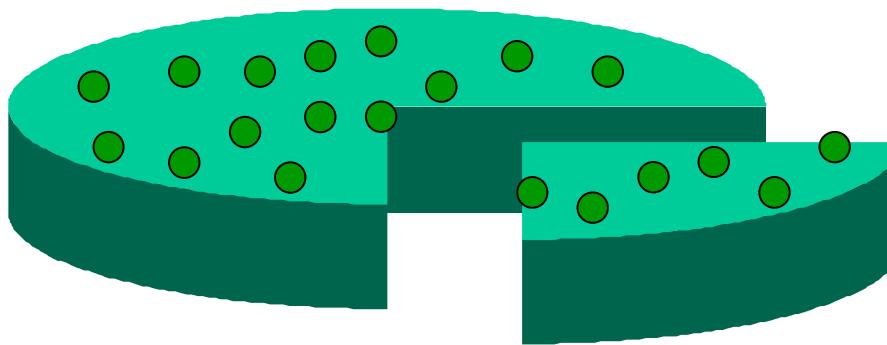
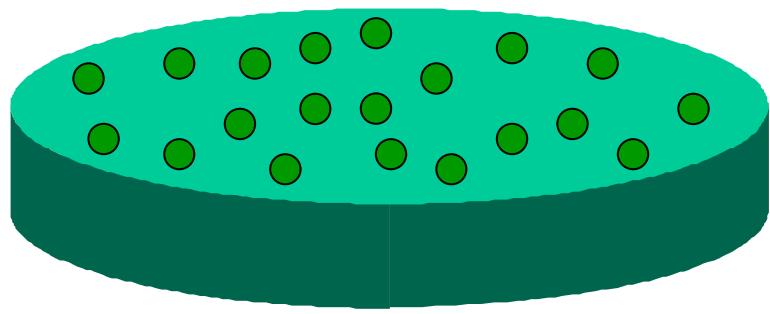
- (how) can we approximate this intuitive notion?
 - first: typical contexts (cover one aspect)
 - collocations
 - second: set of significant typical contexts
(covers most prominent aspects)
 - collocation profile
 - third: investigating the internal structure of the set by contrasting it to other sets
-

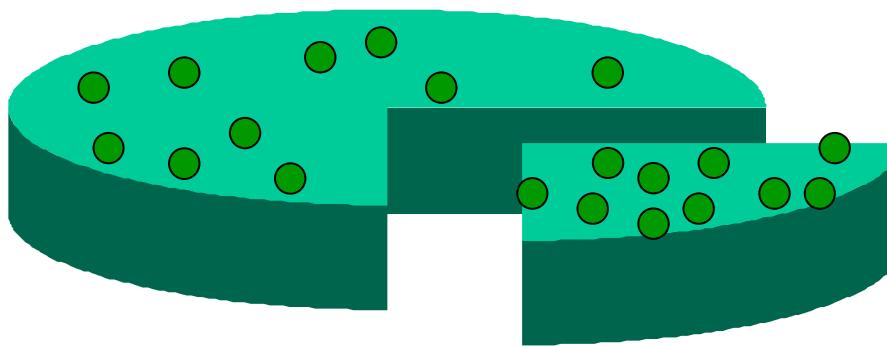
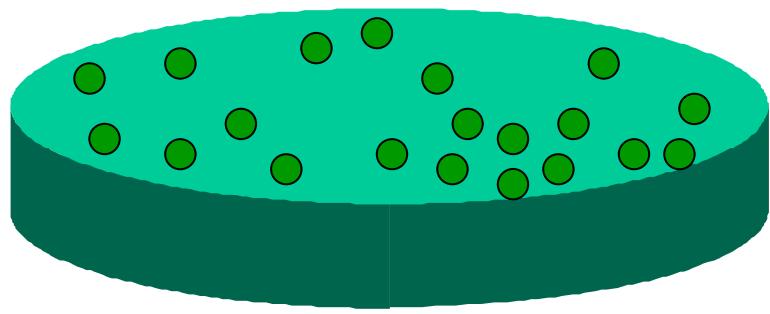
Motivation Reconsidered

- due to focus on „use“
 - empirical foundation necessary
- »*more data is better data*« (R. Mercer)
- thus feasible but also necessary
 - corpus-linguistic (corpus-driven) approach
 - huge corpora
 - sophisticated corpus analysis methods

Correlation Analysis

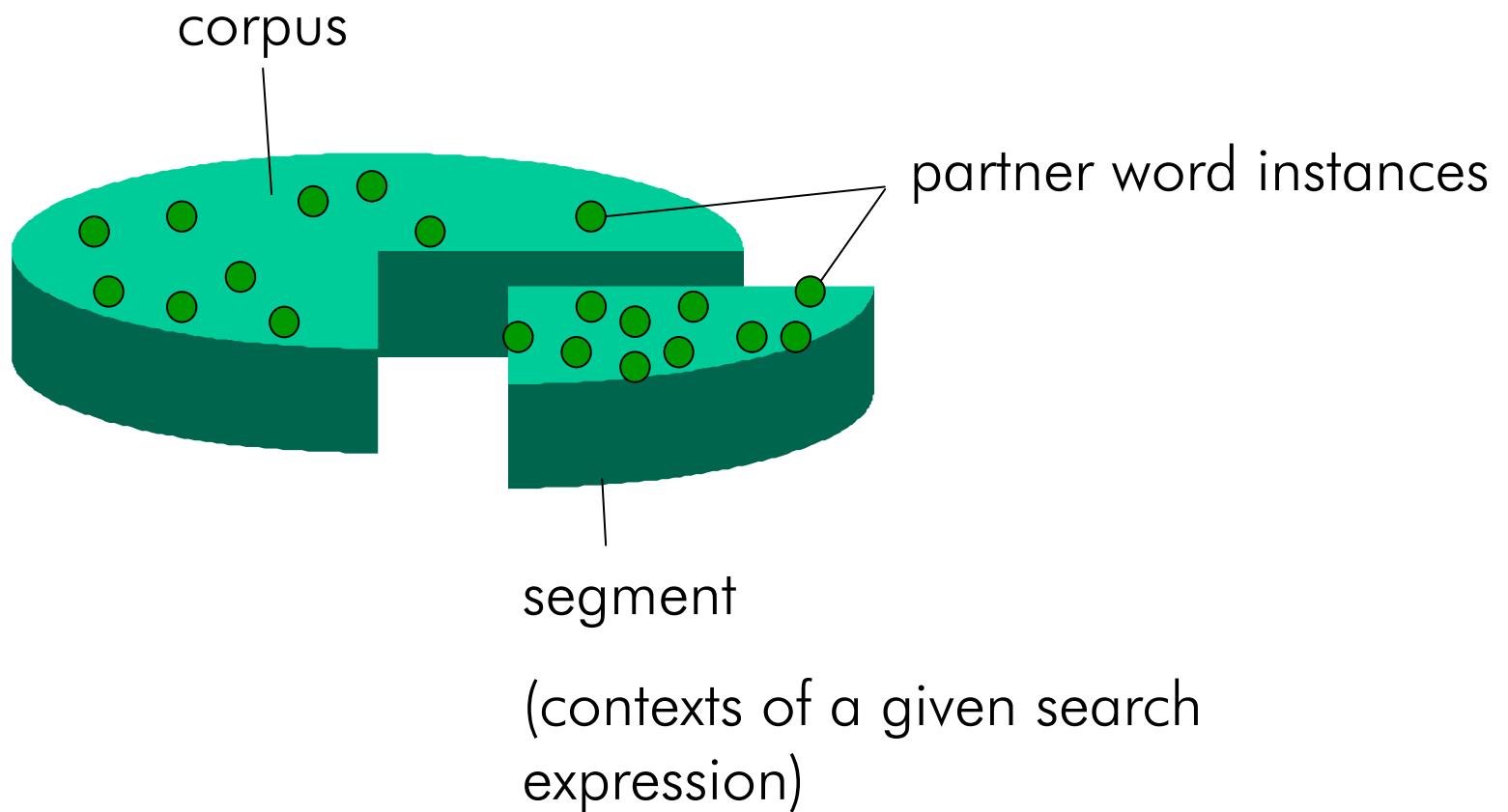
- reveals regularities in (common) occurrences of events
 - analysis indicates potential relations between “potatoes, sauerkraut, eisbein”
(without any knowledge of the German meal or the way to prepare it ...)
-

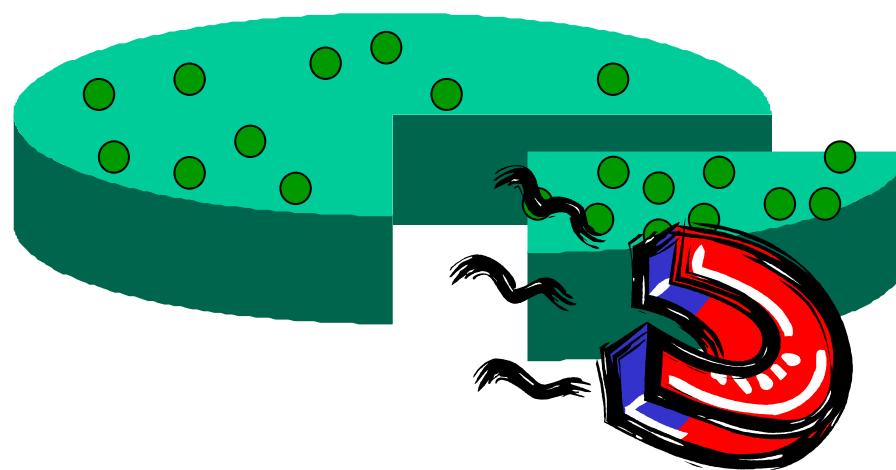




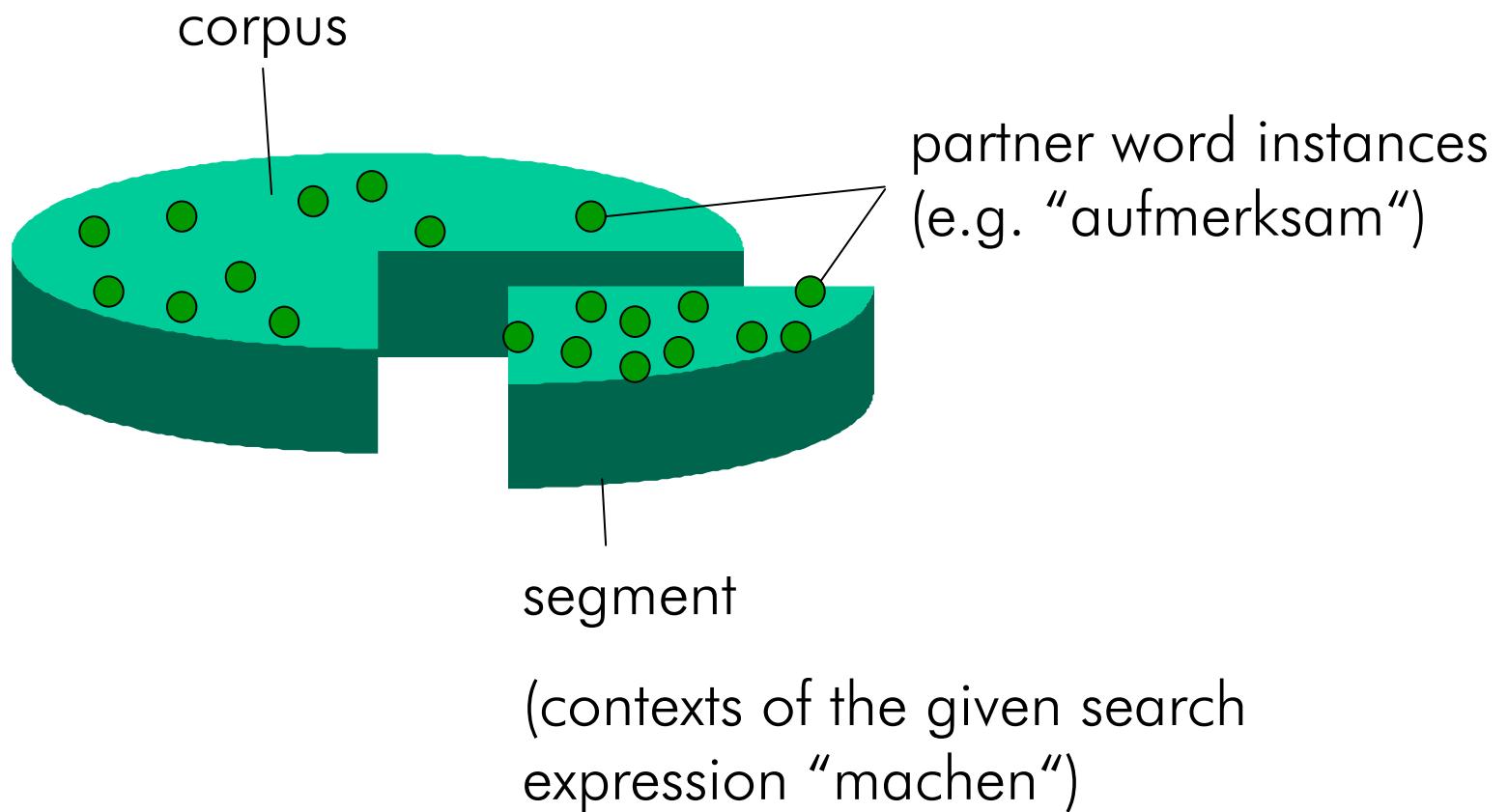
Collocation Analysis

- especially dimension '(lexical) context'
 - reveals *significant regularities* in the *use of word combinations*
 - this version developed by Cyril Belica 1994–1995
 - worldwide online available since 1995
 - can be applied to any virtual corpus
 - currently integrated in COSMAS II
 - developing of further refinements since 2003
-





☒ -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 der
☒ -2 -1 11238 aufmerksam		1617	45% aufmerksam [zu] machen
☒ -1 5 6970 Spaß	richtig viel	11	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	hattent	12	83% hatten [in ...] geltend gemacht daß ...
☒ -2 -1 6051 geltend		810	44% geltend [zu] machen
☒ -2 -1 4188 rückgängig	werden kann wird	11	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	haben	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
☒ -2 -1 3721 Gebrauch	Angebot	25	40% von dem diesem Angebot [...] Gebrauch gemacht
☒ -2 -1 3721 Gebrauch		595	42% Gebrauch [zu] machen
☒ 1 3715 verantwortlich	werden	106	90% verantwortlich [...] gemacht [...] werden
☒ 1 3715 verantwortlich	wird	29	68% verantwortlich gemacht wird
☒ 1 3715 verantwortlich		7	50% verantwortlich [...] gemacht werden



Collocation Analysis

- arbitrary context window size
 - auto focus (positional binding)
 - accuracy, granularity
 - optional lemmatization
 - unique or multiple assigning to clusters
 - optional synsemantic (function) words
 - different statistics
(log-likelihood ratio, mutual information...)
 - syntagmatic patterns
-

Search query

xxxxxxxx xxxxxxxx xxxx xxxx xxxx
xxxxxxxx xxxx xxxx xxxx xxxx xxxx xx
xx xx xxxx xxxx xxxx xxx xx xxxx
xxxxxx xxxx xxxx xxxx xxxx xxxx xx.
xxx xx xxxx xxxx xxxx xxxx xxxx
xxxx. xxxx xxxx xxxx xxxx xx
xxxxxxxxxx.



xxxxxxxx xxxx xxxx xxxx xxxx xxxx
xxxxxx xxxx xxxx xx xx xxxx xxxx
xxxx xx xxxx xxxx xxxx xxxx
xxxxxxxx xxxx xx. xxxx xxxx
xxx xx xxxx xxxx xxxx xxxx xxxx
xxxx. xxxx xxxx xxxx xxxx xx
xxxxxxxxxx.

xxxxxxxx xxxxxxxx xxxx xxxx xxxx
xxxxxxxx xxxx xx xx xxxx xxxx xxxx
xx xx xxxx xxxx xxxx xxxx xxxx
xxxxxxxx xxxx xx. xx xxxx xxxx
xxxx xxxx xxxx xxxx xxxx xxxx. xxxx
xxxx xxxx xxxx xxxx xxxx xxxx xxxx
xxx xxxx.

xxxx xxxx xxxx xxxx xxxx xxxx
xxxx xx xxxx. xxxx xxxx xxxx
xxxxxxxx xx xxxx xxxx xxxx xxxx
xx xx xx xxxx xxxx xxxx xxxx
xxxxxx xxxx xxxx xxxx xxxx xxxx
xxxxxxxx xx. xx xxxx xxxx xxxx
xxxxxxxxxxxx xxxx.

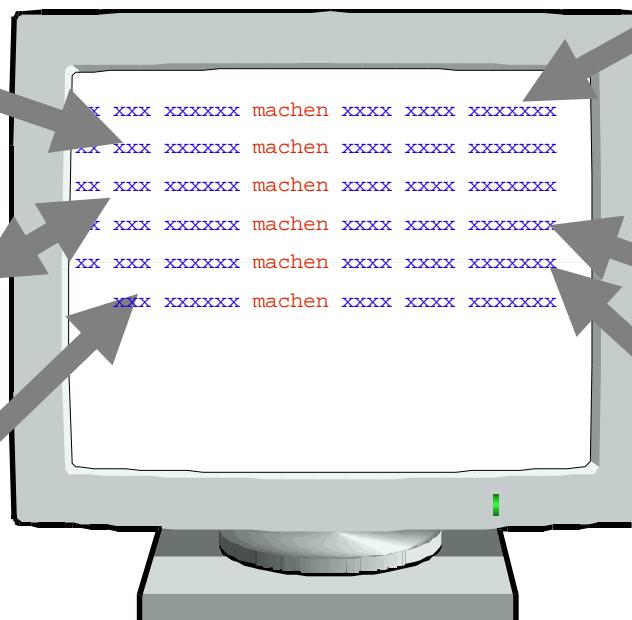
xxx xxxx xxxx xxxx xxxx xxxx
xxxxxxxx xx. xx xxxx xxxx xxxx
xxxxxxxxxx xxxx xxxx. xxxx xxxx
xxxx xx xxxx xxxx xxxx xxxx
xxxxxxxx. xxxx xxxx xxxx xxxx
xxxx xxxx xxxx xxxx xxxx xx xx
xxxx.

Result of search query of “machen”

xxxxxxxx xxxxxxxx xxxx xxxx xxxx
xxxxxxxx xxxx xxxx xxxx xxxx xxxx xx
xxx xx xxxx xxxx xxxx xxxx xxxx xxxx
machen xxxx xxxx xxxx xxxx xxxx xxxx
xxxx. xxxx xxxx xxxx xxxx xxxx xxxx
xxxxxxxxxx.

xxxxxxxx xxxxxxxx xxxx xxxx xxxx
xxxxxxxx xxxx xxxx xxxx xxxx xxxx xx
xxx xxxx xxxx xxxx xxxx xxxx xxxx
machen
xxxx xxxx xxxx xxxx xxxx xx. xxx xxxx
xxxx xxxx xxxx xxxx xxxx xxxx.
xxxx xxxx xxxx xxxx xxxx xxxx.

xxxxxxxx xxx xxxx xxxx xxxx xxxx xxxx
xxxx xxxx xxxx xxxx xx xxxx xxxx xxxx
xxxx xxxx xxxx xxxx
xxxxxxxx xxxx xxxx xxxx xxxx xxxx
xxxx xxxx xxxx xxxx xxxx xxxx
xxxx. xxxx xxxx xxxx xxxx xxxx
xxxxxxxxxx.



xxxxxxxx xxxxxxxx xxxx xxxx xxxx
xxxxxxxx xxxx xx xxx xx xxxx xxxx xxxx
xxx xx xxxx xxxx **machen** xxxx xxxx
xxxxxxxx xxxx xx. xxx xxxx xxxx xxxx
xxxx xxxx xxxx xxxx xxxx xxxx. xxxx
xxxx xxxx xxxx xxxx xxxx xxxx xxxx
xxxx xxxx xxxx.

xxxxxxxx xxxxxxxx xxxx xxxx xxxx
xxxxxxxx xxxx xxxx xxxx xxxx xxxx
xxxxxxxx xxxx xxxx xxxx xxxx xxxx
xx xxxx xx xxxx xxxx xxxx xxxx
machen xxxx xxxx xxxx xxxx
xxxxxxxx xx. xxx xxxx xxxx xxxx
xxxxxxxxxxxx xxxx.

xxx xxxx xxxx **machen** xxxx xxxx xxxx
xxxxxxxx xx. xxx xxxx xxxx xxxx
xxxxxxxxxxxx xxxx xxxx. xxxx xxxx
xxxx xxxx xxxx xxxx xxxx xxxx
xxxxxxxx. xxxx xxxx xxxx xxxx xxxx
xxxx xxxx xxxx xxxx xxxx xx xxxx xx
xxxxxx.

xxxxxxxx xxxxxxxx xxxx xxxx xxxx
xxxxxxxx xxxx xxxx xxxx xxxx xxxx xx
xxx xx xxxx xxxx xxxx xxxx xxxx xx
machen xxxx xxxx xxxx xxxx xxxx xxxx
xxx xxxx xxxx xxxx xxxx xxxx xxxx xxxx
xxxxxx. xxxx xxxx xxxx xxxx xxxx xxxx
xxxxxxxxxx.

xxxxxxxx xxxx xxxx xxxx xxxx xxxx xxxx
xxxxxxxx xxxx xxxx xxxx xxxx xxxx xxxx
xxx xxxx xxxx xxxx xxxx xxxx xxxx xxxx
xxx xxxx xxxx xxxx xxxx xxxx xxxx xxxx
xxxxxx. xxxx xxxx xxxx xxxx xxxx xxxx
xxxxxx.

xxxxxxxx xxxx xxxx xxxx xxxx xxxx
xxxxxxxx xxxx xxxx xxxx xxxx xxxx
xxx xxxx xxxx xxxx xxxx xxxx xxxx
xxx xxxx xxxx xxxx xxxx xxxx xxxx
xxxxxx. xxxx xxxx xxxx xxxx xxxx xxxx
xxxxxx.

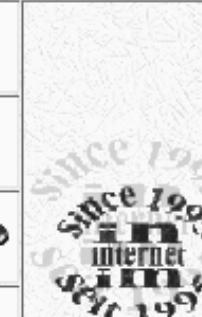
COSMAS-Kollokationsanalyse

Analyseparameter

zu analysierender Kontext	<input checked="" type="checkbox"/> JS	Wörter links
	<input checked="" type="checkbox"/> JS	Wörter rechts
	<input type="checkbox"/>	höchstens 1 Satz <input checked="" type="radio"/>
	<input type="checkbox"/>	Autofokus <input checked="" type="radio"/>
Granularität <input checked="" type="radio"/>	<input checked="" type="checkbox"/> grob <input type="checkbox"/> <input checked="" type="radio"/>	
Zuverlässigkeit <input checked="" type="radio"/>	<input checked="" type="checkbox"/> hoch <input type="checkbox"/> <input checked="" type="radio"/>	
Clusterzuordnung <input checked="" type="radio"/>	<input checked="" type="checkbox"/> eindeutig <input type="checkbox"/> <input checked="" type="radio"/>	
Lemmatisierung <input checked="" type="radio"/>	<input checked="" type="checkbox"/> verwenden	
Funktionswörter <input checked="" type="radio"/>	<input checked="" type="checkbox"/> ignorieren	
Gamma-Wert <input checked="" type="radio"/>	<input checked="" type="checkbox"/> anzeigen	

Alle Parameter auf Standardwerte [zurücksetzen](#)

Analyse starten



xxxxxxxx xxxxxxxx xxxx xxxx xxxx
xxxx xx xxx xx xxxx xxxx xxxx
xxxxxxxx **machen** xxxx xxxx
xxxxxx xx. xxx xxxx xxxx
xxxxxx xxxx xxxx xxxx xxxx xxxx
xxx xxxx xxxx xxxx xxxx xxxx xxxx
xxxxxx.

xxxxxx xxxx xxxx xxxx xxxx xxxx
xxxxxx xxxx xxxx xxxx xxxx xxxx
xxx xxxx xxxx xxxx xxxx xxxx xxxx
xx xxxx xxxx xxxx xxxx xxxx
machen xxxx xxxx xxxx xxxx
xx. xxx xxxx xxxx xxxx
xx xxxx xxxx xxxx xxxx.

xx **machen** xxxx xxxx xxxx
xx. xxx xxxx xxxx xxxx
xx xxxx xxxx. xxxx xxxx
xx xxxx xxxx xxxx xxxx xxxx
xxx. xxxx xxxx xxxx xxxx xxxx
xxxxxxxxxxxx xxxx xx xxxx xx

xxxxxxxx xxxx xxxx xxxx xxx xxxx xxxx xxxx xxxx xxxx xxxx xxxx xxxx
xxxx xx xxx xx xxxx **xxxx xxxx xxx xxx xxxx machen** **xxxx xxxx xxxx xxxx**
xxxxxx xx. xxx xxx xxxx xxxx xxxxxxxxx xxxx xxxx. xxxx xxxx
xxxxx xxx xxxxxxxxxx.

- algorithm counts the occurrences of every (blue coloured) word in the defined context of the hit of the search query
- the calculated observed frequency is compared to the expected frequency (according to the random probability due the distribution of the words in the language)
[with known frequency in the complete data set and this size of the segment]

M01 Der Fraktionschef macht außerdem darauf aufmerksam, dass a
M01 bot in der Innenstadt aufmerksam zu machen. Besondere Bedeutung messen die E
M02 lt auf dieses Versäumnis aufmerksam gemacht. Warum haben Sie bis heute nicht
M03 en im Galopprennsport aufmerksam zu machen. Sein Ziel ist bis heute aktuell:
M03stärkt. Im Umkreis von 20 Kilometern machen Plakatwände auf die Messe aufmerks
M04 brennendes Tuch in den Auspuff und machen auf die angebliche Panne aufmerks
R97 für viele Hattersheimer Bürger: Sie machten SPD-Stadtverordnete darauf aufme
R97 u+ Tennis-Cracks der Eintracht machen auf sich aufmerksam / Multifunktions
R97 li-cher Verschuldung aufmerksam" zu machen. Die rote IGM- Flagge im Wind, ei
R98 wenn wir sie auf Objekte aufmerksam machen, die sie möglicherweise noch nich
R98 nt Horst Wolff. Der Vize-Amtsleiter macht darauf aufmerksam, daß das Abladen
R99 der Yanomami-Indianer aufmerksam zu machen. Die Expo-Verantwortlichen wollte
R99 gung der Fahrbahn darauf aufmerksam gemacht, teilte die Stadt mit. prsb
R99 insam auf ihre Situation aufmerksam machen. Unter dem Motto "Arbeit für die
A97 die Verkehrsberuhigungen aufmerksam gemacht. Neu wird die Herrenhofstrasse d
A97 rken Leistungen auf sich aufmerksam gemacht», sagt Müller, «darum ist es kla
A98 der Haltung eines Hundes aufmerksam gemacht werden, weil aus der Nachbarscha
A98 a Schmid auf ihr Angebot aufmerksam macht. Die Kurse Schwangerschaftsgymnast
A98 auch auf Schwierigkeiten aufmerksam gemacht. Viel zu schnell war der interes
A99 auf die drohende Gefahr aufmerksam machen, innerhalb von acht Tagen bereits
A99 auf das Bild im «Blick» aufmerksam gemacht worden. Beim Abbrechen habe ich
A99 ern für diese Grabkerzen aufmerksam gemacht: Glasbehälter für Grabkerzen sin
A00 usbildung im Rheintal aufmerksam zu machen.
A00 it, auf ihre Angebote aufmerksam zu machen.

A01 ünstige Gelegenheiten aufmerksam zu machen. Und als die Jagd die Vorfahren d
A01 sweise werden sie darauf aufmerksam gemacht, dass unter anderem mit bewusste
E96 usikalischen Fähigkeiten aufmerksam gemacht, sagt ihre Mutter, eine gebürtig
E97 wurde freundlich darauf aufmerksam gemacht, dass an diesem Tag das Parlamen

E99 Amokläufers Eric Harris aufmerksam gemacht. +L

E99 uf sein Fehlverhalten aufmerksam zu machen.

I96 n recht, wenn sie darauf aufmerksam machen, daß mit dem Erwerb eines Zeugnis

I97 h Spaziergänger auf sich aufmerksam machen. Zwei Personen stiegen zu dem Ver

I98 Bedeutung der Prophylaxe aufmerksam machen. Auch in Innsbruck bieten Tirols

I00 u L+ Mit einer spektakulären Aktion macht das Rote Kreuz Schwaz aufmerksam,

K96 beiging, auf den Vorfall aufmerksam gemacht. Als der Täter den Beobachter be

K97 rmals auf diesen Umstand aufmerksam machte.

K98 iive Folgen für die Natur aufmerksam gemacht werden soll.

Der	
Fraktionschef	
außerdem	
darauf	
aufmerksam	
in	
der	
Innenstadt	
auf	
dieses	
Versäumnis	
zu	
von	
den	
Auspuff	
Sie	
...	

K98 ssesprecher der Technik Siemens AG, **machte** darauf aufmerksam, daß die Entwick

K99 sgesetzte Zeit" auf sich aufmerksam **machte**, zeichnet die Silhouette der beid

K00 cksal der Obdachlosen aufmerksam zu **machen**.

L99 ertung von Tierversuchen aufmerksam **machen**. Denn während der millionenfache

Z95 ewachsenen Rechtsordnung aufmerksam **machen?** Oder jene eifrigen Pioniere, die

Z00 partistischer Herrschaft aufmerksam **gemacht**: "Bonaparte möchte als der patri

B98 ie Folgen einer Diktatur aufmerksam **gemacht** hat, gilt an diesem Nachmittag i

B99 Um auf das Reisebüro aufmerksam zu **machen**, haben die Mitarbeiter Plakate au

B00 die vermisste Nadine M. aufmerksam **gemacht**", sagte der Polizeisprecher. Fam

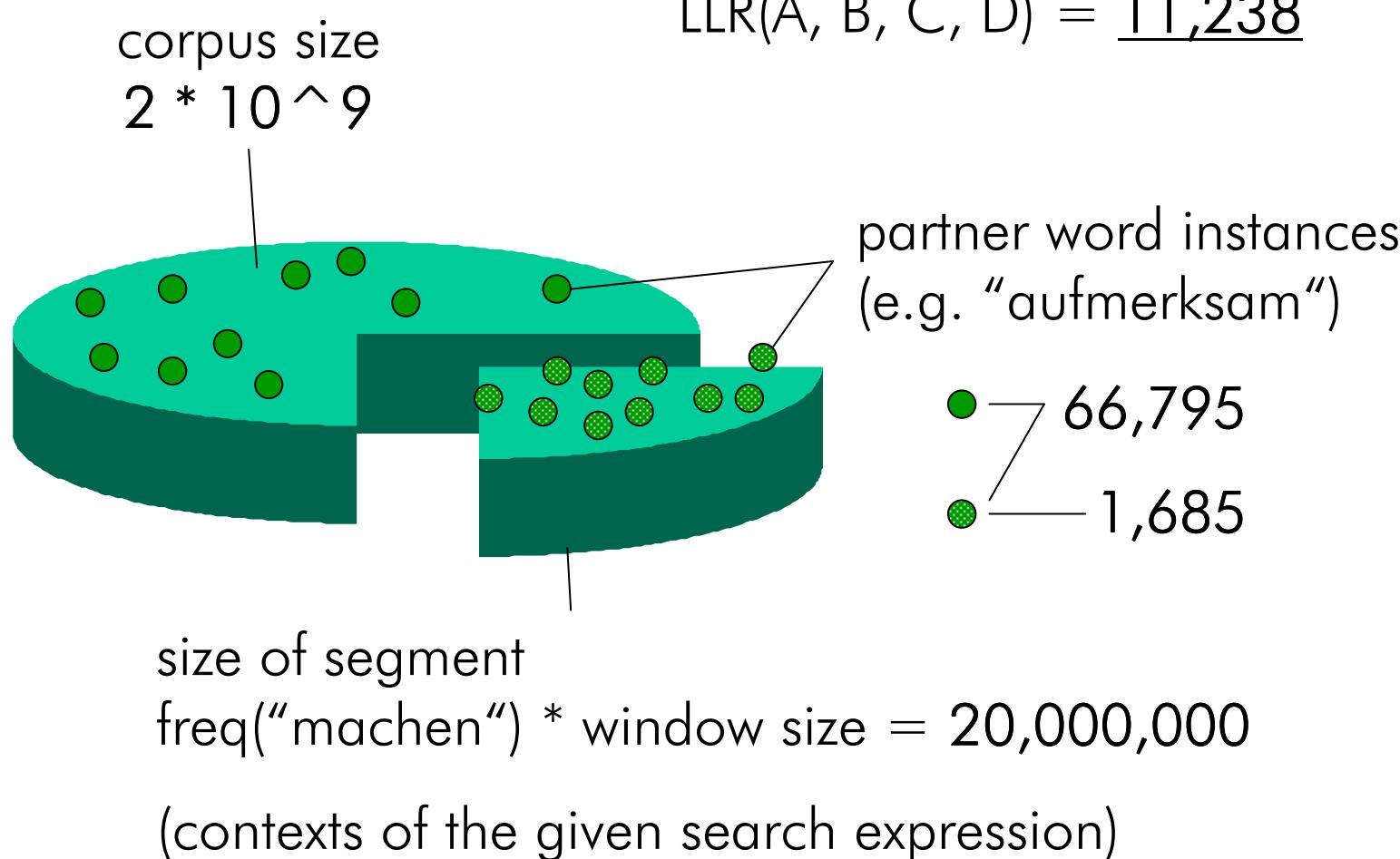
B01 n weiter auf das Problem aufmerksam **machen**. "Ich tret auch mal gern auf und

B01 ken bei deren Verwendung aufmerksam **gemacht**. Die meisten Abbeizmittel schnit

B03 affigen Gesten auf sich aufmerksam **machen** will, schmilzt der evolutionäre V

B04 hicksal seines Hauses aufmerksam zu **machen**. Er verlässt es nicht gern, denn

T88 and herrschende Armut aufmerksam zu **machen** und um Spenden zu bitten.

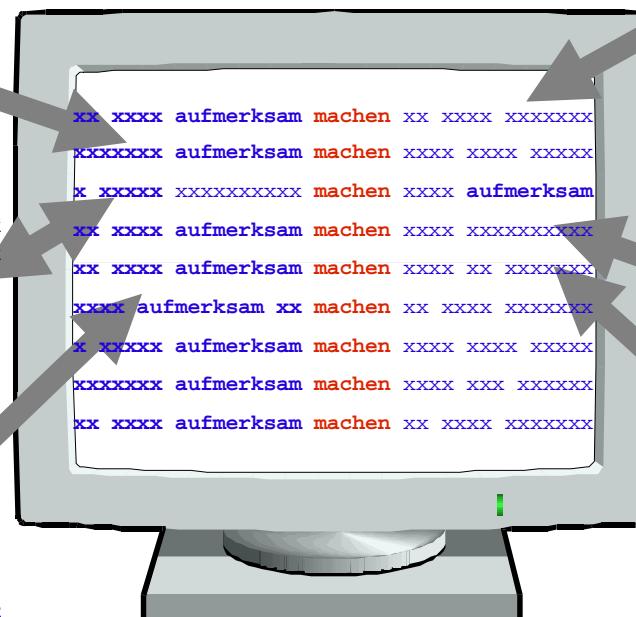


Collocation Analysis “in a nutshell”

xxxxxxxx xxxxxxxx xxxx xxxx xxxx
xxxxxxxx xxxx xxxx xxxx xxxx xxxx xx
xxx xx xxxx **aufmerksam machen** xxxx xxxx xxxx
aufmerksam xxxx. xxxx xxxx xxxx
xxxxxx xxxx xxxx. xxxx xxxx
xxxx xxxx xxxx.

xxxxxxxx xxxx xxxx xxxx xxxx xxxx
xxxxxxxx xxxx xxxx xxxx xxxx xxxx xx
xxx xxxx xxxx xxxx xxxx xxxx
machen xxxx xxxx **aufmerksam** xxxx
aufmerksam xxxx xxxx xxxx xxxx
xxxx xxxx. xxxx xxxx xxxx xxxx
xxxx xxxx. xxxx xxxx xxxx xxxx
xxxxxx.

xxxxxx xxxx xxxx xxxx xxxx xxxx
xxxxx xxxx xxxx xx **aufmerksam machen** xxxx
aufmerksam xxxx xxxx xxxx xxxx
xxxxxx xxxx xxxx. xxxx xxxx
xxxx. xxxx xxxx xxxx xxxx
xxxxxx.



xxxxxxxx xxxxxxxx xxxx xxxx xxxx
xxxxxxxx xxxx xx xxx xx xxxx **aufmerksam machen**
aufmerksam xxxx xxxx xx. xxxx xxxx
xxxxxxxx xxxx xxxx xxxx xxxx xxxx.
xxxxxx xxxx xxxx xxxx xxxx xxxx
xxxx xxxx xxxx.

xxxxxx xxxx xxxx xxxx xxxx xxxx
xxxxxxxx xxxx xxxx xxxx xxxx xxxx
xxxxxx xxxx xxxx xxxx xxxx xxxx
xx xxxx xx xxxx **aufmerksam machen** xxxx xxxx xxxx
aufmerksam machen xxxx xxxx xxxx
xxxxxx xxxx. xxxx xxxx xxxx
xxxxxx xxxx xxxx.

xxxxx **aufmerksam machen** xxxx xxxx
xxxxxxxx xxxx xxxx xx. xxxx xxxx
xxxx xxxx xxxx xxxx xxxx. xxxx
xxxx xxxx xxxx xxxx xxxx xxxx
xxx xxxx xxxx xxxx. xxxx xxxx xxxx
xxxx xxxx xxxx xxxx xxxx xxxx xx xxxx
xx xxxx.

Collocation Profile

- captures a lot of aspects, nuances of the word's use, i.e. characteristics of the discourse, in which its 'meaning' has grown/was established
 - characterizes the target expression, here: a word
-

<input checked="" type="checkbox"/> -2-1 11238 aufmerksam worden	31	96% darauf aufmerksam gemacht [...] worden
<input checked="" type="checkbox"/> -2-1 11238 aufmerksam wollte	17	52% aufmerksam machen wollte
<input checked="" type="checkbox"/> -2-1 11238 aufmerksam Öffentlichkeit	12	66% die Öffentlichkeit [darauf] auf aufmerksam [zu] machen daß der
<input checked="" type="checkbox"/> -2-1 11238 aufmerksam	1617	45% aufmerksam [zu] machen

Sigle

KWIC

P95 Tiroler Landeshauptstadt **aufmerksam gemacht**. Schließlich wird dagegen protestiert Patienten allerdings **aufmerksam machen**", erwähnte Pöldinger, "daß es sich tmeyer, habe auch darauf **aufmerksam gemacht**, daß es bedenklich sei, wenn die rieren wie vor 30 Jahren **aufmerksam gemacht** that.

THM en auf unser Institut **aufmerksam zu machen**, am äußeren Hause anzubringen se F93 Die Katastrophe **machte** weltweit auf eine Gemäldegalerie F95 ter Versuch, auf sich **aufmerksam zu machen** und eineverständnisvollere Behan Z95 ewachsenen Rechtsordnung **aufmerksam machen?** Oder jene eifrigen Pioniere, die Z00 partistischer Herrschaft **aufmerksam gemacht**: "Bonaparte möchte als der patri B98 ie Folgen einer Diktatur **aufmerksam gemacht** hat, gilt an diesem Nachmittag im B99 Um auf das Reisebüro **aufmerksam zu machen**, haben die Mitarbeiter Plakate auf B00 die vermisste Nadine M. **aufmerksam gemacht**", sagte der Polizeisprecher. Fann B01 n weiter auf das Problem **aufmerksam machen**. "Ich tret auch mal gern auf und B01 ken bei deren Verwendung **aufmerksam gemacht**. Die meisten Abbeizmittel schnit B03 affigen Gesten auf sich **aufmerksam machen** will, schmilzt der evolutionäre V B04 hicksal seines Hauses **aufmerksam zu machen**. Er verlässt es nicht gern, denn T88 and herrschende Armut **aufmerksam zu machen** und um Spenden zu bitten.

T90 elegung des Frauenhauses **aufmerksam gemacht** worden. Im November hatte die Üb T90 si meint es gut: Er will **aufmerksam machen** auf die sozialen Probleme in Sisi T91 an und fand schon auf diesen Gehindernis nicht an Studentenaktionen das Tandem an

<input checked="" type="checkbox"/> -2-1 11238 aufmerksam worden	31	96% darauf aufmerksam gemacht [...] worden
<input checked="" type="checkbox"/> -2-1 11238 aufmerksam wollte	17	52% aufmerksam machen wollte
<input checked="" type="checkbox"/> -2-1 11238 aufmerksam Öffentlichkeit	12	66% die Öffentlichkeit [darauf] auf aufmerksam [zu] machen daß der
<input checked="" type="checkbox"/> -2-1 11238 aufmerksam	1617	45% aufmerksam [zu] machen

Sigle

KWIC

T98 llegen hätten ihn darauf **aufmerksam gemacht**, "aber denn dachte ich mir auch:

T99 flötenorchester auf sich **aufmerksam machen**.

T00 gewaltpräventiven Arbeit **aufmerksam gemacht** werden. "Gewaltprävention ist e

MK1 Siegel : Ochse und Krone. der Baron **machte** Bienkopp darauf **aufmerksam** , daß

GOE , durch welche das Volk **aufmerksam gemacht** wird , daß es Große in der Welt

H87 g und die " neue Armut " **aufmerksam machen** sollten , erregten bei den Bundes

M97 ulag" in der Sowjetunion **aufmerksam machte**, ist gestorben. In den 50er Jahre

M98 ntensiven Landwirtschaft **aufmerksam machen**. "Großflächige Felder ohne Hecke

M99 n Mann auf das Geschehen **aufmerksam gemacht**. Dieser Zeuge habe sich leider n

M00 resse und Mundpropaganda **aufmerksam gemacht** hatten, keiner kam, wollen sie i

M01 Der Fraktionschef **macht** außerdem darauf **aufmerksam**, dass a

M01 bot in der Innenstadt **aufmerksam zu machen**. Besondere Bedeutung messen die E

M02 lt auf dieses Versäumnis **aufmerksam gemacht**. Warum haben Sie bis heute nicht

M03 en im Galopprennsport **aufmerksam zu machen**. Sein Ziel ist bis heute aktuell:

M03 tärkt. Im Umkreis von 20 Kilometern **machen** Plakatwände auf die Messe aufmerk

M04 brennendes Tuch in den Auspuff und **machen** auf die angebliche Panne aufmerks

R97 für viele Hattersheimer Bürger: Sie **machten** SPD-Stadtverordnete darauf aufme

R97 u+ Tennis-Cracks der Eintracht **machen** auf sich **aufmerksam / Multifunktions**

R97 li-cher Verschuldung **aufmerksam**" zu **machen**. Die rote IGM- Flagge im Wind, ei

R98 wenn wir sie auf Objekte **aufmerksam machen**, die sie möglicherweise noch nich

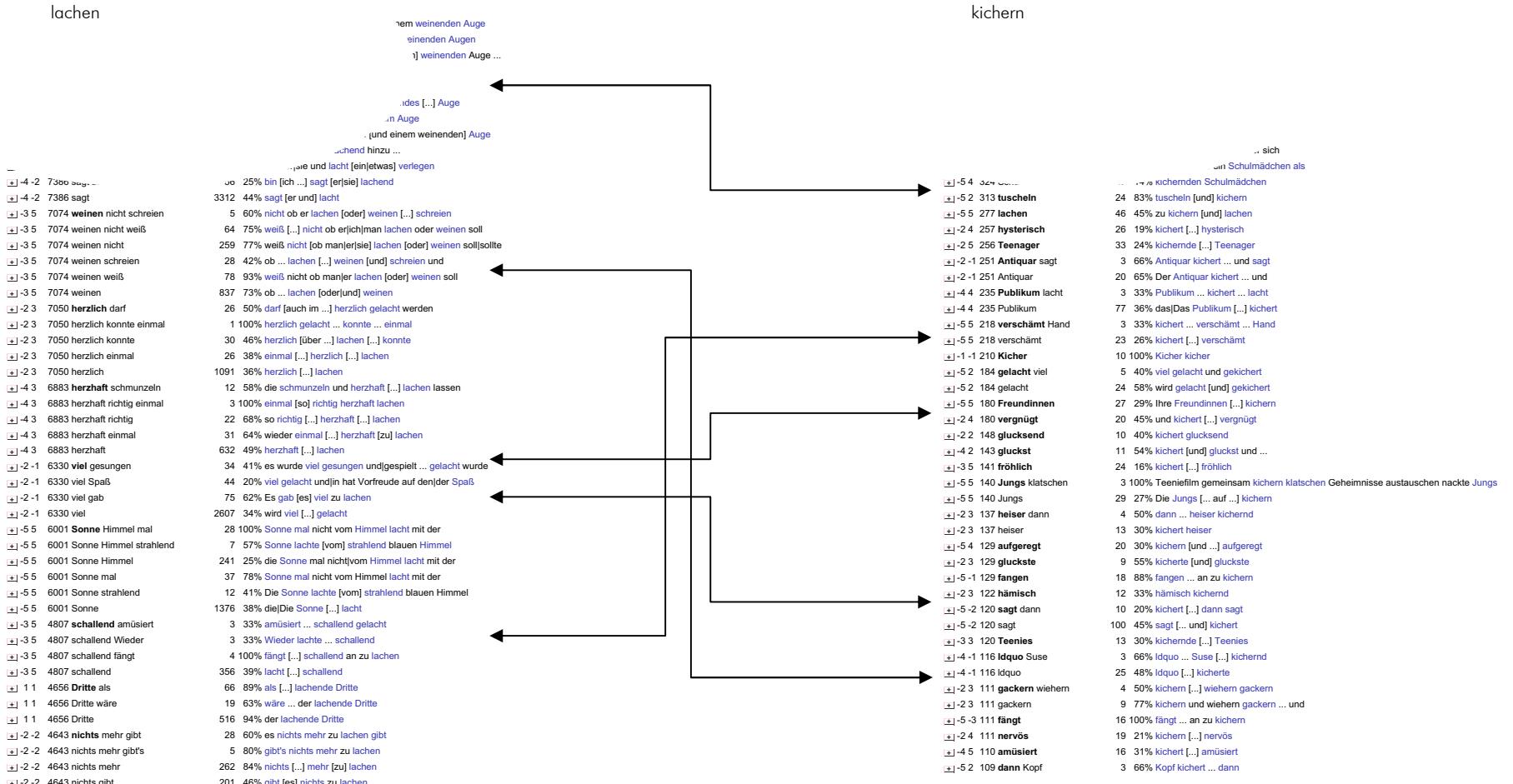
CCDB

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

- investigative and experimental platform
 - approx. 220.000 words (w and w/o synsemantic words)
 - analyses results (collocation profiles) become subject of further investigation
 - explorativ
 - algorithmic
 - e.g. local grammars derived from schemata as generalized syntagmatic patterns
 - intrarelational
 - interrelational
-

Comparison of Collocation Profiles

- relations between profiles (as word characterizations) indicate relations between words
 - if two profiles have many items in common then a lot of usage contexts are comparable
 - similar profiles indicate similarities in usage aspects (e.g. in ‘meaning potential’)
-



Analysewort: ***lachen***, Analysetyp 0

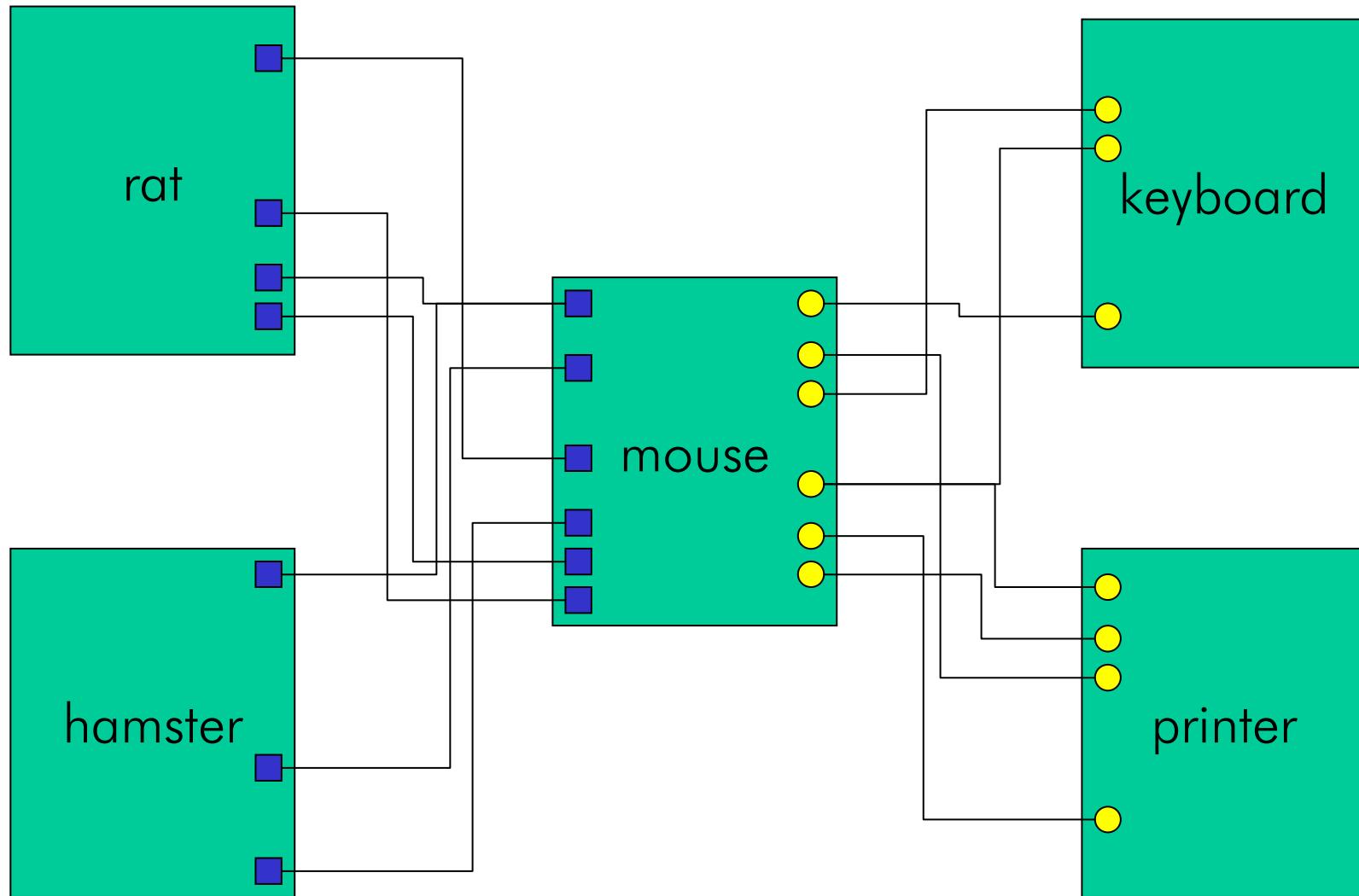
- ± 3 4 17055 **weinenden Auge**
- ± 3 4 17055 weinenden Augen
- ± 3 4 17055 weinenden
- ± 2 5 11830 **Fäustchen** können
- ± 2 5 11830 Fäustchen
- ± 1 4 10182 **Auge** weinendes
- ± 1 4 10182 Auge weinendem
- ± 1 4 10182 Auge
- ± 4 -2 7386 **sagt** fügt
- ± 4 -2 7386 sagt verlegen
- ± 4 -2 7386 sagt bin
- ± 4 -2 7386 sagt
- ± 3 5 7074 **weinen** nicht schreien
- ± 3 5 7074 weinen nicht weiß
- ± 3 5 7074 weinen nicht
- ± 3 5 7074 weinen schreien
- ± 3 5 7074 weinen weiß
- ± 3 5 7074 weinen
- ± 2 3 7050 **herzlich** darf
- ± 2 3 7050 herzlich konnte einmal
- ± 2 3 7050 herzlich konnte
- ± 2 3 7050 herzlich einmal
- ± 2 3 7050 herzlich
- ± 4 3 6883 **herhaft** schmunzeln
- ± 4 3 6883 herhaft richtig einmal
- ± 4 3 6883 herhaft richtig
- ± 4 3 6883 herhaft einmal
- ± 4 3 6883 herhaft
- ± 2 -1 6330 viel gesungen
- ± 2 -1 6330 viel Spaß
- ± 2 -1 6330 viel gab
- ± 2 -1 6330 viel
- ± 5 5 6001 **Sonne** Himmel mal
- ± 5 5 6001 Sonne Himmel strahlend
- ± 5 5 6001 Sonne Himmel
- ± 5 5 6001 Sonne mal
- ± 5 5 6001 Sonne strahlend
- ± 5 5 6001 Sonne
- ± 3 5 4807 **schallend** amüsiert
- ± 3 5 4807 schallend Wieder
- ± 3 5 4807 schallend fängt
- ± 3 5 4807 schallend
- ± 1 1 4656 **Dritte** als
- ± 1 1 4656 Dritte wäre
- ± 1 1 4656 Dritte
- ± 2 -2 4643 **nichts** mehr gibt
- ± 2 -2 4643 nichts mehr gib's
- ± 2 -2 4643 nichts mehr
- ± 2 -2 4643 nichts gibt

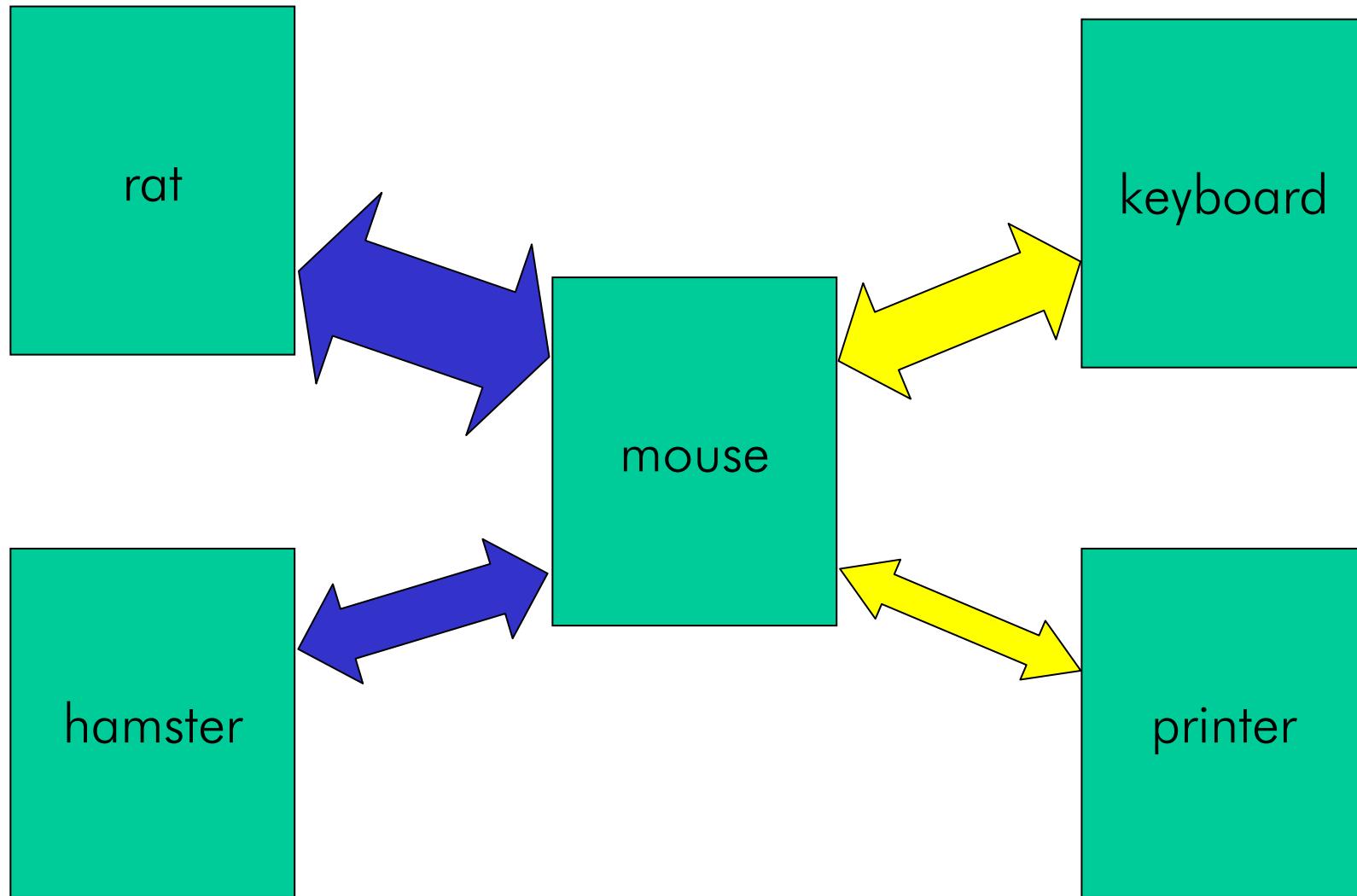
profiles similar
to
lachen
ordered wrt.
similarity meas.

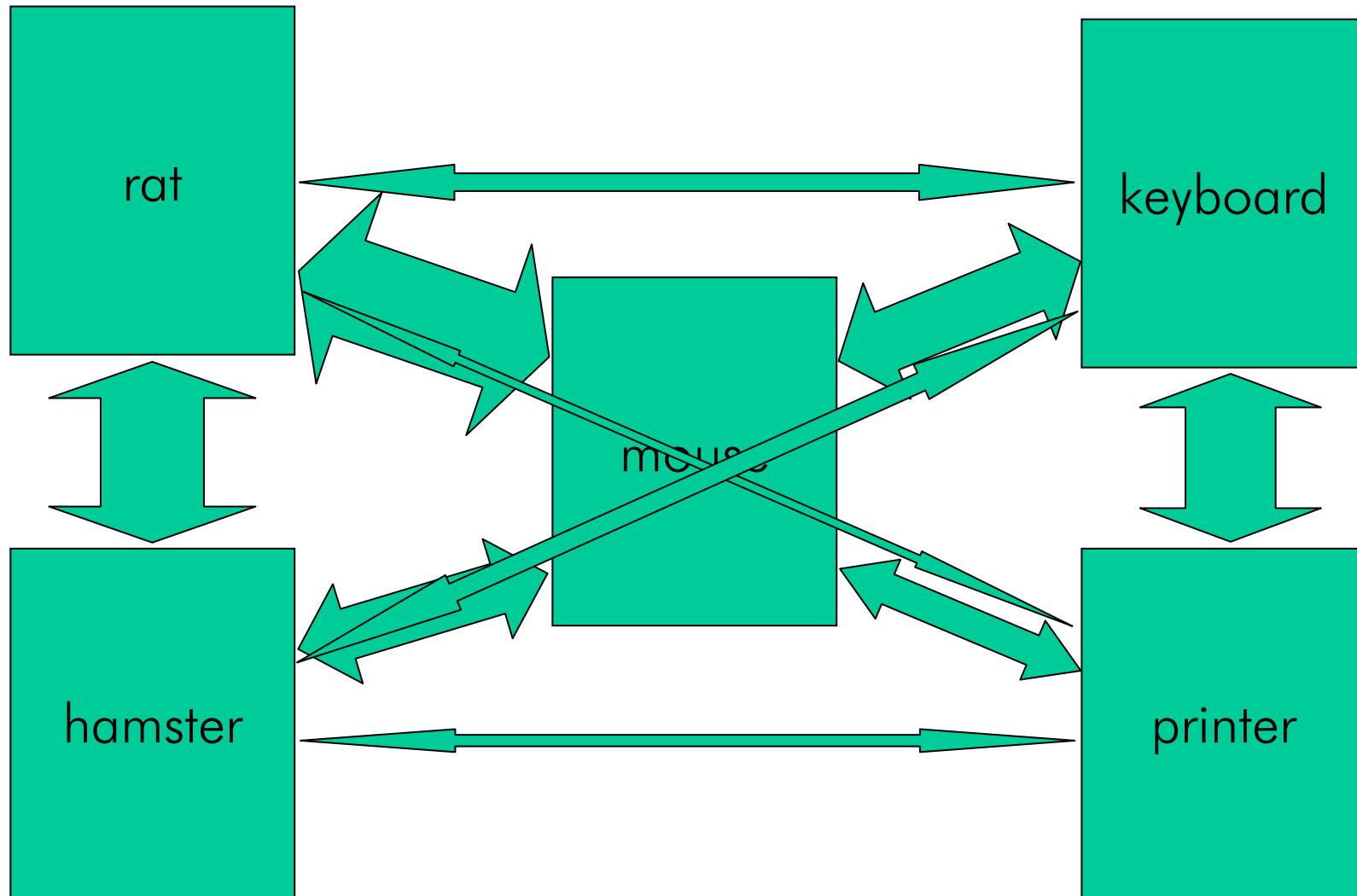
gelacht
kichern
grinsen
lächeln
schreien
johlen
zwinkern
schmunzeln
schluchzen
klatschen
jubeln
brüllen
nicken
amüsieren
feixen
scherzen
plaudern
schunkeln
jauchzen
geklatscht
schwatzten
schauen
glucksen
blinzeln
amüsiert
flirten
umarmt
umarmen

Contrasting

- contrasting profiles yields
 - areas that overlap
(common part of ‘meaning potential’)
 - areas that differ
(distinguishing part of ‘meaning potential’)
-







rat

keyboard

hamster

printer

rat
hamster

keyboard
printer

rat
hamster

keyboard
printer

rat
hamster

keyboard
printer

rat
hamster

2

dog
cat

5

4

keyboard
printer

rat
hamster

dog
cat

keyboard
printer

rat
hamster

dog
cat

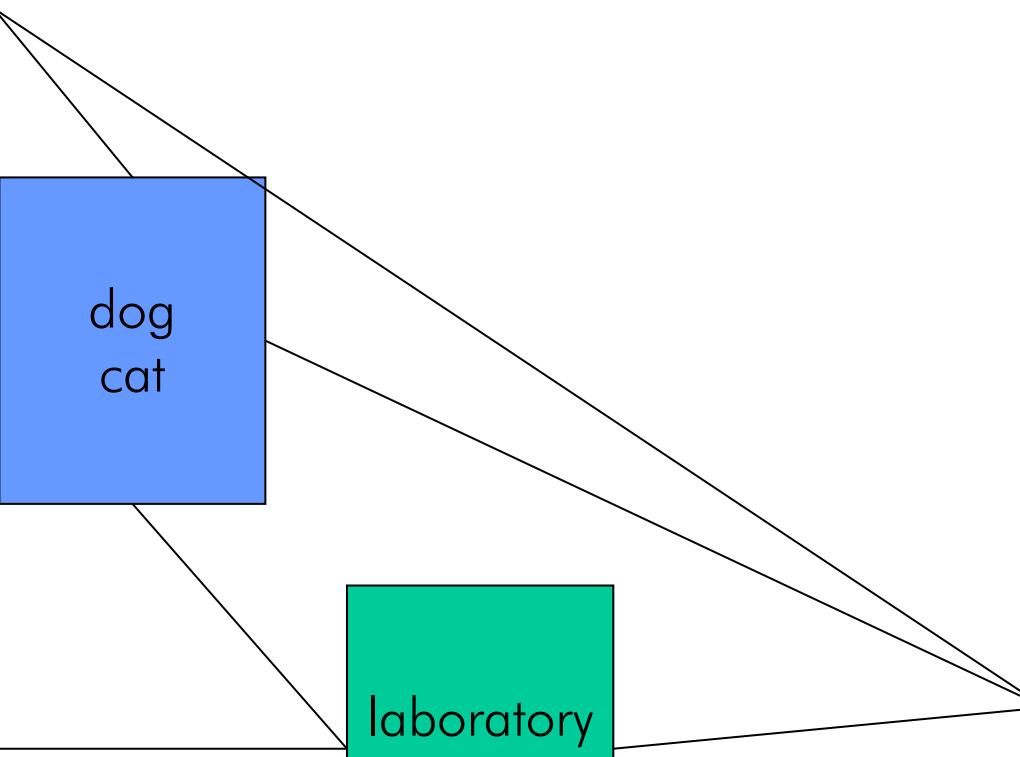
keyboard
printer

rat
hamster

dog
cat

laboratory
researcher

keyboard
printer

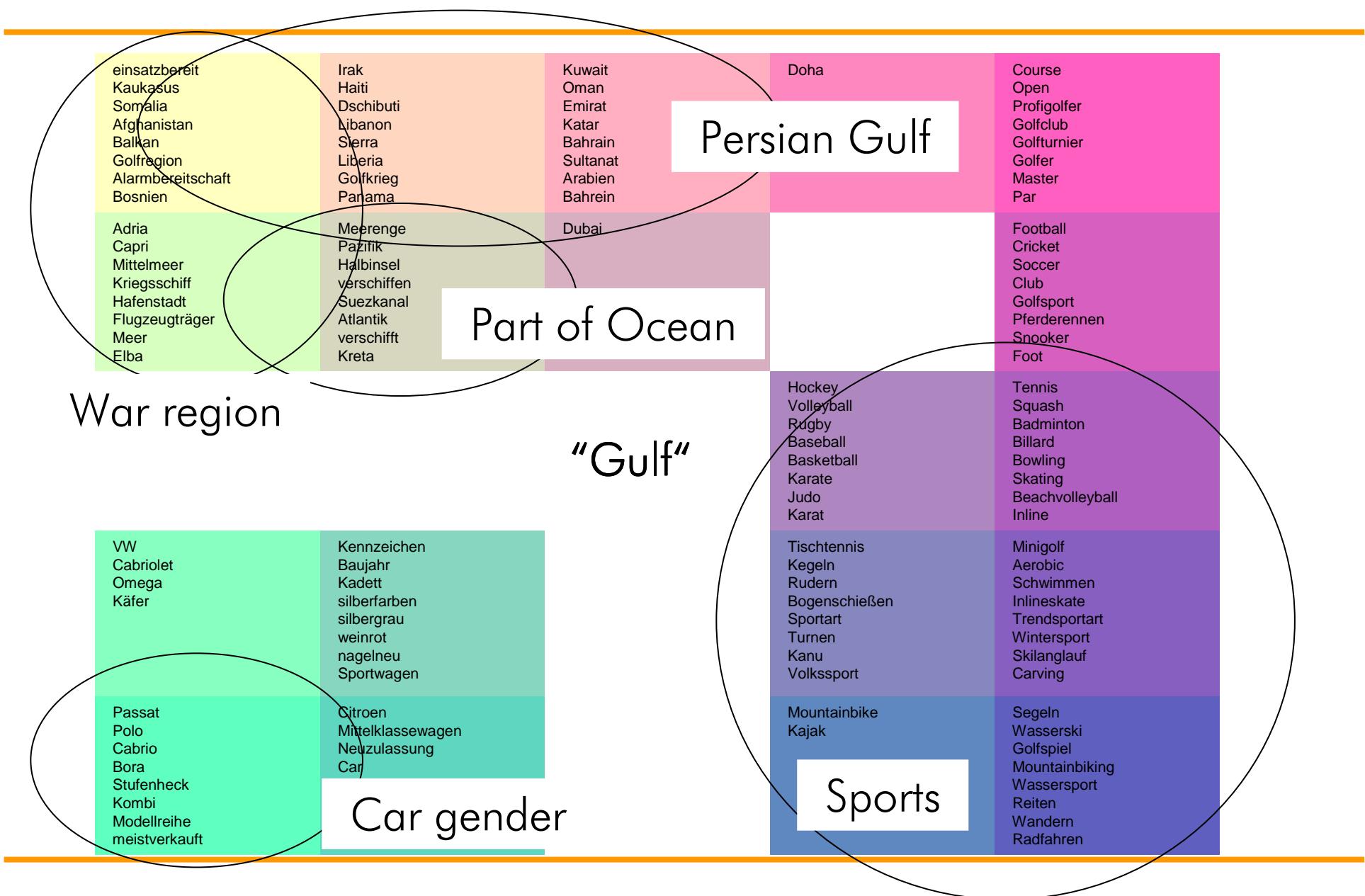


rat
hamster

dog
cat

laboratory
researcher

keyboard
printer



Conclusion

- methodology yields impressive results
 - without linguistic assumptions
 - data-driven by a well-designed empirical base
 - by sophisticated analysis methods
 - helps uncovering usage aspects that contribute to the meaning potential
 - not suited for operationalization, but for interpretation
 - hermeneutically for ‘meaning exploration’
-

Corpus Linguistics is ...

- a science-theoretic coherent approach to the investigation of language
 - requires huge amounts of data and a sophisticated methodology
 - enables new insights in language phenomena
 - [despite elaborated methodological support:] expects the systematic investigation of large data sets
-

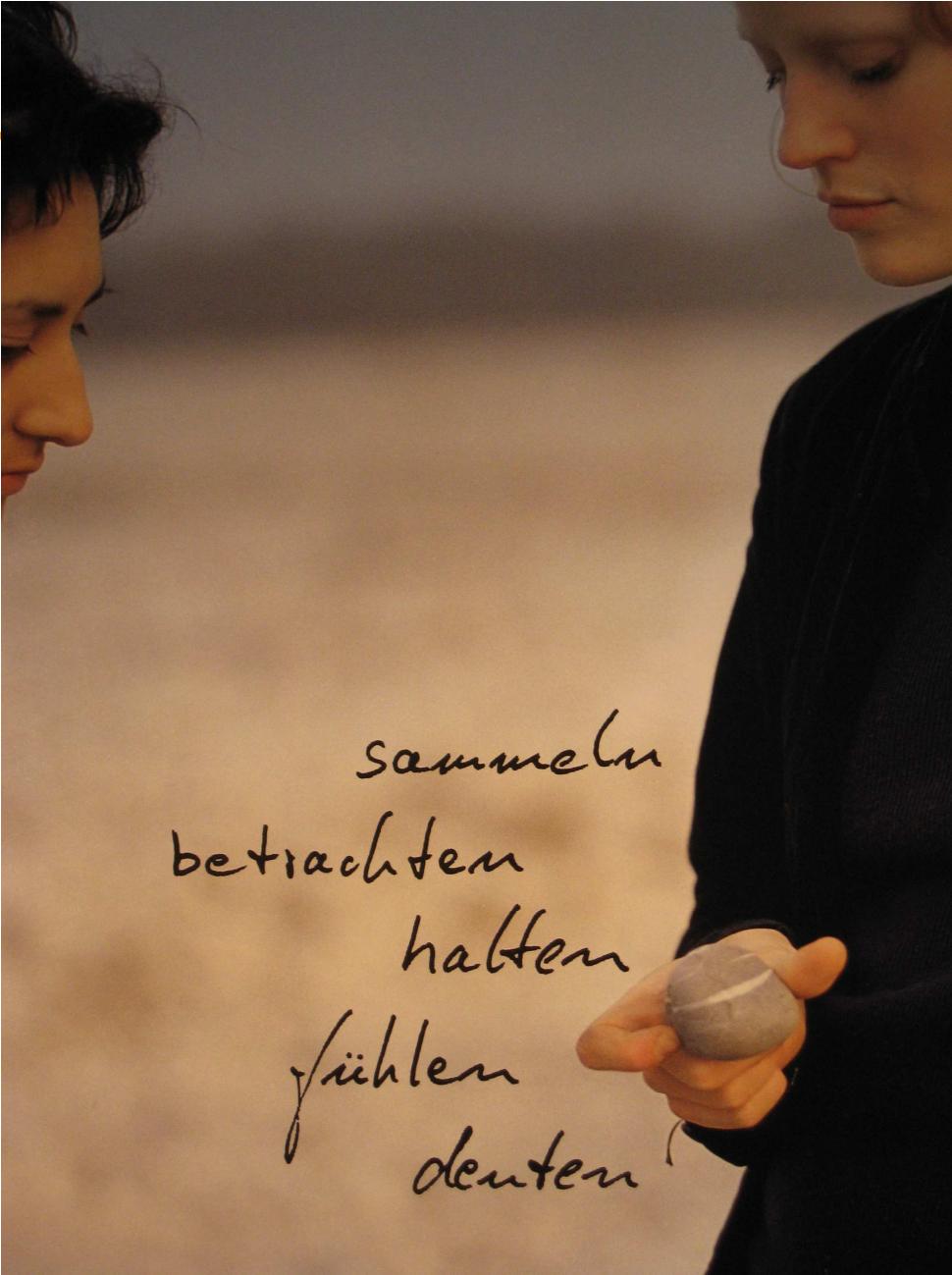
Scenarios

- Description of Multi-Word Expressions/Units
 - Categories and Dictionaries
 - Lexicography
 - Detection of ‘Senses’ as Categories on Usage Aspects
 - Foreign Language Teaching (DaF)
 - condensed input for self-learning environment
 - simulated late partial immersion
-



... in a picture

© Goethe-Institut



... in a picture

© Goethe-Institut

Thank you ...

... very much for your attention.

Further information:

<http://www.ids-mannheim.de/kl/>
korpuslinguistik@ids-mannheim.de
