

# Getan – gesagt?

*Pragmatische und lexikalisierte  
Erklärungen zur Besetzung von  
Argumentstellen mit neuem Material*

Amir Zeldes  
Korpuslinguistik – IDSL HU Berlin  
[amir.zeldes@rz.hu-berlin.de](mailto:amir.zeldes@rz.hu-berlin.de)

# Manche Konstruktionen sind repetitiver als andere...

Part.-Pf. (Pkt.) Part.-Pf.	[N] <i>essen</i>	[N] <i>trinken</i>
<i>Gesagt , gebucht !</i>	<i>Klebreis</i>	<i>Alkohol</i>
<i>Gesagt , getan !</i>	<i>Bratwurst</i>	<i>Milch</i>
<i>Gesagt , getan !</i>	<i>eine Kuh</i>	<i>Alkohol</i>
<i>Versprochen – gehalten !</i>	<i>Menschenfleisch</i>	<i>Sekt</i>
<i>Gesagt getan .</i>	<i>Fleisch</i>	<i>Bier</i>
<i>Gesagt , getan :</i>	<i>Pizza</i>	<i>Bier</i>
<i>Gesagt , getan .</i>	<i>Fleisch</i>	<i>Alkohol</i>
<i>Gesagt , getan .</i>	<i>Fleisch</i>	<i>Alkohol</i>
<i>Gesagt getan .</i>	<i>Kleinigkeit</i>	<i>Kaffee</i>
<i>Gesagt , getan .</i>	<i>süße Sachen</i>	<i>Haarshampoo</i>

# Semantische/pragmatische Erklärungsansätze

- Die Menge der essbaren Gegenstände ist größer als die der Getränke
  - Nicht wirklich – beide Klassen sind offen, nicht aufzählbar:  $|\text{trinkbar}| = |\text{essbar}| = \aleph_0$
- Wir wiederholen Getränke öfter als Gerichte im Alltagsleben
  - Getanes führt zu Gesagtem?
  - Gilt das für alle Fälle?
  - Was ist mit den *gesagt-getan*-Sätzen?

# Exkurs:

## Morphologische Produktivität

- Ein ähnliches Phänomen wird in der morphologischen Produktivitätsforschung thematisiert
- Manche morphologische Prozesse sind „produktiver“ als andere (s. Bauer 2001)
  - größere Häufigkeit
  - größeres Vokabular
  - größere Bereitschaft für Neubildungen

# Sind diese Wortbildungsmuster gleich produktiv?

- Komparativbildung *-er* sehr
- Nomenbildung auf *-ung* sehr
- Nomenbildung auf *-tum* wenig
- Adjektivbildung mit *-lich* mittel
- Adjektivbildung mit *-bar* hoch
- Adjektivbildung mit *-sam* gar nicht?

# Konsequenz: nicht binäre Produktivität

- $n+1$  leichter als  $m+1$ :
  - *machbar, dehnbar, miniaturisierbar ...n*
  - *natürlich, löslich, landessprachlich ... m*
  - *langsam, grausam, ... ?*
- Wie kann man das messen?
- Können wir in jedem Fall beurteilen, ob eine produktive Bildung vorliegt?
  - *Miniaturisierbarkeit*  
(Neologismus? Schon mal gehört?)

# Produktivität messen?

- Baayens (1992, 2001, 2009 i.a.) Vorschläge basieren auf:
  - Tokenanzahl –  $N(C)$
  - Typenanzahl –  $V(N, C)$
  - Anzahl der *Hapax Legomena* –  $V1(N, C)$
- Implizite (falsche) Annahmen:
  - Neologismen  $\subseteq$  Hapax Legomena
  - Konstantes Verhältnis

# Ergebnisse für *-lich/-bar/-sam*

- Korpus: 5 Jahre  
C'T Magazin
- 14.596.537 Token
- 595.022 Typen
- 356.075 hapax

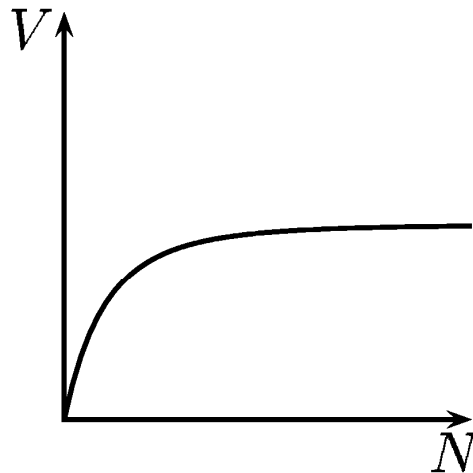
	<i>-lich</i>	<i>-bar</i>	<i>-sam</i>
N	120458	26797	7672
V	829	716	43
V1	201	260	7
$\mathcal{P}=V1/N$	0.001668	0.009702	0.000087

- *-sam* ist in jeder Hinsicht am unproduktivsten
- *-lich* war ggf. bisher produktiver, breitet sich aber nun langsamer aus als *-bar*

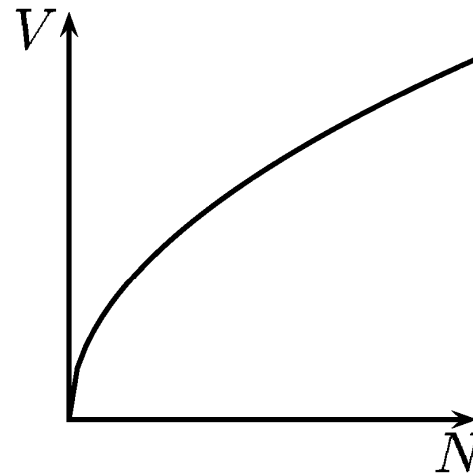


# Vocabulary Growth Curves (VGC)

- wie oft sieht man ein neues Wort, wenn man schon einige gesehen hat?
  - Kurve steigt: Prozess produktiv
  - Kurve wird flach: Prozess unproduktiv



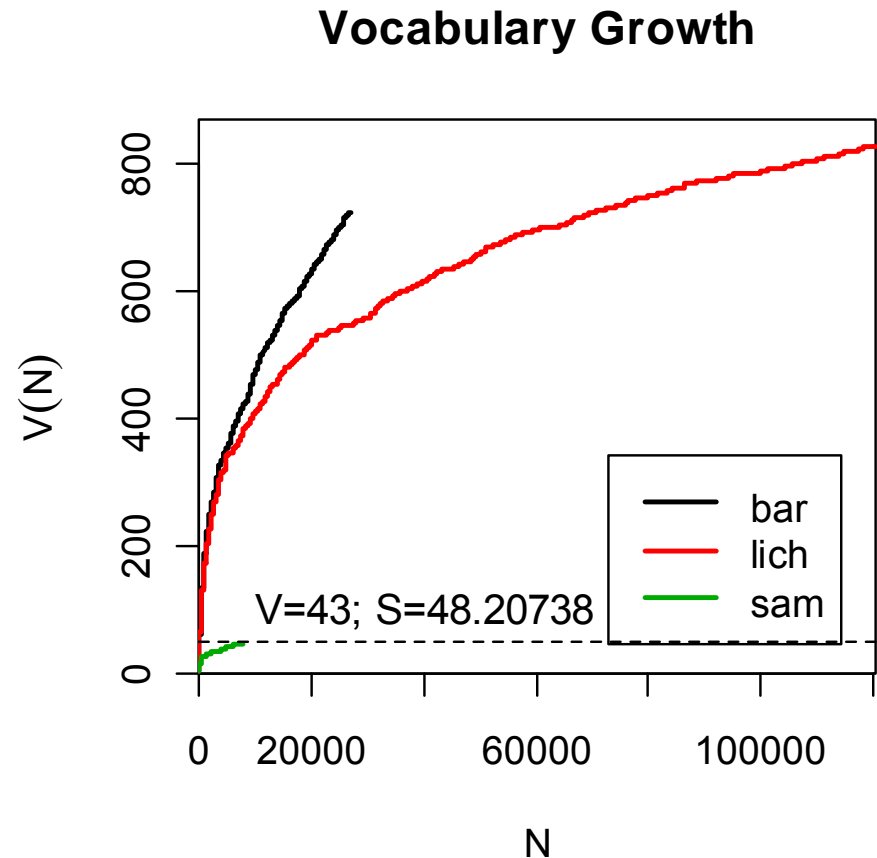
unproductive



productive

# Vocabulary Growth Curves (VGC)

- Visualisierung des Sättigungsgrades
- Statistisches Modell für jede Kurve berechenbar (fZM, Evert 2004)
- Gesamtwortschatz-Vorhersage durch Asymptote ( $S = \textit{Species}$ )
- Vorsicht – unterschiedliche Mengen, Fehler, Dispersion... (vgl. Evert & Lüdeling 2001)

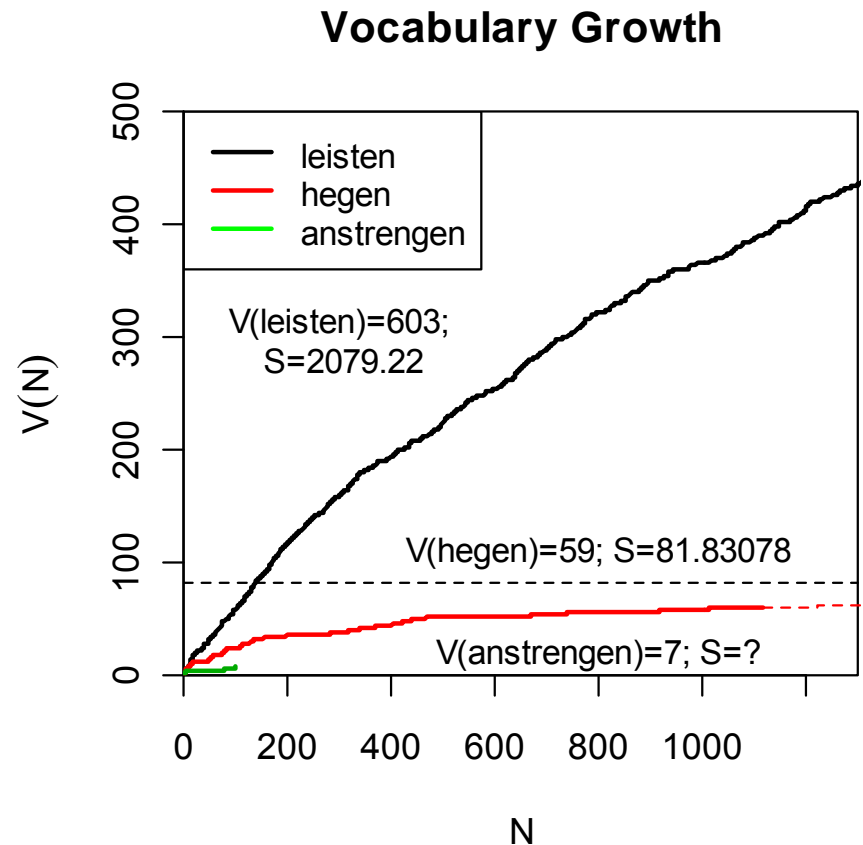


(C'T, von Hand bereinigt)

<http://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/institutkorpora>

# Was hat das mit der Argumentwahl zu tun?

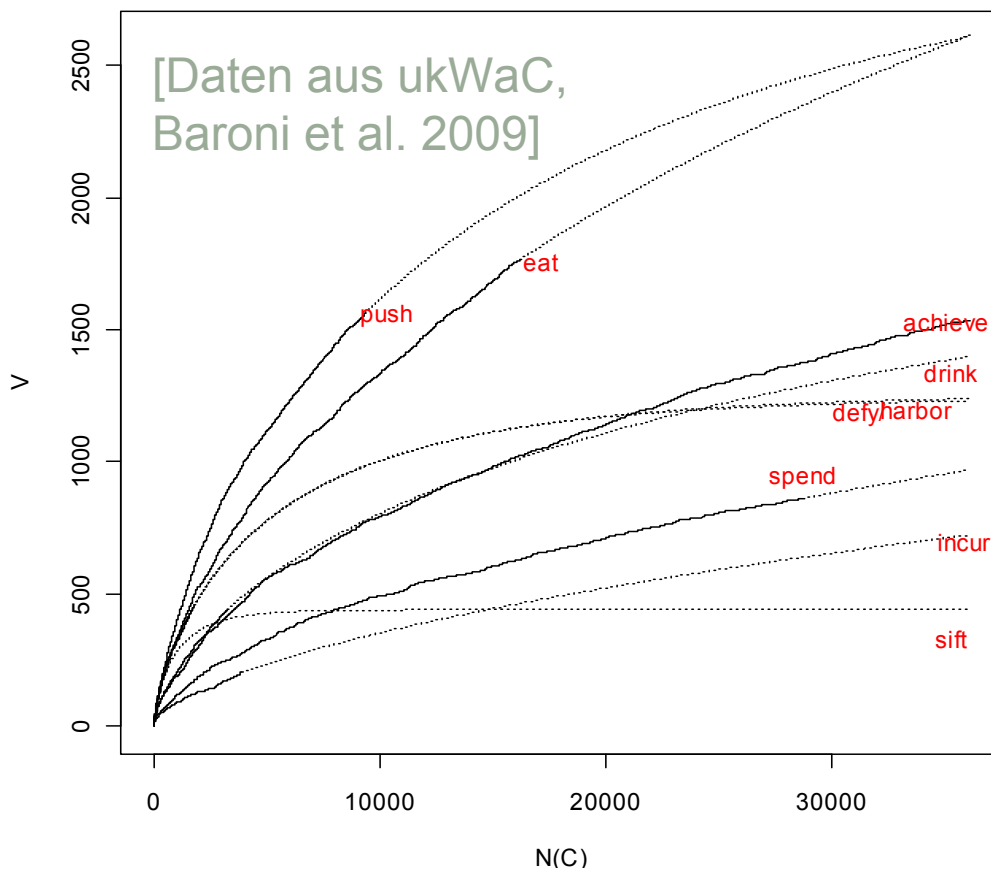
- Argumente kann man genauso messen
- Verhalten sehr ähnlich
- Ist *anstrengen* (im juristischen Sinne) das „-sam der Verben“?
- Objekt-Köpfe:
  - *Prozess*,
  - *Verfahren*,
  - *Klage*
  - ... ?
- $\mathcal{P}(\textit{hinterziehen})=0$  !!



(deWaC, Baroni et al. 2009,  
von Hand bereinigt)

# Wer kann was mehr?

- Produktivität als Phänomenenkomplex (PC)



$$\left\{ \begin{array}{l} PC_N = \{f, V_N, \mathcal{P}_N, S\} \\ V_N = vgc(N) \\ \mathcal{P}_N = vgc'(N) \\ S = \lim_{N \rightarrow \infty} V_N = \lim_{N \rightarrow \infty} vgc(N) \end{array} \right.$$

$$PC = \{f, vgc(N)\}$$

# Ranking nach unterschiedlichen Kriterien

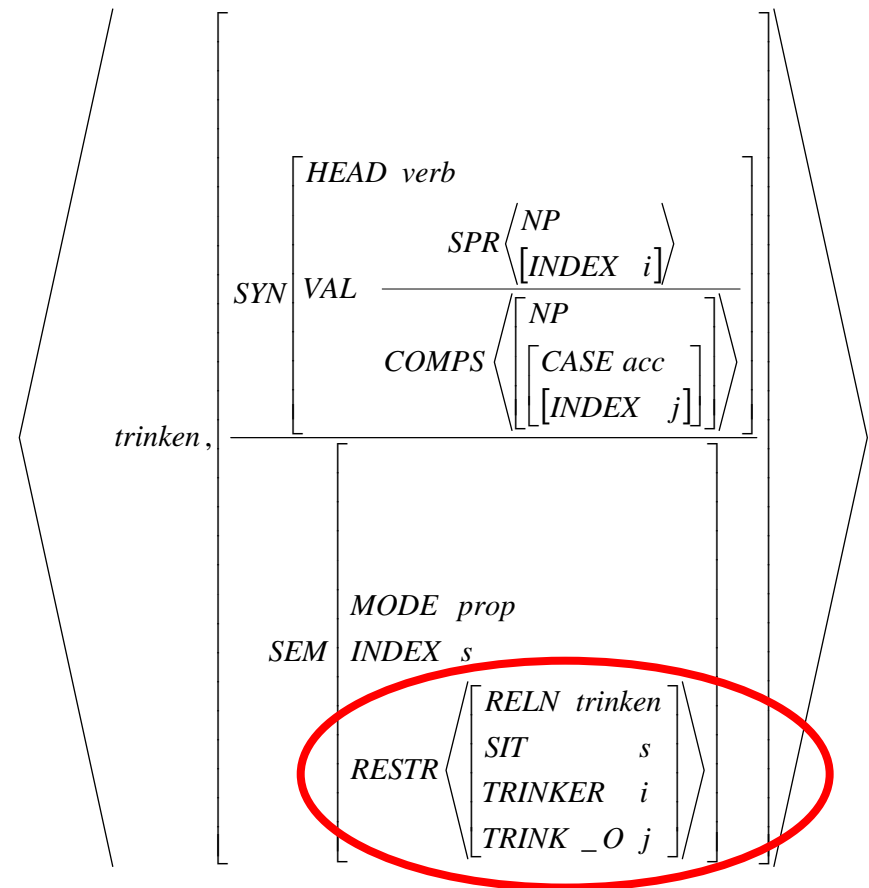
Rank	<b>S</b>	<b>V</b>	<b>N</b>	$P_{N(C)=1000}$
1	<i>eat</i>	<i>eat</i>	<i>achieve</i>	<i>push</i>
2	<i>achieve</i>	<i>push</i>	<i>spend</i>	<i>eat</i>
3	<i>incur</i>	<i>achieve</i>	<i>eat</i>	<i>harbor</i>
4	<i>push</i>	<i>spend</i>	<i>push</i>	<i>defy</i>
5	<i>spend</i>	<i>drink</i>	<i>incur</i>	<i>achieve</i>
6	<i>drink</i>	<i>harbor</i>	<i>drink</i>	<i>drink</i>
7	<i>harbor</i>	<i>defy</i>	<i>harbor</i>	<i>spend</i>
8	<i>defy</i>	<i>incur</i>	<i>defy</i>	<i>incur</i>
9	<i>sift</i>	<i>sift</i>	<i>sift</i>	<i>(sift)</i>

# Lexikalische Semantik als Modell

- Argumentstellen verlangen unterschiedliche semantische Klassen (Jackendoff 1987, 1990)
- Verben spezifizieren Präferenzen bspw. durch sog. Entailments (Weinreich 1966, Dowty 1991)
  - *trinken* verlangt [+liquid]  
*Wasser*<sub>[+liquid]</sub> *trinken*  
# *Steine*<sub>[-liquid]</sub> *trinken*  
? *Lava*<sub>[+liquid]</sub> *trinken*
- Welche Klasse verlangt *anstrengen*?

# Lexikalische Semantik als Modell

- Man muss oft von „Individual thematic roles“ (Dowty 1991) ausgehen
- Verb-spezifische Restriktionen
  - [+trinkbar]
  - [+anstrengbar]
  - ...
- Zirkelschluss – keine Vorhersagekraft?
- Es sei denn, mehrere Stellen teilen dieselbe Klasse



Nach Bosch (2007)

# Hypothese 1 – Semantik

- Wenn ausschließlich semantische Klassen für Argumentwahl verantwortlich sind:
  - i. müssten Argumentstrukturen mit gleicher Bedeutung ähnliche Vokabeln realisieren
    - schwierig, weil zufällig unterschiedliche **Lexikalisierungen** möglich sind
  - ii. sie müssen wenigstens **gleich produktiv** sein, denn neue Fälle können nicht lexikalisiert sein
    - Synonyme als ‚Minimalpaar‘

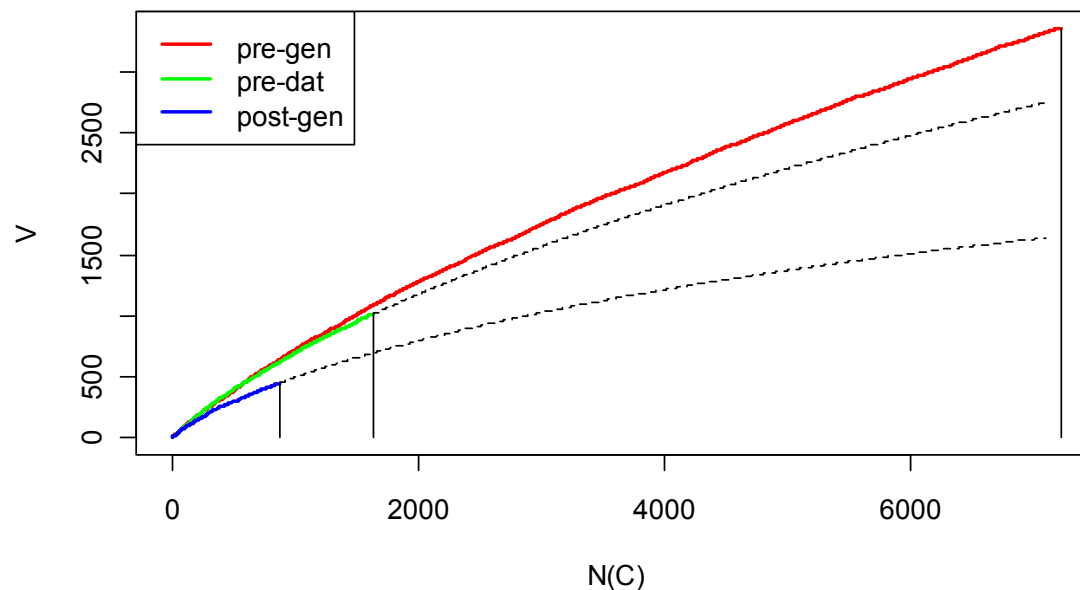


# Fall 1 – *wegen X* oder *X wegen*?

- Wegen tritt mit 3 syntaktischen Argumentstrukturen auf:  
(Braunmüller 1982:200; Zifonun et al. 1997:2080ff; Petig 1997)
    - *wegen des Vaters* = *pre-gen*
    - *des Vaters wegen* = *post-gen*
    - *wegen dem Vater* = *pre-dat*
  - Alle Varianten sind semantisch synonym\*  
*salva veritate*
  - Gibt es Intuitionen über ihre Produktivität?
- \* abgesehen von Register/Pragmatik – mehr dazu später

# *des X wegen* ist weniger produktiv

- Kein signifikanter Unterschied bei pre-dat/gen
- post-gen ist schwächer ( $p < 2.2e-16$ ,  $\phi = 0.2871$ )
- 95% Interval um 20-25% weniger Wortschatz (10-Fold Cross-Validation in ukWaC)

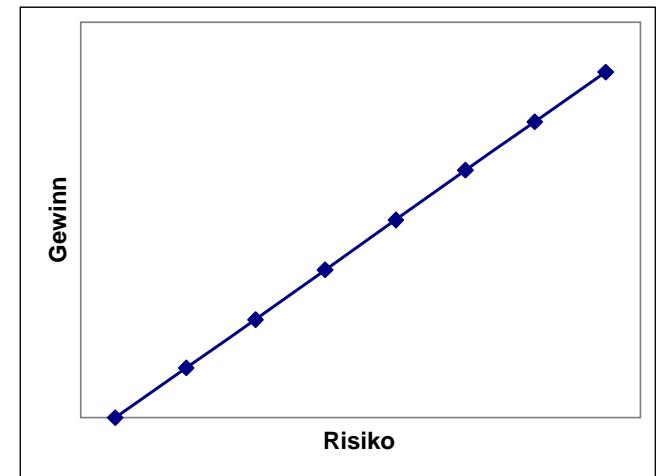


# *des X wegen* auf dem Weg raus?

- Wird post-gen eines Tages auf ihre häufigsten Vertreter beschränkt?
  - *des Geldes wegen*
  - *der Liebe wegen*
  - *der Kinder wegen*
  - *der Übersichtlichkeit wegen*
  - ...?
- Vgl. *von Amts wegen, an Eides statt,*  
Eng. *the world over, in Christ's stead*

# Fall 2 – *je ... desto*

- Monotone Korrelation von Eigenschaften  
(comparative correlatives - CC)
- Präferenzen für beide Teilsätze: (Zeldes 2009,2011)
  - *je* (c1): spatiotemporal
    - *länger*
    - *höher*
    - *älter*
  - *desto* (c2): evaluativ
    - *besser*
    - *schwieriger*
    - *wahrscheinlicher*
- *Je größer das Risiko, desto höher der Gewinn*



=besser

# Fall 2 – *je ... desto*

type	je	desto	kurz-je	kurz-desto	freq
<i>besser</i>	70	212	0	37	18270
<i>später</i>	22	5	0	0	7983
<i>stärker</i>	65	56	0	0	6844
<i>länger</i>	179	23	2	0	4659
...					
<i>höher</i>	179	76	0	1	3330
<i>größer</i>	195	120	4	2	3126
...					
<i>schwieriger</i>	7	24	0	0	1344
...					
<i>wahrscheinlicher</i>	0	10	0	0	103

[c't-Korpus +  
Parlamentsreden]

# Varianten unterschiedlich produktiv

DP VP:

**Je tiefgründiger**  
*der emotionale  
Mehrwert der  
zukünftigen Marke  
ausgelotet ist*

<

**desto prägnanter**  
*lässt sie sich  
benennen*

∨

∨

Nur DP:

**Je größer die**  
*Oberfläche*

<

**desto effizienter**  
*die chemische  
Wirkung*

∨

∨

Ohne  
DP VP:

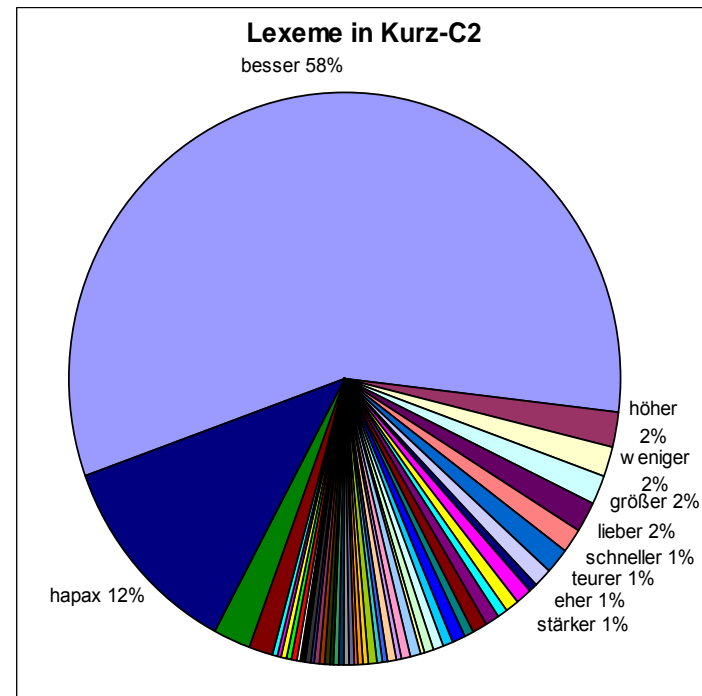
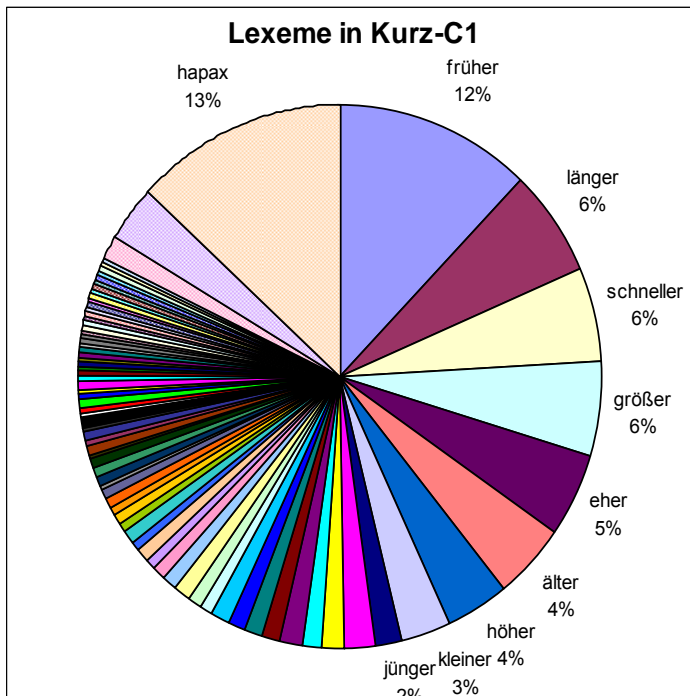
**Je früher**

>

**desto besser**

# Semantische Erklärung?

- Warum ist *desto* generell produktiver?
- Mehr Evaluationen nötig als Bedingungen?
- Warum umgekehrt bei Kurz-CC?



# Ungleichheiten bei weiteren “near synonyms”

- help (O) to V > help (O) V
  - start(O) > begin(O) > commence(O)
  - start Ving > start to V
  - totally A ... > absolutely A
- ...



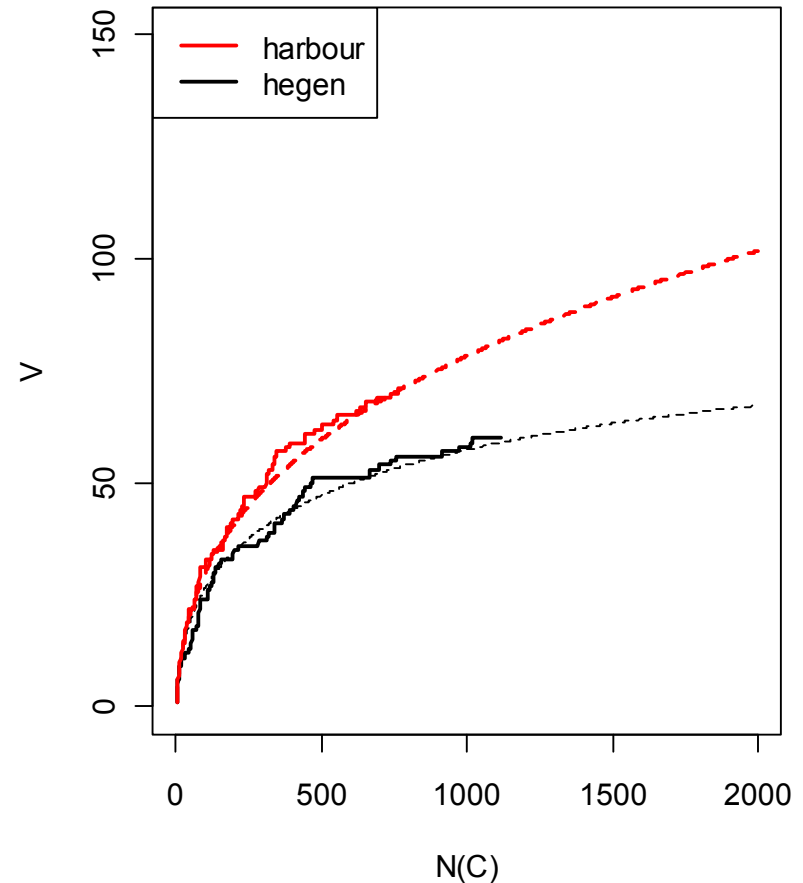
# Hypothese 2 – Pragmatik

- wenn nur Weltwissen die Argumentwahl motiviert...
  - sollten Übersetzungspaare aus unterschiedlichen Sprachen ähnlich sein
- Nichtübereinstimmung als Hinweis auf sprachspezifische Tendenzen
- Produktivität wäre dann sprachspezifisch

# Fall 3 – *hegen* / *harbour*

([+mentaler Zustand])

- En. *harbour* + ment. > De. *hegen* + ment.
- Verbergen englische Muttersprachler mehr mentale Zustände?
- Oder wählen Deutsche lieber andere Verben / Strukturen für unvertraute Argumente?



## *hegen / harbour* ([+mentaler Zustand])

■ Zweifel	19%	■ ambition	12%
■ Verdacht	13%	■ feeling	9%
■ Hoffnung	13%	■ doubt	9%
■ Wunsch	10%	■ hope	8%
■ ...		■ ...	
■ Missachtung	<0.1%	■ Judgment	<0.1%
■ Tapferkeit	<0.1%	■ longing	<0.1%

# Beobachtungen

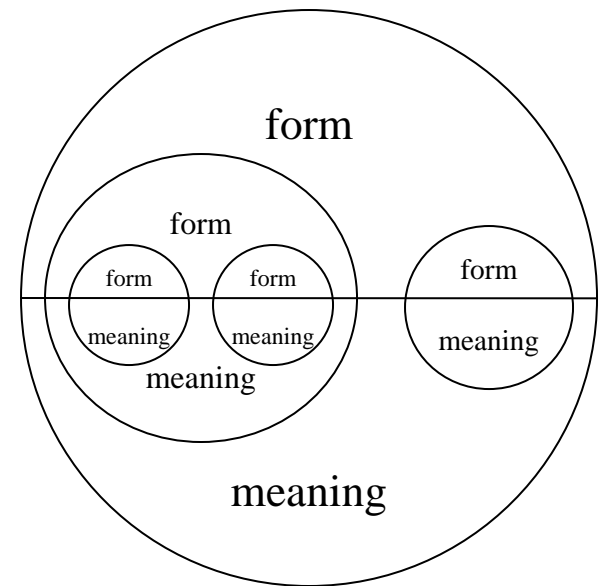
- Konstruktionen mit demselben semantischen Gehalt verhalten sich unterschiedlich und sprachspezifisch
  - Semantische Klassen können Produktivität nicht vorhersagen
  - Sprachübergreifendes Weltwissen genau so wenig (getan  $\neq$  gesagt)
- Sprach- und konstruktionsspezifische Bereitschaftsgrade, neue Argumente zuzulassen

# Wo kommt das her?

- Sprecher geben die Verteilung wieder, die sie empfangen:
  - Unproduktivere Verteilungen (repetitiv, wenig Hapax Legomena) werden als solche erworben
  - Produktive Verteilungen (zipfisch verteilte) verewigen sich im produktiven Gebrauch
- Wir brauchen ein Modell, das Produktivitätswerte speichert bzw. vorhersagt
- Weltwissen mit Syntax verbindet (Zusammenhang mit Gebrauch, semantischen Klassen)

# Construction Grammar (CxG)

- **Gebrauchsbasierte Theorie**  
(Goldberg 1995, 2006; Kay & Fillmore 1999; Croft 2001)
- **hierarchische ‚Form-Bedeutung‘-Kopplungen**
- **Bedeutungsbeitrag gleichzeitig von Bestandteilen und Gesamtheit**



# Das Constructicon (nach Goldberg 2006)

0.7	4236123	Morphem	<i>pre-, -ing</i>
0.00001	105	Wort	<i>avocado, anaconda, and</i>
0.75	15138533	Komplex Wort (teilweise spez.)	[N-s] (for regular plurals)
0	1	Idiom (voll spezifiziert)	<i>Devil may care</i>
~0	55	Idiom (teilweise spezifiziert)	<i>jog &lt;someone's&gt; memory</i>
0.007	123403	Komparativ	[A-er] / more A
0.0001	109	CC	The Xer the Yer
0.002	213382	Ditransitives Verb	Subj Obj1 Obj2

- Entrenchment für jede Konstruktion
- Wächst mit Frequenz
- Gleichermaßen Produktivität?

# Kontextabhängigkeit

- Es gibt keinen einheitlichen Produktivitätswert für jede Konstruktion:

$\mathcal{P}(\text{___er}) >$

$\mathcal{P}(\text{desto ___er VP DP}) >$

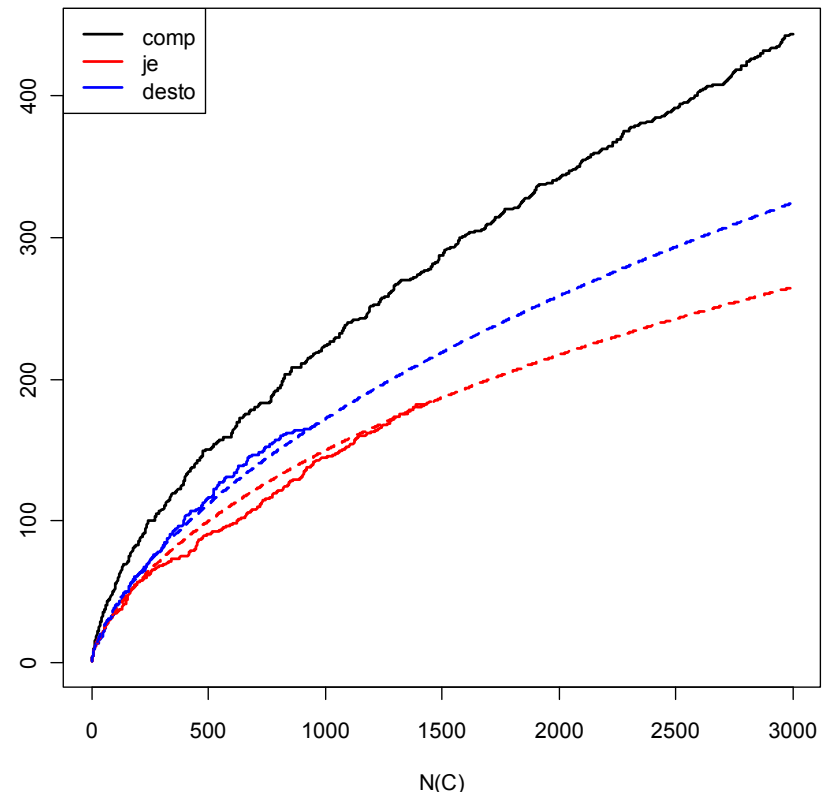
$\mathcal{P}(\text{je ___er VP DP}) >$

$\mathcal{P}(\text{je ___er}) >$

$\mathcal{P}(\text{desto ___er})$

- Unterschiedliche Aspekte des **PC**

- Produktivität als Epiphänomen?  
(vgl. Plag 2006)



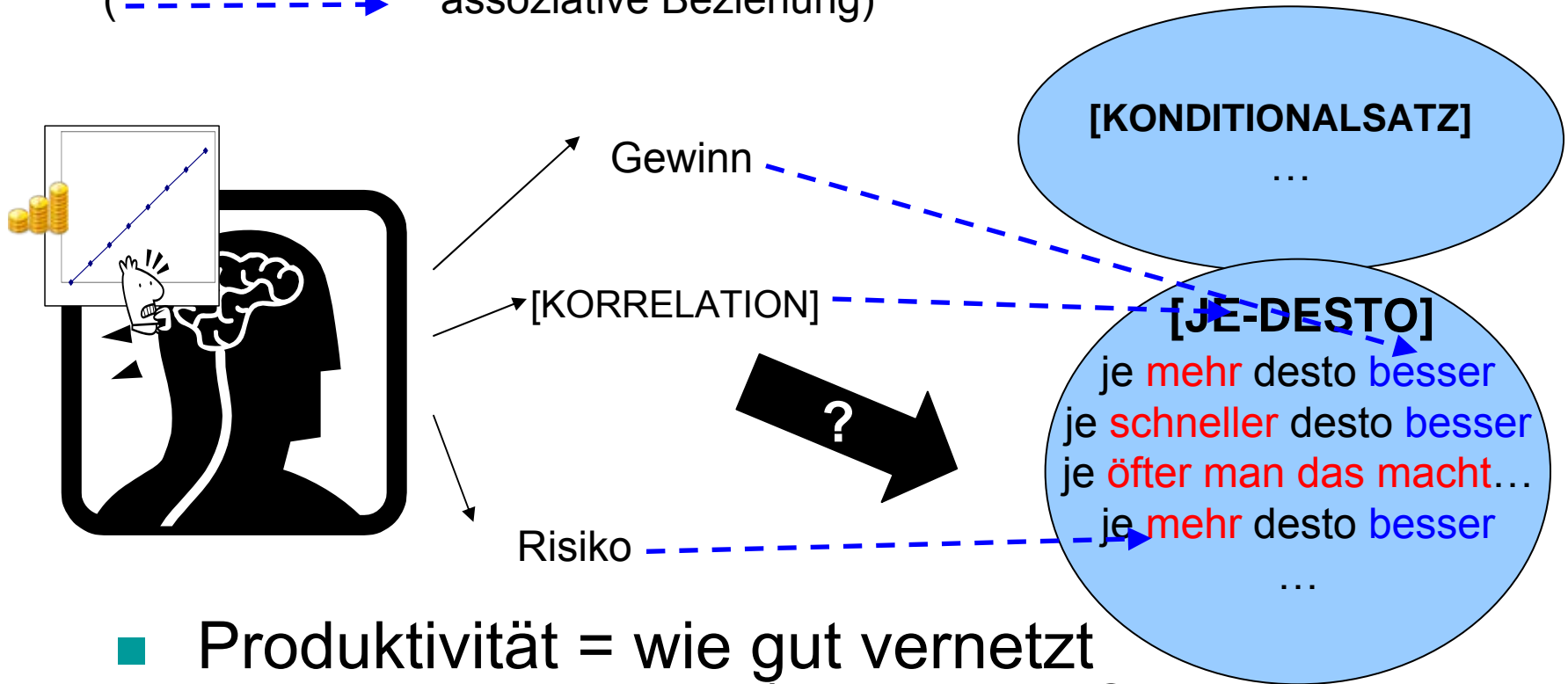


# Mögliche Lösungen? (WIP)

- Werte für jeden Slot?
  - Lösung bei Instanzen mehrerer Konstruktionen unklar
- Folge von prototypenbasierter Grammatik?
  - Hinter Konstruktionen stehen Exemplare  
(Goldberg et al. 2004, Casenhiser & Goldberg 2005, Bod 2006)
  - Verhältnismäßig häufige Vertreter der Konstruktion werden zu Prototypen
  - Semantik als Modell für Erweiterungen
  - Konstruktionen entstehen durch Generalisierung von Prototypen

# Wahl einer Konstruktion

( assoziative Beziehung)



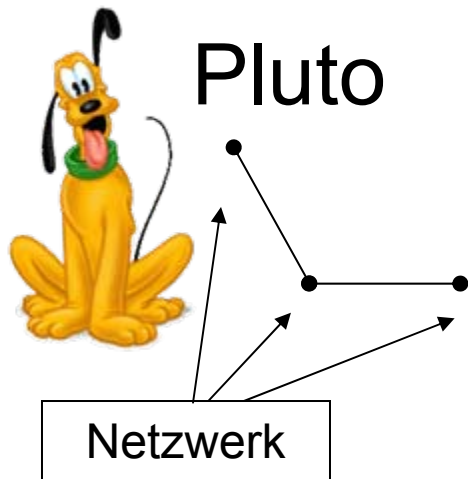
- Produktivität = wie gut vernetzt sind die Exemplare/Prototypen?

# Das Hebb'sche Gesetz

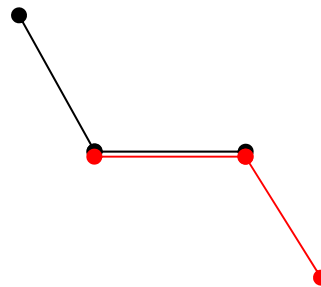
- „Schief“ verteilte Daten führen zu besserer Lernbarkeit (Casenhiser & Goldberg 2005)
- Das Hebb'sche Gesetz als kognitive Grundlage für Produktivitätserwerb:
  - ~„*cells that fire together, wire together*“
  - Und umgekehrt (seltene gemeinsame Aktivierung führt zu Inhibition)

# Klassenbildung durch Exemplare

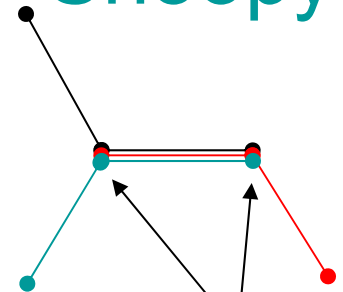
(vgl. Hopfield-Netzwerke, Hopfield 1982)



+Lassie

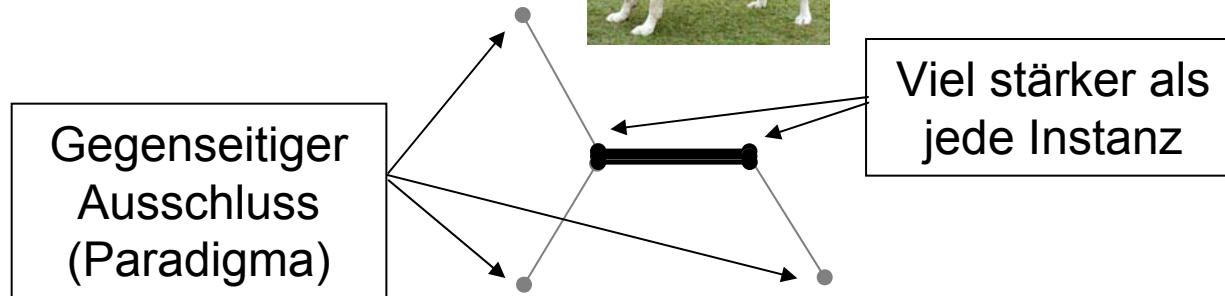


+Snoopy



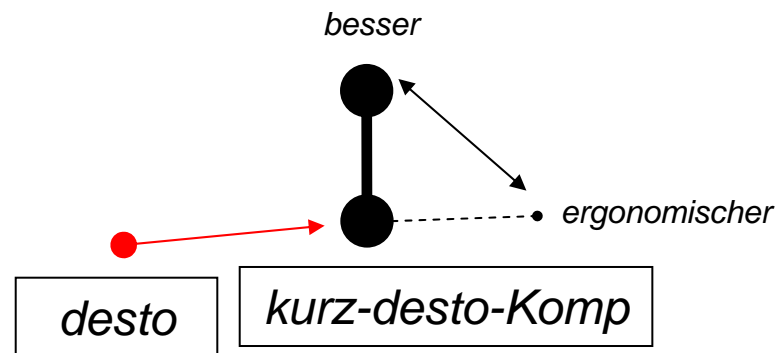
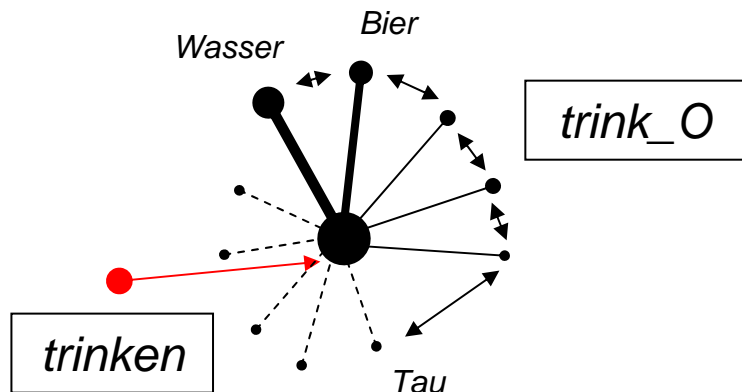
Gemeinsame  
Merkmale

Klasse: [+Hund]



# Produktive und nicht so produktive Netzwerke

- Zu mächtige Exemplare ziehen alle Verwendungen auf sich, schließen andere aus: sie **sind** die Konst.
- Viele Hapax Legomena hingegen machen die Struktur mächtiger als ihre Vertreter, leicht erweiterbar



→ Syntagmatische Beziehung (exzitatorisch)

↔ Paradigmatische Beziehung (inhibitorisch)

# Beziehungen zwischen Netzwerken

- Assoziative Beziehungen suggerieren Konstruktionen
- Produktivität entsteht durch:
  - Assoziative Vernetztheit (einschließlich für Mitglieder der syntagmatischen Achse)
  - Zipfsche Verteilung des Paradigmas (erhöht Schwerkraft der Konstruktion)
  - Mäßige Stärke der häufigsten Vertreter (sonst Inhibition über paradigmatische Beziehungen)

# Zusammenfassung

- Produktivität lässt sich rein semantisch/pragmatisch nicht vorhersagen
- Sprach- bzw. konstruktionsspezifische Tendenzen
- Gebrauch erfordert Erklärung im mentalen Lexikon bzw. Grammatik
- Der Unterschied dazwischen ist für CxG gering:
  - Regelsein ist nicht binär
  - ‚Gute‘ Regeln sind sehr produktive Konstruktionen, marginale Regeln wie **gesagt – getan** sind weniger produktive Konstruktionen

# Desiderata

- Formalisierung der Beziehungen im mentalen Lexikon (vgl. Goldberg 1995)
- Angemessener Formalismus für die Modellierung von Produktivität, besonders interessant:
  - Data Oriented Parsing/LFG-DOP (Bod 2006)
  - Fluid Construction Grammar (de Beule/Steels 2005)
- Auswertung durch Korpora und psycholinguistische Experimente



# Literatur I

- Baayen, R. H. 1992. Quantitative Aspects of Morphological Productivity. In Booij, G. E./van Marle, J. (eds.) *Yearbook of Morphology 1991*. Dordrecht: Kluwer, 109–149.
- Baayen, R. H. 2001. *Word Frequency Distributions*. Dordrecht: Kluwer.
- Baayen, R. H. 2009. Corpus Linguistics in Morphology: Morphological Productivity. In Lüdeling, A./Kytö, M. (eds.) *Corpus Linguistics. An International Handbook*. Berlin: Mouton de Gruyter, 899-919.
- Baroni, M./Bernardini, S./Ferraresi, A./Zanchetta, E. 2009. The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora. In *LRE* 43(3), 209-226.
- Bauer, L. 2001. *Morphological Productivity*. Cambridge: CUP.
- de Beule, J./Steels, L. 2005. Hierarchy in Fluid Construction Grammar. In Furbach, U. (ed.) *KI 2005: Advances In Artificial Intelligence. Proc. of the 28th German Conference on AI*. Berlin: Springer, 1-15.
- Bod, R. 2006. Exemplar-Based Syntax: How to Get Productivity from Examples. *The Linguistic Review* 23(3), 291-320.
- Bosch, P. 2007. Productivity, Polysemy, and Predicate Indexicality. In ten Cate, B./Zeevat, H. (eds.) *Proc. of the 6th Intl. Tbilisi Symposium on Language, Logic and Computation*. Berlin: Springer, 58-71.
- Braunmüller, K. 1982. *Syntaxtypologische Studien zum Germanischen*. Tübingen: Gunter Narr.
- Casenhiser, D./Goldberg, A. 2005. Fast Mapping between a Phrasal Form and Meaning. *Developmental Science* 8(6), 500-508.
- Croft, W. 2001. *Radical Construction Grammar*. Oxford: Oxford University Press.
- Dowty, D. R. 1991. Thematic Proto-Roles and Argument Selection. *Language* 67(3), 547-619.
- Evert, S. 2004. A Simple LNRE Model for Random Character Sequences. *Proc. JADT 2004*. 411-422.
- Evert, S./Lüdeling, A. 2001. Measuring Morphological Productivity: Is Automatic Preprocessing Sufficient? In Rayson, P./Wilson, A./McEnery, T./Hardie, A./Khoja, S. (eds.) *Proc. Corpus Linguistics 2001*. Lancaster, 167-175.

# Literatur II

- Goldberg, A. E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Goldberg, A. E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: OUP.
- Goldberg, A. E./Casenhiser, D. M./Sethuraman, N. 2004. Learning Argument Structure Generalizations. *Cognitive Linguistics* 15(3), 289-316.
- Hopfield, J. J. 1982. Neural Networks and Physical Systems with Emergent Collective Computational Abilities. *Proceedings of the National Academy of Sciences of the USA* 79(8), 2554-2558.
- Jackendoff, R. 1987. The Status of Thematic Relations in Linguistic Theory. *Linguistic Inquiry* 18(3), 369-411.
- Jackendoff, R. 1990. *Semantic Structures*. Current Studies in Linguistics 18. Cambridge, MA: MIT Press.
- Kay, P./Fillmore, C. J. 1999. Grammatical Constructions and Linguistic Generalizations: The What's X Doing Y? Construction. *Language* 75(1), 1-33.
- Petig, W. E. 1997. Genitive prepositions used with the dative in spoken German. *Unterrichtspraxis* 30, 36-39.
- Plag, I. 2006. Productivity. In Aarts, B./M.McMahon, A. S. (eds.) *The handbook of English linguistics*. Blackwell Handbooks in Linguistics. Malden, MA: Blackwell, 537-556.
- Weinreich, U. 1966. Explorations in Semantic Theory. *Current Trends in Linguistics* 3, 395-477.
- Zeldes, A. 2009. Quantifying Constructional Productivity with Unseen Slot Members. *Proc. NAACL HLT Workshop on Computational Approaches to Linguistic Creativity*. Stroudsburg, PA: ACL, 47-54.
- Zeldes, A. 2011. On the Productivity and Variability of the Slots in German Comparative Correlative Constructions. In Konopka, M./Kubczak, J./Mair, C./Stícha, F./Waßner, U. H. (eds.) *Grammar & Corpora, 3rd International Conference*. Tübingen: Narr, 429-449.
- Zifonun, G./Hoffmann, L./Strecker, B. (eds.) 1997. *Grammatik der deutschen Sprache*. Berlin: de Gruyter.