Gebrauchsbasierte Graphematik

Prof. Dr. Roland Schäfer

Institut für Deutsche Sprachwisseschaft
Friedrich-Schiller-Universität Jena

FU Berlin, 6. Juli 2023

Graphematik

Graphematik | Kodierung von Grammatik in Schriftartefakten

Phonologische Prinzipien

- /r/ → <r> | *Rat* [ʁaːt], *Bar* [bɑ͡ɐ̯ə]
- /ṭ/ → <tt> | *Matte* [maṭə]

Morphologische/Lexikologische Prinzipien

- Stammkonstanz | <Tritt> [tʁɪt] ← <Tritte> [tʁɪṭə]

Syntax

- syntaktisches Wort → < > | *Haustür*, *die Tür*
- Nebensatzeinbettung → <,> | *Ich weiß, dass es regnet.*

Das Primat der gesprochenen Sprache … Canard extraordinaire!

Schrift ist nicht natürlich, sondern ausgedacht!

- Siehe Geschichte der Schrift.

Gesprochene Sprache ist im Erwerb primär!

- Wieso sollte das heißen, Schrift wäre irrelevant?
- Einfluss der Entwicklung der motorischen Fähigkeiten?
  Stift mit drei Fingern halten: ab 3 Jahre ; Kreis malen: ab 3,5 Jahre

Es gibt Sprachen ohne Schrift!

- Es gibt Atome ohne Elektronen, vergesst Atome mit Elektronen!

Schrift wird nicht spontan produziert!

- Bildungssprache ist immer überformt, auch im gesprochenen Modus.
- Wie wir sehen werden, wird oft sehr spontan geschrieben.

Gebrauchsbasierte Grammatik

Grammatik | Strukturalistisch und Generativ

Aristotelische Kategorien

- Kategorien sind scharf getrennt.
- Wortklassen, Phrasen, Bedeutungen, ...

Sprache als reines System

- Strukturalismus | Sprache als Verteilungsklassen
- Generativismus | Kognitiver Anspruch, de facto algebraisches Regelsystem

Kompetenz und Performanz

- Kompetenz | Reines, aristotelisches Systemwissen
- Performanz | Alle Arten von Fehlproduktion
- Alle anderen Quellen von Variation | Nicht Teil des Sprachsystems

Universalgrammatik ... Öhm?

Gebrauchsbasierte Grammatik

Kognitiv ausgerichtete Gebrauchsbasierte Grammatik:

Bybee & Beckner (2009), Divjak (2016), Kapatsinski (2014), Tomasello (2003)

- Individueller Input und allgemeine kognitive Mechanismen als einzige Faktoren im Spracherwerb

- Lernen auf Basis von Ähnlichkeitsbeziehungen
- Folglich unscharfe Kategorien, z. B. Flexionsklassen, Satztypen
  Schäfer (2019), Schäfer & Sayatz (2016)

- Ungleich erfolgreicher bei der Erklärung von Variation
- Zum Beispiel Einflussfaktoren bei der Wahl von Alternanten (*Zweifelsfälle*)

Unschärfen und Ähnlichkeiten

Sind Adjektive Eigenschaftswörter?

- Systemgrammatische Antwort | Nein, nicht alle Adjektive sind mit *Wie?* erfragbar.

Was sind denn (proto-)typische Adjektive?

- *grün*, *spannend*, *schön*, *sparsam*

Und weniger (proto-)typische Adjektive?

- *gläsern*
- *zahlreich*
- *fiktiv*, *gestrig*
- *obig*, *drei*

Graduelle semantische/grammatische Eigenschaften | Der Prototyp ist das Wie-Wort!

Vermitteln Sie das mal korrekt in der Schule!

Beispiele aus dem Deutschen

Schwache Maskulina | Manchmal stark (Schäfer 2019)

- *des Automaten – des Automats*
- Neigung zur starken Flexion steigt mit
  Entfernung vom Prototypen der schwachen Maskulina
  Menschenbezeichnungen, Nicht-Erstsilbenbetonung usw.

Maßangaben | Enge Apposition oder pseudo-partitiver Genitiv (Schäfer 2018)

- *eine Flasche guten Öls – eine Flasche gutes Öl*
- Neigung zur Kasusidentität steigt, umso weniger die Maßangabe referentiell interpretierbar ist.

Gebrauchsbasierte Graphematik

Was leisten Schrift und Schreibung?

Grammatische Kategorien

- Wortklassen, phonologische Segmente, Silben, Wörter, Phrasen, Nebensätze/Clauses, Sätze, ...

Abbildung auf graphematische Muster

- Großbuchstaben und Kleinbuchstaben, Formen von Buchstaben (wie runde Vokale <e,o,a> und nicht-runde Vokale <i,u>), Abfolgen von Ober- und Unterlängen, Doppelschreibungen, Spatien, Kommata, Bindestriche, Punkte, ...

Achtung: Sowohl die grammatischen Kategorien als auch die Abbildungen werden als Ähnlichkeitsklassen erlernt und sind intrinsisch unscharf!

Ziemlich eindeutige Schreibprinzipien des Deutschen

- Phonologisches Schreibprinzip
  - Jedes zugrundeliegendes Segment wird durch einen Buchstaben kodiert.
- Prinzip der Silbengelenkschreibung
  - Silbengelenke werden ausnahmslos durch Gelenkschreibungen markiert.
- Prinzip der Konstantschreibung
  - Die Stämme eines lexikalischen Wortes werden so ähnlich geschrieben, wie es angesichts der anderen Prinzipien möglich ist.
- Prinzip der Spatienschreibung
  - Syntaktische Wörter werden durch Spatium getrennt.

Was bleibt bei diesen Prinzipien noch offen?

- Prinzipien | Weitreichende Generalisierungen, oft (fast) ausnahmslos
- Sie erklären nicht jede Schreibung. Es gibt viele weitere Muster bzw. „Regeln“.
  - Ist die Substantivgroßschreibung ein Prinzip?
  - Gibt es ein Kommaprinzip? Oder ein Punktprinzip? Ein Bindestrichprinzip?
  - Gibt es ein Klitisierungsschreibprinzip?
- Wenn die grammatische Kategorie unscharf ist, schwankt auch die Schreibung.
  - Siehe die Beispiele im nächsten Abschnitt.

Welche Generalisierungen kommen in Lernergehirnen an?

Völlig gesteuerter Schrift- und Schreibungserwerb?

- Die Prinzipien werden in der Schule nicht oder nur indirekt gelehrt.
- Realer Unterricht | Viel musterbasiertes Lernen mit Reflexionskomponente
- Gebrauchsbasiertes Lernen von Mustern aus (in der Schule normkonformem) Input

  Zumal seit die Methoden von Jürgen Reichen in vielen Bundesländern endlich verboten sind.

  In meiner Kindheit gab es noch den unfassbaren Fehlerteufel.

Die Variation im nicht immer normkonformen ungesteuerten Input führt über das gesamte Leben zu Modifikationen der erlernten Muster.

- So wird *hamma* zur akzeptierten Schreibung.
- Es bilden sich Interpunktionsmuster für *obwohl* und *weil* mit V2-Stellung heraus.
- Ich habe 2022 mein erstes Vorfeldkomma unbewusst gesetzt.
- Manches davon geht mit einer Änderung der grammatischen Kategorie einher.

Gebrauchsbasierte Graphematik

Die Gebrauchsbasierte Graphematik untersucht spontanen Schreibgebrauch (oft, aber nicht immer in Situationen mit geringem Normdruck) und schließt aus Mustern in der Schreibung auf Muster in den verschrifteten grammatischen Kategorien.

Fallstudien

Klitisierte Artikel

Phänomen und Forschungsfragen

Klitisierte Indefinitartikel (Schäfer & Sayatz 2014)

- Verkürzung des Indefinitartikels zu *n*, *nen*, *nem*, *ne* usw.
- Einschließlich erweiterter Formen wie *nen Auto*

Forschungsfragen

- Gibt es Minimalanforderungen an graphematische Wörter?
- Gibt es paradigmatische Effekte?

Beispiele

(1) a. is doch nur n Spiel.
b. Claudia hat nen Schäferhund ...
c. Frag doch mal in ner Werkstatt nach, ob 125 Grad noch ok sind ...

(2) a. Ansonsten ists nen schönes Spiel :)
b. War mal nen guter Tip vom Sattler.

Welche Formen sind strukturell möglich?

Tabelle 1: Strukturell mögliche Kurzformen des deutschen Indefinitartikels

	Femininum	**Maskulinum**	**Neutrum**
Nominativ	*ne*	*n / nen*	*n / nen*
Akkusativ	*ne*	*n / nen*	*n / nen*
Dativ	*ner*	*nem*	*nem*
Genitiv	*ner*	*nes*	*nes*

Verteilung der Kurzformen auf das Paradigma

Tabelle 9: Summierte Koeffizienten des Modells aus Tabelle 8 für Genus und Kasus. Negative Koeffizienten sind grau hinterlegt und zeigen eine Präferenz für Vollformen. Die schwach signifikante Interaktion GenNeut:KasNom wurde eingerechnet.

	Femininum	**Maskulinum**	**Neutrum**
Nominativ	0,630	−1,021	−1,288
Akkusativ	0,862	0,571	−0,630
Dativ	0,250	0,697	0,452

Einfluss des linken Kontexts

Tabelle 8: Koeffizienten und Chancenverhältnisse für das Modell zu Studie 3.

Unabhängige	Signifikanz	Koeffizient	Chancenverhältnis
Intercept (GenFem, KasAkk, Adj0, LinkskAnd)	***	0,862	2,369
GenMask	***	−0,291	0,747
GenNeut	***	−1,492	0,225
KasNom	*	−0,232	0,793
KasDat	***	−0,612	0,542
Adj1	***	−0,444	0,641
LinkskPraep	***	−0,666	0,514
LinkskInt	***	−0,994	0,370
LinkskSo		0,216	1,241
GenMask:KasNom	***	−1,360	0,257
GenNeut:KasNom	·	−0,426	0,653
GenMask:KasDat	***	0,738	2,092
GenNeut:KasDat	***	1,694	5,439

Verteilung der erweiterten Formen im Paradigma

Tabelle 12: Summierte Koeffizienten des Modells aus Tabelle 11 für Genus und Kasus. Positive Koeffizienten zeigen eine Präferenz für *nen*.

	Maskulinum	**Neutrum**
Nominativ	−1,192	−1,135
Akkusativ	1,192	−1,108

Volle graphematische Enklise

Tabelle 13: Verteilung der Zusammen- und Getrenntschreibung in DECOW2012 für die zehn Wörter, die am häufigsten links des getrennt geschriebenen klitisierten Artikels *ne* auftreten (außer Präpositionen und Fälle, in denen Ergebnisse zu erwarten sind, die nicht der Zielstruktur entsprechen, z. B. *wien*). Verbleibende Fehlbelege wurden nicht aussortiert.

	n			**ne**			**nen**		
Linkskontext	**mit Spatium**	**ohne Spatium**	**% ohne**	**mit Spatium**	**ohne Spatium**	**% ohne**	**mit Spatium**	**ohne Spatium**	**% ohne**
noch	19.097	8.694	0,31	122.728	313	< 0,01	62.970	64	< 0,01
so	11.123	110.846	0,91	103.996	21.923	0,17	42.847	4.927	0,10
auch	10.242	2.084	0,17	80.215	89	< 0,01	38.040	31	< 0,01
ist	4.496	1.295	0,22	58.406	25	< 0,01	9.161	5	< 0,01
und	10.500	1.188	0,10	51.310	13	< 0,01	40.338	8	< 0,01
nur	7.077	2.766	0,28	46.348	36	< 0,01	18.750	10	< 0,01
ich	6.574	1.697	0,21	44.388	30	< 0,01	34.417	28	< 0,01
mir	3.669	637	0,15	27.485	13	< 0,01	20.187	8	< 0,01
wieder	3.905	359	0,08	24.199	9	< 0,01	13.517	2	< 0,01
hat	2.234	1.054	0,32	23.698	22	< 0,01	18.073	8	< 0,01

Ergebnisse

- Die Reduktion ist eine Strategie, einen einheitlich einsilbigen Artikel zu realisieren.
- Als Klitisierung ist die Reduktion am Satzanfang dispräferiert.
- Die erweiterte Form *nen* ist nicht funktional motiviert (Kasus), sondern durch eine allgemeine Tendenz, *n* als graphematisch defizitäres Wort zu vermeiden.
- Volle graphematische Enklise tritt vor allem bei unzureichendem graphematischen Wortstatus des Klitikons ein.

- Es gibt Minimalanforderungen an graphematische Wörter.

obwohl und *weil*

Phänomen und Forschungsfragen

Komplementierer wechseln die Wortklasse

- *obwohl* und *weil* können statt VL- auch V2-Sätze folgen.

Forschungsfragen

- Korrektheit von Annahmen über Funktion von Komma, Punkt usw.
- Wortklassenstatus der neuen Partikeln

obwohl und *weil* mit V2

Schäfer & Sayatz (2016)

(1) a. Also ich bleib bei meinen George , obwohl Arashi auch ziemlich
well I stay with my George , although Arashi also rather
lustig ist !
funny is !
I still prefer George although Arashi is also rather funny!

b. Ich habs mir gegeben , obwohl am Sonntag kamen manchmal
I have.it me given , although on.the Sunday came sometimes
wiederholungen vom Samstag …
repeats of.the Saturday …
I watched [all of] it, even though on Sunday they also showed some repeats from Saturday.

obwohl und *weil* mit V2

(2) a. Verschenken geht nur bedingt , weil das ja nicht jedem
make.present goes just limited , because that yes not everybody
gefällt ;-)
pleases ;-)
It doesn't make a good present either because many people don't like it.

b. Ich dachte nur ich komm an den DSLAM da beim alten Kino ,
I thought only I come to the DSLAM there at.the old cinema ,
weil sonst steht hier näher keiner .
because else stands here closer none .
I just thought I could get reception from the DSLAM by the old cinema.
After all, there is no other access point in the vicinity.

Variation der Interpunktion (Beispiele)

(4) a. Oder ich könnte das Altmetall verwerten , obwohl … viel Metall
or I could the scrap.metal use , although … much metal
ist da nicht dran .
is there not at .
Or I could recycle it as scrap metal. But then again, it doesn't contain much metal.

b. wohin , das sag ich nicht , weil : das weiß ich noch nicht .
where , that say I not , because : that know I yet not .
I'm not going to say where, [*simply*] *because I don't know yet.*

Graphematische Unabhängigkeit von Sätzen

- Kein Satzzeichen | Volle Integration
  Phrasengrenze unterhalb der (Neben-)Satzebene
- Komma | Partielle Integration
  Nebensatzgrenze, Satzkoordination mit reduzierter Unabhängigkeit
- Punkt, Ausrufungszeichen, Fragezeichen | Volle Unbahängigkeit
  Satzgrenze

Empirischer Befund I | Satzinitiale Partikeln

(9) a. Klar , der Patient kann auch einfach 2 Tabletten nehmen […]
sure , the patient can also simply 2 pills take
Sure, the patient could equally well just take two pills […]

b. Andererseits , dieses Tuch ist umstritten .
on.the.other.hand , this shroud is debated .
On the other hand, [*the authenticity of*] *this shroud is under debate.*

c. Nun , dieser Anblick beweist , dass der männliche Penis eigentlich
well , this sight proves , that the male penis actually
potthässlich ist .
butt-ugly is .
Well, this sight proves that the male penis is actually butt-ugly.

d. Zugegeben , das sind die Highlights des Religionsunterrichts .
admittedly , that are the highlights of.the religious.education .
Admittedly, these are the highlights of religious education.

Empirischer Befund II/1 | Wortverteilung bei Doppelpunkt

Colon (total 1,244,898)			
Word	**Translation**	**%**	**Count**
PS	P.S.	6.84	85,147
Zitat	quote	5.51	68,600
Edit	edit	4.03	50,203
EDIT	edit	2.38	29,595
Wohnort	place of residence	2.23	27,719
Fazit	summary	2.12	26,364
P.S.	P.S.	1.91	23,725
Also	well	1.4	17,369
Beruf	profession	1.12	13,952
Aber	but, however	1.01	12,611

Empirischer Befund II/2 | Wortverteilung bei Komma

Comma (total 3,191,317)			
Word	**Translation**	**%**	**Count**
Ja	well, yes	7.5	239,380
Naja	well	6.21	198,089
Also	well, now	3.8	121,348
So	now	3.72	118,625
Nein	no	3.51	111,866
Tja	well	1.99	63,381
Sorry	sorry	1.83	58,403
Klar	obviously, yeah	1.64	52,447
Ok	okay	1.46	46,489
Gut	well	1.4	44,729

Empirischer Befund II/3 | Wortverteilung bei Bindestrich

Dash (total 170,789)			
Word	**Translation**	**%**	**Count**
Und	and, furthermore	1.96	3,353
Also	well, now	1.72	2,940
Aber	but, however	1.59	2,711
Naja	well	1.49	2,548
Ja	well, yes	1.38	2,355
So	now	1.09	1,858
Nein	no	0.99	1,698
YouTube	YouTube	0.97	1,664
Tja	well	0.66	1,121
Klar	obviously, yeah	0.65	1,118

Empirischer Befund II/4 | Wortverteilung bei Dreipunkt

Ellipsis (total 210,593)			
Word	**Translation**	**%**	**Count**
Naja	well	5.46	11,488
Hm	hm	4.23	8,916
Hmm	hm	4.2	8,844
Also	well, now	2.91	6,119
Hmmm	hm	2.87	6,039
So	now	2.28	4,796
Aber	but, however	1.83	3,854
Ja	well, yes	1.82	3,823
Tja	well	1.76	3,700
Ähm	um	1.53	3,219

Empirischer Befund III/1 | Links von *obwohl*/*weil*

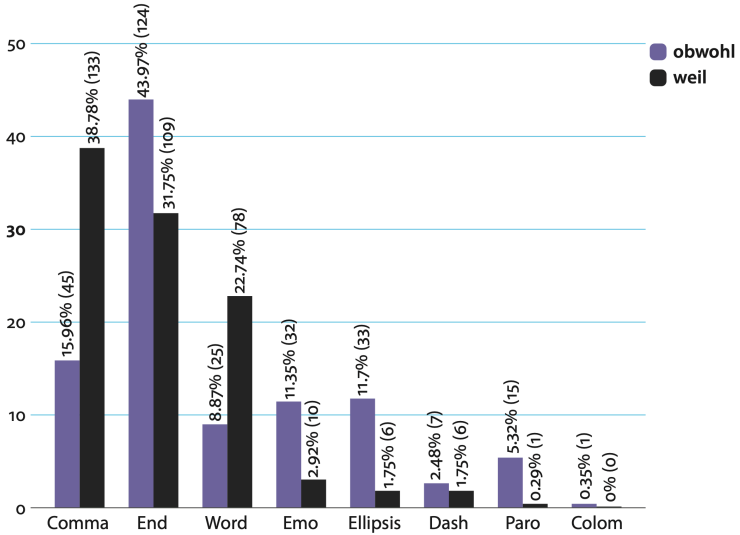


Figure 3. From the dataset for the V2 GLM (Table 6): Counts for the response variable *Subjunctor* (*obwohl* or *weil*) and the regressor *Left*.

Empirischer Befund III/2 | Rechts von *obwohl*/*weil*

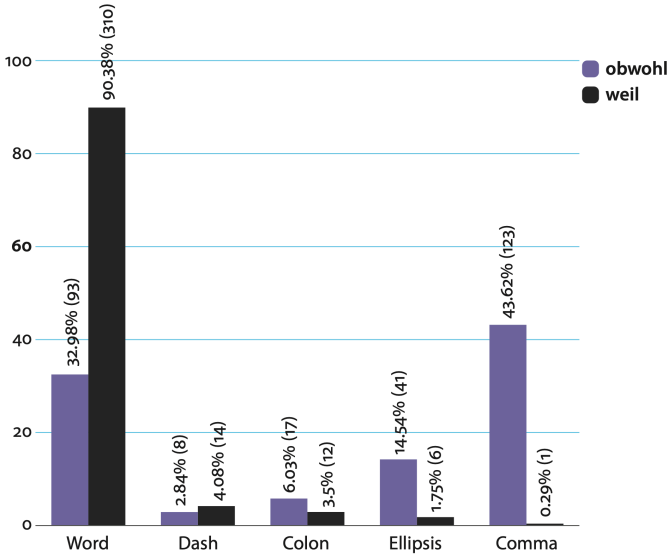


Figure 4. From the dataset for the V2 GLM (Table 6): Counts for the response variable *Subjunctor* (*obwohl* or *weil*) and the regressor *Right*.

Ergebnisse

obwohl und *weil* mit V2-Satzstellung

- *obwohl*
  - Leitet mehr unabhängige Sätze ein
  - Wird öfter vom Komma gefolgt
  - Status | Eher Diskurspartikel außerhalb des Satzes
  - Ähnlich *ja*, *naja*, *also*, *klar*
- *weil*
  - Folgt auch bei V2 eher einem Komma
  - Oder ganz ohne Interpunktion links
  - Komma folgt seltener
  - Status | Eher Konnektor im Konnektorfeld
  - Ähnlich *denn*

Pränominale Herausstellung

Phänomen und Forschungsfrage

Nicht integrierte APs zwischen Artikel und Substantiv in der NP

- *Diese – offensichtlich sinnlosen – Diskussionen über Mutliversen nerven.*

Forschungsfrage

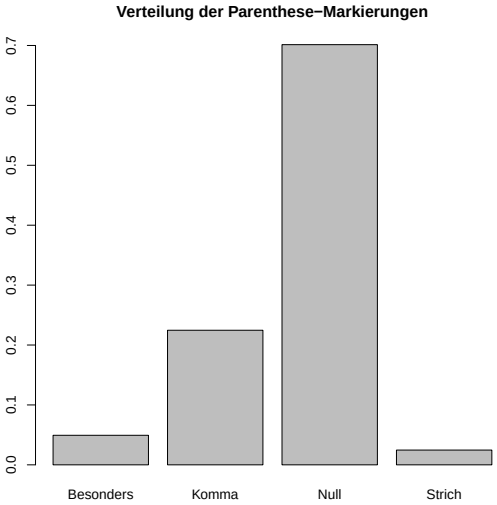
- Genauere Bestimmung der Funktion von Komma, Bindestrich usw.
- Wie strikt wird Herausstellung durch Interpunktion markiert?

Diktatexperiment | Experimentdesign

Sayatz & Schäfer (o. D.)

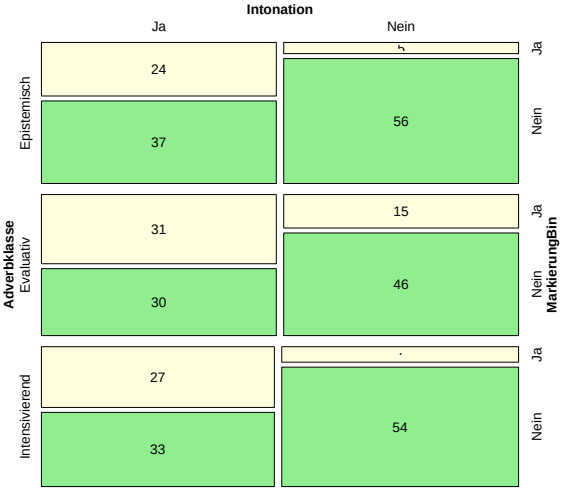
	A	B	C	G	H	M	N	O	P	Q
1	**Index**	**Positio**	**Call**	**Target**	**Experiment**	**Expoadv**	**Expoadj**	**Expoadvtype**	**Intonation**	**Sentence**
2	18	1	Satz 1		FillerWritten					Unter Umständen muss Henrike ihren Ford in die Werkstatt bringen.
3	34	2			FillerNonWritten					Auf Leinwand malt Robin ungern, weil die Technik schwieriger ist.
4	32	3			FillerNonWritten					Saskia hat Montag und Dienstag vergeblich versucht, Tofu zu kaufen.
5	5	**48**	Satz 2	A31	Univerbation					Simone hört ein Hörbuch, während sie am Bogenschießen ist.
6	14	5	Satz 3	B32	Extraposition	Oft	Weiß	Intens	0	**Laura trägt gerne diese oft weißen Tennis-Tops.**
7	20	6	Satz 4		FillerWritten					Felix hat mal wieder seinen Rucksack auf der Anrichte vergessen.
8	40	7			FillerNonWritten					Jan und Jonas sind auf dem Weg nach hause, während Anne und Annika noch diskutieren.
9	29	8			FillerNonWritten					Frank fehlten mehr als einhundert Euro, um einen Biedermeierschrank zu kaufen.
10		9	Frage 1		Question					Hat Saskia versucht, am Donnerstag Tofu zu kaufen?
11	9	10	Satz 5	B11	Extraposition	Endlich	Hart	Eval	1	**Samuel trägt die — endlich harte — Tonvase nach draußen.**
23	12	22	Satz 9	B22	Extraposition	Vermutlich	Wahr	Evid	1	**Nora hörte von Oskar eine — vermutlich wahre — Geschichte.**
32	13	31	Satz 13	B31	Extraposition	Sehr	Nett	Intens	1	**Fabian wird später eine — sehr nette — Freundin treffen.**
46	11	45	Satz 19	B21	Extraposition	Sicherlich	Tief	Evid	0	**Paula und Frida baden in einem sicherlich tiefen See.**

Diktatexperiment | Ergebnisse (1)



Diktatexperiment | Ergebnisse (2)

Verteilung der Markierungen nach Bedingungen



Ergebnisse

- Markierung mit Komma und Gedankenstrich, meistens aber nichts
- Komma im Experiment präferiert, in der Korpusstudie aber über 75% Gedankenstrich
- Intonation verstärkt Tendenz zu Markierung
- Die Adverbklasse wirkt auch als Einflussfaktor
- Was ist denn nun die Funktion des Kommas?

Probleme und Ausblick

Alles auf einmal | Eine Methode, zwei Theorien

Wir testen immer Mengen von Hypothesenfamilien!

- Was ist die relevante grammatische Kategorie?
- Haben wir das grammatische Kategorienspektrum korrekt erfasst?

- Sind grammatische Kategorien unscharf und ähnlichkeitsbasiert?
- Stimmen die Grundannahmen der Gebrauchsbasierten Graphematik?

- Passt unser experimentelles Paradigma zu den Hypothesen?
- Repräsentiert unser Korpus einen konsistenten und relevanten Schreibgebrauch?

- Das muss man alles irgendwie voneinander trennen.

Ausblick

- Mehr Phänomene
- Mehr unabhängige Tests (grammatische Kategorie vs. graphematische Abbildung)
- Replikationen von Studien
- Individuelle Variation (korreliert mit bildungssprachlichen Fähigkeiten usw.)
- Mehr Experimente, weniger Korpora

Anhang

Literatur I

Bybee, Joan L. & Clay Beckner. 2009. Usage-based theory. In Bernd Heine & Heiko Narrog (Hrsg.), *The Oxford Handbook of Linguistic Analysis*, 827–856. Oxford: Oxford University Press.

Divjak, Dagmar. 2016. Four challenges for usage-based linguistics. In Jocelyne Daems, Eline Zenner, Kris Heylen, Dirk Speelman & Hubert Cuyckens (Hrsg.), *Change of Paradigms – New Paradoxes – Recontextualizing Language and Linguistics*, 297–309. Berlin/Boston: De Gruyter Mouton.

Kapatsinski, Vsevolod. 2014. What is grammar like? A usage-based constructionist perspective. *Linguistic Issues in Language Technology* 11, 1–41.

Sayatz, Ulrike & Roland Schäfer. o. D. Gebrauchsbasierte Analyse der Interpunktion bei pränominaler Herausstellung. in Vorbereitung.

Schäfer, Roland & Ulrike Sayatz. 2014. Die Kurzformen des Indefinitartikels im Deutschen. *Zeitschrift für Sprachwissenschaft* 33(2), 215–250.

Schäfer, Roland & Ulrike Sayatz. 2016. Punctuation and Syntactic Structure in “Obwohl” and “Weil” Clauses in Nonstandard Written German. *Written Language and Literacy* 19(2), 212–245.

Schäfer, Roland. 2018. Abstractions and exemplars: The measure noun phrase alternation in German. *Cognitive Linguistics* 29(4), 729–771.

Schäfer, Roland. 2019. Prototype-Driven Alternations: The Case of German Weak Nouns. *Corpus Linguistics and Linguistic Theory* 15(2), 383–418.

Tomasello, Michael. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard: Harvard University Press.