

Corpus-Based Discovery Learning in an Advanced L2 Grammar Course

Nina Vyatkina

University of Kansas

CALICO May 21, 2011
Victoria, BC



Purpose

- Report on the design and implementation of an advanced German grammar review course that followed a usage-based approach and included authentic materials

Usage-Based Grammar Theories

- Bybee (2006):
- grammar is an abstract cognitive representation of language
- yet strongly tied to the experience that a speaker has had with language
- the use of language is lexically particular; certain words tend to be used in certain collocations or constructions
- => inseparability of lexis and grammar
- => language corpora for enhancing experience

Benefits of corpora

- Empirical foundation (data, examples)
- Distribution and frequency
- Text types, register, mode
- Wide range of vocabulary options, alternatives
- Phraseology and collocation
- Authenticity

Potential benefits for pedagogy

- data-driven learning - learning by discovery - inductive learning (Johns, 1986; Bernardini, 2002; Mishan, 2004)
- explicit learning (Bernardini, 2000)
- analytical skills (Dodd, 1997; Fox, 1998)
- consciousness-raising (e.g., noticing semantic preference and semantic prosody: Bernardini, 2001)
- problem-solving abilities (Johns, 1991, 1997; Qiao & Sussex, 1996; Kennedy & Miceli, 2002)
- metacognitive knowledge (O'Sullivan, 2007)
- independent language learning strategies (Bernardini, 2000; Qiao and Sussex, 1996)

Corpus literacy

(Mukherjee, 2006; O'Sullivan, 2007)

- can enhance a process-oriented approach to language teaching and learning (O' Sullivan, 2007; O' Sullivan & Chambers, 2006)
- cognitive skills: predicting, observing, noticing, thinking, reasoning, analyzing, interpreting, reflecting, exploring, making inferences (inductively or deductively), focusing, guessing, comparing, differentiating, theorizing, hypothesizing, and verifying

Methodological difficulties

- Authenticity vs. prototypicality
 - corpus examples are ‘genuine’ but not ‘authentic’ for language learners (Widdowson, 2000; 2003)
 - “authenticity is in the eyes of the participants” (Prodromou, 1996)
 - frequency does not reflect prototypicality (cognitive salience)
- Representativeness, generalizability
 - the ‘attested’, not the ‘possible’
- Language learners are not expert corpus users
 - “most widely accessible corpora were created as tools for linguistic research and not with pedagogical goals in mind” (Brown, 2007)
- => **importance of pedagogical mediation**

Available teaching materials

- Dictionaries and student grammars:
 - Collins COBUILD Advanced Dictionary
 - Macmillan English dictionary for advanced learners (Rundell, 2007)
 - Dodd (2003)
- Teacher guides:
 - McCarthy, McCarten, & Sandiford, 2005; Bennett (2010); Vyatkina & Johnson (2007); Dushku & Thompson (in preparation)
- Articles:
 - McCarten, 2010; McCarthy (2004); Greaves and Warren (2007); Dodd (1997)
- **Scarcity of actual examples of exercises!**

Empirical studies

- Corpus-based exercises a viable alternative to gap-filling exercises? Yes (Cobb, 1997; Stevens, 1991)
- Improved learning? Inconclusive (Brown, 2007; Vannestal & Lindquist, 2007)
 - factors: proficiency levels; cognitive maturity; learning styles
- Student reactions (O'Sullivan, 2007; V & L, 2007)
 - + corpus consultation as a means of improving writing
 - formulating queries; technological difficulties
- Teacher reactions (Breyer, 2009; Heather & Helt, 2011)
 - + enthusiasm about exposing learners to authentic language
 - + improved understanding of reading and analyzing corpus output
 - difficult to find corpora and formulate tasks
 - difficult to create teaching materials

Corpus-based grammar teaching

- mostly neglected in corpus-based teaching (see, however, Dodd, 1997; Greaves & Warren, 2007; Vannestal and Lindquist, 2007)
- Nelson et al. (2002): a corpus is “a wondrous resource for teaching grammar”
- Schulze (2010)
- Kaltenboeck & Mehlmauer-Larcher (2005):
 - knowing basic word classes may be enough
 - even less sophisticated grammar knowledge needed than for consulting reference grammars

Suggested goals and activities

- Comparison of textbook materials and ‘real-life’ language use (Gavioli and Aston 2001; Tognini-Bonelli 2001; Sinclair 2004; O’Keeffe et al. 2007)
 - validation of lexico-grammatical syllabus choices
 - verification of teacher and learner intuitions
- Autonomous research by advanced learners (Mauranen, 2004; Bennett, 2010)
- Establish appropriateness instead of rules (Seeger, 2010)
- Frequent L1-L2 incongruences (Seeger, 2010)

Direct and indirect applications

- Indirect applications:
 - Selective concordances (“clear-cut examples”)
 - Ready-made frequency lists
- Direct applications:
 - Learners search corpora directly
 - under teacher’s guidance
 - on their own
- => **various degrees of teacher mediation and learner autonomy**

The course

- 5 graduate students and 2 advanced undergraduate students (seniors) in German
- 3 hrs a week, conducted in German
- Structural syllabus, followed textbook
- Materials
 - Rug & Tomaszewski, *Grammatik mit (Un)Sinn und Verstand* ('Grammar with common (non)sense')
 - Sick, *Der Dativ ist dem Genitiv sein Tod* ('The dative is to the genitive its death').
 - additional materials
 - **Native speaker German corpora**

Course goals

- understand grammar as “the cognitive organization of one’s experience with language” (Bybee, 2006) and not only an abstract system of rules
- discover patterns of use and establish meanings of lexico-grammatical means of expression in different contexts
- understand language variability and choose appropriate linguistic patterns from the options available to speakers of German

Course goals (cont.)

- compare grammar as presented in course books and as used in authentic texts
- independently work with authentic German texts to analyze grammar usage and to create teaching materials
- formulate and solve linguistic problems
- develop, explore, and reflect on using strategies of learning a language at an advanced level

Corpus literacy

- Mukherjee, 2006:
 - Understand what can & can't be done
 - Know how to “read” and analyze concordances
 - Make sense of basic frequency info
 - (Use output to generate materials)

Corpus: DWDS

- Digital Dictionary of the 20th/21st Century German Language
- Berlin-Brandenburg Academy of Sciences
- Several corpora: Core corpus (Kernkorpus); newspaper corpora; special corpora
- <http://www.dwds.de/>

The core corpus

- 100 million words: balanced chronologically and by text genre
- Time: 10 mill. each decade
- Genres: journalism, literary texts, scientific literature, other non-fiction (cookbooks, guides) – ca. 25% each
- Annotated for parts-of-speech
- Lemmatized
- In-built search and analysis tools

Pedagogical sequence

(Breyer, 2009; Vannestal & Lindquist, 2007)

- introduction to *DWDS* and corpus research, query syntax
- introductory exercises initially based on printed-out concordance lists (indirect approach)
- hands-on exercises on the computer: explore problematic grammatical areas (direct approach)
- formulate one's own corpus queries
- develop teaching activities
- post-course survey (based on Heather & Helt, 2011)

1. Modal verb frequency

- Go to <http://www.dwds.de/>
- Type *sollen* in the search box; perform search; write down the number of hits
- Repeat these steps for *dürfen*, *müssen*, *wollen*, and „*brauchen nicht zu*“
- Rank the modal verbs according to their frequency in the corpus

2. Modality: concordances

1	Ze	1999	... Das ist knefisch - und einmalig. Davor	<u>sind</u>	sämtliche Hüte zu ziehen . Der ...
2	Ze	1999	... mich an einen "Ostsender" wenden. Von derlei	<u>wäre</u>	noch immer viel zu berichten
3	Ze	1999	... zusammenhängen sollen. Was	<u>ist</u>	von den Warnungen zu halten ? ...
4	Ze	1999	... glücklich macht. Das können sein: - Letztere	<u>sind</u>	nur als Halbwüchsige zu gebrauchen ; ...
5	Ze	1999	... die geistige Opposition trieb. So viel Dummheit	<u>war</u>	einfach nicht zu ertragen . Wie ...
6	Ze	1999	... extra auf gelbem Papier drucken ließ. Zunächst	<u>ist</u>	wichtig zu wissen : "Heimfahrten ...
7	Ze	1999	... mehr in diesem Ausmaß weiterzuentwickeln. Was	<u>ist</u>	zu tun ? Abspecken. ...
8	Ze	1999	... noch verpönte Selbstbedienung der Altaktionäre	<u>sei</u>	seit Jahresbeginn häufiger zu ...
9	Ze	1999	... Zone zurückzuziehen. Nach weiteren drei Tagen	<u>sei</u>	Zentralkosovo zu räumen und für den ...
10	Ze	1999	... Umbruchzeiten sondergleichen. Ohne harte Arbeit	<u>sind</u>	solche Zeiten nicht zu bestehen
11	Ze	1999	... Was aber bringt das Ganze? Diese Frage	<u>ist</u>	meist nur abstrakt zu beantworten ; ...
12	Ze	1999	... Sun-Express-Maschine nach Antalya fort. Viel	<u>ist</u>	da nicht zu holen . Ungefähr ...
13	Ze	1999	... die Touristen den ganzen Tag. Nichts	<u>war</u>	zu sehen von der in den Prospekten ...
14	Ze	1999	... das längst vorgemacht. Nehmen Sie das Handy: Es	<u>ist</u>	sehr leicht zu bedienen . Um die ...
15	Ze	1999	... lieben und von ihnen geliebt werden wollen. Sie	<u>sind</u>	nicht zu verwechseln mit sexuell ...
16	Ze	1999	... und Internet-Präsentationen für Studenten	<u>ist</u>	bei Etudiants-Internet-Emplois zu ...
17	Ze	1999	... Nitrosamine oder Benzo (a) pyrene. Diese	<u>wären</u>	relativ leicht zu reduzieren
18	Ze	1999	... zur Arbeit motiviert werden. Da	<u>ist</u>	tatsächlich viel zu verändern
19	Ze	1999	... im Frankfurter Krankenhaus Nordwest. Dabei	<u>ist</u>	der Säurerückfluß gut zu ...
20	Ze	1999	... man kein Vertrauen bei der Bevölkerung. Was	<u>ist</u>	zu tun ? Kompromisse müssen ...
21	Ze	1999	... der Wüste; der Dissens zwischen Metz und Ratzinger	<u>war</u>	nicht zu schlichten . Beide ...
22	Ze	1998	... zu verkaufen, für unter ihrer Würde. Die Kluft	<u>ist</u>	kaum zu überbrücken
23	Ze	1998	... Metallanteil anzusprechen. Vom Auto aus	<u>sei</u>	das kaum zu schaffen . Vermintes ...
24	Ze	1998	... Hemd herumzubürsten und zu klopfen. Da	<u>ist</u>	nichts zu machen . Sascha findet ...
25	Ze	1998	... vor denen Brennholz liegt. Kein Mensch	<u>ist</u>	zu sehen . Der Kopilot gibt mir ...

2. Modality: concordances

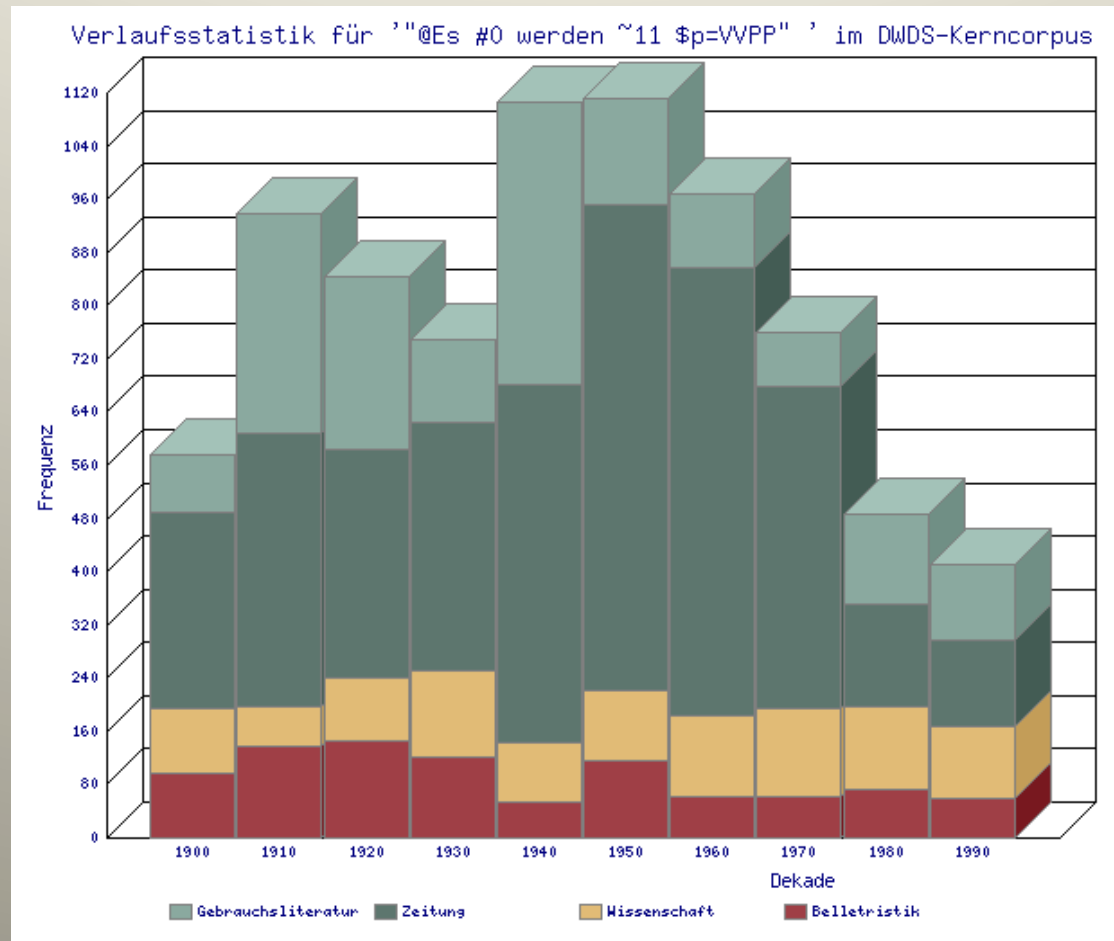
- Read the sentences with the construction *sein + zu + Infinitiv* in lines 5, 10, 12, 14, 21.
- Do they express possibility or necessity?
- Reformulate the sentences, replacing this construction with modal verbs.
- Read the sentences with the construction *sein + zu + Infinitiv* in lines 30, 31, 32, 39.
- Do they express possibility or necessity?
- Reformulate the sentences, replacing this construction with modal verbs.
- In what style / register can these constructions occur (written language, conversation, neutral register)?

3. Impersonal constructions: frequency and distribution

- Search DWDS for each construction from ex. 9-12 (grammar book) and fill out the table.
- How often is each construction used? (hits)
- In what text types is it used more frequently?
- How does the use of each construction change over time?

construction	search string	hits	text type	dynamic	peaks
passive voice	werden && \$p=VVPP	285173	jour/non-fi/ sci	up and down	1940s
<i>man</i> ('one') constructions	man	225244	equal	down, 1990s up	-
<i>sich lassen +inf</i>	lassen && @sich && \$p=VVINF	14604	jour/non-fi/ sci	stable, 1990s up	1990s in jour
<i>sein + zu +inf</i>		5707	jour/fi	down, 1990s up	

3. Impersonal constructions: frequency and distribution



4. Colligations

- *da*-compounds: pronominal adverbs (*davon, dabei, daraus, daran...* - ‘thereof’, ‘thereby’)
- important cohesive devices in German texts
- textbooks: examples of anaphoric uses:
 - Es sei unmöglich, sie auf den Abnehmer abzuwälzen, denn **dadurch** würde eine Preiserhöhung eintreten, welche den deutschen Film teurer - machen würde
- cataphoric uses mostly ignored:
 - Diese aber wurde letztlich determiniert **dadurch**, daß eine dem Denken über den Sinn der Welt um seiner selbst willen nachgehende Literatenschicht der Träger der Soteriologie war.
- each student exported 10 random examples for one compound and analyzed both anaphoric and cataphoric uses; in-class presentation and discussion

5. word order – punctuation: connectors

... Eva Schmiedinger warf ihm wütende Blicke zu	↳	doch Kronbecher erteilte ...
... doch schneller bearbeitet werden"	↳	und Krause repliziert: ...
... Daß Globalisierung nicht nur De-Lokalisierung meint	↳	sondern Re-Lokalisierung ...
... prinzipiell am Phänomen der Kommunikation vorbei	↳	denn Kommunikation findet ...
... daß dies zu einer errechneten Kultur führen wird	↳	denn Sinnformen kondensieren ...
... eine Grenze zu ziehen, deren Kreuzen möglich ist	↳	aber Zeit erfordert. ...

... verwandelt. Wir saßen noch über den Vorspeisen	↳	als die Tür aufging und Eve ...
... während des Krieges kam, in Amerika, in New York	↳	als er anfang , Jazz zu hören, ...
... wahrnahm, weil es in der Küche - wie immer	↳	wenn der Ofen lief - furchtbar heiß ...
... Tonfall, den sein Sohn Gary so haßte	↳	weil er dann nie wußte , ...

... und Vereinbarungen schließen. Das ist der Weg	↳	so hoffen wir, der auch in ...
... (Bundesministerium des Innern) weisen darauf hin	↳	bisher gebe es in der zweiten Phase ...
... was sie eigentlich wollten. AM Baker wirft ein	↳	also handele es sich aus der Sicht ...
... SU ins Feld führen. Der Bundeskanzler erwidert	↳	möglicherweise handele es sich hier ...
... Schiene, die Brummi-Lawine rollt	↳	schon mehr sich Staus ...
... tun. Wenn Probleme auftauchten	↳	dann gab man denen ...

6. Register: connectors

	Hits per million	
	Core corpus (written)	Wende corpus (spoken)
<i>trotz</i> (prep, 'in spite of')	108	50
<i>obwohl</i> (conj, 'although')	81	280
<i>infolge</i> (prep, 'as a result of')	62	0
<i>so dass</i> (conj, 'so that')	4.5	430
<i>im Fall(e)</i> (prep, 'in case of')	30	0
<i>falls</i> (conj, 'if')	49	25

⇒ “nominal style” – NPs + prepositions – written/formal

⇒ “verbal style” – clauses + conjunctions – spoken/neutral

7. Spelling: compound verbs

- One word or two words?
- *kennenlernen, spazierengehen, stehenbleiben, verlorengelassen*
- *kennen lernen, spazieren gehen, stehen bleiben, verloren gehen*
- (get to know, go for a walk, stop / stand still, get lost)

7. Spelling: compound verbs

- Grammar book:
 - personal preference of native speakers
- One word
 - more frequent + rising dynamic
- Two words
 - less frequent + falling dynamic

8. Integrated searches; word profile

DWDS-Wortprofil 2010

Wortprofil für **Tendenz** als **Substantiv** Frequenz:6550 [Relationenfilter](#)

Alle | [Akkusativobjekt](#) | [PP](#) | [Attribut](#) | [Beiordnung](#) | [Infinitivobjekt](#) | [Dativobjekt](#) | [Subjekt \(Aktiv\)](#) | [Subjekt \(Passiv\)](#) | [Genitivmodifizierer](#)

Gegentendenzen Latenz Richtungen **Strömungen** Trends
 abnehmender anhält ansteigende antidemokratische
 antisemitische apologetischen aufweist beobachtende
 destruktiven entgegenwirken **fallende** faschistoide feste

freundliche **gegenläufige**
inflationistischen inflationären
 inflatorischen innewohnende isolationistischen judenfeindliche

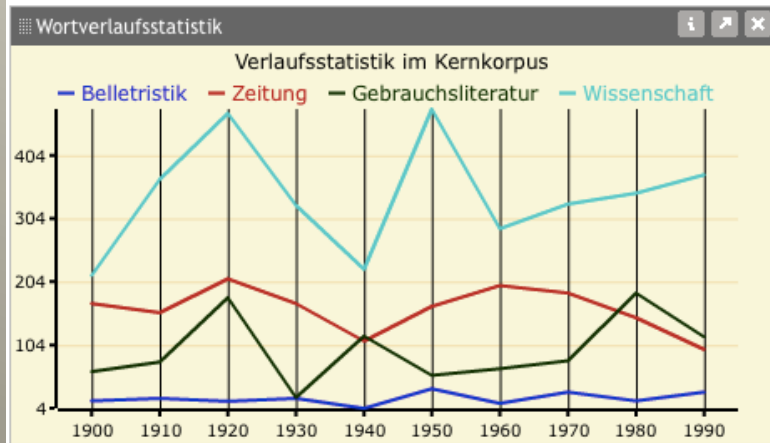
Wortart: **Substantiv** [Zeige Tags](#) | [Tabellenansicht](#)

Die ZEIT & ZEIT Online

Treffer: 18275 [1](#) [2](#) [3](#) [4](#) [>>](#)

- ...em Roman Neue Leben diese **Tendenz** sehr bissig auf den Punkt...
- ... der deutschen Politiker. **Tendenz** zunehmend. Die Leute m...
- ...zu jenen zeitgenössischen **Tendenzen**, die Design als hochprei...
- ...hrzeuge. Besonders die **Tendenz** ist erschreckend. Im D...
- ...bekannt, sondern nur eine **Tendenz**. Insgesamt habe das K...
- ...hinesen - mit zunehmender **Tendenz**.
- ...ür, die nationalistischen **Tendenzen** unter den Chinesen zu för...
- ...der das Ziel von Hackern, **Tendenz** steigend. Im vergangen...
- ...an Niggemeier sieht diese **Tendenz** kritisch. Die Diskussi...
- ...ür, die nationalistischen **Tendenzen** unter den Chinesen zu för...

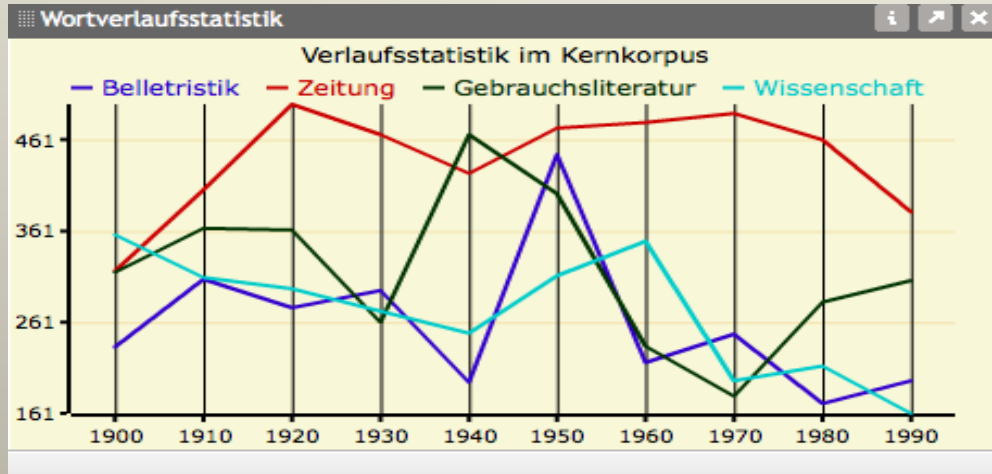
[DDC-Query](#) | [Darstellungsoptionen](#) | [Suchfilter](#)



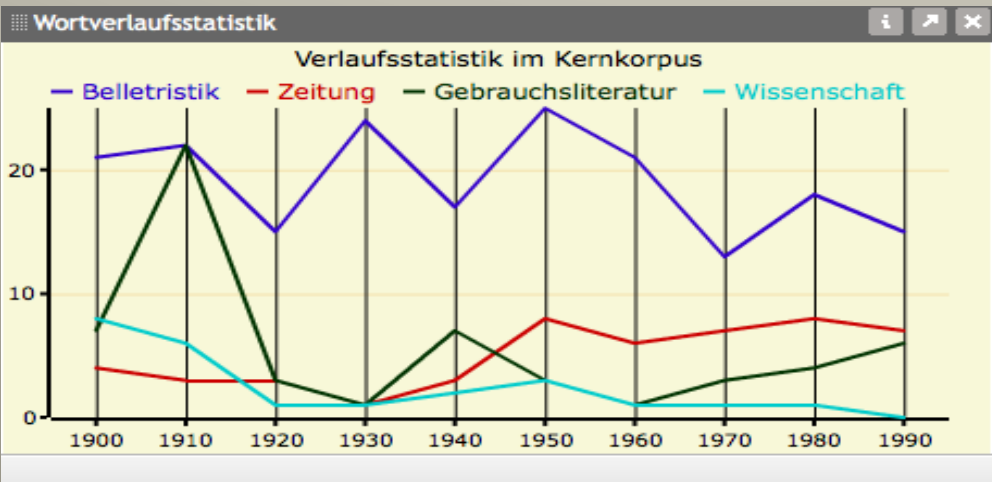
Korpusfrequenzen

Korpus	Hits	Hits [ppm]	Korpusgröße [Mill. Token]
Berliner Tagesspiegel	3617	21.8963	165.19
Berliner Zeitung	5006	19.7579	253.37
C4-Korpus	1842	23.0250	80.00
Compact Memory Corpus	1974	75.2075	26.25
DDR-Korpus	498	57.5281	8.66
Die ZEIT & ZEIT Online	18275	39.7283	460.00
DWDS-Kernkorpus	5321	52.8921	100.60
DWDS-Korpus21	46	24.6169	1.87
Juilland-Korpus	26	52.0000	0.50
Potsdamer NN	272	17.2247	15.79
Wendekorpus normalisiert	3	10.6583	0.28

9. Genre: subjunctive (student-created)



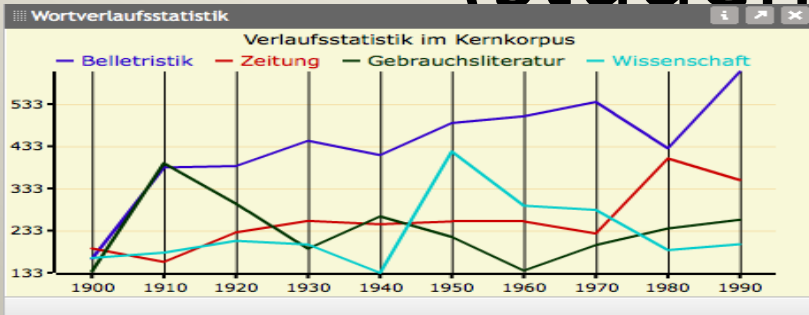
gäbe
(subjunctive I)



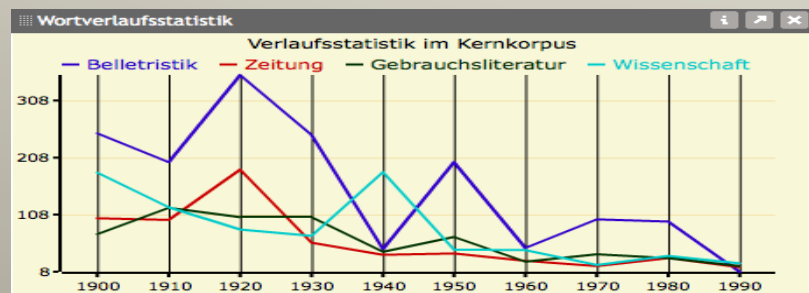
würde geben
(subjunctive II)

10. Time: synonyms ('although')

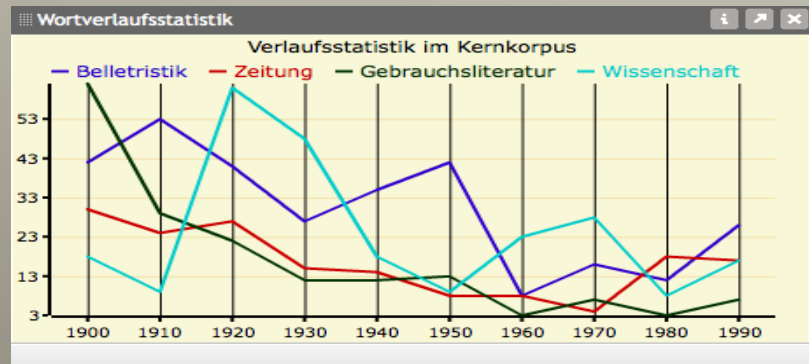
(student-created)



obwohl



obgleich



obschon

How confident are you about your ability to:

	highly	conf.	unsure	not conf.
Understand frequency and distribution data	4	3		
Read and analyze concordance lines	4	3		
Read and analyze “word clouds”	6	1		
Perform a search for specific words	4	2	1	
Perform a search for specific gram. constructions	1	3	2	1
Identify suitable features for corpus search	3	3	1	
Use corpus data to evaluate teaching materials	2	5		
Create corpus-based teaching materials	2	1	2	
Apply frequency and distribution data to teaching	1	4	1	

Will you use corpora in the future to:

	most likely	maybe	unsure	not likely
explore DWDS and/or other corpus sites for better understanding of German language use	6	1		
consult DWDS and/or other corpus sites when you write in German	6	1		
evaluate published teaching materials	2	4		1
generate authentic language samples for teaching	4			
Teach your students to answer their questions about language using language corpora	1	3		

Benefits of corpus-based activities

- Check intuitions
- Formulate queries for grammatical searches
- Insights into frequency of grammar forms and nuances of meaning
- See actual examples of usage
- Resource for teaching (materials, answering student questions)
- Learn about phraseological expressions
- “Collect historical and linguistically analytical data”
- Statistics

Challenges / limitations

- formulating grammatical queries
- mostly concentrated on written language, need spoken corpora
- “hard to listen to other people presenting data when it is only facts and not reasons”
- Need high proficiency in German, not suitable for beginners
- statistics

Summary of survey results

- All students achieved 'corpus literacy' in reading and analyzing distribution, frequency, and concordance data (=Heather & Helt, 2011)
- Difficulties: formulating grammar queries, finding appropriate questions
- All students will engage with corpora in the future
- All who plan to teach German will use corpora in teaching and are confident in doing so (better than H&H, 2011)

Conclusion

- Course goals accomplished
- “corpora are great sources of serendipitous findings” (Stewart, Bernardini, & Aston, 2004)
- Corpus-based advanced grammar courses have great potential (including teacher training), provided
 - careful preparation
 - attention to handouts and worksheets
 - a lot of modeling; logical progression
 - homogenous student population
 - “needs-driven corpora, activities and methodologies” (Brown, 2007)

Large annotated NS corpora

- English
 - <http://corpus.byu.edu/bnc/>
 - <http://corpus.byu.edu/coca/>
- Spanish
 - <http://www.corpusdelespanol.org/>
- Russian
 - <http://www.ruscorpora.ru/en/>
- German
 - <http://www.ids-mannheim.de/cosmas2/>
 - <http://www.dwds.de/>