

ICAME 37
"Corpus Linguistics across Cultures",
Hong Kong, China
26-29 May 2016

Towards Non-Native Metalanguage Standards in MA theses:
an Empirical Comparison of
German, (South) African and Chinese Corpora

Josef Schmied
English Language & Linguistics
Chemnitz University of Technology
https://www.tu-chemnitz.de/phil/english/sections/ling/presentations_js.php
josef.schmied@phil.tu-chemnitz.de



1 Introduction

1.1 Research Team: project in progress

China: Jessica Küchler/Dheskali, Sven Albrecht, Guo Ya, Zhang Xinlei

Cameroon: Daniel Nkemleke, Gabriela Cosmina Djelle, Jenniver Zschocke

Chemnitz: Jessica & Vincenzo Dheskali, et al.

1.2 Combination of research methodologies

- corpus-linguistic comparisons
- (socio-linguistic/attitude) on-line questionnaires
- eye-tracking (for text processing of crucial features)

1.3 Wider context of empirical corpus research on usage standards

1.4 MA theses as 2nd/1st original/research texts with individual and disciplinary conventions

1.5 Academic writing “across cultures” (EU, Asia, Africa?)

popular web definition by James Spradley:

“Culture is the acquired knowledge people use to interpret experience and generate behavior.”

Where does “cultural variation” (in metadiscourse) come from in early academic knowledge presentation?

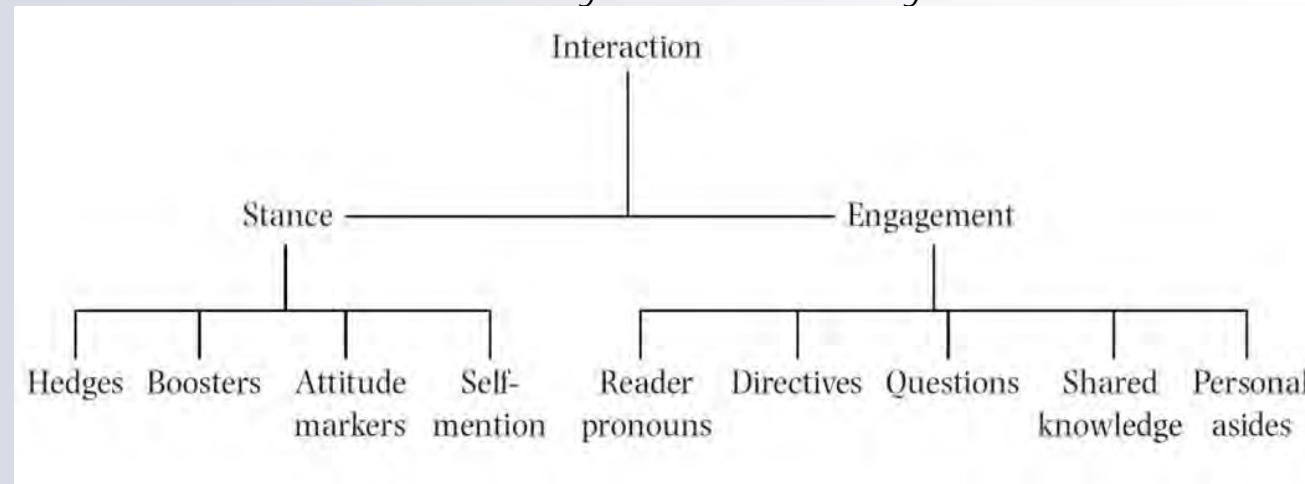
academics’ “socialization/acculturation” as

- gender?
- mother-tongue culture?
- national culture?
- institutional acculturation in the discipline/department

2 Concepts

2.1 Metalanguage

- metalanguage=metadiscourse (Hyland 2005a): writer-reader interaction; linguistic devices that assist writers to organize propositions and present them so that will be easily understood by readers/convince readers



types of reader-writer interaction (Hyland 2005b: 177)

- academic writing in advanced language/*lingua franca* learning (MA)
- international (Swales/Feak 2012) and national? (Siepmann et al. 2011)
textbooks for teaching?

2.2 Metalanguage features discussed here

coherence: cohesion

structuring devices (e.g. conjuncts *and*, *but*, *then*, *because*)

author commitment/stance: hedges – boosters

modal auxiliaries/2 (only epistemic!)

author-reader interaction: personal pronouns (*I/we* – *you*)

2.3 From usage standards to functional standards

different from eWAVE 235 features in L1 and L2 varieties (incl. HK, Singapore)

Hong Kong English

Hong Kong English is the variety of English spoken as a second language by 187,281 people (quoted from the latest population by-census in 2006) in the Hong Kong Special Administrative Region. Hong Kong is basically a monoethnic society with over 95 percent of its total population being Chinese. Cantonese is spoken as a first language and has long been viewed as the language of solidarity and community ties. However, the use of English is increasingly common in certain socially conditioned contexts from the 1980s and 1990s onwards. There are a large number of Filipino domestic helpers in Hong Kong, who speak English with their employers, which makes it necessary to use English in the home. In fact, the 2006 By-Census shows that the percentage of the population claiming to speak English as either their usual language or as an additional language rose from 38.1% in 1996, to 43% in 2001, to 44.7% in 2006. Now that English is being increasingly put in contact with the languages of the indigenous populations in domestic environments, it appears true enough to suggest that Hong Kong English is a variety with its own norms and its own local speech community.

Showing 1 to 5 of 5 entries (filtered from 235 total entries)

← Previous 1 Next →



No.	Feature	Value
Search	modal	Search
121	Double modals	D - attested absence of feature
123	Present tense forms of modals used where StE has past tense forms	A - feature is pervasive or obligatory
125	New quasi-modals: core modal meanings	D - attested absence of feature
126	New quasi-modals: aspectual meanings	D - attested absence of feature
127	Non-standard use of modals for politeness reasons	A - feature is pervasive or obligatory

2.4 Genres in academic discourses

research "output"

- research article
- book reviews
- project proposals
- conference presentations

instructional disc./e-learning

- ppt presentations
 - lectures
 - student presentations
- textbooks
- Wikis
- www pages (HTML, php)

discipline-specific
culture-specific

science "journalism"

- popular science articles
- popular blogs (David Crystal)
- popular science films (Horizon)
- popular science books
- science slam

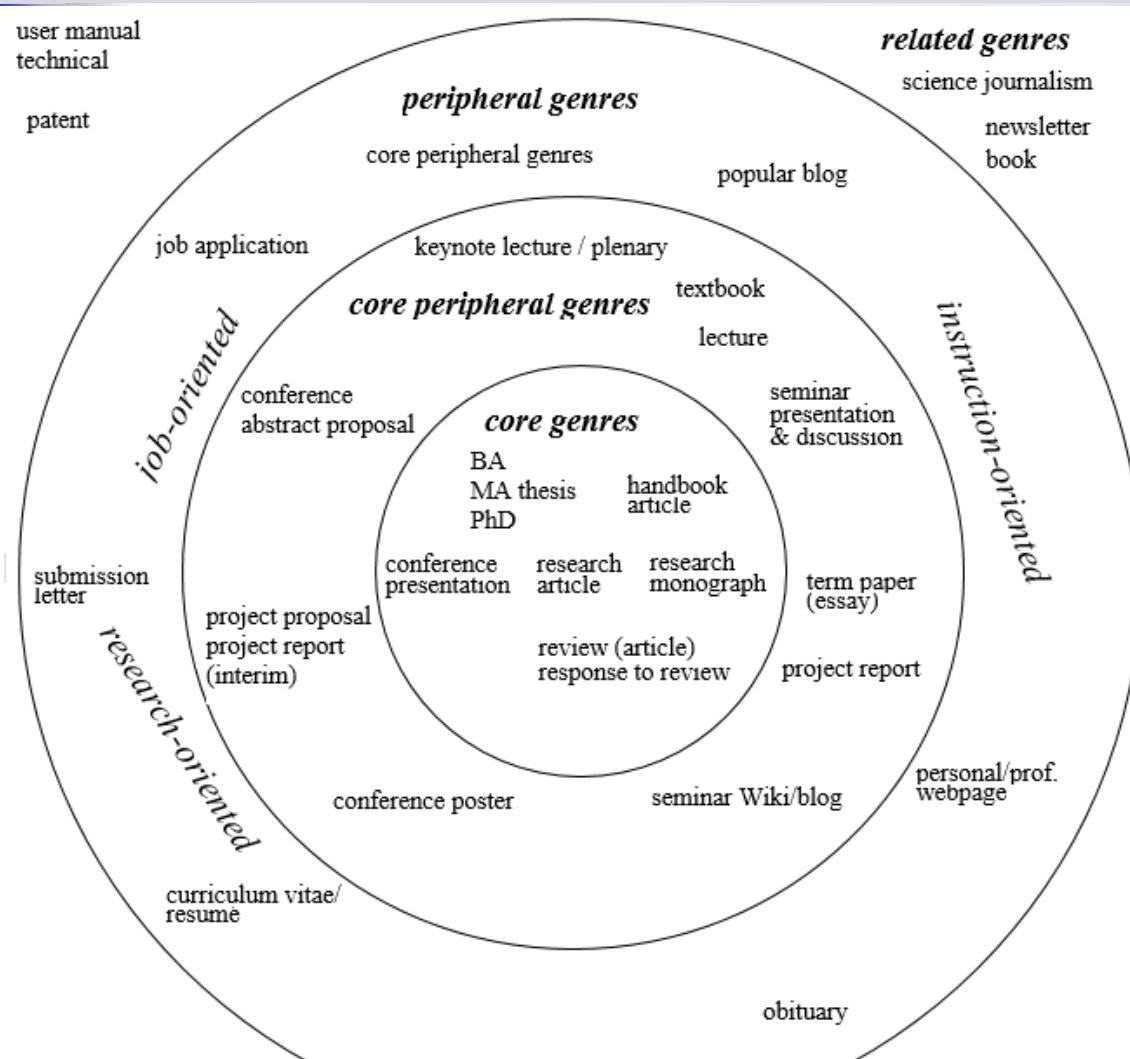
student "literacy": norm-developing

- fieldwork notes, reports
- essays / term papers
- **MA/BA/PhD thesis**
- seminar presentations, disc.

"Novice Academic English"

author-specific
culture-specific

2.5 Prototype approach to genres in academic writing



Schmied, J. (2015: 11)

3 Data

3.1 Compatible stratified corpora (2004-16ff)

	ChemCorpus	ChAcE Corpus		Africa		total
<i>country</i>	Germany	China		South Africa	Cameroon	
		mainland	HK			
<i>university-stratification</i>	Chemnitz	+	+	Stellenbosch	Yaoundé	
<i>discipline-stratification</i>	- (English)	- (English)	+	+	- (English)	
<i>BA theses</i>	80 + 80					
<i>MA theses</i>	40 + 40	304	100?	150	120	1000
<i>PhD theses</i>		106				
<i>other: term paper +</i>	100 + 100	(SYSU 50)				
<i>total</i>	?	?	?	?	?	

3.2 Compilation and processing

3.2.1 Repositories (cf. CLARIN)

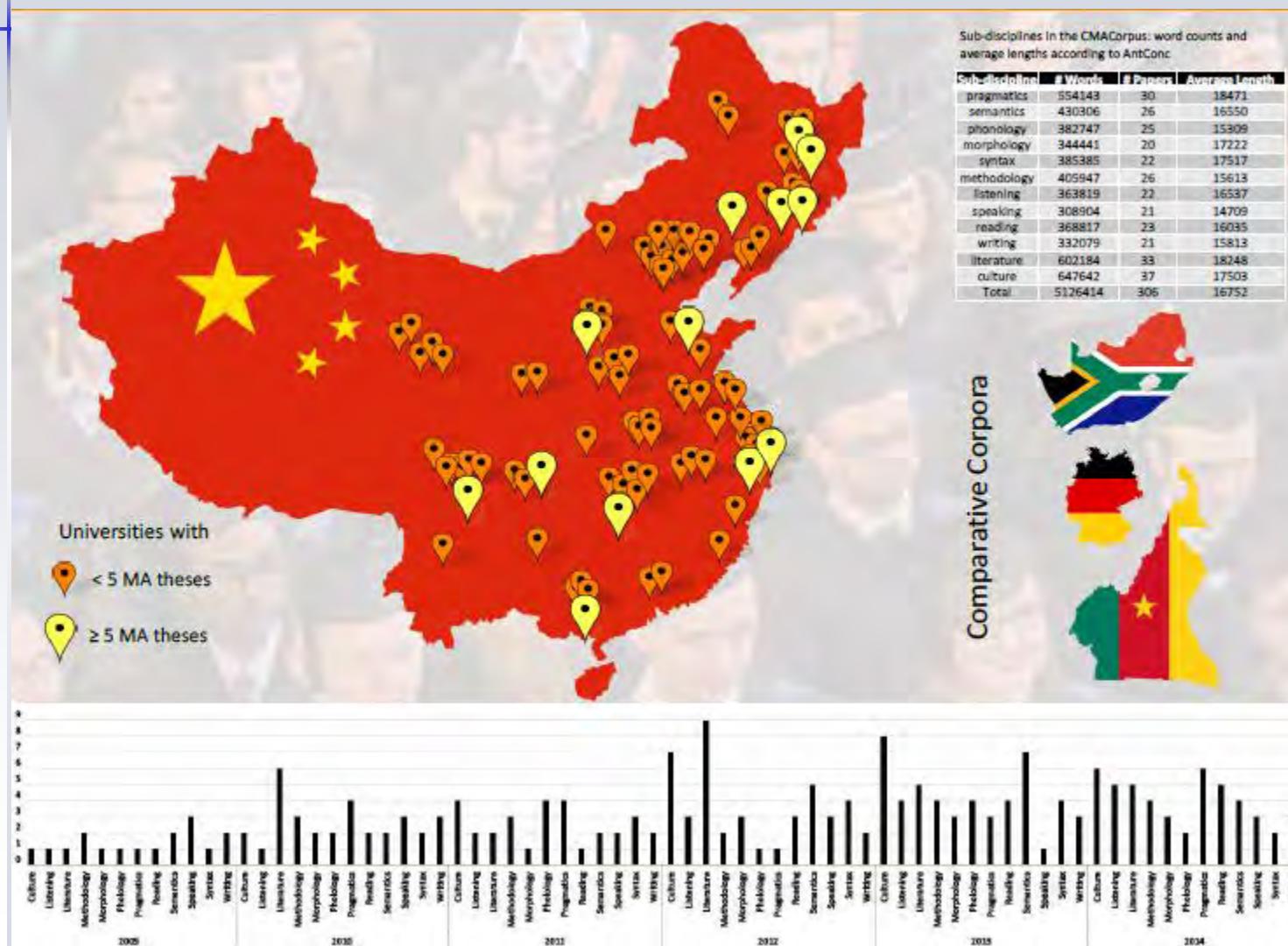
Problems:

- select
- download
- transform



The image shows a screenshot of the Hong Kong Institutional Repositories (HKIR) website. At the top, there is a logo consisting of a circular arrangement of colorful, abstract shapes resembling people or leaves. To the right of the logo, the text "Hong Kong Institutional Repositories" is displayed in a large, dark blue serif font. Below this, there is a row of eight smaller logos, each representing a different institution. From left to right, the logos are: The Chinese University of Hong Kong (CUHK), The University of Hong Kong (HKU), The Hong Kong Polytechnic University (PolyU), The Hong Kong Institute of Education (KNU), The Chinese University of Hong Kong (CUHK) again, The University of Hong Kong (HKU) again, The Hong Kong Polytechnic University (PolyU) again, and The Chinese University of Hong Kong (CUHK) once more. Below these logos, the text "About HKIR" is centered. At the bottom of the screenshot, there is a search interface with a white search bar, a dropdown menu labeled "All Fields", a "Find" button, and a "Advanced" link.

3.2.2 ChAcE(MA)Corpus set-up by university, sub-discipline and year



2.3.3 ChemCorpus set-up by genre and specialisation

genre	specialisation	number of texts	average length	total words
<i>term paper BA</i>	language/linguistics	100	4,200	0.5 Mill.
	culture/literature	100	4,700	0.5 Mill.
<i>project report</i>	(cultural)	120	4,000	0.5 Mill.
<i>BA thesis</i>	language/linguistics	80	12,000	1 Mill.
	culture/literature	80	16,000	1 Mill.
<i>term paper MA</i>	language/linguistics	80	5,700	0.5 Mill.
	culture/literature	80	6,600	0.5 Mill.
<i>MA thesis</i>	language/linguistics	40	25,000	1 Mill.
	culture/literature	40	25,000	1 Mill.
total		720		6.5 Mill.

3.2.4 Corpus texts coding scheme and spreadsheet (ZAMA)

Text	MT	gender	year	dept	DisTyp	words	types	ttr
BZ05AF	A	f	2005	BZ	5NS	35239	4030	11,44
BZ05AM	A	m	2005	BZ	5NS	29077	4004	13,77
BZ06AF	A	f	2006	BZ	5NS	14514	2795	19,26
BZ06GM	G	m	2006	BZ	5NS	28438	4244	14,92
BZ06XM	X	m	2006	BZ	5NS	16416	2801	17,06
BZ07IF	I	f	2007	BZ	5NS	13733	2624	19,11
C05EM	E	m	2005	C	5NS	23413	2393	10,22
C07AM	A	m	2007	C	5NS	20985	2809	13,39
C08CM	C	m	2008	C	5NS	29102	2737	9,40
C08EM	E	m	2008	C	5NS	4373	796	18,20
C09XF	X	f	2009	C	5NS	36100	3802	10,53
C12AM	A	m	2012	C	5NS	18991	2753	14,50
CE04EM	E	m	2004	CE	6EG	49398	3315	6,71
CE10X2M	X	m	2010	CE	6EG	40857	3247	7,95
CE10XM	X	m	2010	CE	6EG	39670	2814	7,09
CE11EM	E	m	2011	CE	6EG	45300	3211	7,09
CE11IF	I	f	2011	CE	6EG	61380	3255	5,30
CE12AM	A	m	2012	CE	6EG	23865	2248	9,42
CS06EF	E	f	2006	CS	1Hu	33604	2820	8,39
CS07AF	A	f	2007	CS	1Hu	57442	4638	8,07
CS09XF	X	f	2009	CS	1Hu	27729	3126	11,27
CS11RM	R	m	2011	CS	1Hu	24962	3336	13,36

3.2.4 Corpus texts in departments and discipline types (ZAMA)

dept.	texts	DisTyp	DisType	texts2	words
BotZo	6	5NS	Humanities (Hu)	27	1417544
Chem	6	5NS	Interdisc. (ID)	26	1000643
CivE	6	6EG	Law (Lw)	18	1146390
CurrS	6	1Hu	SocialS (SS)	20	723460
EngLit	7	1Hu	NaturalS (NS)	39	1115830
Econ	7	2ID	Engineer (EG)	20	637334
EIEng	8	6EG			
Geo	6	2ID	total	150	6041201
Genet	6	5NS			
Hist	8	1Hu			
Journ	6	2ID			
Ling	7	2ID			
LawI	6	3Lw			
LawM	6	3Lw			
LawP	6	3Lw			
MecE	6	6EG			
MedPh	6	5NS			
MedS	6	5NS			
Pol	6	1Hu			
Phys	9	5NS			
Psych	6	4SS			
SocAn	8	4SS			
Sport	6	4SS			
21	150				

4 Metalanguage Features:

definitions and previous comparisons

4.1 Coherence

4.1.1 Indicators: cohesive devices

formal:

conjunctions: *but, while*

adverbs: *first, then, finally*

functional:

sentence adverbials, e.g. clause-initial adverb **ly,*

function prototype

listing/enumerative/additive and

adversative/contrastive/concessive but

sequential/temporal/transitional then

causal/resultative/inferential because

corpuslinguistic comparison by frequency, type,
specificity/complexity and text-/sociolinguistic distribution

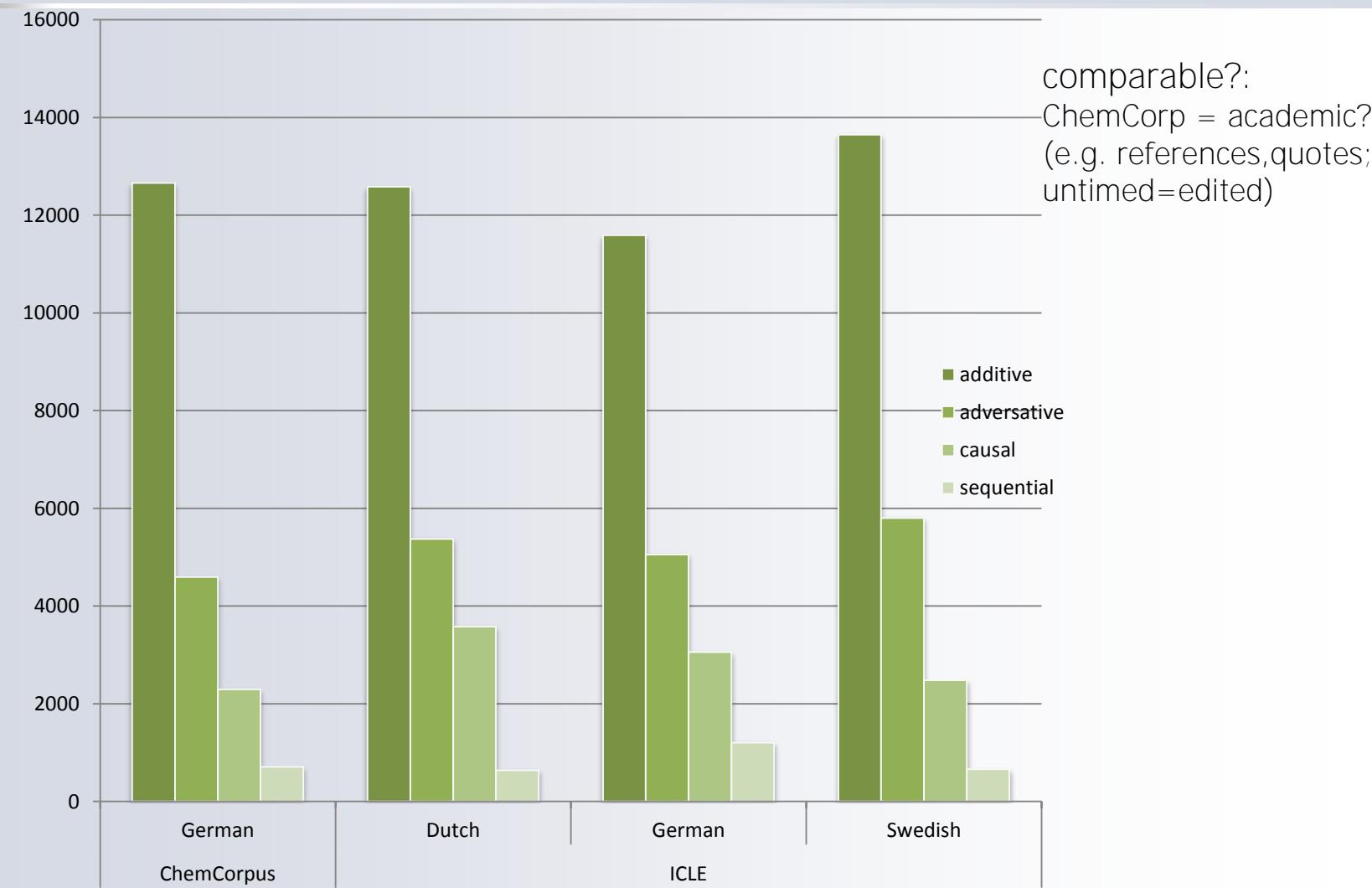
4.1.2 AntConc concordance (KWIC) of *unfortunately* in BA term papers in the ChemCorpus

author - reader interaction = contrary to what one might expect, hope

Concordance | Concordance Plot | File View | Clusters | Collocates | Word List | Keyword List |

Hit	KWIC
1	n the general South African elections this year. Unfortunately, there exists hardly any empirical work on 'borners about the ethnic share of the population do, unfortunately, not exist. <fig> Figure 1 Share of South African
2	frican because in all honesty that is what I am. Unfortunately, the legacy of apartheid as such instilled race a
3	position towards this 'New South Africa'? I-ee: Unfortunately, it's been said so often it's lost its meaning. A
4	n awareness. Working further on this issue goes, unfortunately, beyond the scope of this work, however, further
5	one, but on several paragraphs of the Agreement. Unfortunately, in most researches, the changes that the GFA has
6	The statistic provided in Tim Pat Coogan's work unfortunately only reaches until the year 1993, however, a diff
7	phey, 2003, p. 75). Although most classrooms are unfortunately not designed after proved principles of learning
8	schools' and teachers' readiness to participate unfortunately remained rare. Once having receive
9	. The incomplete questionnaires - of which there unfortunately were quite a lot, both among the paper and online
10	g inclusion of student opinions and perceptions. Unfortunately, this, on the one hand, intricate and complex but
11	corpus including the Tweets regarded to the PL (unfortunately no differentiation between personal and official
12	requent abbreviations in the PL Twitter corpora. Unfortunately, the AntConc toolkit does not work smoothly in co
13	tion of Twitter Corpora Abbreviations </tab> Unfortunately, some abbreviations from Table 8 could not be cle
14	anation could be that v is even shorter than vs. Unfortunately, in the frame of the present thesis no clear evid
15	slation: But this LTI-like abbreviation mania is unfortunately quite convenient for Twitter) (Twitter, 2013).
16	ions. Otherwise, this thesis would be redundant. Unfortunately, within the frame of the PL Twitter analysis it w
17	h. Regarding another filtering option, it is unfortunately not possible to separate official Tweets, which a
18	e some problems regarding these scales. First, I unfortunately could not find a current probability scale. It is
19	er words serves to distinguish between speakers. Unfortunately, films often create a misleading impression of th
20	ent, thus I decided to use older books which are unfortunately part of a series. Furthermore, despite the fact t
21	Mystic River which was released in 2001. Unfortunately, all five novels by Stephen J. Cannell are part o
22	r he also became producer, mainly of CBS series. Unfortunately, Straw wrote and published just one novel that ca
23	crime related words which is generally accepted. Unfortunately their is no such word list. It is conceivable tha
24	

4.1.3 Conjuncts by function in European learner corpora (Albrecht 2013: 36, figure 11)



4.2 Modal auxiliaries

4.2.1 Terminology: stance > hedging/boosting > modality > auxiliaries/2

stance definitions:

- “personal feelings, attitudes, value judgments, or assessments” (Biber et al. 1999: 966)
- “subjective” evaluation on the basis of own knowledge, experience, etc.
- context-dependency

→ academic culture in the discourse community determines how stance is expressed!

author stance and engagement are crucial variables in academic interaction

“Mood is a category of grammar, modality a category of meaning.”

(Huddleston/Pullum 2002: 172)

“Epistemic is subjective, has a strength and is comparable with probability, but it differs from it [deontic] in that it involves inference.”

She must have left yesterday.

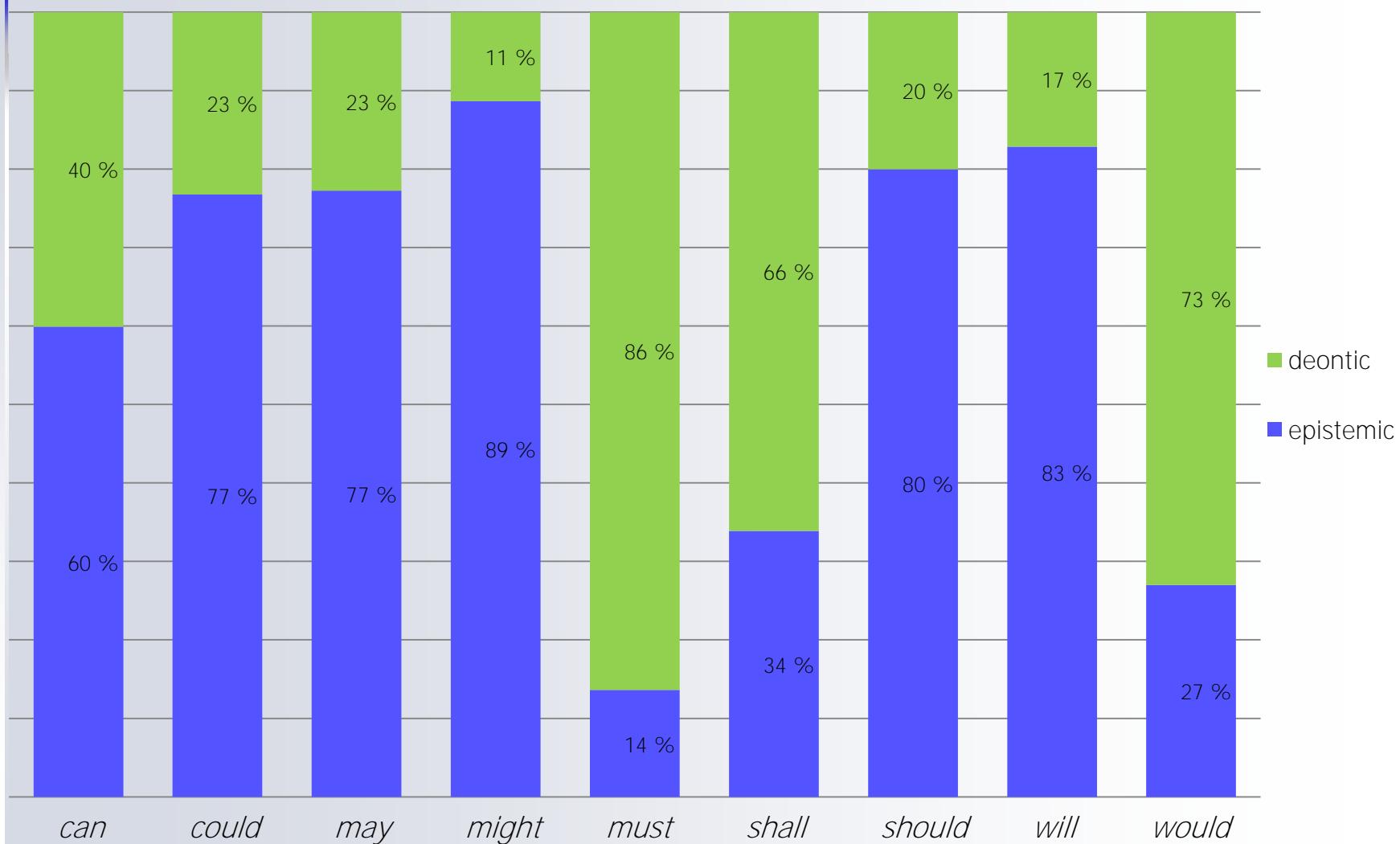
“Deontic is usually subjective, indicating what the speaker considers ‘right’ whether morally or as a matter of expediency.”

One should always tell the truth.

We should buy now while the market is depressed.

(Huddleston/Pullum 2002: 186)

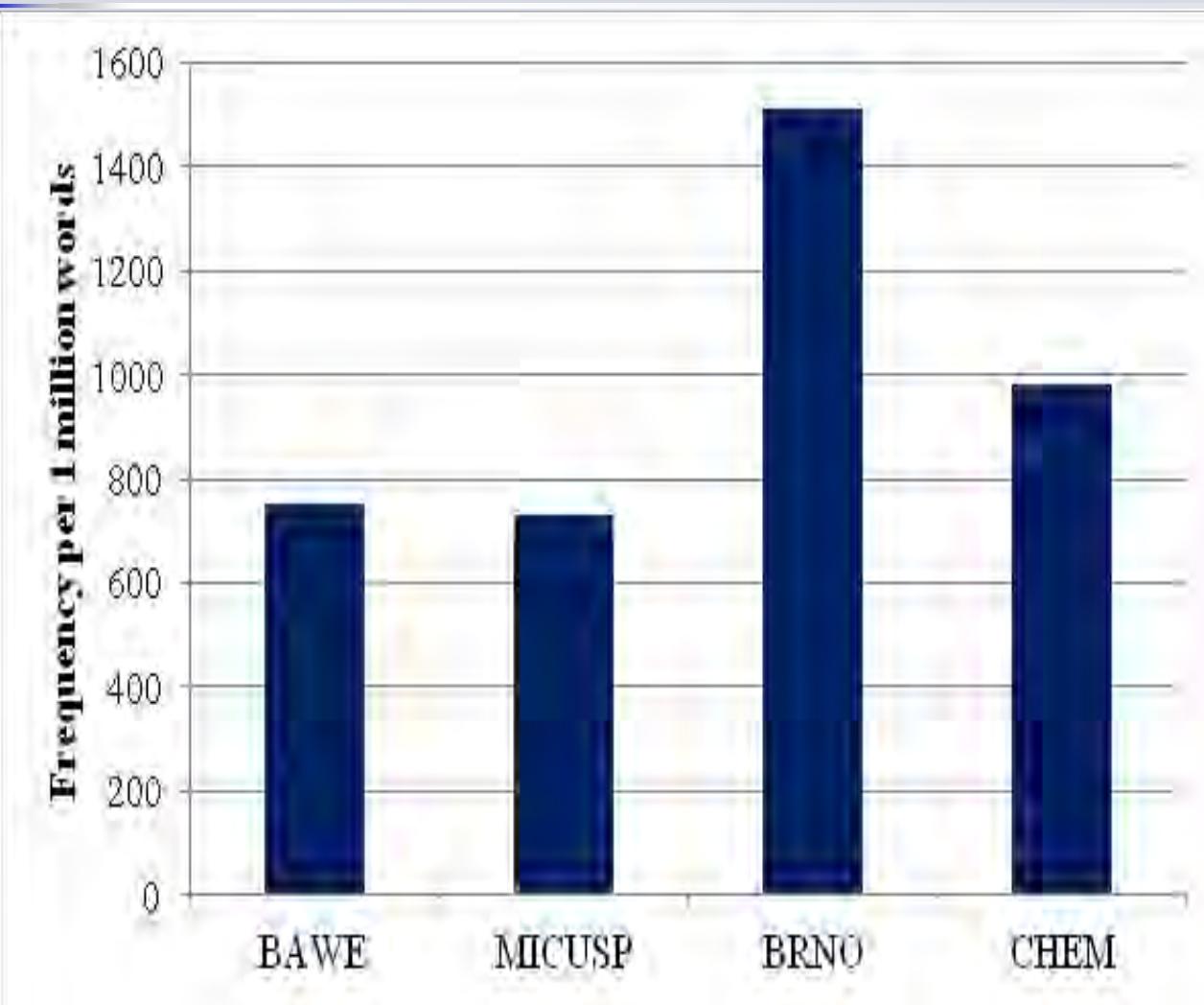
4.2.2 Deontic and epistemic modal auxiliaries in ICE-EA



4.2.3 Hedges/boosters per 1 million words in the BrnoCorpus and ChemCorpus (Bräuer 2013: 49, table 17)

	Brno (CR)	Chemnitz	p-value
Amplifier	2,105.07	1,932.36	
Booster	1,713.69	1,303.09	
<i>considerably</i>	30.40	102.79	< 0.001
<i>highly</i>	136.79	134.76	> 0.05
<i>strongly</i>	34.20	76.52	0.001 ≤ p < 0.01
<i>tremendously</i>	1.90	7.99	> 0.05
<i>very</i>	1,510.40	981.03	0.01 ≤ p < 0.05
Maximizer	391.38	629.27	
<i>absolutely</i>	58.90	31.98	< 0.001
<i>clearly</i>	148.19	462.53	< 0.001
<i>extremely</i>	74.10	57.10	0.01 ≤ p < 0.05
<i>fully</i>	83.59	63.96	0.001 ≤ p < 0.01
<i>thoroughly</i>	26.60	13.70	> 0.05
Downtoner	609.86	785.74	
Approximator	317.28	404.29	
<i>almost</i>	237.49	340.33	> 0.05
<i>nearly</i>	72.20	36.55	0.001 ≤ p < 0.01
<i>virtually</i>	7.60	27.41	< 0.05
Diminisher	199.49	229.55	
<i>a bit</i>	76.00	15.99	< 0.001
<i>slightly</i>	93.09	143.90	> 0.05
<i>somewhat</i>	24.70	41.11	> 0.05
<i>to some extent</i>	5.70	28.55	> 0.05
Minimizer	93.09	151.89	
<i>barely</i>	7.60	11.42	> 0.05
<i>hardly</i>	76.00	132.48	> 0.05
<i>scarcely</i>	9.50	7.99	> 0.05
Total	2,714.93	2,718.10	

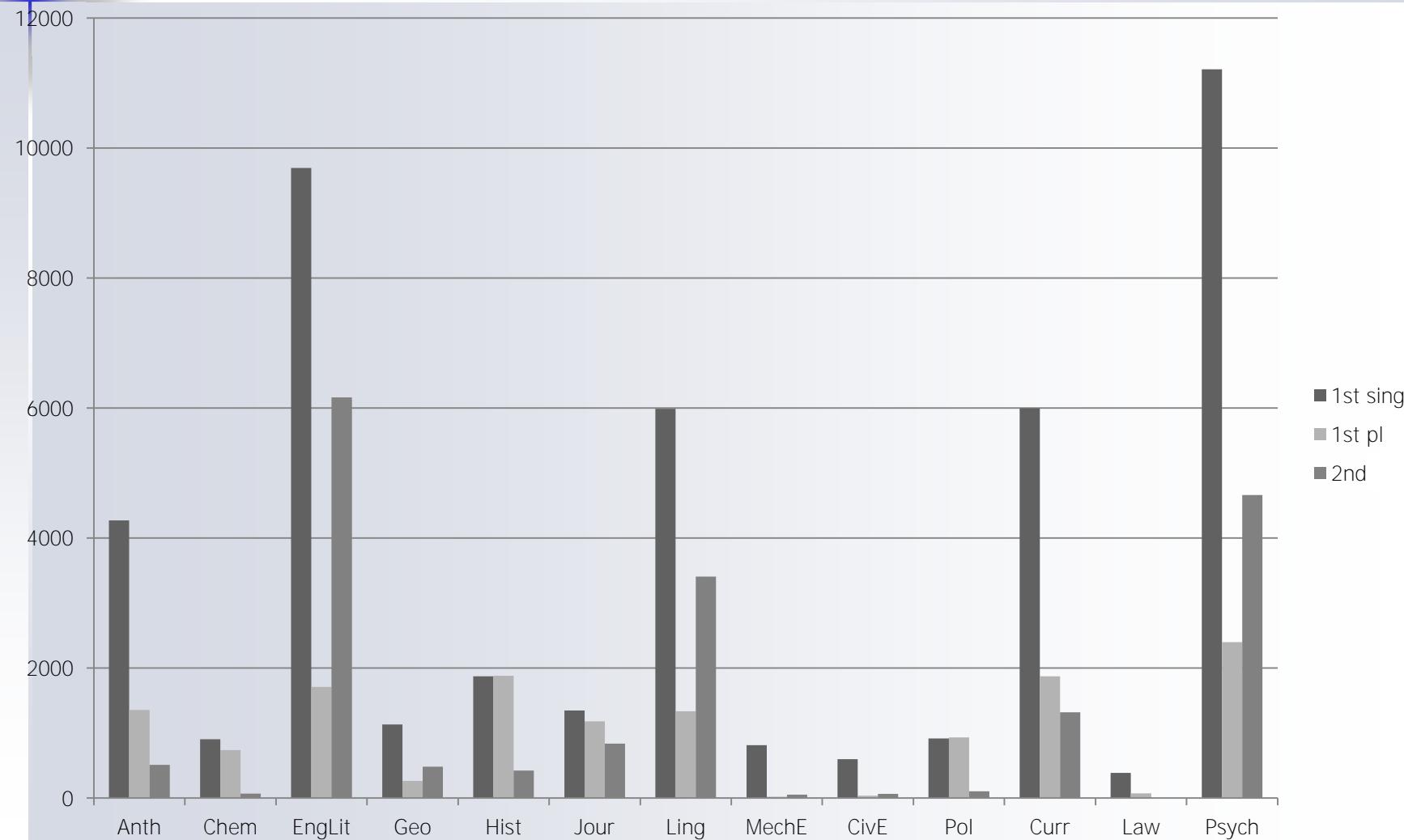
4.2.4 Relative frequencies of *very* in BAWE, MICUSP, the BrnoCorpus and ChemCorpus (Bräuer 2013: 63, figure 3)



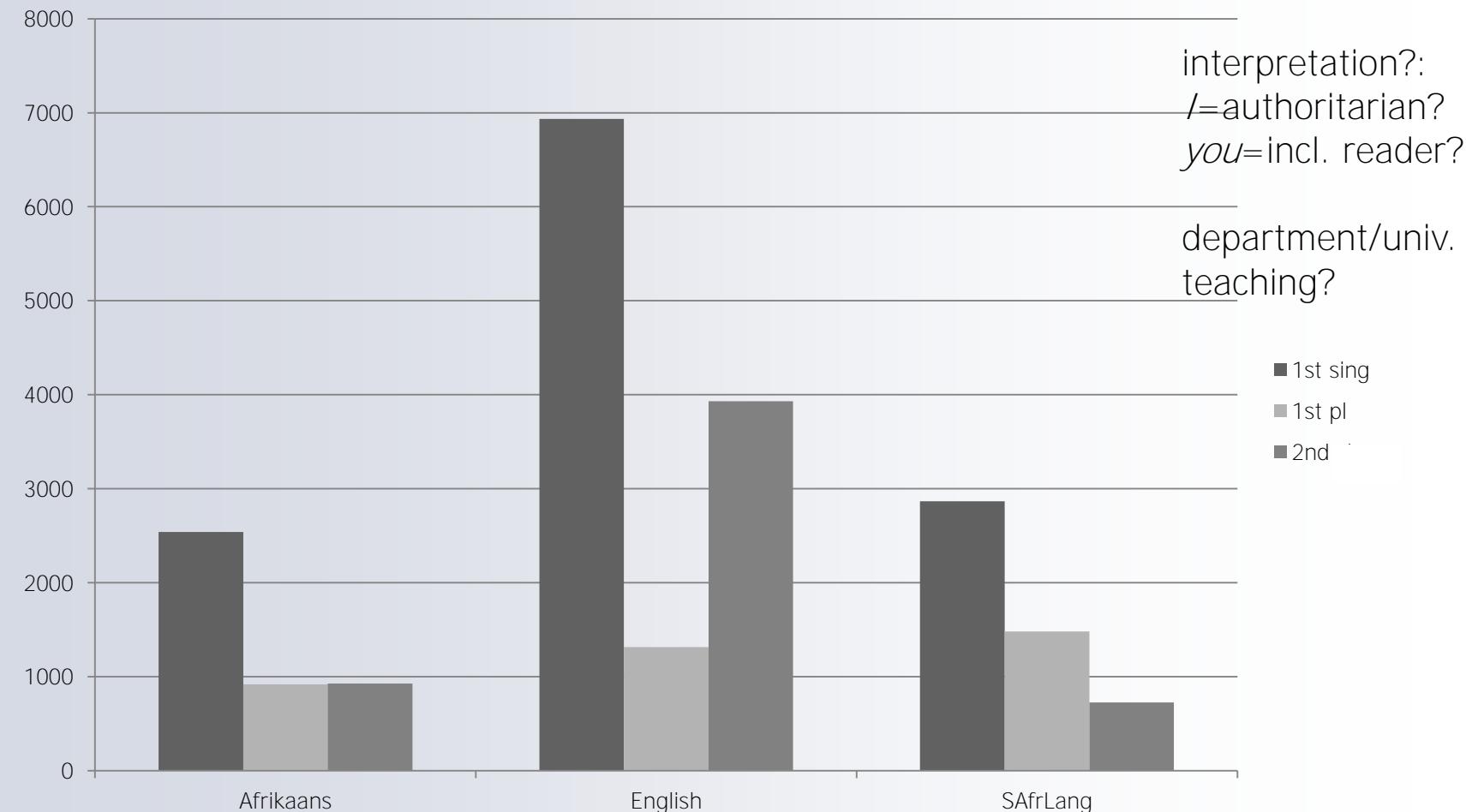
learner-related?
simple vs. specific –
explicit

4.3 Personal Pronouns in ZAMA (Schmied 2013)

4.3.1 Variation in personal pronoun usage (/1 M words) by department



4.3.2 Variation in personal pronoun usage (/1 M words) by mother tongue (names)



5 Preliminary results

5.1 Conjuncts 5.1.1 Resultative as a result (in ChAcEMA)

Concordance Hits 708

Hit KWIC

1 *prejudice, he cannot improve the cultural learning effectively. As a result, it is sincerely hoped tha*

2 *, explain and comment on these behaviors based on their own culture. As a result, unconsciously, their own be*

3 *ential power or ability in the learner that may be developed step by step as a result of internal change through I*

4 *who are strongly influenced by the English language tend to use it. As a result, whenever it is uttered it*

5 *cation of the basic verbs that indicate the desired action directly. As a result, Chinese finds imperative*

6 *of the major languages in the world, reflecting two main cultures. As a result of the reform and open doo*

7 *in apricot forest) to pay tribute to the distinguished doctors, and as a result, “杏林” (apricot forest) ha*

8 *taught to construct an inferential mode in listening comprehension. As a result, we propose that the theori*

9 *vide as an input to the hearer’s inferential comprehension process. As a result, verbal communication can*

10 *the students lack inferential knowledge in their listening practice. As a result, we suggest that the theorie*

11 *prehension but inferential questions were very difficult for them. As a result, we put forward an infer*

12 *special -- but winning as a black American is a knockout.” 36 As a result of her literary and artist*

13 *n who wrote The Pleasure of Enchantment. 60 Moreover, perhaps as a result of the burgeoning field of At*

14 *ts presence within the narrative serves incessantly to disrupt authority. As a result, Beloved makes overt the o*

15 *d it hard to continue the topic, which would make the class mechanical as a result. b) Teachers must be awa*

16 *experience, that they cannot be adequately taught in a classroom. As a result, little attention was direc*

17 *xperiment class they had in fact two more classes than control class. As a result data showed in the inve*

18 *: message or the iconic message can’t avoid ambiguity and subjectivity as a result of over-emphasis on the co*

19 *ategy will be traced out in the direction of conversational implicature as a result of flouting the four conv*

20 *informative as is required”. The deliberate withholding of information as a result of flouting of this sub-maxin*

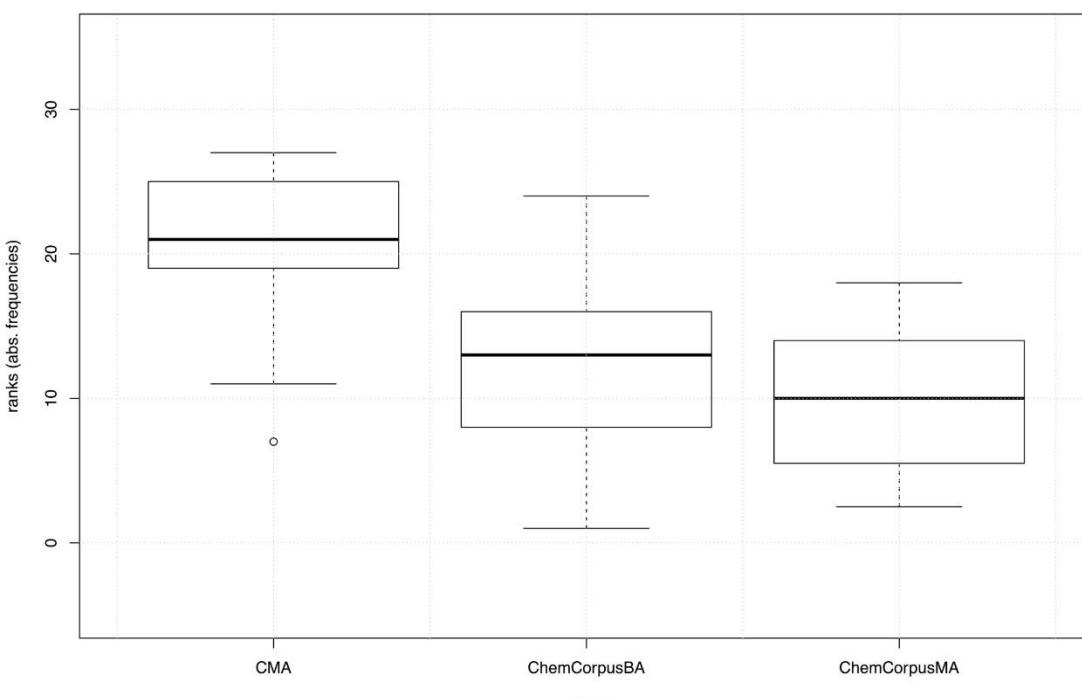
5.1.2 Frequency of semantic conjunct classes in ChAcEMA vs. ChemCorp

conjunct type	ChAcEMA (total frequency)	ChAcEMA (per 1M)	ChemCorpBA (total frequency)	ChemCorpBA (per 1M)	ChemCorpMA (total frequency)	ChemCorpMA (per 1M)
Listing/enumerative	21090	4149	3275	4551	1772	3337
Listing/additive	17223	3388	3711	5157	2584	4867
Summative	797	156	268	372	233	438
Appositive	10508	2067	9580	13315 <i>that is</i>	1275	2401
Resultative/inferential	27423	5395	3131	4351	2287	4307
Contrastive/concessive	16768	3299	4376	6082	2900	5462
Transitional	1516	298	216	300	152	286
Total	95325	18756	24557	34132	11203	21101
Average	13617	2679	3508	4876	1600	3014

5.1.3 Frequency of resultative/inferential conjuncts in ChAcEMA vs. ChemCorp

conjunct	ChAcEMA (total frequency)	ChAcEMA (per 1M)	ChemCorpBA (total frequency)	ChemCorpBA (per 1M)	ChemCorpMA (total frequency)	ChemCorpMA (per 1M)
<i>(and) so</i>	10660	2097	646	897	439	826
<i>accordingly</i>	327	64	43	59	48	90
<i>as a result</i>	708	139	70	97	51	96
<i>consequently</i>	356	70	173	240	111	209
<i>for this reason</i>	75	14	24	33	9	16
<i>hence</i>	700	137	229	318	162	305
<i>in this case</i>	235	46	89	123	69	129
<i>in this respect</i>	47	9	58	80	10	18
<i>in this way</i>	659	129	19	26	19	35
<i>of course</i>	408	80	111	154	65	122
<i>so that</i>	1066	209	86	119	39	73
<i>then</i>	4685	921	282	391	206	388
<i>therefore</i>	4229	832	747	1038	397	747
<i>thus</i>	3268	643	554	770	662	1246
TOTAL	27423	5395	3131	4351	2287	4307

5.1.4 Conjuncts (enumerative) quantitative: ChemCorp vs. ChAcEMA



detailed data distribution not good for hard tests, so (courtesy M. Hofmann):

- frequencies transformed into ranks (Sheskin 2011)
- assumptions: gvlma (=Global Validation of Linear Models Assumptions; Peña & Slate 2006)
- Kruskal-Wallis test (non-parametric equivalent to ANOVA):
 $\chi^2 = 7.86$, $df = 2$, $p = 0.02$
- Dunn's posthoc test (incl. adjustment of p-values in multiple comparisons according to Holms (1979) :

corpus	Z statistic	p-value
ChemBA – ChemMA:	0.653504	(0.26)
ChemBA – ChAcEMA:	-2.034776	(0.04)?
ChemMA – ChAcEMA:	-2.688280	(0.01)

5.2 Modal auxiliaries

5.2.1 Modal auxiliaries qualitative: *may* (in ChAcEMA)

5.2.1 Modal auxiliaries qualitative: *may* (in ChAcEMA)

5.2.1 Modal auxiliaries qualitative: *may* (in ChAcEMA)

5.2.2 Modal auxiliaries qualitative: *might* (in ChAcEMA)

5.2.2 Modal auxiliaries qualitative: *might* (in ChAcEMA)

AntConc 3.2.4w (Windows) 2011

File Global Settings Tool Preferences About

Corpus Files

Concordance | Concordance Plot | File View | Clusters | Collocates | Word List | Keyword List

Hit KWIC File

1 As a result, it is sincerely hoped that the research **might** provide Chinese English teachers some implications on cultural teaching and h CMAC05CU_28.txt

2 Through developing these basic skills, the communication **might** be able to overcome ethnocentric tendencies, and become more sensitive to the needs, va CMAC05CU_28.txt

3 unreasonable. Obviously, ethnocentric people or group **might** cause the failure of communication, even the conflict of cultures (Jia, 1997:111). CMAC05CU_28.txt

4 ion of all the things in western culture. While in theory this **might** seem viable, it is rather impossible to achieve in today's context. It is an obvi CMAC05CU_31.txt

5 Realization, and what the characteristics of those universals **might** be. Concerning requests, one of the most significant findings of CCSARP was that all : CMAC05CU_31.txt

6 terances are both nonliteral and indirect. For example, one **might** utter ÁI love the sound of your voiceÁ to tell someone nonliterally (ironically) : CMAC05CU_31.txt

7 this way. Impersonal: Ái mèáqàdáit-á, Ái wíx-áÁ. So, it **might** not be a bad idea to get it cleaned up. The speaker is not sure about the possibility of realiz: CMAC05CU_31.txt

8 as in English, that the requester expresses a wish which **might** be fulfilled from the requester's point of view, which may also be ÁsubjunctiveÁ. Therefore: CMAC05CU_31.txt

9 their space. Therefore, these automatically assume that there **might** be some social distance or awkwardness in the situation. (4) Off-Record (Indirect) To p CMAC05CU_31.txt

10 on providing possible strategies and principles the translator **might** consider. However, since the process of translation is actually CMAC05CU_36.txt

11 obviously the best policy in this circumstance **might** be to translate the sentence into ÁI only know the most basic moves.Á When th CMAC05CU_36.txt

12 ese expectations of relevance amount to, and how they **might** contribute to an empirically plausible account of comprehension. A detailed elab CMAC05LI_21.txt

13 tively. For example, when I saw the train arriving, I **might** look at my watch, searching from my knowledge of the train timetable, and der CMAC05LI_21.txt

14 assumptions. For example, the sight of my train arriving late **might** confirm my impression that the service is deteriorating, or make me alter my plan CMAC05LI_21.txt

15 of degree. There is no shortage of potential inputs which **might** have at least some relevance for us, but we cannot attend to them all. Relevance theory cl CMAC05LI_21.txt

16 ight, intending you to notice and make the conclusion that I **might** like another drink. Inferential communication is not just a matter of intending to a CMAC05LI_21.txt

17 cognition to a basic layer of information that the audience **might** have picked up anyway. In other cases, the communicator's behavior doesn't provide CMAC05LI_21.txt

18 nd specify strategies such as Álunch togetherÁ, but **might** start from an activity which enables the communicators to actually build a trust relationsh CMAC05LI_21.txt

19 necessary to convince teachers that learning strategies **might** be effective for their students as well as for themselves. Teachers require adequate ex: CMAC05LI_21.txt

20 ividuals we had value, irrespective of what the future **might** hold for us. The women of the black community, whether aunt, grandmother, or neighbors, CMAC05LI_28.txt

21 did it again and then againÁ; It went on that way and **might** have stayed that way, but one evening, after supper, after Sethe, he came downstairs, sat in CMAC05LIT_28.txt

22 rocker and didn't want to be thereÁ; It went on that way and **might** have stayed that way, except one evening, after supper, after Sethe, he came downstairs and la CMAC05LIT_28.txt

23 wo: An Analysis of Beloved thereÁ; It went on that way and **might** have stayed that way except one evening, after supper, after Sethe, he lay on a pallet in the s CMAC05LIT_28.txt

24 ty slavery. ÁThe best thing she was, was her children. Whites **might** dirty her all right, but not her best thing, her beautiful, magical best thing-----the part CMAC05LIT_28.txt

25 periences, a history that even African American communities **might** have begun to forget, or **might** not want to remember.Á 193 As critic Wahneema CMAC05LIT_28.txt

26 an American communities **might** have begun to forget, or **might** not want to remember.Á 193 As critic Wahneema Lubiano said about her Beloved: CMAC05LIT_28.txt

27 Áa fictional account of the interior life of a former slave **might** be more historically ÁrealÁ than actual documents, which were often written from the CMAC05LIT_28.txt

28 counts memories that correspond to those that SetheÁs mother **might** have had of her passage from Africa to America. By Chapter 26, Beloved and Sethe ha CMAC05LIT_28.txt

29 udience. Slave narrators had to hold back memories which **might** alienate their white readers. 253 Different from traditional slave narrative, Morrison: CMAC05LIT_28.txt

30 ges teachers to find out their own teaching problems that they **might** not be aware of before, to help teachers to look for a new way in the process of devel CMAC05ME_19.txt

31 ls with the notion of feasibility. Systematic potential **might** allow for a sentence like Áthe mouse the cat the dog the man the woman mar: CMAC05ME_21.txt

< >

Search Term Words Case Regex

Concordance Hits

Search Window Size

might

Advanced 1193 105

Start Stop Sort

Files Processed Total No. 304

Kwic Sort

Reset 

Save Window

Exit

Level 1 1R

Level 2 2R

Level 3 3R

5.2.2 Modal auxiliaries qualitative: *might* (in ChAcEMA)

5.2.2 Modal auxiliaries qualitative: *might* (in ChAcEMA)

5.2.3 Modal auxiliaries qualitative: *would* (in ChAcEMA)

AntConc 3.2.4w (Windows) 2011

File Global Settings Tool Preferences About

Corpus Files

Concordance Concordance Plot File View Clusters Collocates Word List Keyword List

Hit KWIC File

72 and poets. Later, laurels began to represent honor, and people would call those who had achieved outstanding accomplishments Poet Laureates. Th CMAC05CU_36.txt
 73 ered Harpocrates, the God of Silence, a rose in case that he would distribute the indecent behavior. To Chinese people, 阿roseà reminds people only o CMAC05CU_36.txt
 74 condition, especially national culture because different cultures would give birth to different languages as they are carriers of the culture which they m CMAC05CU_36.txt
 75 imperial times deeply loved, and the scholars of ancient China would liken the so-called four gentlemen (namely plum, orchid, bamboo and 39 ch CMAC05CU_36.txt
 76 while in the English-speaking countries, chrysanthemum as a flower would be exclusively used in the funerals, therefore, resulting in a quite different impl CMAC05CU_36.txt
 77 his method is used in simile or metaphor translation, the target text would be quite fluent so the average target language readers have no difficulty to continue CMAC05CU_36.txt
 78 e heavily colored in the target language culture. Otherwise the version would sound strange and inaccurate. By substitution, we can make the version more CMAC05CU_36.txt
 79 d a special historical setting. Without further annotations, it would not make any sense to the English readers at all. The following is a cas CMAC05CU_36.txt
 80 nvey the meaning of it. Therefore, the strategy of explanation would be applied alternatively with annotations in translating cultural specialities. S CMAC05CU_36.txt
 81 language users to pair sentences with the contexts in which they would be appropriate. (Stephen C. Levinson, 1983:24) This definition makes the CMAC05LI_21.txt
 82 d proposition, I will choose another stronger expression which would so commit me. Then, I may get the implication that I am not in the position CMAC05LI_21.txt
 83 chosen to utter (8) in preference. Yet, the expression of (9) would commit me to Maryàs passing. (9) Mary and Jane have passed the exam. Therefo CMAC05LI_21.txt
 84 eràs meaning. Chapter Three A Survey Study In this chapter, we would present brief introduction of the way by which the proposed study is conducted CMAC05LI_21.txt
 85 In the second part, in order to make sure that the students would answer the questions seriously, all the data collection of the study was CMAC05LI_21.txt
 86 g comprehension and offer them efficient instructions. Then, we would use these theories of pragmatic inference to solve practical problems, i.e. to CMAC05LI_21.txt
 87 of this case. If the woman was hanging about the store, it would be possible that the woman had participated in the robbery. This can be CMAC05LI_21.txt
 88 er sweater. The woman assumes that when she asks for help, the man would surely give her help. Nevertheless, her assumption was wrong: the man refuses CMAC05LI_21.txt
 89 exception era of African American history, for the first time, blacks would have access to a range of economic, educational, and social opportunities that CMAC05LIT_28.txt
 90 istorical and philosophical attention as their polar opposites, which would be wealthy, white, heterosexual men. Secondly, this attention to the marginaliz CMAC05LIT_28.txt
 91 that àWest is bestà and that all other cultures of the world would eventually adopt Western values which would, with the passage of time, becom CMAC05LIT_28.txt
 92 res of the world would eventually adopt Western values which would, with the passage of time, become universal. There was a built-in cultural CMAC05LIT_28.txt
 93 consciously exposes those narrative strategies, or the process, which would lead to totality, that would advocate an essence, a metaphysics, or a meta-na CMAC05LIT_28.txt
 94 narrative strategies, or the process, which would lead to totality, that would advocate an essence, a metaphysics, or a meta-narrative. Finally, we can CMAC05LIT_28.txt
 95 tamount to imposing the wrong tradition on her novels, Morrison would remark in a 1983 interview to Nellie McKay, àI am not like James Joyce; I am not li CMAC05LIT_28.txt
 96 e had been caught as a fugitive. And she had made up her mind that they would not suffer the way that she had and it was better for them to die. And her moti CMAC05LIT_28.txt
 97 k toothà, Beloved looked at the tooth and thought, This is it. Next would be her arm, her hand, a toe. Pieces of her would drop maybe one at a time, maybe all CMAC05LIT_28.txt
 98 ought, This is it. Next would be her arm, her hand, a toe. Pieces of her would drop maybe one at a time, maybe all at once. Or on one of those mornings before Denv CMAC05LIT_28.txt
 99 Or on one of those mornings before Denver woke and after Sethe left she would fly apart. It is difficult keeping her head on her neck, her legs attached to her h CMAC05LIT_28.txt
 100 e----death and the Middle Passage----is the same. Her yearning would be the same, the love and yearning for that face that was going to smile at her. CMAC05LIT_28.txt
 101 6Chapter Two: An Analysis of Beloved them deep and unselfish love. She would even sacrifice herself to save her children from the dirty slavery. àThe best thing CMAC05LIT_28.txt
 102 place her. àIf that was the case, than there was no stopping now. She would get that milk to her baby girl if she had to swim.à 116 Back split, p CMAC05LIT_28.txt

Search Term Words Case Regex Concordance Hits 3731 Search Window Size 105

would

Total No. 304

Files Processed

Reset 

Kwic Sort

Level 1 1R Level 2 2R Level 3 3R

5.2.3 Modal auxiliaries qualitative: *would* (in ChAcEMA)

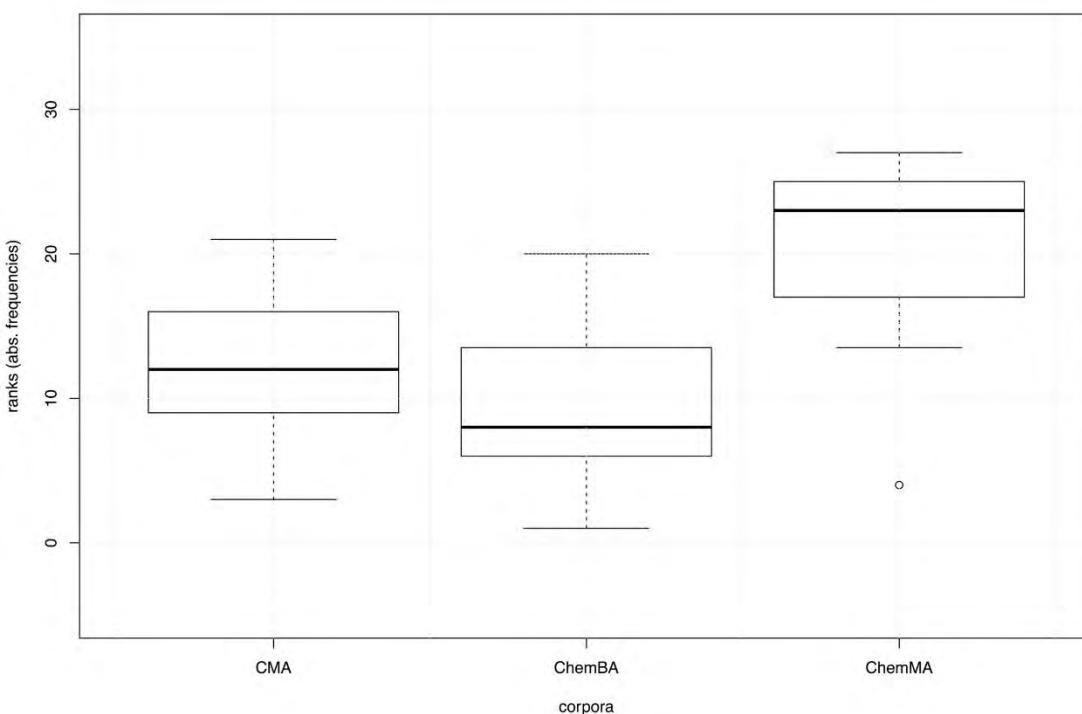
AntConc 3.2.4w (Windows) 2011

File Global Settings Tool Preferences About

5.2.4 Modal auxiliaries quantitative: ChemCorp vs. ChAcEMA

modal	ChAcEMA (absolute frequency)	ChAcEMA (per 1M)	ChemCorp BA (absolute frequency)	ChemCorp BA (per 1M)	ChemCorp MA (absolute frequency)	ChemCorp MA (per 1M)
can	2345	3259	1563	2943	23298	4584
could	1045	1452	591	1113	3978	782
may	648	900	675	1271	6065	1193
might	687	954	522	983	1195	235
must	232	322	96	180	2355	463
shall	275	382	23	43	240	47
should	626	870	369	695	8152	1604
will	1400	1945	1200	2260	10528	2071
would	1019	1416	807	1520	3821	751

5.2.4 Modal auxiliaries quantitative: ChemCorp vs. ChAcEMA



detailed data distribution not good for hard tests, so (courtesy M. Hofmann):

- frequencies transformed into ranks (Sheskin 2011)
- assumptions: gvlma (=Global Validation of Linear Models Assumptions; Peña & Slate 2006)
- Kruskal-Wallis test (non-parametric equivalent to ANOVA):
 $\chi^2 = 8.64$, $df = 2$, $p = 0.01$
- Dunn's posthoc test (incl. adjustment of p-values in multiple comparisons according to Holms (1979) :

corpus	z statistic	p-value
ChemBA – ChemMA:	-2.821521	(0.007)
ChemBA – ChAcEMA:	-0.697955	(0.242)
ChemMA – ChAcEMA:	2.123566	(0.034)

6 Functional standards

6.1 New definition

non-native standards should be

- frequent
 - = in different genres, disciplines, educational levels, etc.
- international
 - = usage in universities in very different countries, e.g. China, Germany, (South) Africa, etc.
- functional
 - = justifiable in logical-semantic terms
- transparent
 - = intelligible as sense elements
- acceptable
 - = **no negative sociolinguistic connotations**, like "uneducated"
- processable
 - = "unnoticed", without unintentional cognitive salience, awareness

→ combine corpus-, socio- and psycholinguistic methodologies, i.e.

- frequencies in a wide range of styles/usages,
- acceptability in one-line questionnaires,
- eye-tracking experiments with appropriate user-groups

*double hedging?
may perhaps
suggest... might
might clearly*

6.2 Eye-tracking experiments

6.2.1 Research applications

widely used in L2 acquisition and L2 processing
(Roberts/Siyanova-Chanturia 2013)

application to African Englishes (van Rooy 2010):
distinguishing between error, innovation, new conventions
(*can be able to*)

measurements:

- total fixations
- regressions
- first fixation duration
- total gaze time

experiments: 2 groups of readers reading real ChE sentences:
13 hedging + modality (Chinese, German)
24 articles + modality (Chinese, Czech/Polish, German)

6.2.2 Eye-tracking: acceptable? vs. unacceptable? usage of *may*

It is hoped that this dissertation may help debaters and the general public to explore the power of the language to effectively argue and persuade.

It soon becomes obvious that the differences across Standard Englishes may be small.

6.2.2 Eye-tracking: acceptable? vs. unacceptable? usage of *may*

It is hoped that this dissertation may help debaters and the general public to explore the power of the language to effectively argue and persuade.

It soon becomes obvious that the differences across Standard Englishes may be small.

problems of interpretation:

- acceptable if concentrated conscious/attention
- unacceptable if processing problem of form, not of meaning (disagreement)

6.2.3 Eye-tracking: acceptable? vs. unacceptable usage of *would*

A heatmap showing eye movements over the sentence. The text is in black, and the heatmap consists of several purple circles of varying sizes connected by lines, indicating the path and focus of the eye as it moves across the sentence. The circles are concentrated on the words 'after', 'graduation', 'would', 'become', and 'English'.

I think after graduation, I would become an English teacher.

A heatmap showing eye movements over the sentence. The text is in black, and the heatmap consists of several yellow circles of varying sizes connected by lines, indicating the path and focus of the eye as it moves across the sentence. The circles are concentrated on the words 'sponsors', 'rich', 'region', 'they', 'support', 'us', 'when', and 'tour'.

We have to get sponsors from rich companies in our region so they would support us when we go on tour.

A heatmap showing eye movements over the sentence. The text is in black, and the heatmap consists of several purple circles of varying sizes connected by lines, indicating the path and focus of the eye as it moves across the sentence. The circles are concentrated on the words 'If', 'I', 'would've', 'had', 'enough', 'money', 'would've', 'traveled', and 'world'.

If I would've had enough money, I would've traveled around the world.

7 Conclusion

7.1 Limitations

- corpuslinguistic:
some features are difficult to search for,
esp. ambiguous forms and 0 articles (a special problem for
speakers on Slavonic languages and Chinese)
- on-line questionnaires:
getting enough respondents
socio-biographical data reliable?
- eye-tracking:
many other factors, incl. linguistic complexities and proficiency
influence reading

7 Conclusion

7.1 Limitations

- corpuslinguistic:
some features are difficult to search for,
esp. ambiguous forms and 0 articles (a special problem for speakers on
Slavonic languages and Chinese)
- on-line questionnaires:
getting enough respondents
socio-biographical data reliable?
- eye-tracking:
many other factors, incl. linguistic complexities and proficiency influence
reading

example: on-line questionnaire
on article choice by SYSU students
(Albrecht 2016: 44, Table 3.9)

	[+definite] (target <i>the</i>)	[-definite] (target <i>a</i>)		
Engineering (n=22)	the	a	the	a
[+specific]	81%	17%	14%	86%
[-specific]	88%	10%	6%	93%

(a) Engineering

	[+definite] (target <i>the</i>)	[-definite] (target <i>a</i>)		
Humanities (n=15)	the	a	the	a
[+specific]	87%	8%	4%	95%
[-specific]	91%	8%	1%	94%

(b) Humanities

7.2 Further work

combination of research methodologies

but impossible to get the same users for writing, reading and answering questionnaire

expansion of ChemCorpus to make it more compatible

(incl. writer information like training and reader expectation like teaching)

for certain usage analyses like semantic prosody or metaphors even

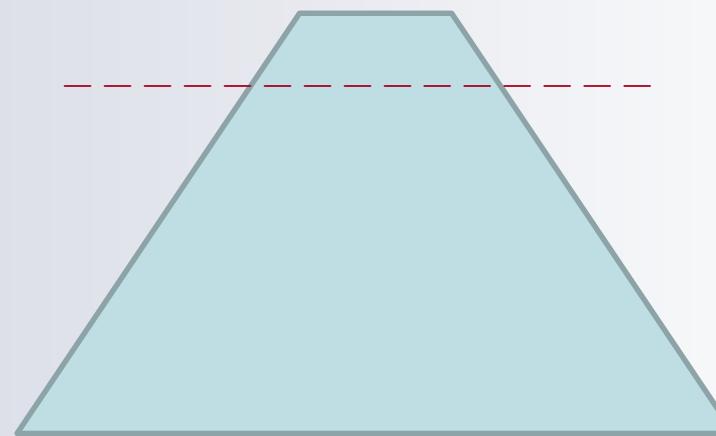
expansion of the ChAcE data-base:

tidy --> bigger

compilation - analysis - compilation - analysis - compilation -

7.3 General international “political discourse”: Convergence or Diversity?

Research English



compromises?
in advanced
lingua franca English?
(Mauranen 2012)

Student/Learner English

Are conventions becoming more and more similar because of Anglo-American dominance (gate keepers and guidebook publications)?

How much convergence do we want in research English?

How much diversity/individual creativity do we accept in student English?

Which features do we want to reduce or increase (teaching)?

References

Albrecht, S. (2013). *Probably it is only a Matter of Time*. An Empirical comparison of Connecting Adverbials in Timed and Untimed Student Writing. BA thesis (Chemnitz).

Albrecht, S. (2016). Corpus-Linguistic and Cognitive Approaches to Determiner Usage in Chinese Student Writing Testing the Fluctuation Hypothesis. MA thesis (Chemnitz).

Bolton, K., Nelson, G., Hung, J. (2002). A corpus-based study of connectors in student writing. *International Journal of Corpus Linguistics* 7:2, 165-182.

Bräuer, M. (2013). Degree Adverbs in Written Academic English: A Corpus-Based Analysis. MA thesis (Chemnitz).

Küchler (Dheskali), Jessica (2015). Usages of *may* and *will* in Chinese and German Student Writings: A corpus-based Comparison. MA thesis (Chemnitz).

Hyland, K. (2002). Authority and invisibility: authorial identity in academic writing. *Journal of Pragmatics* 34, 1091–1112.

Hyland, K. (2005a). *Metadiscourse*. London: Continuum.

Hyland, K. (2005b). Stance and engagement: a model of interaction in academic discourse. *Discourse Studies*. 7/2, 173-191.

Hyland, K. (2012). *Disciplinary identities: individuality and community in academic writing*. Cambridge: CUP.

Mauranen, A. (2012). *Exploring ELF: Academic English Shaped by Non-native Speakers*. Cambridge: CUP.

Pérez-Llantada, C. (2012). *Scientific Discourse and the Rhetoric of Globalization*. London: Continuum.

Roberts, Leah/Anna Siyanova-Chanturia (2013). Using eye-tracking to investigate topics in L2 acquisition and L2 processing. *Studies in Second Language Acquisition* 35 , 213 – 235.

Schmied, J. (2011). Academic Writing in Europe: a Survey of Approaches and Problems. In Schmied, J. (ed. 2011), *Academic Writing in Europe: Empirical Perspectives*. Göttingen: Cuvillier, 1-22.

Schmied, J. (2013). Academic Knowledge Presentation in MA theses: from Corpus Compilation to Case Studies of Disciplinary Conventions. *Brno Studies in English* 38, 149-164.

Schmied, J. (2015). Graduate Academic Writing in Europe in Comparison: a Research-based Approach to Metalanguage and Genre. In Schmied, J. (ed. 2015). *Academic Writing for South Eastern Europe: Practical and Theoretical Perspectives*. Göttingen: Cuvillier. 1-24.

Siepmann D., J.D. Gallagher, M. Hannay, J. L. Mackenzie (2011). *Writing in English: A Guide for Advanced Learners*. Tübingen: A. Francke.

Swales, J.M. (1990). *Genre Analysis. English in Academic and Research Settings*. Cambridge: CUP

Swales, J. M., Ch. B. Feak (2012, 3rd ed.). *Academic Writing for Graduate Students: Essential Tasks and Skills*. U. of Michigan Press.

van Rooy, Bertus (2010). Error, innovation and new conventions in African Englishes. Presentation Chemnitz, July 2010.