The ChemCorpus as a Reference Corpus for Comparative Studies in Academic Writing

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

1.1 Concepts in academic discourse/academic writing

o   meta discourse =
linguistic devices that assist writers to organize propositions and present them in a way that will be easily understood by readers

Interaction

Stance ———————————————— Engagement

Hedges   Boosters   Attitude   Self-   Reader   Directives   Questions   Shared   Personal
                    markers   mention   pronouns   knowledge   asides

types of reader-writer interaction (Hyland 2005: 177)
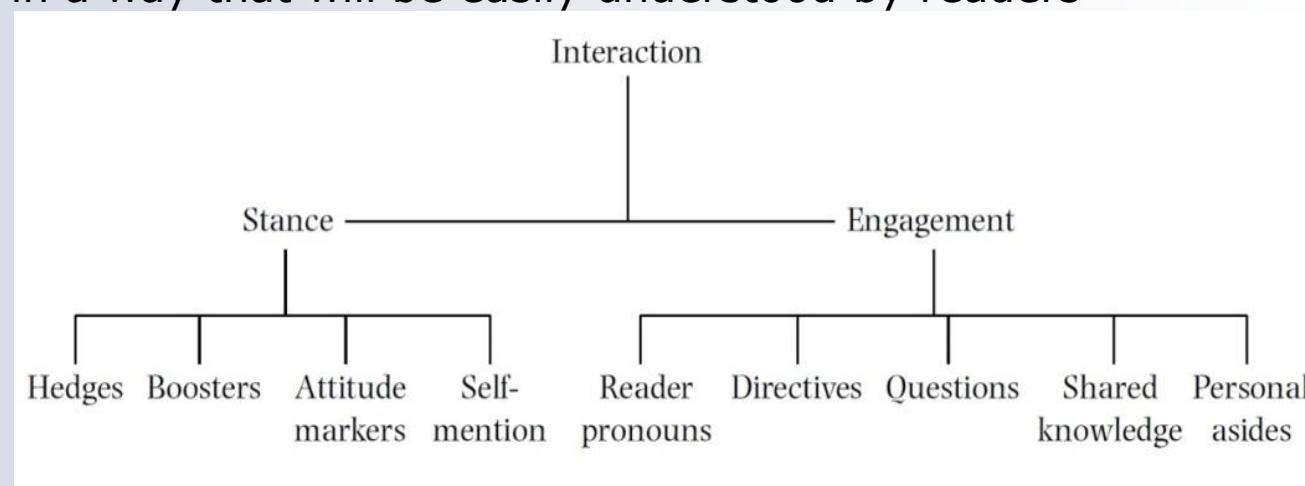
o   discourse community

o   genre

international (Swales/Feak 2012) and national? (Siepmann et al. 2011) textbooks for teaching?
included in MA (PhD) courses world-wide (e.g. Chemnitz)

1.2 Discourses in science communication

discourse community approach = a writer/speaker appeals to shared knowledge to create a community of discourse: *of course, as we know from* …

research discourse →	**instructional discourse**
↓	↓
popular discourse	**student discourse**
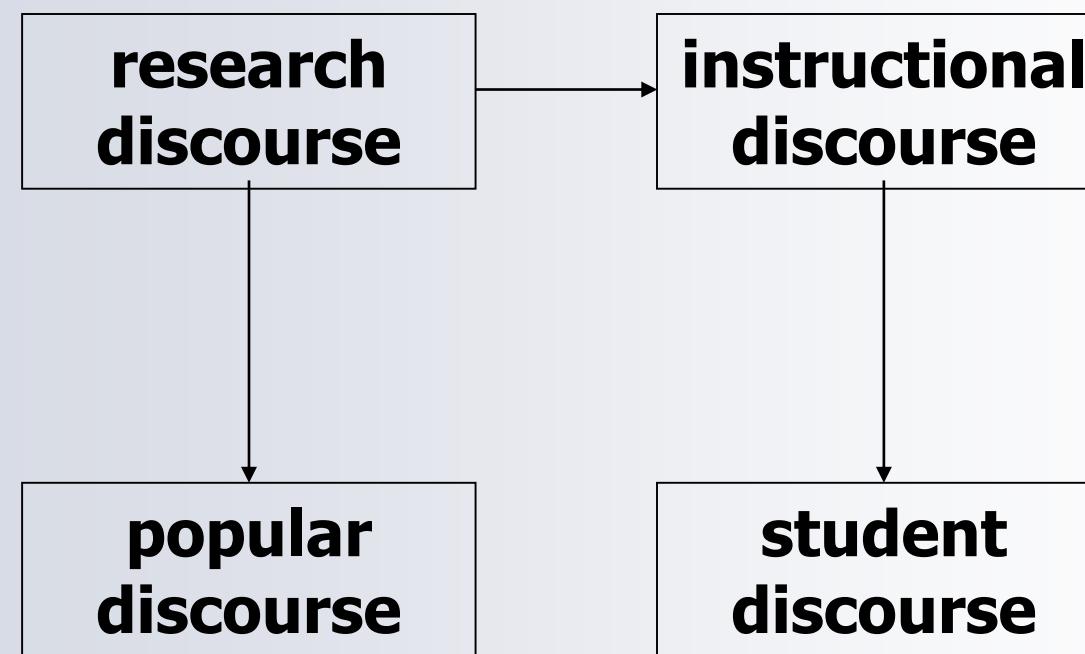
1.3 Def. Genre Approach

(1) Genres are recognizable communicative events, characterized by a set of communicative purposes identified and mutually understood by members of the professional or academic community in which they regularly occur.

(2) Genres are highly structured and conventionalised constructs, with constraints on allowable contributions not only in terms of the intentions one would like to give expression to and the shape they often take, but also in terms of the lexico-grammatical resources one can employ to give discoursal values to such formal features.

(3) Established members of a particular professional community will have a much greater knowledge and understanding of the use and exploitation of genres than those who are apprentices, new members or outsiders.

(4) Although genres are viewed as conventionalised constructs, expert members of the disciplinary and professional communities often exploit generic resources to express not only 'private' but also organizational intentions within the constructs of 'socially recognized communicative purposes'.

(5) Genres are reflections of disciplinary and organizational cultures, and in that sense, they focus on social actions embedded within disciplinary, professional and other institutional practices.

(6) All disciplinary and professional genres have integrity of their own, which is often identified with reference to a combination of textual, discursive and contextual factors. (Bhatia 2004: 23)

1.4 Genres in discourses

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

discipline-specific
culture-specific

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

instructional disc./e-learning
- ppt presentations
  - lectures
  - student presentations
- textbooks
  - Wikis
  - www pages (HTML, php)

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

"Novice Academic English"

author-specific
culture-specific

1.5 Matrix of genre types in Academic Writing

CONTEXT	**GENRE**	type	medium	audience/ readership	expertise in years	length	purpose	*comments*
article	**research** (scholarly) **article**	r	w	specialist	5	5,000+	drive research	in double-blind peer-reviewed journal with impact factor
book	research book	r	w	specialist	6	10.000	drive research	also monograph, trend towards IMRAD
	textbook	r	w	novice	8	6-10,000	introduce novices	also course book, intoduction?
	handbook	r	w	specialist	10+	10,000+	state-of-the-art	honour+reliable?
	book reviews	r	w	specialist	4	1.000	evaluate/critique	also section in book/thesis
	state-of-the-art review	r	w	specialist	10+	2.000	start project?	research field survey
	article collection (ed)	r	w	specialist	10	150+p	state-of-the-art?	multiple authors,1 topic/area;conference?
project doc.	project proposals	r	w	specialist?		5.000	accepted for qualif./funding	a promise with milestones to check progress
	BA/MA project proposal	t	w	learner	3	1.000	qualification	trend to IMRAD, research questions/hypotheses
	PhD project application	r	w	learner	5	5.000	qualification/funding	trend to IMRAD, research questions/hypotheses
	(thesis/PhD) defense	r	s	learner	3-5y	20+20-40m	qualification	trend to IMRAD, research questions/hypotheses; open discussion
	project/progress report	r	w	specialist		6-10,000	justify expense	multiple authors,1 topic/area
conference	conference presentations	r	s	specialist	5	15+5m	drive research?	protect area? ask advice?
	key-note (lecture)	r	s	specialist	10+	45-60+15-30m	state-of the-art?	famous scholar,personality
	plenary (lecture)	r	s	specialist	10+	30-60m+15	research overview	broad topic?
	progress report	r	s	specialist	5	10m	demonstrate research? ask advice?	
	(conference) abstract, proposal	r	s	specialist	5	300-600w	acceptance	
	poster introduction	r	s	specialist	5	3m	view poster	very consice+attractive?
	conference poster	r	w	specialist	5	1 A0	introduce research,popularise?	Rtrend to IMRAD; fig/tables+ref.s
	conference proceedings (ed)	r	w	specialist	10+	150+p	document research	multiple authors,1 topic/area
	conference report	r?	w	public	3	1.000	drive research?	popular? dissemination
university	lecture	t	s	learner	10+	45-90m	disseminate knowledge	competes with textbook? problem attention span
teaching	student/seminar presentation	t	s	learner	1	10-20m	qualification	problem discussion? self-protective?
	Wiki	t/r?	w	learner	1		collaborate in Knowledge creation	groupwork/multiple authors
	classroom discussion	t		learner	1		collaborate in knowledge creation	problem interaction
	field notes	t	w	self?	1		collect information	data collection methodologies?
	BA thesis	t/r?	w	specialist	3	40+p	qualification	
	MA thesis	r	w	specialist	5	60+p	qualification	
	PhD thesis	r	w	specialist	8	200+p	qualification	conventions,not a research book
	habilitation/postdoctoral thesis	r	w	specialist	10+	200+p	qualification	tend to small; replaced by articles in p-r j.?
subsidiary?	(article) abstract	r	w	specialist	5	1-300w	read? full article	part (free; to decide worth paying?)
	handout	t/r?	w	l/sp	1	1-2p	support,take-home	large diagrams,figures,statistics; examples;references
"valorisation"	university journal,newsletter	r	w	public	5	1-2p	demonstrate "value"	untrained?
	popular blog	r	w	public?	1	1+1+1	time-line of development (projects,carrer)	technical platform for old genres?
	popular science book	t/r?	w	public	8	80-200?	create interest in research?	

1.6 Prototype approach to genres in academic writing
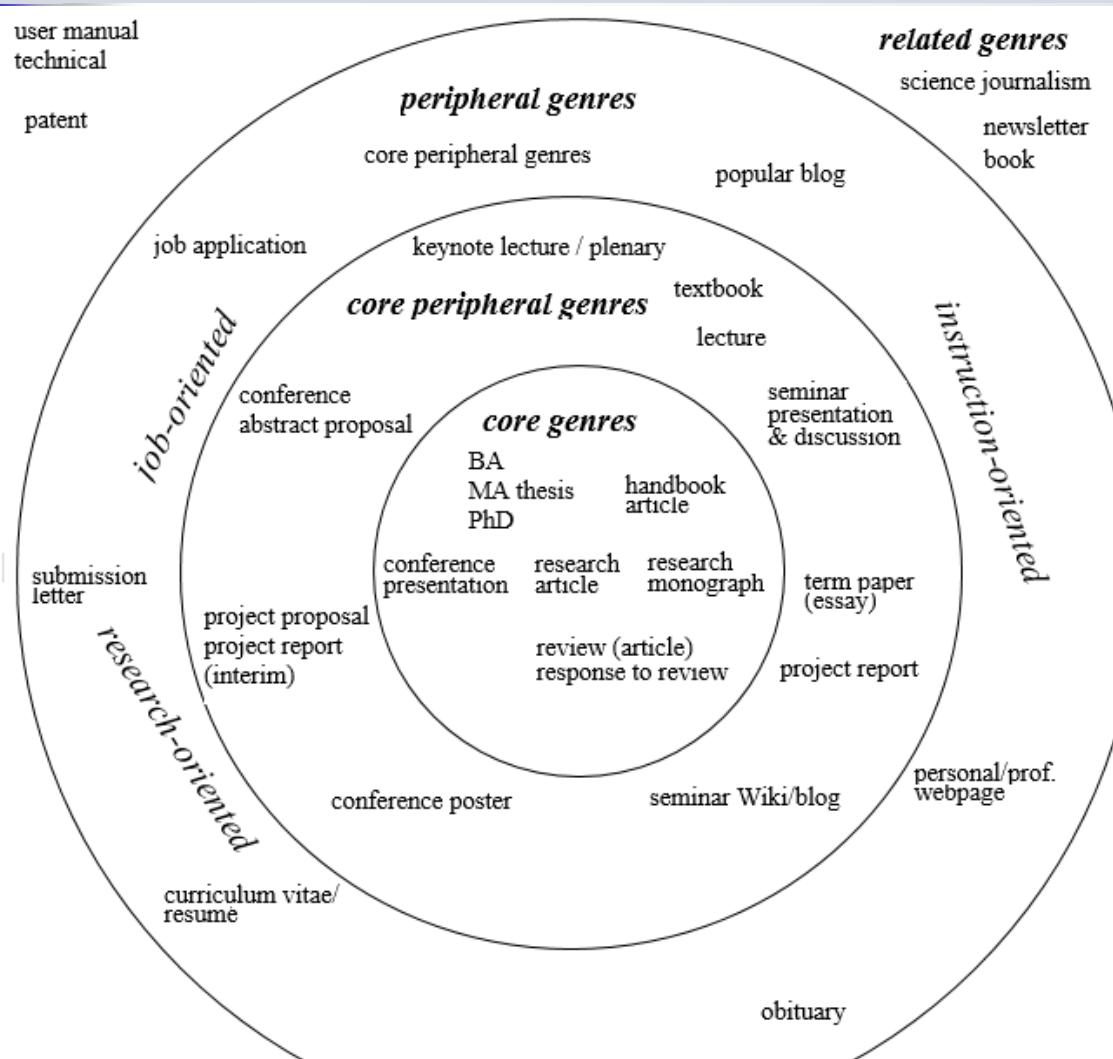
1.7 CARS model for research paper introductions
(Swales 1990: 141)

Move 1: Establishing a territory
        Step 1 Claiming importance and/or
        Step 2 Making topic generalizations and/or
        Step 3 Reviewing items of previous research
Move 2: Establishing a niche
        Step 1a Counter-claiming or
        Step 1b Indicating a gap or
        Step 1c Question-raising or
        Step 1d Continuing a tradition
Move 3: Occupying the niche
        Step 1a Outlining purposes or
        Step 1b Announcing present research
        Step 2 Announcing principle findings
        Step 3 Indicating article structure

1.8 AIMAC organization: Revised IMRAD structure

structure	substructure (tactical alternative)	key lexemes/phrases (as indicators)
A = abstract	keywords in context	*focus, discuss, approach*
I = issue	new	*not enough research yet*
	relevant	*important, academic discourse, practical application*
	focussed	*concentrate, emphasise, purpose*
M = methodology	previous research, i.e. lit. review incl. evaluation	*concept developed, review, refer to, proceed to, claim*
	hypotheses possible?	*research question*
	data base	*corpus, data collection*
	tests/procedure	*calculate*
A = analysis	examples as evidence	*illustrate, show, prove*
	statistical tables as summaries	*table, figure, diagram, graph, bar*
	significance to generalise	*significant, chi^2*
C = conclusion	summary	*in conclusion, finally/at last, we have shown, discussed above*
	interpretation	*this proves that*
	contextualisation	*in a wider perspective, apply*
	limitations	*more data, beyond the scope*
	outlook	*further research is necessary, predict, dissemination/application of results*

2. ChemCorpus as a reference corpus
2.1 Comparative Studies issues: why?

research questions:

o   What is effective writing? - reader-oriented?

o   How much culture-specific writing is acceptable?

o   What is good *lingua franca*/non-native writing?

o   Can we accept features more easily when they are also common in other cultural academic traditions?

o   Are some features functionally motivated (explicitness) and thus acceptable?

discover "more grammar than meets the eye"
Tool: Antconc (like WordSmith)
→ relative frequencies ←comparable corpora? subcorpora

2.2 ChemCorpus set-up by genre and specialisation

genre	specialisation	total words	average length
term paper BA	language/linguistics	0.5 Mill.	4,200
	culture/literature	0.5 Mill.	4,700
project report	(cultural)	0.5 Mill.	4,000
BA thesis	language/linguistics	1 Mill.	12,000
	culture/literature	1 Mill.	16,000
term paper MA	language/linguistics	0.5 Mill.	5,700
	culture/literature	0.5 Mill.	6,600
MA thesis	language/linguistics	1 Mill.	25-30,000
	culture/literature	1 Mill.	25-30,000
total		3.5 Mill.	

3. Organisation, argumentative structure
3.1 AIMAC from natural sciences into humanities?

structure	substructure (tactical alternative)	key lexemes/phrases (as indicators)
A = abstract	keywords in context	focus, discuss, approach
I = issue	new	not enough research yet
	relevant	important, academic discourse, practical application
	focussed	concentrate, emphasise, purpose
M = methodology	previous research, i.e. lit. review incl. evaluation	concept developed, review, refer to, proceed to, claim
	hypotheses possible?	research question
	data base	corpus, data collection
	tests/procedure	calculate
A = analysis	examples as evidence	illustrate, show, prove
	statistical tables as summaries	table, figure, diagram, graph, bar
	significance to generalise	significant, chi^2
C = conclusion	summary	in conclusion, finally/at last, we have shown, discussed above
	interpretation	this proves that
	contextualisation	in a wider perspective, apply
	limitations	more data, beyond the scope
	outlook	further research is necessary, predict, dissemination/application of results

3.2 Word outline view of a BA thesis indicating an IMRAD structure

Table of Contents

3.3 Table of Contents of a BA thesis indicating a non-IMRAD structure

Polish Migration to the UK

Contents

3.4 Table of Contents of a linguistic BA term paper indicating an IMRAD structure

Table of Content

3.5 Table of Contents of a literature BA term paper indicating a non-IMRAD structure

Table of Contents

4. Linking

Indicators: cohesive devices

formal:
        conjunctions: *but*, *while*
        adverbs: *first, then, finally*
functional:
        sentence adverbials, e.g. clause-initial adverb **ly,*

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

4.4 Functional Categories by L1 (Albrecht 2013: 36, figure 11)
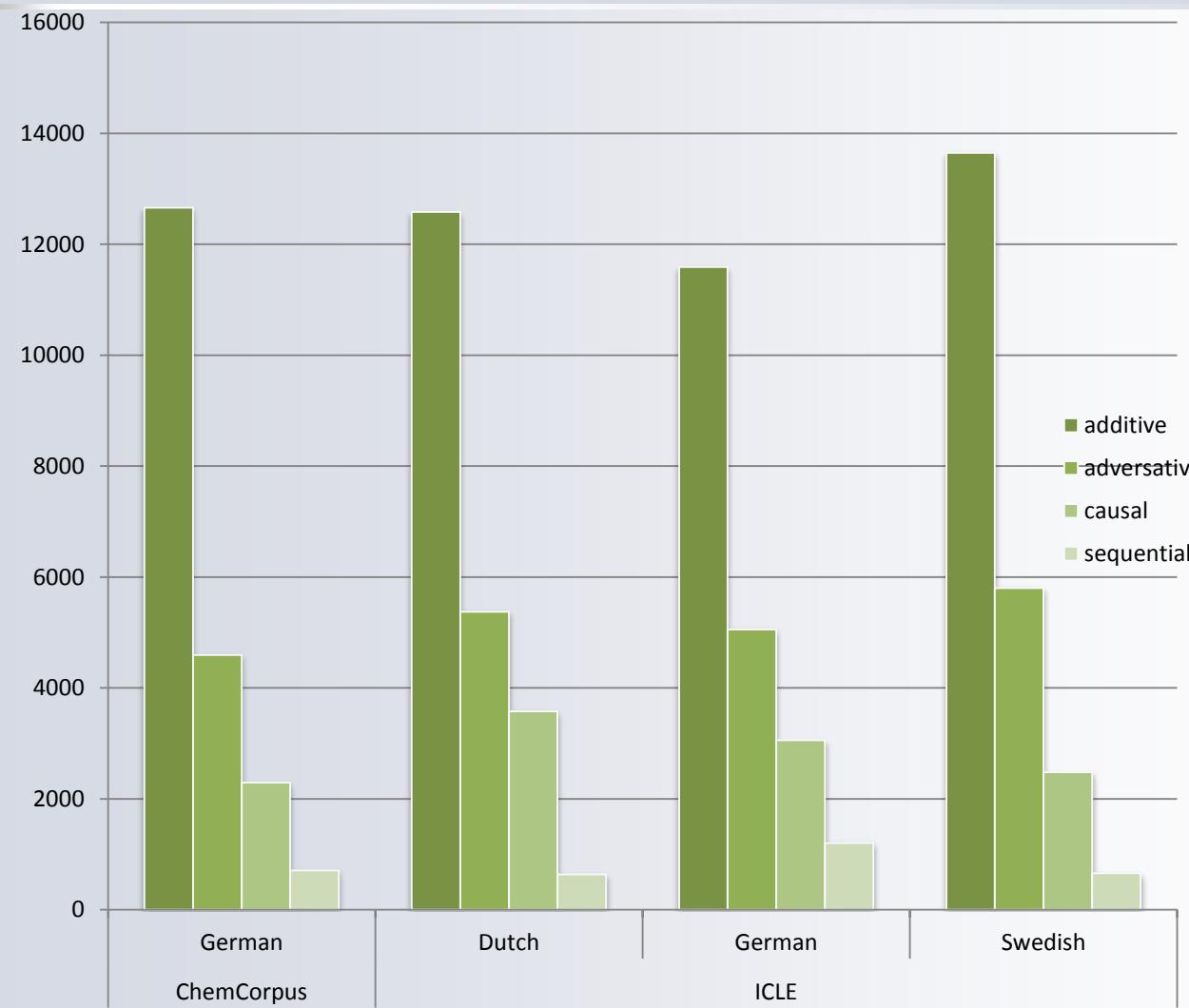
5. Stance

key terms: stance > hedging > modality
  - ᵒ "personal feelings, attitudes, value judgments, or assessments" (Biber et al. 1999: 966)
  - ᵒ "subjective" evaluation on the basis of own knowledge, experience, etc.
  - ᵒ context-dependency
  - → academic culture in the discourse community determines how stance is expressed!

author stance and engagement are crucial variables in academic interaction:
"writing is always a personal and socio-cultural act of identity whereby writers both signal their membership in a range of communities as well as express their own creative presence" (Hyland 2006: 35)

5.1 AntConc concordance (KWIC) of *really* in BA term papers in the ChemCorpus

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

Hit	KWIC
1	zed definition of what morphological productivity really is. Several linguists have already dealt with thi
2	eak down. This is what everyone says, but is this really true? This will be the question this paper is goi
3	outcomes so it is possible to judge if devolution really can break up Britain. The last chapter "Conclus
4	political developing of the states. If it is not really possible to co-ordinate the states it is not real
5	ally possible to co-ordinate the states it is not really possible to find a common denominator. As a matte
6	quence this problem with not having the chance to really co-ordinate the states the danger of the UK drift
7	itain? The answer is very simple. Yes, devolution really can break up Britain. Devolution can lead to a
8	: 124) Sade needs urgently answers but no one she really trusts in is near her to talk. She needs someone
9	her to talk. She needs someone to confide in, she really needs her father but she does not know where he i
10	annot come back anymore. Nevertheless, she is not really gone. She will always stay in Sade's mind and giv
11	sometimes without directly saying it. She cannot really understand why her uncle and her father had send
12	r Side of Truth is a very brilliant story which I really enjoyed to read. <ref>< References
13	of course I wished for sons, but having daughters really involves you in issues regarding sexism and patri
14	tries to make the girl understand that it is not really her fault that the beloved grandfather does not r
15	y her fault that the beloved grandfather does not really want to have anything to do with Kahu.
16	s the so called "Bloody Sunday". This incident is really well-known and there are several movies and songs
17	to Hidge's party. So Rodney's relaxed attitude is really a good example for living a life without any worr
18	elong. However, the colours might not help by really figuring out where the house is located or in wha
19	ratings, has debased notions of what the country really is. Historically, it is fair to assert that news
20	ich we have very little control". Hence the truth really lies hidden under the surface. (Guerin 127) And t
21	looking back at their sources or origin. We don't really care why things are the way they are and rely on
22	I know, it has not always been a curse word, but really only described a person that is dark- skinned.
23	me utterance, but we can never be sure if this is really the right answer or not. Going back in history an
24	His uncle lived in the countryside, which Joseph really enjoyed. His father became very ill and was freed

5.2 AntConc concordance (KWIC) of *definitely* in BA term papers in the ChemCorpus

5.3 Relative frequencies of *very* in BAWE, MICUSP, the BrnoCorpus and ChemCorpus (Bräuer 2013: 63, figure 3)
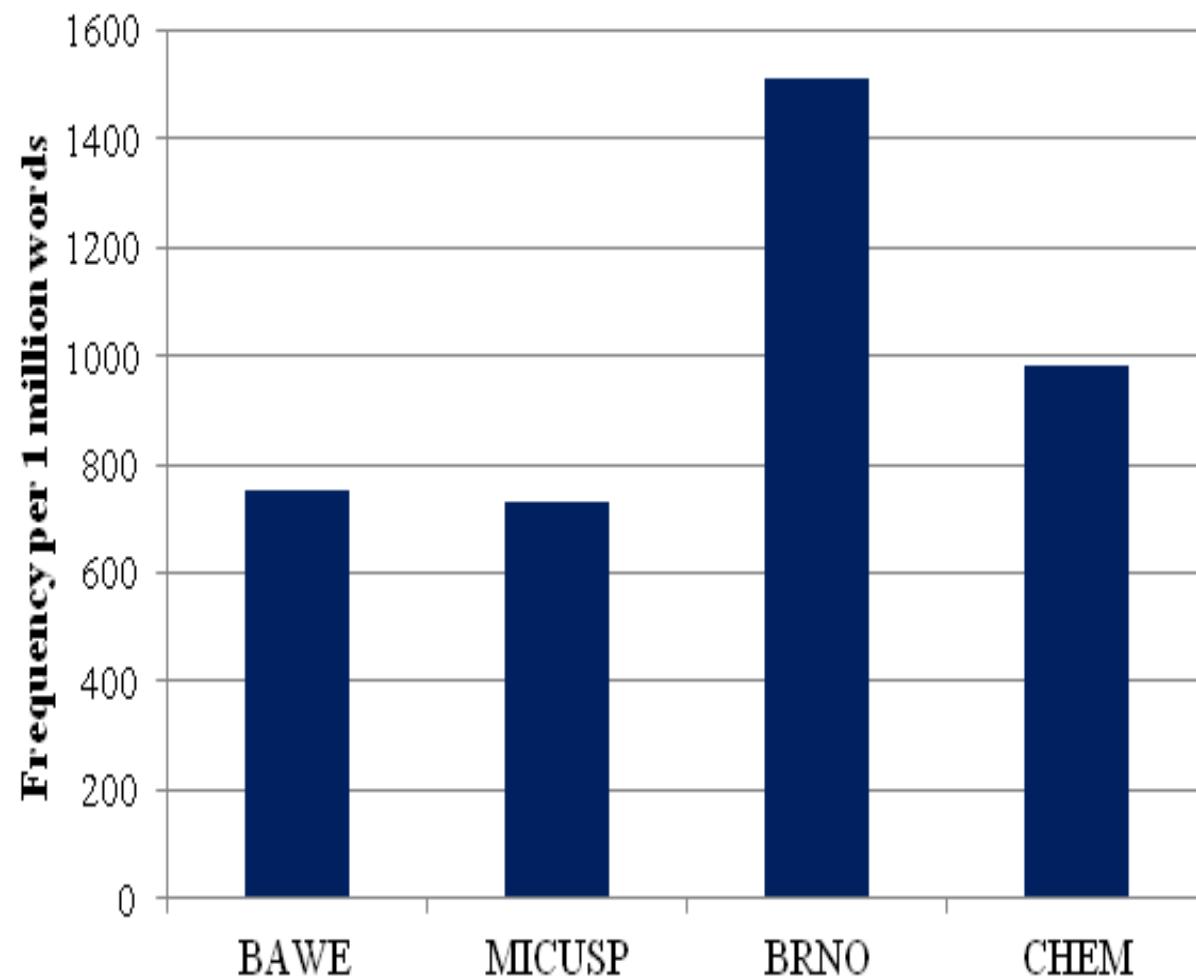
5.4 Rel. frequencies per 1 million words in the BrnoCorpus and ChemCorpus (Bräuer 2013:49, table 17)

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

6. Conclusion

Academic English is an accessible and useful topic for students of applied linguistics (including TESOL)

metadiscourse features can be learnt easily and inductively

- junior writers, MA students become aware of reader-writer interaction – and can use this in their own thesis writing, and so on
- through explorative investigation of model texts they find their own stance between personal identity and disciplinary conventions
- Can corpus databases replace native-speaker introspection?
  - objective – subjective?
  - frequency + attitudes?!
- Are conventions becoming more and more similar because of Anglo-American dominance (gate keepers and guidebook publications)?
- Can we establish an academic *lingua franca* norm on a functional basis? – even against Anglo-American traditions?
- Can the (non-native) ChemCorpus serve as a model?
- Remember there are no native speakers of Academic English?

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).

Bolton, K., G. Nelson, J. Hung (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).

Gao, Quoping (2012). Interpersonal Functions of Epistemic Modality in Academic English Writing. *Chinese Journal of Applied Linguistics* 35, 352–364.

Hůlková, I. (2011). Conjunctive Adverbials in Academic Written Discourse: Conjunctive Adverbials in Academic Written Discourse. In Schmied, J. (ed. 2011), 129-142.

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

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

Schmied, J. (2015). Graduate Academic Writing in Europe in Comparison: a Research-based Approach to Metalanguage and Genre. In Schmied, J. (ed. 2015), 1-24.

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

Schmied, J. ed. (2015). *Academic Writing for South Eastern Europe: Practical and Theoretical Perspectives.* Göttingen: Cuvillier Verlag.

Schmied, J. ed. (2011). *Academic Writing in Europe: Empirical Perspectives.* Göttingen: Cuvillier Verlag.

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

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