

**Analysing our own students' Writing:
Differences & parallels as a basis for
establishing functional standards?**
(Josef Schmied)
Preparing & using the Modena Corpus
(Matteo de Christoforo)

Josef Schmied
English Language & Linguistics
Chemnitz University of Technology
http://www.tu-chemnitz.de/phil/english/ling/presentations_js.php



Introduction Why What How Conclusion 3/30

1. Concepts in teaching academic discourse and writing

academic vs. student writing
discourse community approach
genre
metadiscourse

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

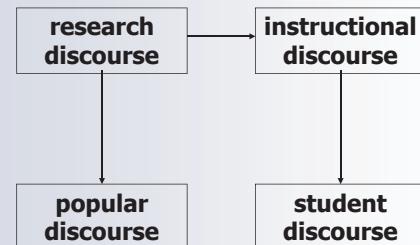
lingua franca, non-native standards developing on a functional basis?
language variation = error + individual style

metadata:

- to follow individual student development from BA to MA
- to follow teaching (conventions) development over 20 years, pre-/post-Bologna MA

Introduction Why What How Conclusion 4/30

1. Discourses in science communication



Introduction Contexts Examples Teaching Conclusion 5/30

1. The Genre Approach

research "output" <ul style="list-style-type: none"> research article book reviews conference presentations project proposals / reports handbooks 	teacher "talk"/e-learning <ul style="list-style-type: none"> ppt presentations <ul style="list-style-type: none"> lectures student presentations textbooks <ul style="list-style-type: none"> Wikis/blogs www pages (HTML, php)
discipline-specific culture-specific	
science "journalism" <ul style="list-style-type: none"> popular science articles (<i>New Scientist</i>) popular blogs (David Crystal) popular science films (<i>Horizon</i>) popular science books science slam 	student "literacy" / "Novice Academic English" <ul style="list-style-type: none"> lecture/fieldwork notes "essays"/"papers" seminar presentations BA/MA/PhD theses
author-specific culture-specific	

Concepts Why What How Conclusion 6/30

2. Comparative Studies data: why

2.1 How can we find comparable data?

CLARIN (=Common Language Resources and Technology Infrastructure
<https://www.clarin.eu/resource-families/L2-corpora>

<https://uclouvain.be/en/research-institutes/lil/cecl/learner-corpora-around-the-world.html>

data collection:

Czech: BrnoCorpus
South Africa: MA Stellenbosch
China: MA, PhD theses
Cameroon: Nkemleke 2011

all vs. ChemCorpus (of Chemnitz student writing): BA, MA ...

2. Comparative Studies

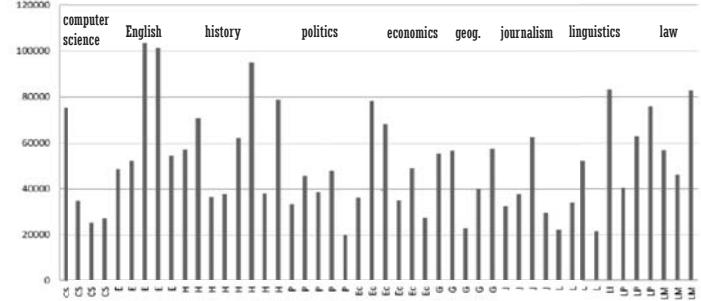
2.2 Examples

	ChemCorpus	ChAcE Corpus	Africa		
country	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
<i>PhD theses</i>		106			
<i>other: term papers +</i>	100 + 100	(SYSU 50)			
total	?	?	?	?	?

2. Comparative Studies: differences between disciplines

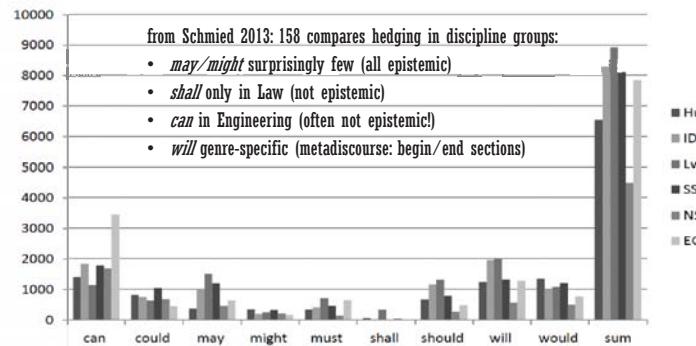
Figure 1a. "Soft science" texts in the ZAMA corpus according to text length (names starting with discipline acronyms)

from Schmied 2013: 153,
100 MA theses across soft and hard sciences



2. Comparative Studies: differences between disciplines

Figure 4. Modals by discipline type (per 1 million words)



2. Comparative Studies discourse: why

2.3 "Non-native" "academic" writing is special
in academic writing nobody is native!

even more:

native conventions may not be functional and
thus may be "expanded", sub-rules "neglected", "simplified",
etc. without risking comprehensibility
(e.g. relative constructions, articles)

"academic" = research-based

at least at "novice"/university level (C1?, MA?)
NOT argumentative essay writing in practical language teaching/learning,
as in ICLE=International Corpus of Learner English

Modena corpus texts

3. Qualitative analyses of student texts

examples from advanced level of non-native English

distinguish between L1 specific and L2 specific?

from "local" (grammatical, formal) errors to "global" (textual, functional) errors and back

Modena
Alessio
24/11/22

Concepts	Why	What	How	Conclusion	19/30
4. Quantitative analyses: teaching					
O'Keeffe, McCarthy, Carter (2007) or https://www.english-corpora.org/language-learning.asp					
Feature	Importance for language learning and teaching				
Alternative phrases	One of the hardest things for language learners is knowing which words sounds good together. For example, which synonym of <i>parent</i> is most common with the word <i>argument</i> ? One simple, fast search in COCA provides this information. Searches like this are either not possible or are very cumbersome and time-consuming with other online corpora, such as Sketch Engine or CQPWeb.				
Word sketches	Learners want to see rich information on specific words (not just collocates). At English-Corpora.org, for every one of the top 60,000 words in a corpus, you can see the definition, synonyms, more specific and more general words, collocates, related topics, clusters, concordance lines, frequency, and links to external resources like pronunciation, images, videos, and translations for 100+ languages.				
Browse	Learners want to search for words, and they want to find words by frequency (so they can see where they might have gaps in their vocabulary). At English-Corpora.org, you can search for words by word form, part of speech, frequency, meaning (e.g. words in a definition), synonym, more specific or more general words, and even pronunciation.				
Find related words	Words are best learned as part of a "system" of related words. For example, if learners can relate <i>telescope</i> to other concepts like <i>Earth</i> , <i>Sun</i> , <i>star</i> , <i>planet</i> , <i>galaxy</i> , <i>universe</i> , <i>scientist</i> , or <i>astronomy</i> , they have a better chance of knowing what <i>telescope</i> means, and of remembering it. Only English-Corpora.org allows learners to find both collocates and related topics (which co-occur <i>anywhere</i> in the text), and which provide great insight into the meaning of a word.				
External resources	Many language learners benefit from multi-modal information for a given word or phrase, such as pronunciation, images, videos, and translation to their native language. English-Corpora.org has the only corpora that link to so many types of external resources, in so many useful ways. And when you're looking at Keyword In Context (KwIC) entries for a word, there are a wide range of "one click" resources that help you to kind information on words that you might not already know.				

Modena
Alessio
24/11/22

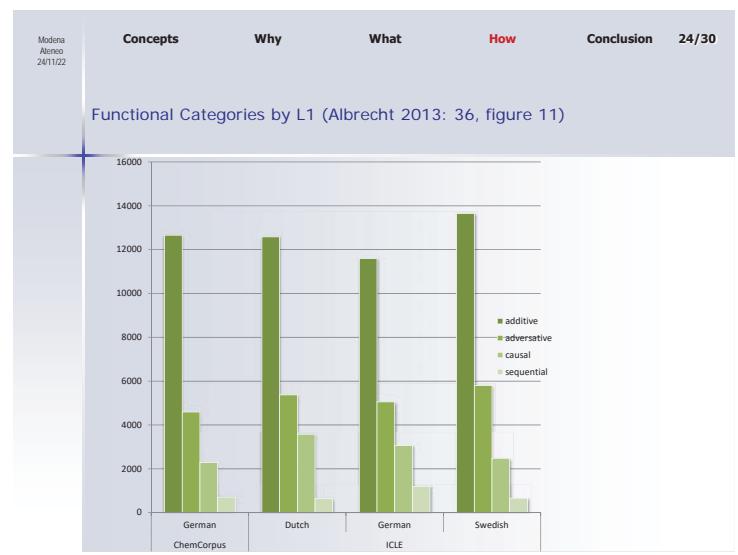
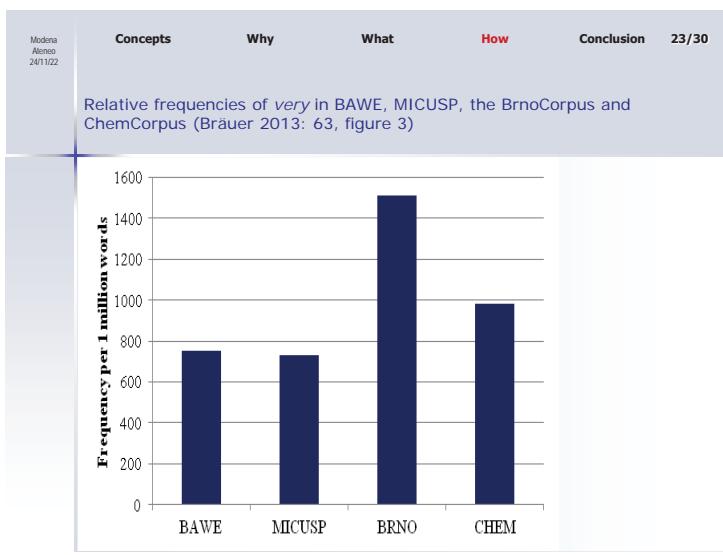
Concepts	Why	What	How	Conclusion	20/30
Cohesive devices: linking, connectors, etc. e.g. clause adverbials, clause-initial <i>so</i> , special case?					
coherence = extralinguistic factors contributing to the creation of texture (=implicit) cohesion = linguistic means which create texture (=explicit)					
1	It is unfortunately not the case that knowledge of the mechanical				00C2AX
2	Unfortunately, an instantaneous comparison was not possible				00C3AX1
3	Unfortunately, exact information on gas usage is unavailable				00C3AX1
4	Unfortunately, one has no a priori way of predicting				00C3AX1
5	Unfortunately, neither easily works.				00C4AX1
6	Unfortunately, there are as yet no reliable calculations that				00C4AX1
7	Unfortunately our experience is confined to an equilibrium ...				00C4AX2
8	Unfortunately, the energy resolution of a neutrino telescope is ...				0015AX
9	... technique to obtain black hole masses which, unfortunately , is unfeasible				0018AX
10	Unfortunately, lack of information about the collective ...				0021AX
11	Unfortunately, the complexity of each subprocess also grows ...				0025AX
12	Unfortunately these theorems tell us practically nothing about ...				0032AX
13	Unfortunately there are very few analytic results available				0040AX
14	Unfortunately, all recent experiments are, in principle,				0043AX
15	Unfortunately, all those exciting recent experiments are				0043AX
16	Unfortunately, those prior art solutions require daunting experiments				0043AX
17	Unfortunately, the NID is uncomputable since the constituent ...				0045AX
18	Unfortunately, in many places such information is getting harder ...				0056NS
19	It's a good protective barrier," says Hildebrand, " unfortunately ."				0069NS
20	Unfortunately , because of the rarity of plant data from this ...				010CPN

Modena
Alessio
24/11/22

Concepts	Why	What	How	Conclusion	21/30
Relative frequencies per 1 million words in the BrnoCorpus and ChemCorpus (Bräuer 2013: 49, table 17): from global to local					
	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 ≤ p < 0.01		
tremendously	1.90	7.99	> 0.05		
very	1,510.40	981.03	0.01 ≤ 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 ≤ p < 0.05		
fully	83.59	63.96	0.001 ≤ 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 ≤ 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			

Modena
Alessio
24/11/22

Concepts	Why	What	How	Conclusion	22/30
Relative frequencies per 1 million words in BrnoCorpus by academic discipline (Bräuer 2013: 54, table 19)					
	Literature	Linguistics	Methodology	p-value	
Amplifier	2,555.75	1,646.22	2,378.79		
Booster	1,743.94	1,228.78	2,106.06		
considerably	25.27	35.28	30.30	> 0.05	
highly	208.51	99.95	111.11	0.001 ≤ p < 0.01	
strongly	0.00	70.55	30.30	> 0.05	
tremendously	0.00	0.00	5.05	> 0.05	
very	1,510.15	102.01	1,929.29	> 0.05	
Maximizer	511.81	417.43	272.73		
absolutely	94.78	64.67	25.25	> 0.05	
clearly	23.79	135.22	90.91	> 0.05	
extremely	120.05	47.03	60.61	> 0.05	
fully	31.59	135.22	80.81	> 0.05	
thoroughly	31.59	35.28	15.15	> 0.05	
Downtoner	916.20	487.99	469.70		
Approximator	445.94	305.73	217.17		
almost	303.29	252.81	171.72	> 0.05	
nearly	145.33	52.91	30.30	> 0.05	
virtually	6.32	0.00	15.15	> 0.05	
Diminisher	131.93	94.07	196.97		
a bit	3.91	35.28	141.41	0.001 ≤ p < 0.01	
slightly	214.83	47.03	35.35	> 0.05	
somewhat	44.23	11.76	20.20	> 0.05	
to some extent	18.96	0.00	0.00	> 0.05	
Minimizer	145.33	88.19	55.56		
barely	12.64	5.88	5.05	> 0.05	
hardly	120.05	70.55	45.45	> 0.05	
scarcely	12.64	11.76	5.05	> 0.05	
Total	3,171.96	2,134.20	2,848.48		



Modena Atheneo 24/11/22	Concepts	Why	What	How	Conclusion	25/30
Diachronic comparisons: Magister vs. MA theses (Schmied 2023)						
Field	Magister theses (2002-2013)		Master's theses (2012-2020)		Total words	
	texts	words	texts	words		
Linguistics	10	321,967	10	265,088	587,055	
Cultural Studies	10	324,137	10	254,106	578,243	
Total	20	646,104	20	519,194	1,165,298	

Table 4: ChemCorpus (Schmied & Dheskali 2015-2020) sample for this study

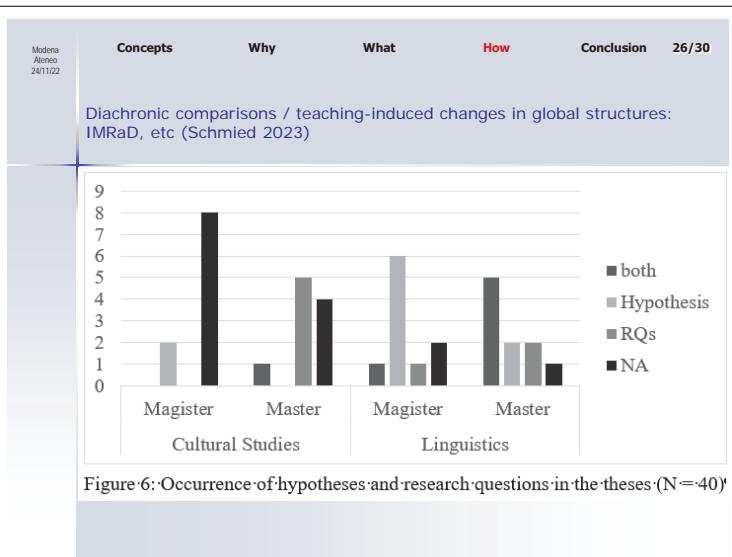


Figure 6: Occurrence of hypotheses and research questions in the theses (N=40)

Modena Atheneo 24/11/22	Concepts	Why	What	How	Conclusion	28/30
Modena corpus examples						

Modena Atheneo 24/11/22	Concepts	Why	What	How	Conclusion	28/30
5. Conclusion 1: critical review						
for natural language/idiomaticity on the local level						
<p>5.1 "local" errors may be 1st language induced and culture-specific ("Chinese does not have articles/determiners")</p> <p>5.2 "global" errors may be target-language-(and culture-)specific ("English uses a complex system of modal auxiliaries")</p> <p>5.3 some errors may be academic-writing related (nominalisations, compounds)</p> <p>5.4 test usage using on-line data-bases/corpora byu-corpora (google scholar)</p> <p>5.5 find good models in area of interest, library, ...</p>						

Modena Atheneo 24/11/22	Concepts	Why	What	How	Conclusion	29/30
5. Conclusion 2: towards functional standards?						
some culture-specific features are stereotyped - but not unique! non-culture-specific features are as interesting as culture-specific ones?						
corpus databases can replace native-speaker inspection: - objective - subjective? - frequency + attitudes?						

Modena Atheneo 24/11/22	Concepts	Why	What	How	Conclusion	30/30
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., Nelson, G. & Hung, J. (2002). A corpus-based study of connectors in student writing. *International Journal of Corpus Linguistics* 7, 2, 165-182.

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

Díaz-Negrillo, A., Fernández-Domínguez, J. (2006). Error Tagging Systems for Learner Corpora. *RESLA* 19, 83-102.

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

Hulíková, I. (2011). Conjunctive Adverbials in Academic Written Discourse: Conjunctive Adverbials in Academic Written Discourse. In Schmied, J. (ed.), 123-142.

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.

O'Keeffe, A., M. McCarthy & E. Carter (2007). *From Corpus to Classroom. Language use and language teaching*. Cambridge U.P.

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

Schmied, J. (2011). Academic Writing in English in Comparison: Linkers and Adverbials in the ChemCorpus and Comparable Data-bases. In R. Plö & C. Pérez-Llantada (eds). *English as a Scientific and Research Language in Europe*. Berlin: De Gruyter Mouton, 159-184.

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

Schmied, J. (2023). English MA Theses at a German University before and after the Bologna reform: Comparing global rhetorical structures and stances in Linguistics and Cultural Studies. *TOKEN Special Issue*.

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

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