

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

Josef Schmied  
English Language & Linguistics  
Chemnitz University of Technology  
[http://www.tu-chemnitz.de/phil/english/ling/presentations\\_js.php](http://www.tu-chemnitz.de/phil/english/ling/presentations_js.php)



**1. Concepts in teaching academic discourse and writing**

academic vs. student writing  
discourse community approach  
genre  
metadiscourse

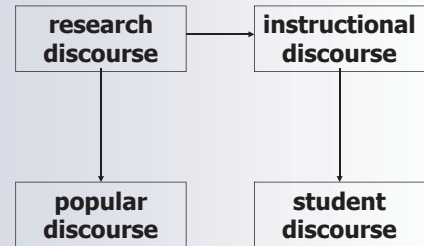
- international (Swales/Feak 2012) and national? (Siepmann et al. 2011) textbooks for teaching?
- included in MA (PhD) courses world-wide (e.g. Chemnitz)

*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

**1. Discourses in science communication**



**1. The Genre Approach**

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

**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/ilic/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

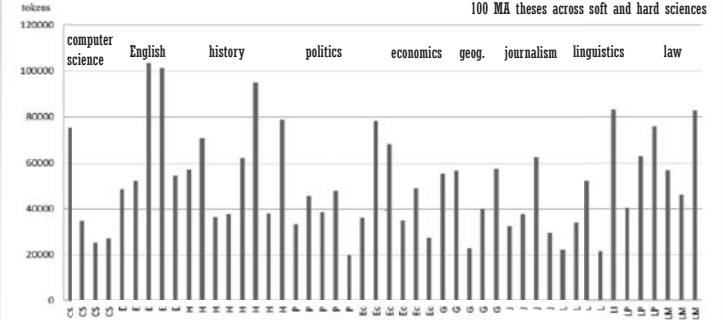
### 2.2 Examples

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

## 2. Comparative Studies: differences between disciplines

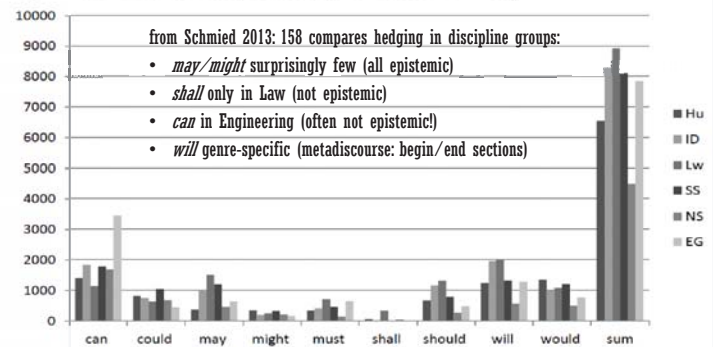
Figure 1a. "Soft science" texts in the ZAMA corpus according to text length (names starting with discipline acronyms)

from Schmieid 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 Aleno 24/11/22

Concepts	Why	What	How	Conclusion	13/30
----------	-----	------	-----	------------	-------

### 3. Sample text book review: Chinese

style marked: **informal** **idiomatic** **elaborate** **explicit**

This book was first published in 1999 by Continuum International Publishing Group **with** its full title *Construing Experience through Meaning: A Language-based Approach to Cognition*, and authored by M. A. K. Halliday and Christian M. I. M. Matthiessen, and republished in 2008 by World Publishing Company. Halliday is a **famous linguist in the world** and the founder of Systemic Functional Linguistics. Matthiessen **got** his Master's degree from University of California, Los Angeles, and mainly conducts researches on Systemic Functional Linguistics at the present.

As its name suggests, this book aims at a study of cognition from the perspective of language. Instead of inspecting the reflection of cognition in human language, it tries to construct a framework of human experience by the use of meaning system. In order to **get** to this objective, the book is divided into five parts.

The first part is a general introduction **of** the theoretical preliminaries of this book. It sees language as functions, with three metafunctions specifically: ideational metafunction, interpersonal metafunction, and textual metafunction. This book is mainly concerned with the first function, ideational metafunction, and an important conception is introduced to the readers—the ideation base, which leads to the second part, a detailed expatiation of the ideation base.

Modena Aleno 24/11/22

Concepts	Why	What	How	Conclusion	14/30
----------	-----	------	-----	------------	-------

### 3. Sample text article abstract: Czech

style marked: **informal** **idiomatic** **elaborate** **explicit**

Terms change meanings as soon as they are used in a specific language, context, discipline, or doctrine. Evidence is one of them. This article claims that evidence is a concept we can communicate universally and consists of three parts for the development of the related discourse. In this article we will in part I examine the conditions and limits the postmodern era offers **for evidence in argumentation** in texts by thinkers of postmodernity, research, and commentators. In part II we describe the place of evidence in a theoretical rational discipline (rhetoric) focusing on the question how evidence refers to the classical model of argumentation in rhetoric. With a topological model of evidence we conclude that the evidential argument as a carrier of argumentation can be an oral, written, or sensual experienced, e.g. visual, means and illustrate this in examples that comprises means of evidence and communicative means. **Despite** the employment of evidence in a variety of fields, we will show that evidence as a concept can be described as reference to sensual experience. In part III we **show in examples** the conditions and practical application of evidence. We claim that when artificial and technical processes dominate **development of** knowledge and limit evidential aspects, the obtained knowledge can only be a knowledge that is a reference to the technical process, but lacks evidential authenticity.

Modena Aleno 24/11/22

Concepts	Why	What	How	Conclusion	15/30
----------	-----	------	-----	------------	-------

### 3. Sample text MA abstract: forensic??

style marked: **informal** **idiomatic** **elaborate** **explicit** "model academic"

The Past Perfect in Discourse Abstract

In Hughes and McCarthy's (1998) view, the grammar is a set of discourse-affected linguistic choices and a grammatical choice will be sufficiently explained if the contextual features are taken into consideration. In this paper, the probabilistic relationship between the past perfect and its context is **observed** by analyzing spoken and written texts, namely, **the** discourse. It is found that the past perfect clause is backgrounded to narrate supporting events in narration and its occurrence in because extension also supplies backgrounding information as justification or explanation following two contrasting claims. The paper illustrates that a particular grammatical form should be analyzed beyond the range of sentence, namely, to analyze it in its context.

Modena Aleno 24/11/22

Concepts	Why	What	How	Conclusion	16/30
----------	-----	------	-----	------------	-------

### 3.2 Examples: abstracts (CMAC05CU\_31)

great academic word list  
prepositions  
articles (specificity+definiteness)  
idiomaticity

This paper **attempts to conduct a contrastive study on** the realization of requests in **0** Chinese context and **0** English context **from** the pragmatic perspective, **with** the framework provided by linguists Blum-Kulka and House. Western scholars have made a multilingual investigation **on** the request strategies and found that all languages studied preferred conventionally **0** indirect strategy. While Chinese scholars claim that Chinese did not fit into **the** frame and developed a new perspective to justify the conclusion. Through data collecting and data coding, this paper, **with** its own proofs and methodology, tries to reach the same conclusion **with** other Chinese researchers in this field.

**A** research is included in this paper in which data is collected by the instrument of observation sheets distributed among altogether 70 university of different majors in two universities. They are required to record all the requests they have made, received and overheard within a week. The data mostly **derived** from **the** authentic conversation and is valid **for** coding and analyzing.

**To** widen the scope of **request strategies studies**, the reasons **for** the differences **in** the choice of the request strategies between the Chinese speakers and English speakers are put forward **with** the consideration of **the** politeness theory, proposed by Brown and Levinson (1978 and 吉林大学硕士学位论文 iii 1987). Since China and other English speaking countries favor **the** different politeness systems, **it is no doubt** that politeness systems will influence the adoption of request strategies. Equally important, cultural variables, such as collectivism and individualism and **the** social power distance are valued with diverse attitudes in different cultural backgrounds, which also account for the differences.

**A** pedagogical suggestion for improving the **appropriateness** of English requests is presented in the last part of the paper **in 0** hope to help English learners avoid **the** cross-cultural communication failure and enhance their communicative competence.

Modena Aleno 24/11/22

Concepts	Why	What	How	Conclusion	17/30
----------	-----	------	-----	------------	-------

### 3.2 Examples: abstracts (CMAC05CU\_36)

great academic word list  
prepositions  
articles  
style? (level)

Culture and language are closely related to each other. **It's** impossible to translate one language into another without considering the cultural factors involved. In this sense, translation is a cross-linguistic and cross-cultural communicative activity. The ultimate purpose of translation is to promote **the** cross-cultural exchange. Translation, in its essence, is an activity of transformation between cultures with their own language as cultural carrier. Then, **how to realize** the transformation of cultural connotation from the source language to the target language has become one of the important criteria to judge whether the translation is successful or not. Both English and Chinese words carry different cultural connotation **respectively**, so we have to master the two languages if we are intent on holding the two cultures. Among the elements of a language, lexical word is a basic one. As the main unit of expressing language meaning, words involve abundant cultural connotation due to the cultural influence in the long process of being used.

Language is a part of culture and is the carrier of culture as well. Language and culture cannot be separated from each other. Language would be meaningless without culture and culture would be unimaginable without language. Translation, therefore, is not only an inter-lingual transformation but also an intercultural transformation. There exist many similarities and differences between the two languages of English and Chinese; the similarity makes general culture, while the dissimilarity forms heterogeneity. Lexical gap refers to the phenomenon of non-equivalence in both of the two languages. **That is to say**, the referent object in the source language does not exist or seldom appears in the target language; or **0** referent meaning in the two languages is the same, but the semantic meaning is different, or the denotative meaning is the same, but the pragmatic meaning is different. Different cultural backgrounds have their own traits and their unique words in expressing cultural heterogeneity, which, when reflected in the target language, will cause the phenomenon of **0** lexical gap or semantic gap...

Modena Aleno 24/11/22

Concepts	Why	What	How	Conclusion	18/30
----------	-----	------	-----	------------	-------

### 4. Quantitative analyses of student texts

corpus compilation and selection  
comparable corpora? subcorpora

corpus annotation:

- text-characteristic (TEI, XML headers, CEFL?)
- POS vs. problem-oriented

survey: Diaz-Negrillo/Fernández-Domínguez (2006)

corpus analysis: Antcon + Sketch engine  
"more grammar than meets the eye"

- personal pronouns
- modal auxiliaries
- cohesion devices (linkers)
- (sentence) adverbs

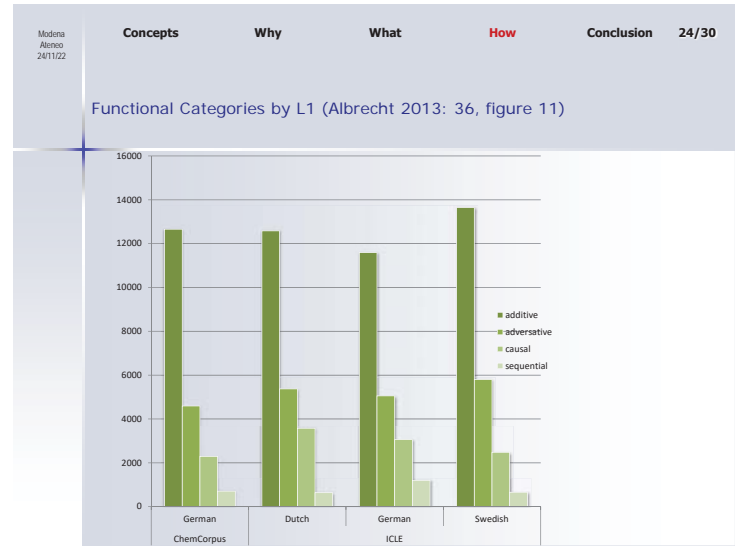
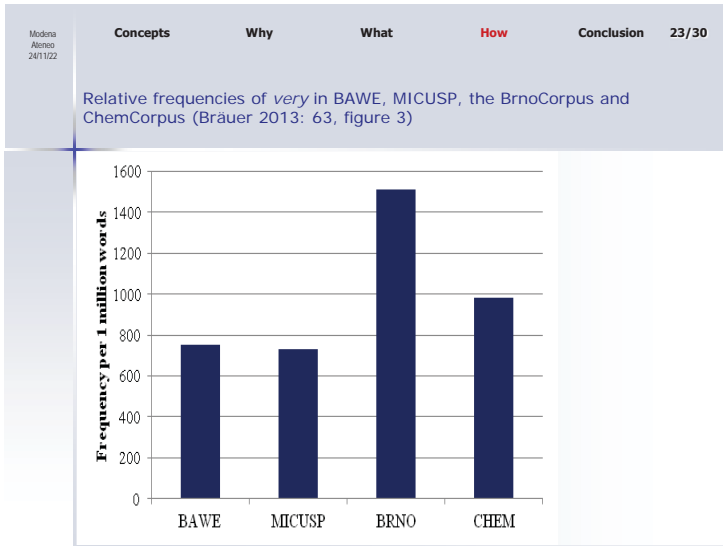
corpus results interpretation? relative frequencies (rather than categorical right or wrong)

Concepts	Why	What	How	Conclusion	19/30
<h2>4. Quantitative analyses: teaching</h2> <p>O'Keeffe, McCarthy, Carter (2007) or <a href="https://www.english-corpora.org/language-learning.asp">https://www.english-corpora.org/language-learning.asp</a></p>					
<b>Feature</b> 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>potent</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 collocations). 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, collocations, 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 telescope 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 telescope means, and of remembering it. Only English-Corpora.org allows learners to find both <i>collocates</i> and <i>related topics</i> (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.				

Concepts	Why	What	How	Conclusion	20/30
Cohesive devices: linking, connectors, etc. e.g. clause adverbials, clause-initial ,, special case?					
coherence = extralinguistic factors contributing to the creation of texture (=implicit) cohesion = linguistic means which create texture (=explicit)					
1	It is <b>unfortunately</b> not the case that knowledge of the mechanical				00C2AX
2	<b>Unfortunately</b> , an instantaneous comparison was not possible				00C3AX1
3	<b>Unfortunately</b> , exact information on gas usage is unavailable				00C3AX1
4	<b>Unfortunately</b> , one has no a priori way of predicting				00C3AX1
5	<b>Unfortunately</b> , neither easily works.				00C3AX1
6	<b>Unfortunately</b> , there are as yet no reliable calculations that				00C4AX1
7	<b>Unfortunately</b> , our experience is confined to an equilibrium ...				00C4AX2
8	<b>Unfortunately</b> , the energy resolution of a neutrino telescope is ...				00I5AX
9	... technique to obtain black hole masses which, <b>unfortunately</b> , is unfeasible				00I8AX
10	<b>Unfortunately</b> , lack of information about the collective ...				00I2AX
11	<b>Unfortunately</b> , the complexity of each subprocess also grows ...				00I5AX
12	<b>Unfortunately</b> these theorems tell us practically nothing about ...				00I5AX
13	<b>Unfortunately</b> there are very few analytic results available				00I5AX
14	<b>Unfortunately</b> , all recent experiments are, in principle,				00I5AX
15	<b>Unfortunately</b> , all those exciting recent experiments are				00I5AX
16	<b>Unfortunately</b> , those prior art solutions require daunting experiments				00I5AX
17	<b>Unfortunately</b> , the NID is uncomputable since the constituent ...				00I5AX
18	<b>Unfortunately</b> , in many places such information is getting harder ...				00I5AX
19	It's a good protective barrier," says Hildebrand, " <b>unfortunately</b> ."				00E9NS
20	<b>Unfortunately</b> , because of the rarity of plant data from this ...				01C0PN

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		
<b>Amplifier</b>	2,105.07	1,932.36			
<b>Booster</b>	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>users</i>	1,510.40	981.03	0.01 ≤ p < 0.05		
<b>Maximizer</b>	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		
<b>Downtoner</b>	609.86	785.74			
<b>Approximator</b>	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		
<b>Diminisher</b>	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		
<b>Minimizer</b>	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		
<b>Total</b>	2,714.93	2,718.10			

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	
<b>Amplifier</b>	2,555.95	1,646.22	2,378.79		
<b>Booster</b>	1,743.94	1,228.78	2,106.06		
<i>considerably</i>	28.27	35.28	30.30	> 0.05	
<i>highly</i>	208.51	99.95	111.11	0.001 ≤ p < 0.01	
<i>strongly</i>	0.00	70.55	30.30	> 0.05	
<i>tremendously</i>	0.00	0.00	5.85	> 0.05	
<i>users</i>	1,510.15	102.01	1,929.29	> 0.05	
<b>Maximizer</b>	511.81	417.43	272.73		
<i>absolutely</i>	94.78	64.67	25.25	> 0.05	
<i>clearly</i>	23.79	135.22	90.91	> 0.05	
<i>extremely</i>	120.05	47.03	60.61	> 0.05	
<i>fully</i>	31.59	135.22	80.81	> 0.05	
<i>thoroughly</i>	31.59	35.28	15.15	> 0.05	
<b>Downtoner</b>	916.28	487.99	469.78		
<b>Approximator</b>	445.94	305.73	217.17		
<i>almost</i>	303.29	252.81	171.72	> 0.05	
<i>nearly</i>	145.33	52.91	30.30	> 0.05	
<i>virtually</i>	6.32	0.00	15.15	> 0.05	
<b>Diminisher</b>	315.93	94.07	196.97		
<i>a bit</i>	3.91	35.28	141.41	0.001 ≤ p < 0.01	
<i>slightly</i>	214.83	47.03	33.35	> 0.05	
<i>somewhat</i>	44.23	11.76	20.30	> 0.05	
<i>to some extent</i>	18.96	0.00	0.00	> 0.05	
<b>Minimizer</b>	145.33	88.19	55.56		
<i>barely</i>	12.64	5.88	5.05	> 0.05	
<i>hardly</i>	120.05	70.55	45.45	> 0.05	
<i>scarcely</i>	12.64	11.76	5.05	> 0.05	
<b>Total</b>	3,171.96	2,134.20	2,848.48		



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
<b>Total</b>	<b>20</b>	<b>646,104</b>	<b>20</b>	<b>519,194</b>	<b>1,165,298</b>

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

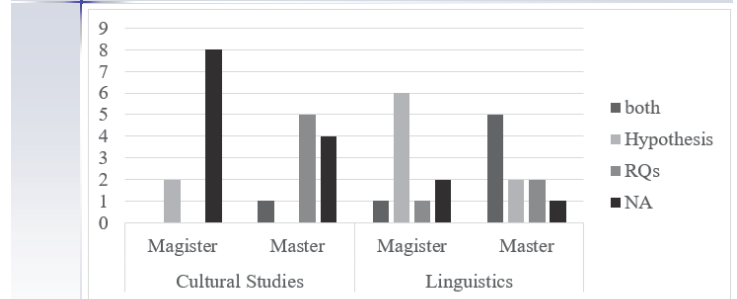


Figure 6: Occurrence of hypotheses and research questions in the theses (N=40)

for natural language/idiomaticity on the local level

**5.1** "local" errors may be 1<sup>st</sup> language induced and culture-specific ("Chinese does not have articles/determiners")

**5.2** "global" errors may be target-language-(and culture-)specific ("English uses a complex system of modal auxiliaries")

**5.3** some errors may be academic-writing related (nominalisations, compounds)

**5.4** test usage using on-line data-bases/corpora byu-corpora (google scholar)

**5.5** find good models in area to interest, library, ...

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?!

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).

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

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

Hólková, I. (2011). Conjunctive Adverbials in Academic Written Discourse: Conjunctive Adverbials in Academic Written Discourse. In Schmied, J. (ed.), 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.

O'Keefe, A., M. McCarthy & R. Carter (2007). *From Corpus to Classroom. Language use and language teaching*. Cambridge UP.

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

Schmied, J. (2011). Academic Writing in English in Comparison: Linkers and Adverbs in the ChemCorpus and Comparable Data-bases. In R. Plo & 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. *Brao Studies 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 stance in Linguistics and Cultural Studies. *TOLENT 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, 3<sup>rd</sup> ed.). *Academic Writing for Graduate Students: Essential Tasks and Skills*. The University of Michigan Press.