Academic Writing World-Wide: Comparing Metadiscourse Conventions, Credibility, and New Functional Standards?[1]

Josef Schmied (Chemnitz University of Technology)

Abstract

This chapter discusses key concepts of academic writing, especially metadiscourse, credibility, and functional standards. It discusses them in theoretical terms, but it also offers practical advice in the two genres, BA/MA/PhD theses and project proposals, which are crucial for young scholars' success in the research community today. It uses examples from comparable corpora from Africa, Europe and China to illustrate writing issues. It compares empirically usage and norm conventions and argues that discrepancies may not be due to mother-tongue interference exclusively, but due to English system problems; they can thus be discussed as possible acceptable deviations in the norm-developing process of non-native academic English, an advanced variety of *lingua franca* English.

Keywords: comparative student writing, comparable corpora, metadiscourse, argumentative patterns, style conventions, credibility, functional standards, conjuncts

1. Introduction

In the context of international networking for academic cooperation, especially between African and European scholars, writing conventions play a decisive role, for establishing contact, for funding applications and for publishing proposals, to name just a few types of cooperation that have become an essential part of international digital communication over the past 20 years. The globalisation of academic cooperation and the further spread of English as a *lingua franca* have also lead to a discussion of conventions and their differences between various academic cultures and between native and non-native speakers. From a constructivist perspective, it has been pointed out that there are no "native speakers" of academic writing and conventions can be negotiated to some extent by individual writers seeking to construct their academic identity in their disciplinary contexts (Hyland 2012). For young scholars from Africa, this raises fundamental issues, which have not been adequately addressed in teaching and research.

This chapter therefore explores three key concepts in international empirical comparison, metalanguage, genre-specific argumentation structure, and functional standards. My main argumentation is that the way to professional academic writing

[1] I wish to thank all colleagues in the Chemnitz Academic Writing Research Group for the continuous discussion that lead to this contribution and this volume, our partners in Cameroon and the Alexander-von-Humboldt Foundation for their longstanding collaboration and support.

is similarly difficult for non-native writers from Europe, China and Africa, and a comparative view may help all writers in their respective communities to develop their specific writing skills more easily, possibly also in contrast to traditional native conventions. All three key concepts are related to rhetorical consciousness raising in a new sense. The writers' pragmatic consideration when constructing texts for the specific discourse community in their disciplinary genres is particularly important for non-native users of English.

In the practical sections, this article discusses similar issues as the practical handbooks by Siepmann et al. (2011: 3), which "is geared specifically towards the needs of German-speaking readers", and by Swales/Feak (2012), which is widely used internationally. In its consistent "from – to" sections, this article tries to help young academic writers to move confidently and successfully from their own individual experience to an awareness of their academic community's conventions either in independent studies or as an initial input in graduate tuitions. My approach is generally functional in three senses: First, I try to give examples of practical guidelines and strategies that will help writers to produce a more effective academic text; second, I try to explain the functions that govern conventions and question them when these functions are not obvious; and third, like most similar text books, I use a functional grammar (like Halliday) as a theoretical basis without taking this theoretical basis for granted.

2. From individual to community-specific metadiscourse

2.1. Metadiscourse definitions

Young researchers often assume that they just have to report "objectively" the "facts", but tend to forget that other researchers can only read these "facts" properly, when they are written in their proper context. This does not only imply all the technical terms that young scholars have been told to master in their respective theoretical contexts, but also what is traditionally often seen as "subjective" elements. Of course, in the history of rhetoric and argumentation theory the contrast between *ad rem* and *ad hominem* (i.e. focus on the object or the recipient of scientific discourse) has been discussed and in practical classes, simple guidelines (like to replace "subjective" *I* by passive constructions) have been given for a long time. The strong focus on the writer - reader relationship is relatively new, however, so that formal exclusion of the scientist seems to be replaced by open and explicit inclusion of the scientific writers in their texts. Hyland (2015: 303) even says "authors are everywhere in their texts, presenting stance towards their topics and readers". This is today often called metalanguage or metadiscourse; both terms suggest literally "beyond" the mere content or proposition, focussing on the pragmatic and communicative contexts. The difference between the two terms is small, except that metalanguage is more used in programming and philosophy, whereas metadiscourse rightly emphasises the pragmatic writer-reader relationship, which is particularly important in our context.

Interestingly, the term metadiscourse does not feature prominently in the practical handbooks by Siepmann et al. (2011) and Swales/Feak (2012). The latter provide at least three references with the useful definition and somewhat play down its importance (ibid: 147):

> sentences or phrases that help readers make their way through the text by revealing such things as organization, referring readers to relevant parts of a text, or establishing logical connections. Metadiscourse is a noticeable feature of academic writing, although its value and frequency of use varies from one writing culture to another.

This definition does not only restrict metadiscourse in cultural terms, it also neglects the "subjective" or rhetorical elements. Hyland (2007) rightly emphasises the interaction in his subtitle *Metadiscourse. Exploring Interaction in Writing*. Since then, his systematic case studies have found many followers, who contributed little to the concept and more to the comparisons world-wide. Kawase (2015: 115), for instance, discusses a number of definitions and summarises carefully in Halliday's theoretical context: "It appears that the majority of metadiscourse theorists […] have adopted the notion that metadiscourse does not serve an ideational function (i.e., to construct propositional content) but textual and interpersonal functions". This is confusing to beginners, since they do not consider non-propositional elements important and writer-reader interaction not objective.

For us, metadiscourse comprises all expressions that organize the content and convey the author's beliefs and attitudes towards it. Researchers do not simply discuss facts or ideas, they also wrap up their content in metadiscourse, i.e. seek to claim solidarity with their readers, evaluate previous research and their own analyses, acknowledge alternative views, etc. As Hyland (2012: 206) wrote:

> Raising student's awareness of the language options available to them in negotiating an identity they feel comfortable with is also important in EAP classes. Once again, teachers can use corpus evidence to help students move beyond the conservative prescriptions of textbooks and style guides and into the preferred patterns of expression of their disciplines. An orientation to instruction based on access to choice through genre teaching and consciousness-raising can help students understand how writing conventions are enabling rather than deterministic. It can reveal the ways that typical patterns provide broad parameters of choice through which they can craft a distinctive self.

2.2. Argumentative structure

For over 20 years, Swales has developed his genre-approach, which lead to the widely-used textbook *Academic Writing for Graduate Students* (Swales/Feak 2012). This "is conceived as providing assistance with writing part-genres (problem-solutions, methods, and discussions) and genres (book reviews and research papers)" (ibid: viii). In this chapter, I focus on the genres theses and project applications, which are particularly important for young scholars from Africa (and beyond). Whereas applications function as scientific offers, hopefully convincing plans to carry out a project in a specific frame (time. budget), theses are the conventionalised reports that are to demonstrate that the candidate is worthy of being admitted to the next level of academic qualification, from BA to MA to PhD to full researcher, who knows their field and the core genres research articles, text books, handbooks, etc. Theses and project applications are less often discussed in terms of Swales' moves and steps (cf. Nkemleke 2016) as the well-known IMRD (Introduction, Methodology, Research, and Discussion; cf. Schmied 2015) macrostructure seems to be expanding from the most central academic genre, the research paper or article. It also seems to spread from the natural sciences into the social sciences and humanities, although in the latter we find many more structures depending on the topic and sub-discipline. Siepmann et al. (2011: 41-56) start from the first academic text genre at universities, the term paper, and distinguish between the traditional "literary essay" and the "linguistic mini-article". The trend in this direction is so clear that international "Writing Services" sometimes segment their offers into "Chapter 1: Introduction", "Chapter 2: Literature Review", "Chapter 3: Methodology", "Chapter 4: Analysis", "Chapter 5: Discussion", "Chapter 6: Conclusion", in addition to offering to write the (more expensive) complete thesis.

At the micro level, a common problem in theoretical - descriptive writing is the (mis-)use of repetition, esp. in sequences like *I am going to show/demonstrate/prove* – [some examples] – *I have shown/demonstrated/proven*. Repetition without convincing evidence (cf. 4 below) does not make claims facts - and illustrative metaphorisation neither. This does not seem to be a technical term in English, but in French métaphorisation and even métaphorism stands for the excessive use of metaphors. Although Siepmann et al. (2011: 450f) list a number of advantages in favour of metaphors ("colour", "reinforce", "facilitate memorisation", popularisation, even "embellish"), it is not always clear to non-native writers to what extent they are effective in the readers' culture.

The development of individual moves in sections is exemplified in Swales/Feak (2012), especially for the research paper, but little for the seminar/term paper or thesis. In contrast, Siepmann et al. (2011: 24-27) discuss in detail that the problem of working from excerpts from a reading list to an individual literature review can be solved by "*interacting*", e.g. grouping and selecting points, establishing a perspective, determining an intention, dividing the material into sections, and entitling sections and paragraphs. Of course, too many quotations disturb the flow of an academic text and may tempt readers to skip sections if the topic is in their well-known field. The hierarchy of quotations seems to be: non-integrated quotation of original (!) first or key definition of a concept, integrated quotations of further steps towards your working definition and paraphrases only for the less important special points (but still properly acknowledged to avoid plagiarism).

Finally, I would like to emphasise that young researchers should be aware of these pattern conventions since they add decidedly to the credibility of academic writers in their research community, i.e. the examiners who read and mark their theses and evaluators who read their research proposals. Even breaking the conventions or playing with them requires a sophisticated awareness of effective handling of metadiscourse features, structural decisions and stylistic choices.

2.3. Usage/style conventions

Modern students who have been taught in English using communicative methods often find it difficult to develop a feeling for style and register. This is, however, particularly important for academic writing, because it adds greatly to the impression the reader and examiner has of the scholarliness of a text. Siepmann et al. (2011) devote a whole "Module" to "Style and stylistic competence" and identify the five principles of aptness, clarity, concision, variety, and elegance "in order to obtain optimal results" (ibid: 415). I do not want to conceal that these principles can be contradictory, i.e. specialised terms may be concise, but it is not always clear whether they are well-known to all readers. For this purpose, (general) academic word lists have been developed (Coxhead 2000). Here, 570 "semantic fields" (or rather lemmata) have been identified and classified into ten subsets of 60 words according to frequency (later discipline-specific lists have been added for finance and agriculture, for instance). This implies that many words are considered quite general, whereas only few are used more or less exclusively in academic writing (like *context* and *environment* and *approach* and *theory*, respectively in the most common list – and even the meaning of *approach* in academic texts is different from the included adjective *unapproachable*). Of course, this does not mean that it is sufficient just to use as many academic words as possible or that readers just scan through a text to judge its professionality on the basis of the quantity of technical terms used (like the Oxford Text Checker; cf. Ochieng/Dheskali 2016).

The academic word list has been developed further into the more extended New Academic Vocabulary List, the Academic Formulas List and the Phrase List (Martinez/Schmitt 2015). This shows the work in progress in the corpus linguistic approach in academic writing research and its application in teaching.

3. From plausibility to professional credibility

Young academics have to learn that their writing has to be convincing and that they have to use appropriate rhetoric to make sure that their readers and listeners perceive what they write as professional. An initial step in this direction is, of course, the argumentative plausibility. The plausibility of an argument can be enhanced by adding references, which may prove that other scholars have pursued similar ideas and thus the current argumentation is not extreme. This is particularly important for African and Chinese scholars who sometimes feel that they stand on the shoulders of important predecessors, not only from Europe and America but also from the first generation of post-independence scholars that have made an impact on the discussion of varieties of English in Africa. In this endeavour, the correct referencing and evaluation of previous work makes a good basis, which can be underlined by different verbs of saying, *claim* vs. *prove* in (1) and *admits+attests+argues+states* in (2), as well as the well-known linguistic features of hedging (3) and boosting (4), the following examples illustrates this:

(1) Besides that, its presence betters the learning environment and enriches the learning experience. (Markovac and Rogulja, 2009) Punie (2007) claims that ICT enables the child to process the learning content in an entertaining and interesting way, while McPake, Stephen, Plowman, Sime and Downes (2000) have proved that the usage of ICT also develops the child's competences. (CAMDIPESII2015FCPA)

(2) Povey (1983:7) admits that it is hard to determine how many languages are spoken in the country. Quoting Mbassi-Manga (1973), he attests that there are 285 existent languages in Cameroon. Kouega (1998) argues that there are some 236 home languages which are spoken natively in Cameroon. Echu (2003) states that "Cameroon is a linguistic paradise comprising 247 indigenous languages". (CAMDIPESII2010FCNM)

(3) Also, by surveying the intelligibility of such learners, we may possibly discover that the supposed teaching of RP in classrooms is still unfeasible (CAMDIPESII2010MCDM).

(4) This certainly explains why Ulysses is considered by many critics as the mirror image of this period. (CAMDIPESII2012MSAV)

The other more data-related feature that enhances the plausibility of results is the compatibility with previous work and in particular with project-specific data. This can be emphasised by a long list of references, but often it is more convincing when this list is well-integrated into a more detailed argumentation so that the parallels and differences become clear (as in the first examples above).

Another, related feature that helps to enhance plausibility is the visualisation of methodologies and in particular results. Here, the differences between tables and different types of diagrams have to be trained in detail because they are the bread and butter of a convincing empirical applied linguist.

In the end, of course, the concept of credibility combines all the features of plausibility and professionality discussed so far. Credibility is particularly difficult to achieve in an age, when through daily experience, especially on the internet, young scholars tend to lose the belief in differences and face the crucial issue of evaluation of scientific writing. Credibility of older scholars seems easy because their position is sometimes even visible in the email address and the previous scholarly achievement speaks for itself, although even older scholars seem to have to learn to market their own successes today more than ever before. Another feature of credibility is the most detailed documentation and stepwise argumentation: This is based on the principle that, whoever can explain what they do in detail, is not afraid of making their thinking available for detailed criticism, but also that reproduction of research results may be possible this way. In this context, it may also increase the writers' credibility if they make their data available in data repositories, as has been propagated by the European CLARIN initiative.

The following argumentation tries to demonstrate why six criteria for professional credibility are logically important, especially for young scholars, who can use them as a checklist to see whether they have used enough of these "formalities" to convince their readers.

3.1. The importance of being formal

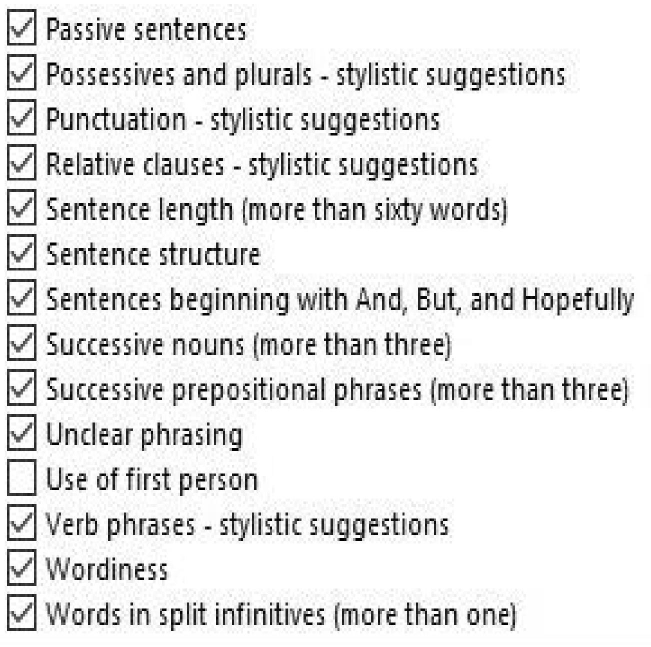
In their short list of style principles, Siepmann et al. (2011: 418) define:

> 'aptness' means that your communicative choices must be in conformity with the norms of contemporary academic style, which is characterized first and foremost by objectivity, economy and precision of expression, and by a comparatively high degree of formality.

They relate formality to professional credibility offering many examples that illustrate much more than the usual avoid list (contracted forms like *don't*, vague expressions like *a lot of*, etc.) and demand to "try to achieve congruence between substance and form" (ibid.: 419). This seems elementary for applied linguists, who are trained language specialists after all, but even for advanced learners of English, the specific richness of English vocabulary and the intricacies of English idiomaticity pose a particular challenge. Young researchers are forced to check whether they have used all technical terms and the discipline-specific multi-word routine formulae consistently, and whether they have replaced all informal words and phrases, so that their reader has the professional impression required. Here, community-specific discourse conventions show the in-group examiner or evaluator immediately whether the writer is part of their group and receives a favourable verdict.

3.2. The importance of formatting and other "formalities"

Formatting academic texts today is relatively easy and well controlled by current word processors, which provide style checks and an outline view (see boxes) that allows writers to see whether their thesis (or rather its table of contents) has a consistent hierarchical structure, if the corresponding formatting is used appropriately. This is particularly important for the professional reader (examiner, referee, etc.), who wishes to decide quickly whether a project proposal, for instance, should be considered seriously.



This article, for instance, shows a clear hierarchical structure:

- Academic Writing World-Wide: Comparing Metadiscourse Conventions, Credibility, and New Functional Standards?
  - Josef Schmied (Chemnitz University of Technology)
    - Abstract
    - 1. Introduction
    - 2. From individual to community-specific metadiscourse
      - 2.1. Metadiscourse definitions
      - 2.2. Argumentative structure
      - 2.3. Usage/style conventions
    - 3. From plausibility to professional credibility
      - 3.1. The importance of being formal
      - 3.2. The importance of formatting and other "formalities"
      - 3.3. The importance of concept discussion
      - 3.4. The importance of references
      - 3.5. The importance of evidence in tables and figures
      - 3.6. The importance of exemplification and evidence in examples
    - 4. From usage standards to functional standards
      - 4.1. A new debate of standards in academic writing
        - 4.1.1. Towards a new definition for non-native academic English
        - 4.1.2. Eye-tracking experiments
      - 4.2. Examples for functional redundancy of British English rules
        - 4.2.1. Grammatical underdifferentiation in unambiguous contexts
        - 4.2.2. Lexical choices in unambiguous context (collocation rules)
    - 5. From comparative studies of features to comparison of text-collections
      - 5.1. Comparing features world-wide
      - 5.2. Comparing academic theses world-wide
      - 5.3. The ChemCorpus as a model
      - 5.4. Quantitative comparative case studies: conjuncts
    - 6. Conclusion
    - References

Since the most important function of academic writing is the discussion of the writer's own work in the context of relevant existing scholarship, the handling of references is crucial. Here, conventions have changed greatly in humanities over the last 50 years, from a seemingly endless number of footnotes to a radical "in" if relevant or "out" if it does not contribute essentially to a conclusive argumentation – or an excursus in a substantial endnote. Of course, providing "masses of references" is not a virtue any more today when references can be extracted from international libraries and journal article collections quickly, it rather suggests that the writers found it difficult to decide on the most relevant references. Or it even suggests that the writer did not take the readers' right seriously to check all references easily in order to distinguish clearly between the writers' own scholarly contributions and others' and between an independent publication (like a hand- or text book) and an article in a refereed journal with impact factor or "only" a collection of conference contributions. Again, the consistent application of the standard reference system (APA for social sciences, including linguistics, or MLA for humanities, including literary studies, see Dunlop/Dheskali 2016) increases the writer's credibility; any deviation may lead to the conclusion that the writer is not yet fully integrated into the research community. This strong plea to take such "formalities" extremely seriously is based on the experience that readers want to find their expectations met and their work as examiners or reviewers made easy. Unintentionally distracting "formalities" may have annoying consequences for both sides in an academic discourse …

3.3. The importance of concept discussion

The large theoretical background is usually difficult to discuss for MA students – it is usually included with a few key-words or references to generally accepted authorities, e.g. "in Halliday's systemic functional grammar"[2]. The smaller theoretical units like concepts, however, have to be discussed in every scholarly text, where authors at least have to give a working definition of their key concepts. Although young scholars see this step often as a necessary long list of boring quotations, this is already a good opportunity to show their scholarship. The advancement of learning can only be achieved and demonstrated through digesting critically what has been argued by others and using the material for further argumentation or expirical analyses by the writer.

The choice of references and direct and indirect quotations clearly shows the seriousness of the work and the expertise of the writer. Arguably, popular reference works, dictionaries and especially Wikipedia are an acceptable starting point for students' concept discussion, when they are used as stepping stones to the real authorities in the sub-discipline. However, such tertiary literature like encylopedias "should generally not be quoted in academic essays, since it is typically intended for popularisation of scholarly findings" (Siepmann et al. 2011: 41 fn). Similarly, ephemeral internet sources are only rarely the appropriate reference - even if the URL and the date are duly added, which makes checking the credibility of the source easy for the reviewer. In a thesis, students usually have to demonstrate their expertise in a small area. This community integration is more important than the actual results of the research – students are often surprised that a credible, professional presentation of "non-results", like the non-confirmation of a hypothesis or negative answer to a research question, can still earn them an excellent grade.

---

[2] This is one of the few cases where the reference to standard literature seems unnecessary, since it is so well-known – and the standard theory developed and first published by Halliday in *An Introduction to Functional Grammar* (1985) has been updated by his students, and thus Matthiessen (4th ed. 2014).

3.4. The importance of references

In the literature review, students often see only the quantitative aspects of their references and forget the qualitative side: For the definition of key concepts, the first mention of a term, the semantic development and the current state of the definitions in current handbook articles are crucial. Handbook articles are for this section much more appropriate than (introductory) textbooks, since they are written by the specialist identified by the editors or the publishers as members of the research community. Textbooks, almost like Wikipedia, are more explanatory texts that rarely show the limits of concepts or emphasise the application problems that are most important at the beginning of a research project. Even within different schools, writers may be forced to side with one approach or the other – and a look at the references alone should indicate this.

Although the importance of references is not the same in all sections of an academic text, the proportion should be balanced: concepts and their references focused on in the literature sections should be taken up again in the discussion and conclusion section, so that the academic reader can see the network structure of the thesis, for instance. In the new, more radical, “in” or “out” decision, readers would like to be guided by clear references and not misguided by irrelevant “red herrings”, to use a dangerous culture-specific term.

3.5. The importance of evidence in tables and figures

Many students underestimate the persuasiveness of figures and tables, which often attract the attention of the fast readers (and examiners and referees) and persuade them to continue. The intuitive effect of these visualisations can be greatly enhanced by appropriate presentation choices between tables, which show more details, and figures, which create a more immediate impression, between pie charts, when the total is 100 %, bar charts, when the values are not directly related, and line charts, when we can follow a real development, over time, for instance (Fig. 5 and Fig. 6 below are examples). The appropriate choice of charts signals professional competence. The choice of stereotypical colours seems elementary; when we distinguish, for instance, between English, Scottish and Irish, the choice of red, blue and green, respectively, is evident, anything else can only confuse.

A special type of figure is necessary for project proposals, a time chart, which is important for time planning and (self)management. The most well-known types are probably Gantt charts, for which several software options are offered[3]. The following example could be adapted to most empirical PhD projects.

[3] https://www.ganttproject.biz/download is a simple example.

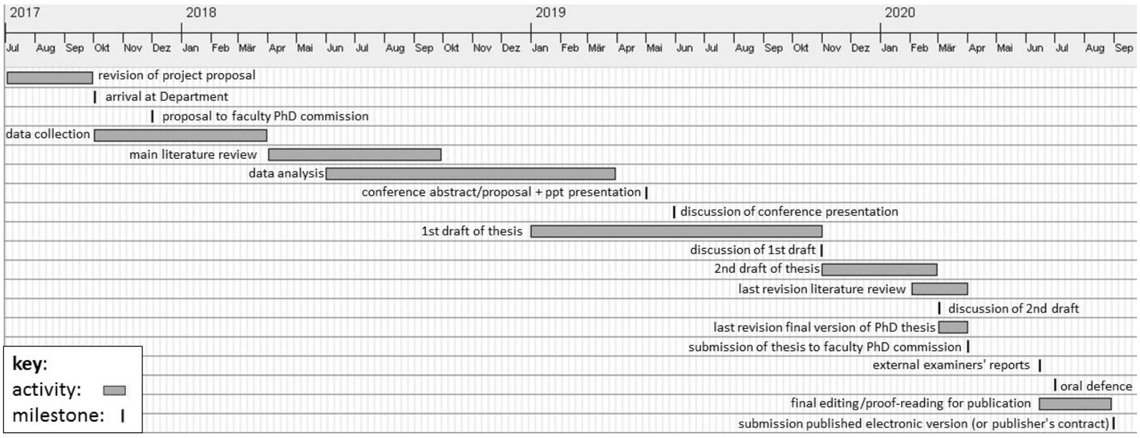


Figure 1: Gantt chart of a PhD proposal

3.6. The importance of exemplification and evidence in examples

The weakest evidence in empirical linguistics and sometimes the only evidence in theoretical linguistics are examples. Whereas in academic texts for beginners, examples serve as illustration to make concepts or language features clear, usually in their prototypical usage. The real art is, of course, examples that show the limits of a classification or argumentation, because in applying concepts in borderline cases real scholarship can be demonstrated. Every concept has its limitations and arguing about non-prototypical cases is not a weakness but a strength – the opportunity to use such cases to show their scholarship should always be taken up by young scholars since it adds greatly to their academic credibility.

4. From usage standards to functional standards

4.1. A new debate of standards in academic writing

4.1.1. *Towards a new definition for non-native academic English*

The discussion of standards of English is almost as old as the discussion about English itself (cf. Hickey 2012). However, the discussion of standards of English has had a unique dimension since Kachru (1982) started the discussion of "New Englishes" and their norm-developing tendencies in the "outer circle" (with English as official language, as in India or Ghana) and since the English as a *lingua franca* movement (Maurenen 2012) added a new dimension (beyond the learner levels) to the old debate of norms for non-native users of English.

Non-native standards of English have the advantage that they can be scrutinised and analysed according to logical and processing usability. When we say that criteria for non-native standards have to be frequency, international usage, functional appropriateness, transparency, acceptability, and processability, all these terms have to be discussed in detail:

Non-native standards have to be frequent means that they should occur in different genres, different disciplines, on educational levels, especially on higher levels and should not be related to individual writers or individual classes if they are taught in particular classes or schools. Non-native standards should be international means that the same usages occur in different countries, for instance in the European Union, in Asia and in different parts of South Africa. When we say non-native standards should be functional, this means that they should be justifiable in logical semantic terms, this means that irregularities may be regularized and sub-rules may be neglected, without undue emphasis on what appears as conventions without function or unnecessary distinctions and sub-rules. This may be seen in the context of the processability hypothesis or theory that states that learners restructure their L2 knowledge systems in an order of which they are capable at their stage of development (Pienemann 1998). Non-native standards should be transparent in the sense that the resulting texts should be intelligible in terms of its sense and pragmatic effect. When we say that non-native standards should be acceptable, this means that there should be no strong negative socio-linguistic connotations so that its users risk to be classified as uneducated, which would go against the academic reputation sought for. When saying that non-native standards should be processable we mean that the forms should go unnoticed, i.e. without intentional cognitive salience according to formal criteria, if it is not, of course, intended as an awareness raising style feature.

In conclusion, we can say that if we combine corpus-, socio- and psycholinguistic methodologies we may be able to find out new standards because of their frequencies in a wide range of styles and academic usages, their acceptability in online questionnaires, their familiarity in eye-tracking experiments with the appropriate user groups.

4.1.2. *Eye-tracking experiments*

Eye-tracking experiments have been used widely in second language acquisition and second language processing research, e.g. Roberts/Siyanova-Chanturia (2013). Of course, such experiments can be disturbed easily, if the conditions or the reader contexts are not ideal. We take for granted the focused and proficient reader and hope that individual problems and distractions are evened out by the different participants. In eye-tracking experiments, we have several measuring points: total fixations, regressions, first fixation duration, and total gaze time (see Figure 2 and 3 below). Since these applications are relatively new, the following examples merely serve to illustrate that a multi-method approach may be fruitful in our discussion of the awareness of non-standard features in academic writing.

Its application to African Englishes has hardly been attempted. The only exception is Van Rooy's (2010) article, where eye-tracking is discussed to distinguish between errors, innovations, and new conventions (as in the example *can be able to*). We conducted only a few experiments with African, European and Chinese English readers and we used modality and (definite) articles as trigger sentences. The following examples show that certain usages of *may* and *would* do not really receive as much attention or even cause confusion as predicted from a Standard English (StE) perspective. These modal verbs have extended their meanings well beyond that in comparable other languages anyway, so that non-native users use the opportunity to go far beyond what is considered normal in traditional native English.

In Figure 2, the modal *may* in the first sentence is usually considered tautological since the hypothetical nature of the statement in the subordinate clause is indicated by *hope* in the main clause already. The *may* in the second example is problematic, since it contradicts the *obvious* in the main clause. However, in both cases, the reader does not stop to gaze at the construction.

In Figure 3, the modal *would* in the first sentence seems to be a polite form of the wish *will* or a short form of *would like to*; the second *would* in the *if*-clause seems to be problematic, since the reader obviously backtracks to look at the unusual (in StE considered tautological) *would* again, even before reading the *would* in the main clause which is considered as a clear signal of the unreal nature of the statement again.

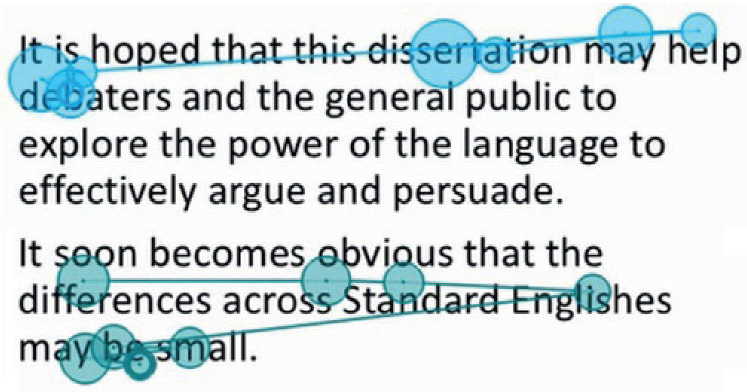


Figure 2: Unusual usage of *may* passes unnoticed in African reading

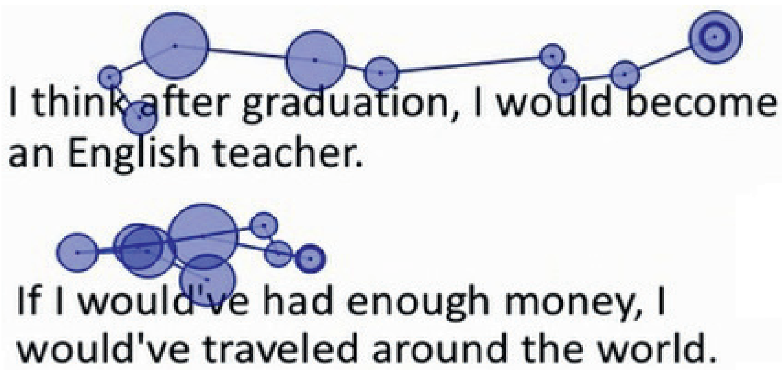


Figure 3: Unusual usage of *would* passes unnoticed in African reading

Since all sentences were actually used in academic writing, they demonstrate that the awareness of these mistakes is low and the users do not feel disturbed by these forms, except if the double signal with *if* introducing the subordinate clause and *would* in the main clause.

4.2. Examples for functional redundancy of British English rules

Most handbooks on academic non-native writing are useful for studying the principles and key concepts, but do not contain too many problematic or even controversial lexical items in their glossaries beyond the technical terms. In the following examples, we have chosen deliberately such cases and propose "neglecting the rules", arguing that the functional necessity of choosing one correct form and deciding against a deviation may not be necessary. However, I also try to indicate where such "generosity" has its functional limits because useful differences, even subtleties, may have such a high functional value in academic discourse that they should be maintained.

Most examples can be classified as problems of "underdifferentiation", which can be found in grammatical and lexical choices, not only in register and stylistic choices (cf. Schmied 1991: 129). If the choice is correlated by the choice of other clause elements (such as conjunctions or prepositions), the functional load is low, the choice is functionally redundant (e.g. *fill in/out/up* and *discuss +/- about* below) – and can thus be accepted as an option, not an error. In some bases, underdifferentiation may make a difference, but the meaning difference may not be important (e.g. *discuss about/on* below). This may even contradict the principle of clarity (cf. 2.3. above), because more explicit usually means clearer to the specialist and, in lexical cases, more concise.

In the following, we discuss examples of what is according to British usage books and even modern grammars an error in different types of categorisations. In our examples, we can see however that all these deviant usages occur in academic writing by German, Chinese and African writers, so they can neither be mother-tongue- nor teaching-induced. This raises the question if such usages simply contradict (often unsystematic) conventions, but could be seen as acceptable (or at least not marked explicitly as incorrect) in academic *lingua franca* English.

4.2.1. *Grammatical underdifferentiation in unambiguous contexts*

Our first example is the rare 100 % rule that demands that after initial *since* as conjunction and *for* as preposition the Present Perfect has to be used to signal that there is no time gap between the time of action and the time of speaking. Since this is a relatively recent rule, it is not considered as strict in American English as in British English. However, many examples make it clear that for most speakers of other languages with a less rigid tense system the distinctions between Past Tense, Present Perfect and Present Tense are not as clear as standard grammar books make us believe. Examples from our three national corpora are numerous, but not easy to find. Example (5) is a little complex with the *before – after* comparison, so that the German MA student was confused and used a simple *is* instead of the correct *has been*. Example (6) illustrates the correct use of the present perfect after *since*, example (7) illustrates the correct use of Past Tense since there is a time gap between 1980 and now:

(5) The idea behind starting in 2002 is to establish a loose relation to the development of postracialism to the time before and since Obama is in office. (MA14Ft_CH)
(6) Since CBI emerged in Canada in the 1960s, many studies have examined the theoretical bases and justifications for the use of CBI. (CMAC)
(7) For several years after this, Conrad was a sea man until 1980 when he went to the Congo. (CAMDIPESII2010FRTAA)

A second grammar rule that seems relatively strict in modern English is the necessity to maintain a "sequence of tenses" in related clauses; usually the matrix clause determines the subordinate clause. This is the case in English clauses subordinated by "epistemic" conjunctions and specially in indirect speech, where the time frame PAST, for instance, has to be carried over from the main clause to all verbs in the subordinated speech, so that any Present Tense including Present Perfect and Future Tense is not appropriate. This "backshifting" from the tenses in the direct speech contradicts the impression of the "natural sequence" (especially when general truths are expressed in the subordinate clause); it is caused by the frame set by the superordinate clause and the writer has to bear that in mind when choosing verb forms in the subordinate clause. Example (8) from the ChemCorpus and (9) from ChAcE are more problematic than example (10), where the clause can be seen as expressing a general, timeless truth. Maybe this argumentation can also be used for example (11):

(8) Among the male participants younger than 35,31% said that they only have basic qualifications, whereas 69% claimed that they have better English skills. (CBA11Ft_DB)
(9) Hence, it is sincerely hoped that this study will attract more attention of others to more and further researches of learning strategies to improve English listening comprehension learning and teaching in the college. (CMAC14LI_8)
(10) Krashen (1982) claimed that learners with high motivation, high self-confidence and a low level of anxiety are better equipped for success in second language acquisition. (CMAC07RE_16)
(11) They believed that the Black man is essentially evil and this makes the society racist. (CAMDIPESII2010FPMA)

A similar case of sequence of tenses is the rule that English *if*-clauses do not contain *would* as the related main clauses do, as in the following cases (12), (13) and (14) by German, Chinese and Cameroonian MA students, respectively:

(12) Thus, to illustrate this with an example, if I would want to become an American, it would not suffice for me to proclaim: "I am American." (MA14Ft_EL)
(13) It was also designed to know if teachers would recommend the teaching of English language through pop music. (CAMDIPESII2010FACNE)
(14) The stress on the second word of the verb is then weakened or lost, especially if it would otherwise be next to the other strong stress. (CMAC11PH_12)

In many languages, there is a parallel between the tense in the conditional clause and the main clause, which is not acceptable in English since it is considered tautological. Similarly, *(just) in case* is followed by a *should* and not a *will/shall* in the hypothetical matrix clause, but this is not always the case in our non-native academic corpora.

A third grammar rule that seems particularly difficult for most learners of English with a non-Germanic background is the fine distinction between specificity and definiteness as a basis for article rules in English. Numerous examples of this, such as (15) and (16), can be found in the ChAcE Corpus, some in the CamCorpus, few in the ChemCorpus:

(15) It was not until __ late 1980s when process-based EFL theories came into being as a result of the universal appealing for regarding second language writing as essentially a cognitive activity (CMAC12WR_13)

(16) However, this model only implies the verbal system and fails to give an account for the significant role of ___ nonverbal system in reading, such as context or mental imagery. (CMAC08RE_22)

These examples are typical cases for the eye-tracking experiments mentioned above. They are puzzling for native speakers in a few cases, but rarely constitute comprehension problems – and all these could be accepted, especially since the Chinese are joined by many Africans and most speakers of Slavonic languages in finding this English article system far too complex.

4.2.2. *Lexical choices in unambiguous context (collocation rules)*

A special feature of the seemingly easy language English is that at advanced levels the idiomatic choices of collocates are very limited, since they were fixed in grammars and usage books often centuries ago, when variability was not tolerated by language specialists. Two types of usage conventions can be distinguished:

a) conventionalised choices

Well-known examples conventionalised choices are British – American differences in preposition choices: Whether you *fill in* or *fill out* or even *fill up a form* is cognitatively slightly different but still similar; maybe *fill up* is more discouraged because it gives the impression of "completeness", which however is also expressed by the synonym *complete a form*. In Standard English dictionaries, however, *fill up* collocates with *tank/car*, *fill out* is marked as American with *forms* and *fill in* is marked as British with *forms*, but not with *hole*, and *fill in (for someone) fill somebody in (on something)* as "to tell somebody about something that has happened" (Oxfordlearnersdictionaries s.v.). In such cases, the choice of related prepositions should not constitute a communication problem and all three variants could be accepted and are indeed used in the same community (17-19).

(17) All the subjects received uniform instructions as to how to <u>fill out</u> the worksheet so as to minimize confusion. (CMAC06PH_22)

(18) The study revealed that boys can be more polite than girls when it comes to euphemistic usage because girls tend to fill up the gap with their mimic and general comportment. (CAMDIPESII2010MAN)

(19) The Authors should equally endeavour to use more meaning-based activities to practice the grammar points treated in the student's book because in real life situations, language is not used to fill in gaps per se, but to convey ideas and to fill in information gaps. (CAMDIPESII2010FECN)

A slightly different case is the much-discussed *discuss about*, which is explicitly marked as wrong in standard dictionaries, but supported by the analogies of *speak/talk about* and it can be found in native usage to introduce not a question but a field like *architecture* or *globalisation*. The analogy here is debate, where *debate about/on* is not an issue. Again, even if this "redundant" usage with *about* is not taught actively, it does not have to be counted as explicitly wrong on account of its transparency and easy processing in cases like (20) and (21).

(20) In the section of revision, students discuss <u>about</u> their first drafts in groups. (CMAC14WR_8
(21) Unlike the conventional plays which discuss <u>about</u> society and its politics Theatre of the Absurd focuses more on the metaphysical aspect of life than the physical. (CAMDIPESII2010MGDT)

A related issue is the distinction between *discussion about* and *discussion on*, where the latter seems to indicate the more scholarly activity, but these conventions indicate a different feature category, where the number of semantic choices is reduced. Although there is a functionally motivated choice here, the genre and situational context make it usually clear: in theses, it should be a serious discussion with *on*, in conversation *about* is more likely.

b) reduction of choices

Another very old usage is *on the other hand* without preceding *on the one hand*, just to express an argumentative contrast, as in (22). In the CMAC, the proportion of *on the one hand* : *on the other hand* is 137 : 430, in the CamMAs 81 : 244. But is this necessary, since English has many other conjuncts that serve the purpose: *alternatively*, *in contrast*, etc.? Here, the specific explicit balanced contrast between *on the one/other hand* and the simple adversatives is lost – unnecessarily?

(22) On the other hand, language is influenced and shaped by culture. (CMAC05CU_28)

Another old example from usage handbooks is the distinction between *because* and *as/since*. Even the corpus-based *Collins COBUILD English Usage* (2013 s.v.) states categorically: "In writing, the reason clause is sometimes put first, and **as** or **since** is used instead of 'because'". All our non-native corpora include dozens of examples (23) that this rule is not known, not taken seriously, or not adhered to by non-native graduates. Of course, the distinction between *since* (+known to reader, as in 24) and *as* (25) is functional and cannot be discarded so easily.

(23) Because even the smallest nations are too vast to allow mutual acquaintance of all members, the community in any nation is an imagined one. (MA11Mt_ES.)
(24) Since language is inextricably tied to culture, language teaching should absolutely include culture teaching. (CMAC05CU_28)
(25) As learners are likely to teach and impress their peers and family members by singing songs, their pride and interest in the language will tend to grow, given that they face limited immediate opportunities to use English. (CAMDIPESII2010FACNE)

Many well-known distinctions have been given up in non-standard British usage, such as *borrow* (dative) vs. *lend* (recipient; passive), *learn* vs. *teach* (active, expert), *bring* (direction of speaker) vs *take* (direction from speaker), etc. In all these cases, the immediate context makes the expression unambiguous: *borrow from* vs. *to*, *learn* +/- *me/him*, *here* vs. *there*. The contrasts are hardly noticed, as in (26):

(26) Italians and other immigrants equally crowded huge steamships that should <u>bring</u> them to the New World. (MG04Ft_SF)

Typical accepted expansions of lexical meanings include *until* (with time due, deadline) instead of *by* (in 27), or the use of the more general *in* instead of *into* (in 28):

(27) As for Elizabeth's falling in love with Darcy, it is something not accomplished until near the end of the book. (CMAC06LIT_31)
(28) I would put it in her hands. (MG08Ft_VL)

The contrast between the determiners *which* and *what* (+limited choice) is rarely found in non-native academic English, although in (29) and (30) the distinction seems logically correct:

(29) Lexicographers are still not sure from <u>which</u> language it originates. (BA13Ft_JZ)
(30) The credibility of an analysis depends on what is intended to be investigated, <u>what</u> language variety will be investigated and what corpus would suit your analysis. (BA16Ft_AK)

Special cultural meanings beyond the expected surface meanings have a few English expressions like *surely*, which does not signal emphasis in agreement (like *definitely*), but rather disagreement or surprise in British contexts. Thus, the two adverbs *surely* and *definitely* may seem tautological in standard English usage (31), whereas (32) is a standard collocation:

(31) Perhaps we can say that comprehensible input is like Chinese medicine, which the [whose] effects may come slowly but surely and definitely. (CMAC09WR_18)
(32) This can come slowly but surely, replacing the classroom teacher as Sadker and Sadker write (1991: 538). (CAMDIPESII2010MOLT)

Good examples can be found in the academic word list (cf. Coxhead 2000[4]). Distinctions that are functionally important and should be maintained are, for instance, *imply* ("he assumes this without actually claiming it; it may be wrong") vs. *infer* ("on the basis of the evidence, I draw this logical conclusion; it cannot be wrong") in argumentative contexts and *comprise* (where all parts are mentioned) vs. *include* (where just a few important parts are mentioned) in statistical descriptions. In the following examples (33) and (34) are problematic and (35) is only correct if these are all the sections.

(33) The above data imply that, although there are minor differences, the two groups are similar in their language backgrounds. (CMAC06SP_14)
(34) In order to teach this novel in a secondary school that is situated out of Banso or that is in Banso, but comprises students from other ethnicities, this approach can be quite useful. (CAMDIPESII2013FELB)
(35) It includes: the background to the study, the statement of the problem, the objectives of the study, significance and scope of the study. (CAMDIPESII2010FACNE)

---

[4] The approach, originally developed from 28 academic disciplines has been developed further, cf. http://www.uefap.com/vocab/select/awl.htm or http://www.newacademic-wordlist.org/.

5. From comparative studies of features to comparison of text-collections

5.1. Comparing features world-wide

In variationist studies so far, English features world-wide have been collected usually on an individual or purpose-specific basis. This has often lead to long lists of grammar features, which can, of course, be summarized into "tendencies" (Schmied 1991: 64-91, for the whole of Africa), e.g. in verb complementation and idiomatic preposition choice. Rarely has the occurrence of features been weighted according to their "frequency" or "evasiveness" in a national context. The best example of this type is eWAVE, which is not only a standard reference work, but also a convenient electronic database obtainable on-line for all (African) scholars (Kortmann/Luckenheimer eds. 2012). For Cameroon, this expert interview approach has been pursued for Cameroon English as well as for Cameroon Pidgin from a specialist perspective each, but the results (Figure 4) cannot always be confirmed in texts by Cameroonian writers. However, it has to be admitted that features, especially those that are perceived in sociolinguistic terms as negative (like feature 78: double comparatives) only need to occur once or twice to make all listeners or readers aware of the (educational) background of the text producers.

No.	Feature	Value
7	Me instead of I in coordinate subjects	A - feature is pervasive or obligatory
8	Myself/meself instead of I in coordinate subjects	A - feature is pervasive or obligatory
9	Benefactive "personal dative" construction	A - feature is pervasive or obligatory
34	Forms or phrases for the second person plural pronoun other than you	A - feature is pervasive or obligatory
38	Specialized plural markers for pronouns	A - feature is pervasive or obligatory
39	Plural forms of interrogative pronouns: using additional elements	A - feature is pervasive or obligatory
55	Different count/mass noun distinctions resulting in use of plural for StE singular	A - feature is pervasive or obligatory
59	Double determiners	A - feature is pervasive or obligatory
62	Use of zero article where StE has definite article	A - feature is pervasive or obligatory
78	Double comparatives and superlatives	A - feature is pervasive or obligatory
84	Comparative marking only with than	A - feature is pervasive or obligatory
85	Comparative marking with more…and	A - feature is pervasive or obligatory
88	Wider range of uses of progressive be + V-ing than in StE: extension to stative verbs	A - feature is pervasive or obligatory
113	Loosening of sequence of tenses rule	A - feature is pervasive or obligatory
119	Would for (distant) future in contrast to will (immediate future)	A - feature is pervasive or obligatory
122	Epistemic mustn't	A - feature is pervasive or obligatory
147	Was for conditional were	A - feature is pervasive or obligatory
154	Multiple negation / negative concord	A - feature is pervasive or obligatory
159	Never as preverbal past tense negator	A - feature is pervasive or obligatory
165	Invariant non-concord tags	A - feature is pervasive or obligatory
169	Non-standard system underlying responses to negative yes/no questions	A - feature is pervasive or obligatory
208	Deletion of to before infinitives	A - feature is pervasive or obligatory
209	Addition of to where StE has bare infinitive	A - feature is pervasive or obligatory
221	Other adverbs have the same form as adjectives	A - feature is pervasive or obligatory
222	Too; too much; very much 'very' as qualifier	A - feature is pervasive or obligatory
229	No inversion/no auxiliaries in main clause yes/no questions	A - feature is pervasive or obligatory

Figure 4: Features of Cameroon English in eWAVE

Unfortunately, such lists lump together very different types of features: non-standard native-speaker features (like 34: 2nd personal plurals other than *you* or 154: multiple negation), non-native simplification (like 113: loosening of sequence of tense rule or 163: invariant non-concord [question] tags), both can be seen as system problems of English; different conceptualisations of grammatical categories (like 55: count/mass noun distinctions or 62: zero article where StE has definite article). Although all these examples have been listed as "Cameroonian" I am inclined to give them different status and turn a blind eye to simplifications and re-conceptualisations more easily than others, especially when they are not confusing to the reader (cf. 4.1.2 above).

Unfortunately, Cameroon is not available for direct on-line quantitative comparison in the Corpus of Global Web-Based English (GloWbe) (cf. Davis 2004-, Davies/Fuchs 2015, Ochieng/Dheskali 2016). Cameroon was not included here because it is usually considered a francophone country first and not perceived as bilingual, it seems. Although there have been various attempts at compiling different Cameroonian corpora (Nkemleke 2016), they are not (yet) directly compatible to the International Corpus of English (ICE). So maybe a domain-specific subset of English usage is a first step in this comparative direction. This is why in our comparison of academic writing a new specific corpus was created (which can be integrated into a larger Cameroon English corpus later).

5.2. Comparing academic theses world-wide

In one of his latest summaries of "findings and gaps" in academic English, Hyland (2015: 303) identified a need for "studies which focus on NNES [non-native English speaker] students and how their academic writing in English is similar and distinct from NESs". On this basis, we could compare "covert codification" (Hickey 2012: 20) in an area where open discussion and conscious decision has not taken place, since the practical guidelines for accepting or correcting non-standard features are not very general (cf. section 4.2 above). A systematic attempt at comparing academic writing across nations is only possible when clearly comparable data are available. Based on the codification of the ChemCorpus (below), we have been able to retrieve from two African universities, Stellenbosch and Yaoundé, and from mainland China and Hong Kong. As Table 3 shows, the focus is on MA theses, although extensions have been tried successfully into BA theses in Chemnitz and PhD theses in China. The major drawback of our comparable corpora is that in non-native contexts like Germany, China and Cameroon, the disciplinary spread is restricted, because most other disciplines write their theses in German, Chinese and French, respectively.

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

Table 1: Compatible stratified corpora (2004-2016ff)

The ChemCorpus can be called a monitor or reference corpus, because it is so big that only sections of it can be used for comparative purposes. In previous publications, for instance, Albrecht (2015) compares adverbials in timed and untimed student writing in the ChemCorpus with the comparative native speaker corpora *BAWE* and *MICUSP*, Beyer (2015) compares hedging in the BA and MA theses by the same students included in the ChemCorpus, Küchler (2015) compares the term papers in the ChemCorpus with a compatible corpus of Chinese student writings and Edusei (2015) compares the German BA theses with the Albanian English. In similar ways, research questions can be derived from the setup, either as culture-specific or as first language induced interference phenomena. Many empirical comparisons are possible using the 129 linguistic features investigated by Biber (2006: 15-18) empirically (based on a relatively small corpus of 2.7 million words from four US universities in spoken and written registers).

5.3. The ChemCorpus as a model

As a base model for these comparisons, the ChemCorpus has been developed (cf. details in Schmied 2015). Its main advantage for comparison is the different degrees (traditional Magister, BA/MA and PhD); unfortunately, different text-types have to be neglected in this comparison, because the global trend is towards more uniform requirements.

The word figures for the genres and number of texts given in Table 2 are just rough estimates to make the corpus coherent and logical in proportion. The average is also variable, although the requirements are standardizing and the figures given are standard in many parts of the world, but the actual original texts submitted (and also stored in their original form) have more words, since in the corpus-processing stage the number of words is reduced. This is necessary to prepare the input suitable for analysis using the standard tools (like *AntConc* in the following examples). Thus bibliographies/reference lists, appendices and even figures and tables in the text are taken out to ensure that only words in context were really included in the quantitative analyses. The figures can, of course, be adapted to local departmental standards, but all this has to be documented in the corpus manual.

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

Table 2: ChemCorpus set-up by genre and specialization

The specific naming of the corpus texts should begin with the logical short file names (cf. the sample sentences above) so that even from a relatively short code we can see all the relevant information, in particular the text type, the educational or professional status of the user, the year of usage, first language and gender, as far as possible.

One big issue that has to be considered when comparing BA and MA texts is that we have to distinguish between countries and universities where English is used as a second language like in some parts of South Africa, where English is used as an international language in English medium universities as is the case in some universities in Turkey, for instance and some universities where English is only used in English-related subjects so that a much more restricted disciplinary spectrum can be included. In both cases, it is not quite clear to what extent the related teaching have an effect on the English used in the thesis. In some cases, the teaching may be through subject specialists who are more familiar with the subject and its specialised vocabulary than with academic writing principles in general. Thus, the distinction between ESP (English for Specific Purposes) and EAP (English for Academic Purposes) has to be discussed again in this context.

Although the subject-specific vocabulary and hardly idiomaticity has to be considered, this is only partly possible in comparative studies like the one proposed here. Details of similar comparisons can be found in Edusei (2015).

5.4. Quantitative comparative case studies: conjuncts

Linking devices have been considered central in text analysis and text production since Halliday & Hasan's seminal book *Cohesion in English* from 1976, and linking devices have been central to all student instruction and a topic in many student writing analyses (cf. Albrecht 2015, Bolten/Nelson/Hung 2002, Wagner 2011). Although linguists do not always agree on the complete list of linking devices, the most central and prototypical indicators are *and* with additive, *then* with sequential, *but* with adversative, *because* with causal function.

Figure 5 shows two types of comparison: on the one hand, a comparison of texts from the International Corpus of Learner English (ICLE) in the Germanic languages Dutch, German and Swedish; on the other hand, two German corpora, the traditional ICLE and the more recent ChemCorpus, which is not only more recent but also includes more argumentative text types. However, the columns at this level show that there are no major differences between link types.

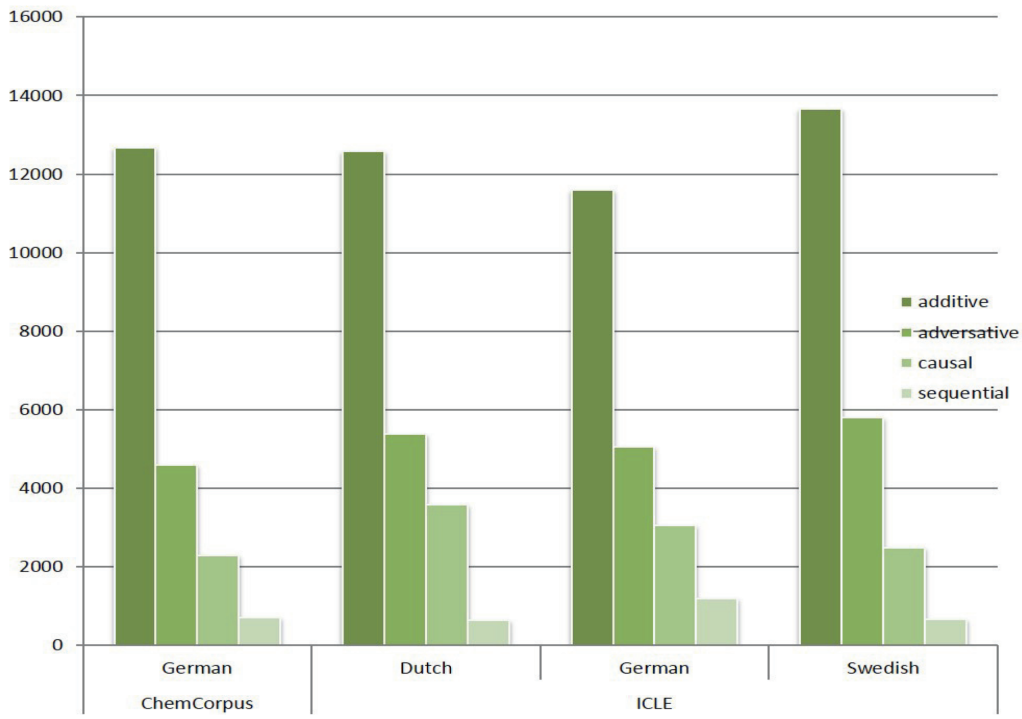


Figure 5: Conjuncts by function in European learner corpora (Albrecht 2015: 76, figure 5)

The comparison individual conjuncts in the BA and MA parts of the ChemCorpus and the CMA corpus has to be normalized (to 1 million words) because the Chinese corpus is much bigger. The individual figures (Figure 6) reveal "national" preferences, such the Chinese *(and) so* vs. the German formal *hence*, the resultative Chinese *as a result* vs. the sequential German *consequently*. Interestingly, *thus* seems to be used excessively more often in German MA theses than in German BA (and Chinese MA) theses.

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

Figure 6: Frequency of resultative/inferential conjuncts in ChAcE-MA vs. BA and MA theses in ChemCorp

The same data can be used for a more sophisticated statistical comparison (Fig. 7). Without going into the details of the statistical tests indicated, the figures show and the boxplots illustrate nicely that in this feature the Chinese MA corpus seems to be more similar to the German BA corpus, whereas the other two are more different from each other. Such calculations, of course, can only be the starting point for more thoughts about possible similarities and differences, which are not always easy to interpret (in particular when the data distribution is not always ideal for such hard tests).

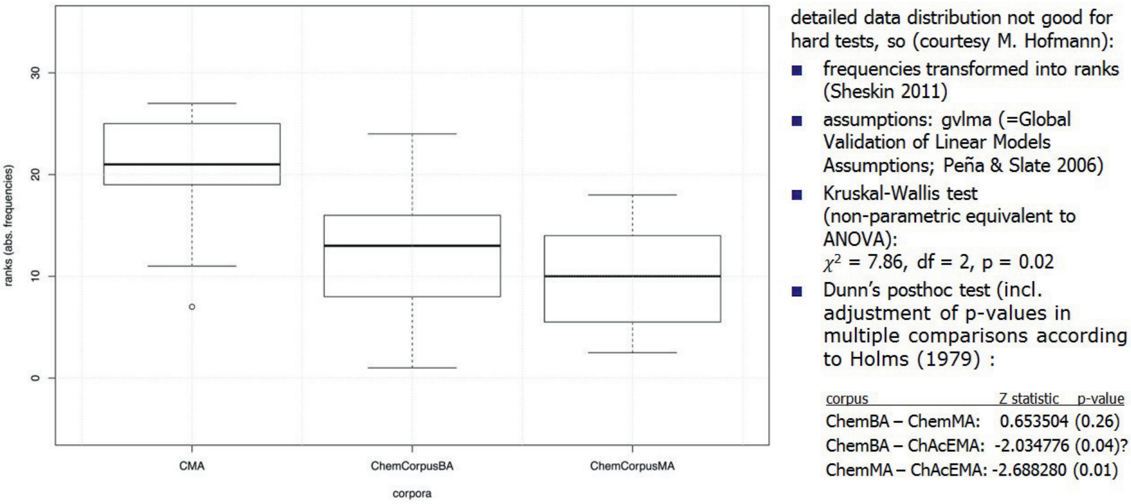


Figure 7: A comparison of conjunct types in German BA and MA theses and Chinese MA theses

6. Conclusion

This contribution has attempted to demonstrate to graduate students (and their teachers) that a few key concepts like metadiscourse (features) and genre (conventions) are enough to make academic writing relatively easily accessible to graduate students. If they follow a practical, research-based and empirical (corpuslinguistic) approach, they can construct their own model collections using the ChemCorpus as a reference corpus that offers a (partial) comparative data-base so that they can find models for their own writing, solutions to formal and concrete queries and their own stance in practical writing challenges.

The examples from our German, Chinese and African corpus analyses illustrate how interactive resources can be used by writers to manage the information flow to persuade their readers to adopt their preferred interpretations. In terms of functional grammar, writers seek to display an interpersonal tenor consistent with the disciplinary identity they wish to project. This can be negotiated with the gate-keepers of international science only to a certain extent (Pérez-Llantada 2012: 151f).

References

Albrecht, S. (2015). Connecting adverbials in student writing: Usage in the ChemCorpus and comparable corpora. In Schmied, J. (ed.) *Academic Writing for South Eastern Europe: Practical and Theoretical Perspectives*. Göttingen: Cuvillie, 67-80.

Beyer, D. (2015).”It is probably the reason why…”. Hedging in BA and MA theses by German ESL students. In Schmied, J. (ed.), *Academic Writing for South Eastern Europe: Practical and Theoretical Perspectives*. Göttingen: Cuvillier, 81-98.

Biber, D. (2006). *University Language: A Corpus-based Study of Spoken and Written Registers*. Amsterdam: John Benjamins.

Biber, D., Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes* 9, 2-20.

Biber, D., Rippon, R. (eds.) (2015). *The Cambridge Handbook of English Corpus Linguistics.* Cambridge: C.U.P.

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.

Collins COBUILD (^{2}2013). *English Usage*. London: HarperCollins.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34/2*, 213-238.

Davies, M. (2004). Brigham Young University Corpora. Available at http://corpus.byu.edu/, Accessed on 09/01/16.

Davies, M., Fuchs, R. (2015). Expanding Horizons in the Study of World Englishes with the 1.9 Billion Word Global Web-Based English Corpus (GloWbE). *English World-Wide,* 36, 1-28.

Edusei, J. (2015). Code glosses in student writing: A comparative study of Albanian and German BA theses. In Schmied (ed.), *Academic Writing for South Eastern Europe: Practical and Theoretical Perspectives*. Göttingen: Cuvillier, 119-130.

Halliday, M. A. K., Hasan, R. (1976). *Cohesion in English*. London: Longman.

Halliday, M.A.K., Ch. Matthiesen (^{4}2014). *Halliday's Introduction to Functional Grammar.* London: Routledge.

Hickey, R. ed. (2012). *Standards of English*: *Codified Varieties Around the World*. Cambridge: C.U.P.

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

Hyland, K. (2007). *Metadiscourse: Exploring Interaction in Writing*. London: Continuum.

Hyland, K. (2012). *Disciplinary Identities: Individuality and Community in Academic Writing*. Cambridge: C.U.P.

Hyland, K. (2015). Corpora and written academic English. In Biber, D., Rippen, R. (eds.) 292-308.

Kachru, B. (1982). *The Other Tongue - English across Cultures*. Urbana, Ill.: University of Illinois Press.

Kawase, T. (2015). Metadiscourse in the introductions of PhD theses and research articles. *Journal of English for Academic Purposes,* 20, 114–124.

Kortmann, B., Luckenheimer, K. (eds.) (2013). The Electronic World Atlas of Varieties of English. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at http://ewave-atlas.org, Accessed on 09/09/2016.

Küchler (Dheskali), J. (2015). Usages of *may* and *will* in Chinese and German student writings: A corpus-based comparison. Unpublished MA Thesis. Chemnitz University of Technology.

Martinez. R., Schmitt, N. (2015). Vocabulary. In Biber, D., Rippen, R. (eds.), 439-459.

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

Nkemleke, D. (2016). Analyzing research article introductions in the Humanities using the CARS Model: How genre knowledge can enhance academic writing skills of junior scientists. This volume, 35-58.

Ochieng, D, Dheskali, J. (2016). Simple tools on the internet for English non-native academic writers in Africa. This volume, 59-80.

Pienemann, M. (1998). *Language Processing and Second Language Development: Processability Theory.* Amsterdam: John Benjamins.

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

Roberts, L., Siyanova-Chanturia, A. (2013). Using eye-tracking to investigate topics in L2 acquisition and l2 processing. *Studies in Second Language Acquisition,* 35, 213-235.

Schmied, J. (1991). *English in Africa: An Introduction.* London: Longman.

Schmied, J. (2015). Graduate academic writing in Europe in comparison: a research-based approach to metalanguage and genre. In Schmied (ed.). *Academic Writing for South Eastern Europe: Practical and Theoretical Perspectives.* Göttingen: Cuvillier. 1-24.

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

Simpson-Vlach, R.,Ellis, N.C. (2010). An academic formulas list: new methods in phraseology research. *Applied Linguistics*, 31/4, 487-512.

Swales, J. M., Feak, C. (2012). *Academic Writing for Graduate Students* (3rd ed.). University of Michigan: University of Michigan Press.

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

Wagner, S. (2011). Concessives and contrastives in student writing. In Schmied. (ed.). *Academic Writing in Europe: Empirical Perspectives. Research in English and Applied Linguistics*. Göttingen: Cuvillier Verlag, 23-49.