



# Graduate Academic Writing in Europe in Comparison: A Research-Based Approach to Metalanguage and Genre<sup>1</sup>

*Josef Schmied (Technische Universität Chemnitz)*

## Abstract

This contribution introduces the key concepts of academic writing, metalanguage and genre. Metalanguage is seen to include all writer-reader interaction, esp. stance and engagement markers. The concept of genres in academic writing is discussed as a core-periphery model with the research article in the centre and the conference presentation, research monograph, handbook article and the chain from BA through MA to PhD thesis as other core genres. All concepts are explained and illustrated by examples from the *ChemCorpus*, which can serve as a (partial) reference corpus to all the other national mini-corpora in the SE European academic writing project and beyond. A research-based approach means that writers do not learn rules, but discover patterns and conventions themselves, either by testing ideas from textbooks or by exploring their own small corpora, even to test whether their linguistic variables are appropriate for their text/genre or socio-biographical variables. They can also use comparisons with similar corpora to position themselves in the spectrum between individual identity and disciplinary convention. Through this approach graduates gain skills that should be useful for their own writings at university and even for their professional life afterwards.

## 1. Introduction

This state-of-the-art introduction tries to show to junior researchers and practitioners in academic writing that this is an interesting and useful field: interesting because it is accessible to young researchers with only a few basic “theoretical” concepts and useful because students may profit from practical knowledge in the field for writing their theses during their studies or even for improving and editing other writers’ work after their studies. Of course, no student of English or writing specialist in general would want to evaluate and discuss research findings in other disciplines, because they would not have the technical and subject-specific knowledge. But if we assume that research should be written in such a way that it can be understood not only by specialists, but by other academics as well; then graduate writers not only impress the direct supervisors, but even demonstrate that they can make a contribution in a sub-discipline or small research area to the academic discourse.

This “introduction” takes only some initial research experience for granted: term papers as first practices and (hopefully) a BA thesis as a first writing project.

---

<sup>1</sup> This contribution is based on the discussion with partners in the Chemnitz Academic Writing research group, and project partners in Macedonia, Montenegro, Kosovo, and Albania. Discussions with Dana Beyer were always helpful to clarify ambiguities or find even better illustrative examples. We all have to thank the German Academic Exchange Service and their committed advisors (not “administrators”) for their continued support during the project including the finances for this final documentation.



It does not discuss learner English below C1 level (in the *Common European Framework of Reference for Languages*) or all problems of general writing, which may also play a part, but concentrates on the special features that make writing “academic”. Our approach can be described as practical, research-based and empirical. It is “practical”, since it encourages young researchers to use a similar hands-on approach and illustrates all concepts by concrete examples from the *ChemCorpus* (see below). It is “research-based”, since it goes beyond practical language exercises, leading the young researcher to take a functional perspective, either by trying to prove the usefulness of conventions, guidelines, “rules”, wherever possible or even by trying to find regular usages and choices in other writers’ academic texts. It is “empirical”, since it encourages readers to test the assumptions and hypotheses they may have had themselves or they may have read in this introduction or other instructional academic texts.

Although young researchers may be painfully aware that academic writing has its discipline-specific vocabulary and idiomaticity, we take for granted that all academic writing today is expected to use an accessible metalanguage to guide the reader and to make the reading process as easy as possible. This is based on the conviction that good academic work does not have to “hide” behind complex vocabulary, discourse and style “to reflect the complexity of the subject” or “impress the reader”. Well-written academic work includes the special attention to the reader all the time, even that the evaluation of the writing (in terms of a good mark, further funding or a positive citation) depends only partly on the “actual work”, but also on the convincing, honest and open presentation in a conscious attempt to use metalinguistic choices for their writer–reader interaction that encourages academic discourse and the advancement of learning. Keeping a balance between individuality and convention is the central challenge in academic writing and every writer has to be aware of this in the writing process. Thus even writers who intend to “construct their academic self” towards the unconventional end of the continuum must be aware of the conventions, either by reading about them or by identifying them through the practical, empirical and research-based approach offered here. This encouragement of “rhetorical consciousness raising” (and related RCRI = “rhetorical consciousness raising instruction”) has been discussed for a long time (since Swales 1990: 213).

This is also necessary since the conventions (as understood by the reader, evaluator, supervisor, etc.) and expectations seem to have become stricter over the last few decades. This can best be seen by the concept of genre, which is used today quite differently from 25 years ago and is very useful as it differentiates according to functional necessities in a modern globalised and digitalised academic world. Today, teachers as well as students can compare academic writing world-wide and see cultural (or language-specific) differences – and finally decide where to “position” themselves in the academic research space (cf. Hyland 2012). If young researchers aim at demonstrating familiarity, expertise and intelligence, they can do that on the basis of the conventions irrespective of how much their readers are aware of them.



## 2. Understanding key principles

Since academic writing specialists today cannot be specialists in all the possible subjects where English is used as an international lingua franca, they have to collaborate with the subject specialist and concentrate on the formal contexts and conventions that are important in all disciplines. Our focus in the following discussion will not be on specific conventions, but rather on the choices that young graduates have to be aware of and make according to the chosen academic identity. This applies in particular to non-native speakers who have not yet developed the necessary feel for the language that they use together with most of their colleagues as an international lingua franca (Mauranen 2013). But this raises the question of how we can develop such *Sprachgefühl* more consciously starting from the experience in finding conventions and pattern through empirical learning from authentic materials. This brings our approach in line with current philosophies of learning: Social constructivism suggests that knowledge and social reality are created through daily interactions between people and particularly through their discourse (cf. Hyland 2012).

### 2.1. Focus on metalanguage

The focus in this contribution will be on metadiscourse or metalanguage, i.e. the language that includes various conventions in author–reader interaction. The basic guideline or the basic principle for all our decisions should be reader-appropriate metalanguage, which depends on various factors, in particular the experience that readers have in the academic levels from BA to postdoctoral theses. The reader-specific perspective is particularly important in cases of hedging and boosting, which are two sides of the same coin of stance (cf. Trajkova 2015). The pragmatic meaning of both types of interaction is clear: In the case of hedging (*the data suggest*), the writer protects himself from the possible attack of other subject specialists by emphasizing the tentative position: writers invite constructive criticism and cooperation; whereas by boosting (*I am convinced*), academic writers indicate that they would be willing to “fight it out”, if other specialists were to attack them in those points.

Figure 1 (from Hyland 2005: 177) summarizes the whole spectrum of metadiscourse with two types of interaction, stance and engagement: Hedges and boosters can be seen as two ends of the stance continuum. Attitude markers can be expressed in verbs (*praise*), adverbs (*appropriately*) and adjectives (*best, excellent*) and allow the authors to align or distance themselves from previous authors. Self-mention through personal pronouns (*I/me/my* vs. *we/us/our*) is a feature of “soft” sciences (Hyland 2009: 76); the choice between plural and singular forms for the author is a disputed feature of discipline conventions.

Engagement can be seen in reader pronouns: *you/your* are used to bridge the communication gap between writer and reader. Directives are mainly expressed by imperatives (*take the following example*) and modals (*must, will*). Personal asides

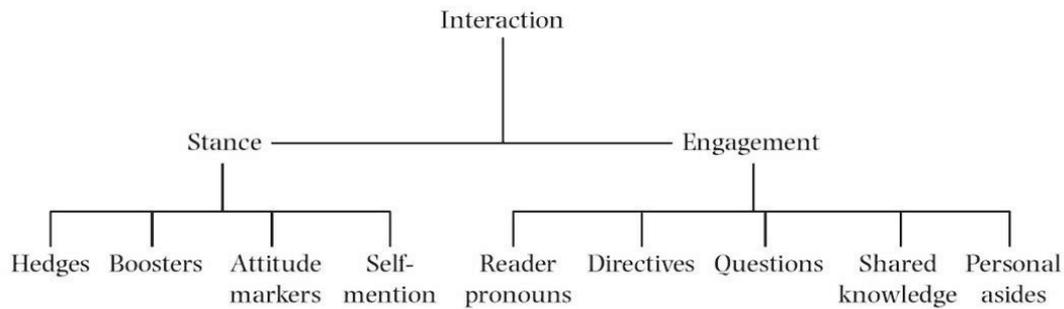


Figure 1: Types of interaction (Hyland 2005: 177)

are rare in most genres (*if you wish*). Appeals to shared knowledge create a community of discourse (*of course, as we know from ...*). Rhetorical questions expect no answer but ask the readers to think and judge for themselves (*what can we do?*).

In this article, we cannot discuss discipline-specific content. We can only discuss metalanguage conventions, i.e. the discourses that have developed in specific sub-disciplines and departments and are applied to the genres that the students have to write first within their own departments and later on beyond (Schmied 2012 on an empirical study from a South African university). Theoretically, we assume that students move from a more individual approach to a department-specific approach and later to a subdiscipline-specific approach, so that their work can be compared world-wide, so that they can compete with other researchers in publications in appropriate journals.

## 2.2. Focus on genre

The concept of genre has made a surprising career in applied linguistics over the past 25 years. Since Swales' 1990 definition, it has replaced earlier concepts (like styles, register or text-type) and has been widely adopted and extended. Swales (2004: 4) sees his own studies in a "continuing phenomenon [...] an increasing generification of administrative and academic life". He himself has added hierarchies (ibid: 12) and genre chains, sets and networks to the system (ibid: 18-25). Bhatia (2004: 23) proposed a widely accepted detailed concept:

- (1) Genres are recognizable communicative events, characterized by a set of communicative purposes identified and mutually understood by members of the professional or academic community in which they regularly occur.
- (2) Genres are highly structured and conventionalised constructs, with constraints on allowable contributions not only in terms of the intentions one would like to give expression to and the shape they often take, but also in terms of the lexico-grammatical resources one can employ to give discursal values to such formal features.
- (3) Established members of a particular professional community will have a much greater knowledge and understanding of the use and exploitation of genres than those who are apprentices, new members or outsiders.
- (4) Although genres are viewed as conventionalised constructs, expert members of the disciplinary and professional communities often exploit generic resources to express not only



‘private’ but also organizational intentions within the constructs of ‘socially recognized communicative purposes’.

- (5) Genres are reflections of disciplinary and organizational cultures, and in that sense, they focus on social actions embedded within disciplinary, professional and other institutional practices.
- (6) All disciplinary and professional genres have integrity of their own, which is often identified with reference to a combination of textual, discursive and contextual factors.

An alternative or complementary perspective is rare: Hyland (2009) distinguishes between research discourses (with the focus on research articles and conference presentations), instructional discourses (with a focus on university lectures, seminars, undergraduate textbooks), student discourses (with a distinction between undergraduate and postgraduate genres) and popular discourses (with a focus on science journalism). In a more complex set-up, this list of genres could be viewed in a central-peripheral frame and presented as in Figure 3 below.

### 2.3. Argumentative structures in genres

In research discourses, the trend towards a uniform argumentative structure seems most obvious: IMRAD/IMRD is an acronym for Introduction, Methods, Results, And Discussion. Originally, research articles in natural sciences were typically structured in this basic order. The IMRAD format has been adopted by a steadily increasing number of academic journals, since the first half of the 20th century. This structure has come to dominate academic writing in the sciences, although it is not suitable for many genres of academic writing, like case reports and meta-analyses, which usually have non-IMRAD structures. Recently, the structure has been expanded into AIMRAD to include a structured summary or abstract (and even, I may add, a list of 4-6 keywords for the index or for specialised searches). For applied linguistics, it may be appropriate to adapt the acronym to AIMAC, which concentrates on the analysis, including the interpretation and possibly even the summary and a complex conclusion (see Table 1 below).

For the beginner’s analysis of such sections, it may be useful to add a substructure wherever possible and some typical lexemes or phrases that can serve as indicators in academic text analyses.

The CARS (=Creating A Research Space) model (Swales 1990) has been an important breakthrough in the field and has stimulated a number of further studies. Swales’ CARS model for introductions is based on his study of articles across a range of disciplines. He identified the following moves as common among most articles (Figure 2):

- Move 1: Establishing a territory
  - Step 1 Claiming importance and/or
  - Step 2 Making topic generalizations and/or
  - Step 3 Reviewing items of previous research
- Move 2: Establishing a niche
  - Step 1a Counter-claiming or
  - Step 1b Indicating a gap or
  - Step 1c Question-raising or
  - Step 1d Continuing a tradition



- Move 3: Occupying the niche  
 Step 1a Outlining purposes or  
 Step 1b Announcing present research  
 Step 2 Announcing principle findings  
 Step 3 Indicating article structure

Figure 2: CARS model (Swales 1990: 141)

A similar structure has been found in book reviews (Motta-Roth 1998: 35) and review articles (Diani 2004: 109). But it may be difficult to present a detailed series of moves and steps as a model for a complete thesis (and maybe it is more appropriate for MA than BA levels). Even Swales & Feak (2012) only mention moves explicitly in the introduction and the methodology sections.

Table 1 attempts to combine the IMRAD and the CARS concepts and adds a few key lexemes or phrases as examples that may indicate the structure or sub-structure to the reader or that may be used consciously by the writer to meet the expectations of the examiners. As in the CARS model, not all steps are necessary and some are alternatives. What is not indicated in this table is the important network system that connects, for instance, the relevance in the issue section and the contextualisation or outlook in the conclusion, the tests indicated in the methodology section and the significance in the analysis, etc. Some of these networks seem elementary, but it is surprising how often thesis writers fail to check whether all research questions raised in the methodology section are answered explicitly or adequately in the summary or whether all hypotheses are rejected or confirmed.

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

Table 1: Revised IMRAD structure: AIMAC



The lexemes and phrases are just examples of what experienced and quick readers (who just scan the text in their routine when they have many other texts to read) expect to find in such sections and hopefully tick off as “familiar” in their mental genre schemata and, as an examiner, even in the printed thesis submitted. Such coherent academic lexemes and phrases help to create a global framework in top-down planning, which is a major headache for young academic writers (cf. p. 17 below). This is useful because for some writers at least, getting this overarching span from the first to the last page is the real difference between writing a short term paper (10 pages/5,000 words) and a BA thesis. Of course, this global AIMAC metalanguage form is just a scaffold that has to be filled with concrete academic content, but it is useful because it helps the writer to concentrate on the local details and thus to create a positive “familiar good standard” impression on the reader/examiner, who notices that the writer is a member of his/her academic discourse community.

There are more theoretical approaches to the argumentative analysis in discourse, but their application to academic genres has hardly begun, although the choice between more empirical and more theoretical, more critical and more epistemological argumentations promises to reveal interesting differences between national academic traditions in Europe and beyond (e.g. Smirnova 2013). Siepman et al. (2011) use the concept of rhetorical moves and give some intriguing examples without developing it into a comprehensive organisation concept.

#### 2.4. Metalanguage features in academic writing

The list of prominent discourse features discussed in textbooks of academic writing is extensive. Swales & Feak (2012) discuss features related to their CARS and genre model (e.g. abstracts, reviews, commentaries, critiques, data commentary, introduction/methodology/discussion/result sections; we can see examples in 5.1. below), but also textual features like claims and hedging/boosting, linking words/phrases and connectors (this is why we will discuss them in 5.2 below). Siepman et al.’s (2011) Guide focuses on construction/paragraph organisation, focus/topic/message/climax sentence, but also on well-known functional concepts like elaboration, exemplification, given/new information or information packaging. Both handbooks emphasise less the qualitative (corpuslinguistic) perspectives on interaction markers (see p. 4 above and 5.3. below) or lexical bundles. But lexical bundles (*it can/could be argued that*) have been analysed in detail elsewhere (e.g. Biber/Barbieri 2007 or Dontcheva-Navratilova 2012), especially to compare native vs. non-native usage. Chen/Baker (2010), for instance, claim that non-native writers prefer VP-based bundles, whereas native writers prefer NP-based bundles. They compare the 1990 FLOB-J (the academic part of the Freiburg-Lancaster-Oslo/Bergen corpus) with BAWE (British Academic Written English) and a (Chinese) non-native corpus. The most frequent bundles are surprisingly different for the two British corpora and many bundles are not really academic (cf. the criticism in Simpson-Vlach/Ellis 2011 or Salazar 2014: 149): Although *on the other hand* is an expected non-native usage (functioning like *by contrast* and not as a counterpart to *on the one hand*), it also



occurs almost half as often in FLOB-J (34 vs. 19, but 0 in BAWE). Similarly, *at the same time* is used very often by non-native and less than half as often by the native writers (24, 5/10 times) and as well as 16 vs. 10 times in BAWE (0 in FLOB-J). Real academic phrases like *in the context of* occur only 5 times in the Chinese corpus and 19 times in FLOB-J, but not at all in BAWE – but does that mean anything? Only *as a result of* is used equally well in all three databases (12, 17 and 9 times). For the analysis of larger bundles, we need very large corpora if we do not want to restrict ourselves to the most frequent constructions only like *can be used/explained/seen/found*, etc. – all clearly hedges. At least, such studies remind us that we need to discuss good academic English on the basis of a good general idiomatic English – and maybe our *ChemCorpus* family is not big enough for this.

Finally, it has to be mentioned that even punctuation is an important element in academic writing and author–reader interaction, because good pronunciation structures the text for the reader and thus makes it easier to process. This is reflected in different guides as Greenbaum’s (1996) *Oxford English Grammar* and Swales & Feak’s *Handbook*, which includes an interesting punctuation “decision” figure (2012: 39), in particular to choose between periods, commas and semicolons. Siepman et al.’s *Guide* (2011) devotes a whole chapter 4 (p. 192-224) to punctuation. Schmied (2013: 32ff) shows how punctuation has been included in good writing training, because it has functional values (i.e. to make life easier for the reader) that go far beyond the narrow-minded application of old style rules.

### 3. Exploring genres in academic writing today

Academic writing skills were explicitly included in many new curricula at European universities (Schmied 2011) and many take a genre approach. Students at Master’s level have written term papers and their Bachelor theses. Thus they are aware of the concept of genre, but they may not see it as directly related to their own work. They have to develop a feel for the usefulness of the concept of genre and the metalanguage genre conventions of their discourse community to move successfully from student to research discourse during their MA studies. Our survey includes two approaches that complement each other:

Table 2 shows a matrix approach to genres in academic writing, trying to complement the most central variables of research/teaching, spoken/written and specialised/learner (and a few others) by concrete figures on writer’s experience and length, and trying to combine them as mega-types in special contexts. The comments column serve as a basis for further discussion. The approach does not produce a neat mosaic and not all possible genres have been included, like a round-table (talk) with statement, comments and question/answer session during conferences.

Figure 3 (below) offers a gradient approach from more central to peripheral genres and even a few examples of academic related genres from journalism, politics, marketing and technology. But even some peripheral genres (like press releases or obituaries) may be debatable. Since it is impossible to put all genres in a two-dimensional space, the dimensions instruction-, research- and job-orientation have been added as further dimensions.



Graduate Academic Writing in Europe in Comparison: Metalanguage & Genre 9

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

Table 2: Matrix of genres and wider contexts in academic writing



For a classification of current academic genres, we have to start with the research article, which is the prototypical type of research texts and well researched since Swales' (1990) analysis. It is written by the specialist for the specialist and serves the purpose of driving research further in the field; it usually requires considerable experience in academic writing, which may start from MA level, but most journals would probably only accept articles written by experienced researchers that have been in the business for considerable time. However, the names of researchers do not play a role in the decision and evaluation process because the standard procedure in renowned journals with a high impact factor is a double-blind peer-review process, so that neither the reviewer knows the author nor the other way round. This does not mean that writer and reviewer are "strangers". A good editor knows which reviewer would be interested in doing the job. To make sure that the review is done quickly and reliably, editors usually choose reviewers that belong to the same discourse (sub-)community, so that the review is not only a burden that the reviewers are usually not paid for, but also interesting and useful for the reviewers' own academic work. Additionally, writing the review should make them feel honoured and well-informed about current developments in their academic field. Of course, reviewers do not only expect to find general academic words and phrases, but also discipline-specific terminology. In case the reviewer has a somewhat different specialisation, a "standard", "familiar" general academic metalanguage may help even more.

Traditionally, the best-known academic genre family are books: The standard qualification at book length is usually the research book, which is also written by the specialist for the specialist and usually after a PhD/Master's thesis. It is considerably longer than the research article, and this is the monograph that also drives research, although it is usually more difficult to publish, and the publication takes more time than smaller research texts, like the scholarly article. It tends to follow, in particular in the natural sciences, an IMRAD structure (see above). In contrast to the research book, a text book is a book for students, a novice readership. Though it is often shorter than a research book, it also requires special expertise: Its purpose is to introduce novices to the area and to make them enthusiastic, so that they go on with their reading. A similar concept is pursued by a course book or even an introductory book, which are also geared towards lower levels at university teaching and sometimes even follow the structure of a university course of 12 to 14 weeks. They all have to be sufficiently different from a lecture series or a conference "paper" selection, which are rarely published. A handbook, by contrast, is written by even more experienced specialists. It is usually quite thick and detailed with many state-of-the-art articles, so that the specialist perspective at a certain point in time is shared with other academics; again it is an honour to be asked to contribute to a handbook, however that implies that editors and publishers expect a reliable contribution, which is the reason why students interested in concept formation may take a conveniently short contribution in handbooks as a reliable starting point for further research. Printed texts that derive from spoken genres like key-note lectures and conference collections do not serve this purpose so clearly.

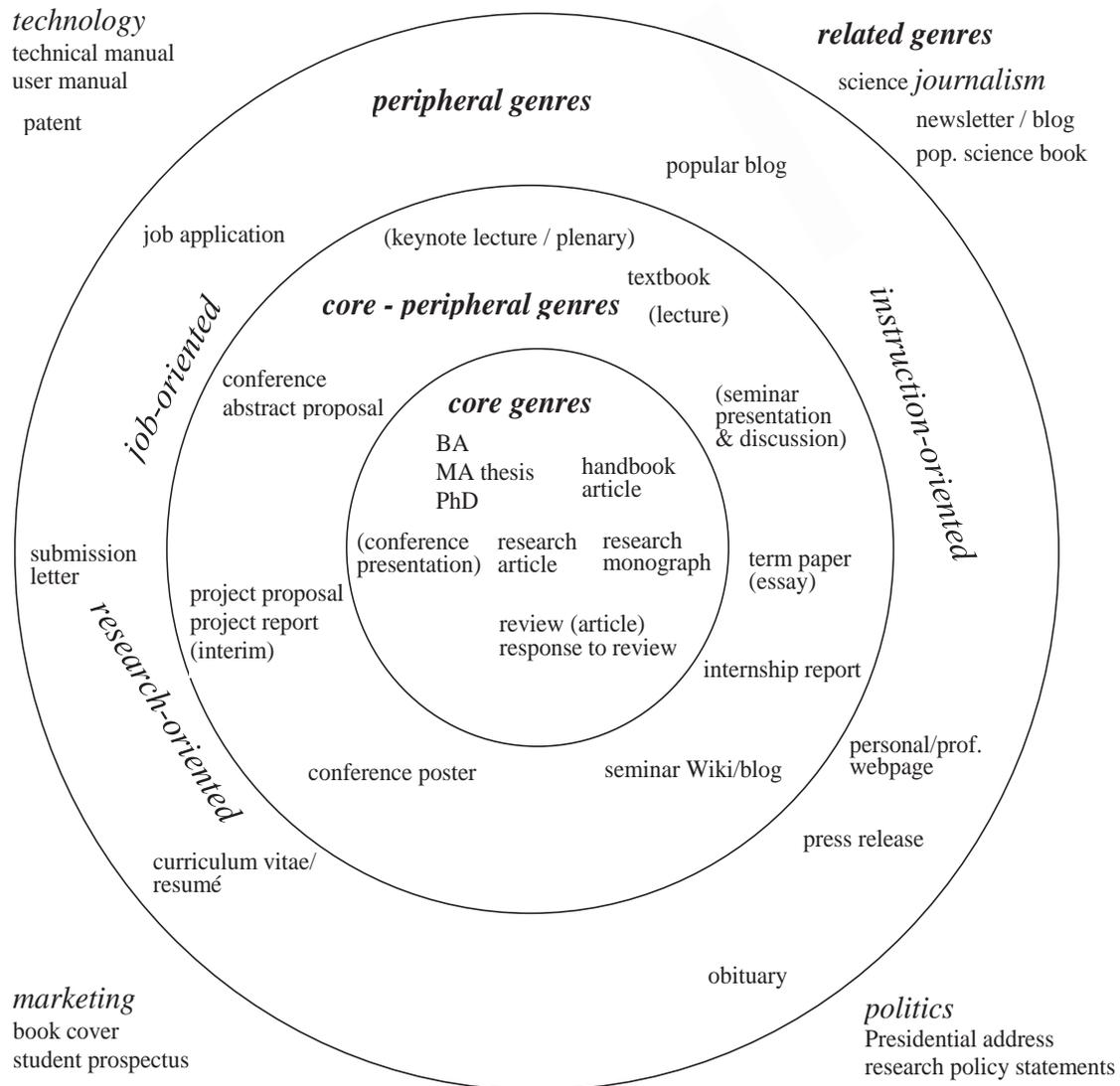


Figure 3: A prototype approach to academic writing

A further genre family are project proposals. They are usually teaching-oriented and may have a length of around 5,000 words, in which the researchers promise to reach certain milestones, so that in the end a project can be completed profitably; these milestones allow not only the researchers themselves, but also the funding agency or supervisors to check whether the progress is adequate to achieve a specific goal in a certain time. At university level, the first academic writing project is usually a BA thesis, which may still be considered a teaching genre, although it is supposed to introduce the students to independent research. Such a project is usually written at the end of a three or five-year period of learning at university and it often has a length of 15-20,000 words. The purpose is, of course, quite elementary, to be awarded a certain degree in the university career. A PhD project application by contrast is really research-oriented and it usually takes some five years of experience to write a proposal that may be around 5,000 words, so that a



funding agency is tempted to spend money on the applicant or a professor is willing to supervise the project. An oral genre that is related to a project application can be a thesis or PhD defence which is usually done after three or five years at university level; the standard format in many universities is a 20 minute presentation and a 20 to 30-minute discussion based on this presentation or the PhD thesis as such. Finally in this category of projects, the project or progress report has to be mentioned, that is usually about 6-10,000 words long. It is often written by multiple authors because of the funding needs and research experience.

An additional family of genres are conference genres: The prototypical conference presentation is spoken, rarely spoken to be written, and it requires a few years of experience although it is only 15 to 20 minutes long plus a 5 to 10-minute discussion afterwards. A more advanced conference contribution is a keynote lecture which requires usually some longer research experience (of more than 10 years) and again, it is a contribution of 45 to 60 minutes first and a discussion of 15 to 30 minutes afterwards. Usually the presenter is a senior scholar, famous for his/her books or personality. Related to the keynote lecture is the plenary lecture, which is often quite similar, sometimes a little shorter, but it also requires some similar experience. The conference abstract or conference proposal has to be written and submitted to the academic committee organizing the conference and 300 to 600 words are usually enough to decide whether the contribution can be accepted or not for the conference. New academic genres at conferences are poster introductions and conference posters. Poster introductions are usually very short and the related conference poster is usually only one A0 sheet, often including several tables and figures. Due to the fact that it is supposed to relate a usually young researcher to more experienced researchers or peers, it also contains an address and email that can be used for suggestions and general academic discussion afterwards. A conference is often made memorable by the conference proceedings, which are an edited collection of articles that are based on a conference presentation. They are usually not too long, especially when they are submitted for electronic publication on a CD (shortly after or even before the conference). Such a volume often includes multiple authors, but should concentrate on merely one topic and area. A conference report is then a more popular research genre which basically serves to disseminate the information on the topic of the conference. It can often be written by relatively young scholars and it is also rarely longer than two pages.

Special cases are genres that relate the academic work to the general public. This type of genre family I have called *valorisation* (in Table 2), because the French term shows more clearly that the purpose is to disseminate information on academic activities and convince a general public that academics pursue useful projects. Traditionally, such genres are considered peripheral, although they may include advanced technology. Thus popular blogs can easily include pictures or sounds and popular science books can also be developed further in this direction.

The central steps of academic qualification genres lead from the first academic attempt in a BA thesis (pursued after three years of study and often little more than 40 pages) to an MA thesis (after five years of study and over 60 pages), but the



expectations differ. The contrast to a PhD thesis is striking in quality and in quantity, since that can only be pursued after around eight years of study altogether and may take up at least 200 pages today. Quite similar is the next career step. If it exists at all, a postdoctoral thesis or *Habilitation* at traditional universities requires more than 10 years of experience, although the length does not have to be more than the PhD thesis. However, all these theses serve the writers' further personal qualification and are integrated into a relatively rigid and well-documented sequence of procedures.

Finally, it is worth mentioning some genre-related issues: ambiguous and supplementary genres. Thus abstracts serve different purposes: conference abstracts can be only promising proposals, the project just needs to be finished before the conference; article abstracts have to be finished and the complete version of the article has to be available immediately after the abstract is read. Oral contributions are traditionally accompanied by handouts, which serve not only as a guideline for the reader in the free speech that is expected in academic speaking today; they may also include figures and tables that do not fit well into a presentation and that the listeners may want to take home for further study. A handout is usually a genre that focusses on research, though it can also be used for teaching purposes at university level.

#### 4. Compiling corpora to analyse conventions

Although students may be well aware of the big reference corpora from the British National Corpus to the Brigham Young corpora (Davies 204-), it is also important that students get used to analysing much smaller and more specific corpora or extract subcorpora that are more comparable with their own materials for analysis. It is extremely important for young researchers to recognize the conventions of their own subdisciplines and compare them to the much broader options that departments and individual researchers are able to choose from. A corpus approach is appropriate not only for non-native speakers of English, who are not sure whether they can rely on their intuitions, but also for novices to base intuitive judgments on real-data evidence.

Finally, many students are inclined to use “Web as corpus” techniques to find out about lexical patterns (e.g. look for Google responses to queries of *different from* vs. *different than*, cf. Schmied 2015).

##### 4.1. Identifying textual, social and linguistic variables

When we compile a new corpus, we usually start with the biggest context variables: these include text variables like written vs. spoken, genre, etc. and social variables like native vs. non-native, gender, age/educational level, etc. The importance of these variables depends on whether we can set up simple logical research hypotheses (e.g. female writers are more “careful”, “tentative” and thus prefer more or different features).



For the linguistic features, it would be most useful if we could refer to standard textbooks (or grammars and dictionaries) or model articles to find a comprehensive list of features, preferably in different functional (sub-) categories. Unfortunately, many young researchers notice that usually there is no uniform opinion on these categorisations in their secondary literature or references and the delimitation of their linguistic concepts and their “ingredients” is their first major task.

An issue closely related to this is the question of how big a research corpus has to be. This of course depends on the frequency of the linguistic variable analysed and the number of textual and social variables included in the analysis.

For a sound statistical analysis not only an absolute number of cases is important, but also their relative distribution in the cells for each variable combination. Of course, such variables can be taken out and the results collapsed later-on during the analysis, but initially at least it would be most useful to have enough cases to exclude variables empirically as unimportant or insignificant at the end and not impressionistically at the beginning.

#### 4.2. Comparing corpora in Academic Writing

The *ChemCorpus* can be compared with other available current academic corpora of student or academic writing, less to the *ICLE* Corpora, which have mostly been compiled in the 1990s of argumentative student essays which are a different genre and not really part of academic writing today, more with the *BAWE* corpus, in particular if we take the English or the linguistics parts and students that have German as a first language only. It can be compared quite well with the *Brno Corpus*, which was developed in a parallel project over the last five years and includes a clear division between linguistics, literature and methodology specialisations. These non-native academic corpora can also be compared to the academic or academic humanities parts of the *British National Corpus* (*BNC*; from the early 1990s) or the *Corpus of Contemporary American English* (*COCA*; or even the *TIME magazine*), bearing in mind that those are of course much more heterogeneous as far as the production date and the socio-geographical data of writers are concerned (Davies 2004-). For explorative purposes, the academic corpora mentioned above can of course be used when the frequency of linguistic phenomena analysed is normalized, per 1 million words, for instance.

#### 4.3. Compiling stratified and *ad-hoc* or test corpora

Corpus compilation methodology and corpus design are discussed in all standard textbooks (e.g. Biber/Conrad/Reppen 1998: 246-250). Usually, “representivity” is considered an unrealistic goal and “diversity” is aimed at through stratified sampling. This means when many texts by female students are compiled already we look specifically for texts by male students until we have half or a representative or a satisfactory number. What satisfactory means in this context depends on the importance of the variable in our research questions or hypotheses. Only a well-designed corpus will bring us reliable data for an analysis of language



variation. An old issue in academic writing corpora is the size of the texts: generally, complete texts are preferable, because the argumentative, lexical or cohesion patterns may be quite different in the sections Introduction, Methods, Analysis, Results or Conclusion discussed in 1.3 above. The size of the corpus necessarily depends on the frequency of the variables investigated: pronunciation variables need only small, idiomatic or collocation variation large samples.

Sometimes it seems reasonable to compile a quick and dirty ad-hoc or test corpus first, just to see whether the (limited, insufficient, and skewed) data promise results. This is comparable to the tradition of disposable corpora from translation studies (Gavioli 2006).

#### 4.4. The *ChemCorpus* as a reference corpus

The following matrix (Table 3) shows that the *ChemCorpus* consists of data compiled from two types of degree programs, the traditional German Magister as well as the modern BA/MA from the Bologna System. The earliest pieces of academic writing in the Bologna System are the so-called term papers, which are specific exercises where students can demonstrate their critical thinking in linguistics, literary criticism or cultural studies and get used to the formal conventions, i.e. MLA versus APA style. The structure in these papers is usually not very strict, yet all and in many cases relatively narrative, based on the available secondary literature with (hopefully) some conclusions of their own. Only little afterwards some students have to write their first report, which are partly related to their ERASMUS studies abroad, partly to other projects where this form is appropriate. Essentially, it is the review of the plans they had before the projects were started and compared to the results which may have digressed more or less from the original plan. The first piece of academic writing students have to submit to qualify for further studies is usually a BA thesis which is in many ways a requirement if students want to change universities to continue in a Master's programme. The BA thesis and the MA thesis are usually quite similar and differ in length as well as in academic rigor. MA theses tend much more towards a conventional structure (like IMRAD style), but deviations are acceptable and depend very much on the subject chosen.

The length and number of texts per section in Table 3 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.