Filozofická fakulta Univerzity Karlovy v Praze

Ústav anglického jazyka a didaktiky

DIPLOMOVÁ PRÁCE

Intertextualita v odborném stylu: způsoby citování v humanitních a přírodních vědách

Intertextuality in Academic Writing: Citation in Soft and Hard Sciences

Vedoucí diplomové práce: **PhDr. Markéta Malá, Ph.D.**

Zpracovala: **Jana Štěpánková**

Obor: **Anglistika-amerikanistika; Dějiny umění**

Praha, červenec 2012

Prohlašuji, že jsem diplomovou práci vypracovala samostatně, že jsem řádně citovala všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

Souhlasím se zapůjčením diplomové práce ke studijním účelům.
I have no objections to the MA thesis being borrowed and used for study purposes.

V Praze dne 15. srpna 2012   podpis

Table of Contents

List of Tables

List of Figures

1. Theory

1.1 Introduction

This thesis focuses on citation practices in academic writing, particularly in the genre of the research article. Citation is one of the central features of academic writing, its importance has been recognized by many authors (e.g. Biber, 2006; Charles, 2006; Hyland, 1999, 2009; Thompson, 1996; Thompson and Tribble, 2001). By providing reference to prior work, scholars show that their research is not only relevant for their discipline, but also integrated into broader knowledge achieved in the particular field. Citation thus plays a significant role in providing context for new claims, constructing facts, integrating them into the current state of knowledge and proving researchers' familiarity with their colleagues' work. This all is achieved by establishing intertextual links. This process helps authors to be regarded as members of academic communities, which is essential for their recognition.

Special attention is paid to the integral type of citation (Swales, 1990), its subtypes and distribution across individual sections of research articles. Both self- and other-citation is included. Results will be compared between hard and soft sciences, each domain being represented by two disciplines: hard sciences by articles from astronomy and biology, soft disciplines by texts from art history and linguistics. For this purpose, a corpus of research articles from academic journals has been created. The thesis aims at describing citation practices adopted in these four disciplines and examining these practices with regard to the different conceptions of knowledge-making followed by hard and soft sciences. For the purpose of this analysis, several theoretical concepts have to be addressed. The individual sections of the theoretical part of this thesis are devoted to particular concepts, proceeding as follows.

Firstly, in sections 1.2.1 – 1.2.3., the notion of academic discourse and its clear preference for texts written in English will be discussed. The concept of academic discourse comprises also the idea of academic communities (e.g. Bizzell, 1992; Halliday, 1978; Swales, 1990) which has to be understood as it is closely connected to specific practices followed by writers within individual academic disciplines. These practices are dependent on the way in which academic communities perceive construction of knowledge, therefore it is also important to describe particular views on knowledge-making adopted in the four examined disciplines.

Secondly, it is necessary to define the genre of research article, one of the sub-genres of

academic discourse. The analysis in section 1.2.4 follows with the description of the structure of research articles, both the widely accepted IMRD structure (Swales, 1990), and its variations. The individual parts of research articles serve different rhetorical purposes, employing thus different strategies and language structures. One of these strategies is also establishing intertextual links for the above mentioned purposes. Considering the different aims of different parts of research articles, it is obvious that citations tend to be used in some parts more often than in others. This thesis also explores the predominant distributional patterns employed in individual disciplines.

Thirdly, particular ways of establishing intertextual links will be reviewed in section 1.3. This last theoretical part describes intertextuality in general as well as its types and classifications proposed by several authors (e.g. Fairclough, 1992; Hyland, 1999; Swales, 1990; Thompson, 1996; Thompson and Tribble, 2001). In contrast to preceding sections, this adopts a syntactical point of view rather than a semantical one. Particular types of citations, integral and non-integral, and their subtypes will be described in terms of grammatical categories. The theoretical part ends with a section focused on reporting verbs used in integral citation and provides a brief review of different points of view on reporting verbs, their tense and voice, as well as several classifications proposed by Thompson and Ye (1991), Hyland (1999) and others.

1.2 Academic Discourse

Traditionally, it has been acknowledged that there is a specific variety of English which can be considered the language of science. As Crystal and Davy point out in their book *Investigating English Style,* scientific prose performs several tasks, for example: reporting experiments, discussing problems, giving instructions, stating laws or defining concepts (Crystal and Davy, 1969: 251). In a number of these tasks the researchers need to report on the work done by other researchers or to refer to their own work. How the researchers working within various academic disciplines deal with reporting constructions, and what lexical choices they make will be discussed later. Furthermore, the choice of reporting verbs is much influenced by the amount of effort which the author puts in defining the subject matter as precisely as possible. Thus, as will be demonstrated in section 3.4, some disciplines allow a greater variety of verbs, whilst others strictly stick to few alternatives.

However, the language of science is only a part of a broader concept of academic discourse and, moreover, it is only one of various demonstrations resulting from the discourse practices. The term academic discourse refers to "the ways of thinking and using language which exist in the academy"

(Hyland 2009:1). Hyland's widely-quoted definition clearly states that discourse does not involve academic writing only, nor does it refer merely to university lectures. It comprises the whole academic environment, attitudes, communication, terminology, world knowledge, etc. The relationship between a member of the discourse community and the discourse itself is multidimensional: the members are constantly being shaped by the discourse to which they contribute (and thus co-create it) and within which they conduct their research. Although the study of academic discourse is relatively young, recently it has gained prominence and filled pages of linguistic journals such as *English for Specific Purposes*, *Journal of Pragmatics*, *Applied Linguistics*, etc., in addition to periodicals focused exclusively on academic writing such as *English for Academic Purposes, Journal of Second Language Writing* and others.

Our society has undergone several significant changes which have led to the "explosion of research on academic discourse […] over the past 20 years" (Biber 2006:6). Two of these changes are worth mentioning here. The most evident reason for this massive interest in academic discourse is the fact that English has become the international language of academic research, science and scholarship. Hyland (2009:4) even claims that English is nowadays becoming "less a language than a basic academic skill for many users around the world", which at the same time directs our attention to the second major change in the situation of academic writing. Most of the research articles, scientific journals, and even books are now available online and thus can be followed worldwide. This simple fact increases pressure on every researcher, regardless of their origin, to publish in English and consequently leads to an increasing need for various handbooks, courses and studies aimed at the style of writing, giving lectures, communicating with peers from elsewhere. Simply said, aimed at academic discourse.

In his *Explorations in Applied Linguistics*, Widdowson (1979) claims that there is a universal rhetoric used by the academia, a scientific discourse which is even independent of its realizations in a particular language. Offering an approach different from Hyland's, Widdowson sees scientific *discourse* as a "mode of communicating, or universal rhetoric, which is realized by scientific *text* in different languages by the process of *textualization*" (Widdowson, 1979:52; original emphasis). In this point of view, a *text* is defined as a formal manifestation of the language and can be examined on a quantitative basis (for example, observing the frequency of particular linguistic forms), whereas *textualization* refers to the functional realization of the language and can be examined qualitatively (for example, if we explore how the forms express specific elements of discourse). Widdowson thus manifests that the features of scientific discourse are primarily universal, although they are *textualized* variously through individual languages. Furthermore, this universal structure,

or, as Widdowson puts it, the patterns of rhetorical organization which establish the structure of scientific exposition are modified also by individual stylistic variations. However, in general the patterns impose "a conformity on members of the scientific community" (Widdowson, 1979: 61). The question of different languages is not considered important given the fact that this thesis deals with English texts only, nonetheless the notion of a rhetoric universal to all academic communities is closely connected to examining the differences and similarities among specific disciplines.

There is no need to engage with the concept of academic discourse in general any further as it serves only as a broad basis to derive from. Academic discourse comprises, roughly speaking, written and spoken discourse, various discourses of specific genres (lecture, research article, letter, etc.) and also, from a different viewpoint, the distinct discourses of different disciplines. This thesis is focused only on a part of academic discourse, namely on the written genre of research articles. The characteristics of this genre is provided in section 1.2.4.

1.2.1 Discourse communities

As mentioned earlier, academic discourse comprises the whole academic environment, however, it is widely acknowledged among the scholars engaged with the issue of discourse, that this general discourse encompasses a multitude of sub-discourses corresponding to particular academic disciplines, even to particular sub-disciplines. For example, Becher and Trowler see academic disciplines as different "tribes" of scholars working inside their "territories" (objects of their study). Although the distinction of particular disciplines raises some problems, basically the disciplines have boundaries and structurally, they are "manifested in the basic organizational components of the higher education system" (Becher and Trowler, 1989, quoted in Messer-Davidow, 1992: 678). Hyland notes that "even in applied linguistics, for example, there are fundamental distinctions in methods, concepts, and forms of argument between, say, cognitivists and post-modernists." (Hyland 2009: 19).

In general, the analyses of discourses of various academic disciplines are based on the assumption that scholars working within the same academic discipline form a kind of social group with particular aims, needs and methods. Members of these groups share not only the whole body of knowledge but also one unique disciplinary discourse. This shared discourse has gradually evolved from sets of conventions which were followed by writers dealing with similar problems. Academic knowledge is thus seen as "embedded in the wider processes of argument, affiliation and consensus-making of members of the discipline" (Hyland 2004:6). The main focus of this thesis is the examination of the differences in citations and reporting structures employed in four distinct

academic disciplines. According to the above mentioned assumption that writers of each discipline form a specific social group, this analysis will compare works of four of these distinct groups. Therefore it is necessary to devote attention to the very concept of the academic groups itself.

They are examples of the so-called discourse communities. Following the extensive concern in discourse analysis and closely associated with the study of genre, the interest in discourse communities has been developed relatively recently. "In the 1980s researchers turned to examining writers and writing in particular settings (e.g. Bazerman, 1988; Becher, 1989; Myers, 1989). These studies showed that writers' plans, goals and other process-based strategies are dependent on the particular purpose, settings and audiences." (Martín-Martín, 2006:197) The notion of "discourse community" has become an integral part of studies of academic discourse and academic writing since. Applied linguistics is aware of the existence of specific conventions adhered to by particular communities. The notion of discourse communities is closely connected to the sociolinguistic theory of language developed by Halliday. He expresses an idea that the members of particular communities "possess a communicative competence", and are able to make sense of a text because they *know* "what the speaker is going to say" and are sensitive to the "particular cultural, situational and verbal context". (Halliday, 1978:61)

The term "discourse community" is also widely used by Swales (1999) who has inherited it from Herzberg and his definition[1] This definition says that the notion of discourse communities is based on the assumption that "discourse operates within conventions defined by communities, be they academic disciplines or social groups" (Swales, 1999:21). This idea implies that language is a social concept and a type of social behaviour. As noted in the previous paragraph, such sociolinguistic approach was taken earlier by Halliday (1978) who claims that we can hardly look at language without taking into account its users and that the social context is the pivotal base from which it is possible to consider language. This approach is followed both by Hyland and Swales. The latter elaborates it further and suggests six defining characteristics that will "[identify] a group of individuals as a discourse community". According to Swales, a social group will be considered a discourse community if it meets the following criteria: it has common public goals; it has mechanisms of intercommunication among its members which it uses primarily to provide information and feedback; it possesses specific genres and specific lexis; it has a threshold level of members (Swales, 1999).

Swales's approach is further discussed by Bizzell. Regarding Swales, she points out that from the

---

[1]   Swales quotes a paragraph from: Herzberg, Bruce. 1986. The politics of discourse communities. Paper presented at the CCC Convention, New Orleans, la, March, 1986.

above quoted criteria it is evident that he uses the notion of discourse communities to "explain how a social group employs discourse to coordinate complex activities, to work together on a large, long term project" (Bizzell, 1992: 226). Apart from this, she offers a broad definition of a discourse community saying that it is a "group of people who share certain language-using practices" (Bizzell, 1992: 222). This "tentative" definition (in her own words) entails that the stylistic conventions used by the particular community govern both their inner interactions and their communication with outside world. Furthermore, Bizzell claims that the members of a discourse community perceive world in line with the canonical knowledge shared by their community.

To sum up the chapter about discourse communities, it is essential to emphasise the consequences which arise from Bizzell's viewpoint. Both the stylistic conventions and the shared view of the world create a fundamental basis determining the behaviour of the community, its research methods, writing practices and strategies and thus also the way it deals with the present state of knowledge and previous research results. Swales's point of view should also be taken into account. Briefly speaking, he claims that all members of a discourse community share the same object of study, the same procedures, interactions, and discourse conventions. These features form together a disciplinary culture which is then adopted back by individual authors. Considering the fact that soft and hard sciences create different discourse communities, impose different requirements on scholars and on the research methods employed, it is supposed that also the writing strategies and style differ to a great extent, including citations.

1.2.2 Soft versus Hard Sciences

This distinction between the so-called "hard" and "soft" sciences is a widely established concept, however, it still encourages many scholars to examine the inherent properties of these two types of disciplines, and to re-define the boundaries between them. For example, the sociologist N. W. Storer explored the relationship between the atmosphere of the researchers' community and the characteristics between soft / hard sciences. He claims that the organization of knowledge affects not only the research practices employed but also the degree of personality among the members of the academic community. Building on the assumption that hard sciences work with more exact methods and the research results can be more easily verified, Storer remarks that "we will find a greater degree of impersonality than in the soft sciences" because "one's colleagues can more easily identify any weaknesses in one's work" (Storer, 1966:79).

Hyland agrees that the concept of soft and hard domains presents problems. He is aware that creating a clear cut boundary between these domains "runs the risk of reductionism [...] by packing

a multitude of complex abstractions into a few simple opposites" (Hyland, 2009:63). He thus sees the hard versus soft distinction rather as a continuum with two opposing poles. Assigning the "hard" pole to sciences and the "soft" to humanities, he further defines contrasting qualities for each pole as we can see in Figure 1. Some of Hyland's characteristics determine also the style of reporting in hard or soft disciplines, respectively. For example, linearly growing knowledge in sciences means that the researchers build upon the level of knowledge achieved so far and pile their results on top of the previous ones, while more extensive disputation and disapproving is expected in humanities. Experimental methods used in sciences suppress the personality of the scholar and also preclude the evaluation of his findings; whereas in humanities, the author of a particular thought is usually considered important. Exactly the opposite, the authors stand in the centre and their claims are confronted with various theories of other authors, evaluated, contradicted or approved.

SCIENCES	SOCIAL SCIENCES	HUMANITIES
HARDER	⟵————————⟶	SOFTER
Empirical and objective		Explicitly interpretive
Linear growth of knowledge		Dispersed knowledge
Experimental methods		Discursive argument
Quantitative		Qualitative
More concentrated readership		More varied audience
Highly structured genres		More fluid discourses

Figure 1: Continuum of academic knowledge (Hyland, 2009:64)

In his more recent comment on educational research, Berliner proposes a different perspective on the "soft" versus "hard" distinction. He argues that "soft" scientists do their science under permanently changing conditions, they have to take into account various aspects of context and interactions. Therefore, he establishes new categories: "Easy-to-do science is what those in physics, chemistry, geology, and some other fields do. Hard-to-do science is what the social scientists do […]" (Berliner, 2002). However, regardless of the names given to these two domains of knowledge, the basic distinction stays the same. It is generally accepted that e.g. biology falls into the category of hard disciplines whereas e.g. history is generally included among the soft disciplines.

The question remains whether these two discourse communities (soft versus hard disciplines) need to be further divided into subgroups. It is presupposed that art history and linguistics as the chosen representatives of humanities; or biology and astronomy representing hard sciences employ different rhetorical strategies, show different needs and impose different requirements. It could be argued that art history is in a way "softer" than linguistics. This argument, supporting Hyland's

notion of continuum mentioned above, will be discussed later with respect to the reporting structures found in particular disciplines. Each discipline perceives the world in a different way which influences the whole style of referencing and citing. Therefore it will be necessary to divide the soft science-community or the hard science-community, respectively, into two sub-communities to avoid too much generalization and provide enough specifications for the four particular rhetorical contexts.

1.2.3 Scientific genres

Of course, academic writing is not one consistent genre. It covers not only discourses of various academic disciplines but also various types of texts participating in the process of gaining, demonstrating, sharing and broadening knowledge. The text types include textbooks, research and popularization articles, reviews, Nobel Prize acceptance speeches, and others. As Halliday points out (Halliday, 1988:140), the term "scientific English" is only a label for a specific register and it does not imply that this register, or functional variety, is homogeneous. Halliday understands the term "register" as a set of features that tend to co-occur almost regularly. Myers, drawing on Fairclough, sees science, "not as a discourse, a single set of social practices around one thing, but as an order of discourse, a terrain of competing discourses and practices" (Myers, 2003:267). Other authors use different terms, for example "scientific discourse" or "scientific register", however, despite the slightly different notions, the idea remains the same.

Among the more recent discussions about genre, a conference held in Madison, Wisconsin in 2005 (International Association of Applied Linguistics Conference) presented 7 linguists and their views on the notion of genre. (Each of the speakers was asked two questions, the first for six of them being invariably to define genre.) For example, Paltridge defined genre as "ways in which people get things done through their use of language in particular contexts. An academic essay [...] is a socially-approved way in which students show what they know, what they can do, and what they have learned in a course of study" (Johns et al., 2006:235). The second speaker, Hyland, supports Paltridge's view on genre as a socially constructed phenomenon by claiming: "For me, genre simply refers to socially recognized ways of using language. It is a term we all use for grouping texts together and representing how writers typically use language to respond to and construct texts for recurring situations." (Ibid, p.237) He also understands that writers need some level of familiarity and experience to be able to produce appropriate texts that would allow them to enter particular "genre-using community" (Ibid.).

Referring to the claim made by Halliday et. al. that language varies in relation to different users

(speakers) and uses (purposes) and, as a result of this variation, distinct varieties of particular languages must be taken into account (Halliday, McIntosh and Strevens, *The Linguistic Sciences and Language Teaching*,1964; summarized in: Widdowson, 1979:54), Widdowson proposes a distinction of dialects (ascribed to different users) and registers (different uses). Moreover, he points out that the term scientific discourse covers a variety of types of texts and adds that "the discourse of scientific instruction, of science as a subject, such as appears in textbooks [...] is different from the discourse of scientific exposition, of science as a discipline, such as appears in research papers" (Widdowson, 1979:52).

As mentioned above, this thesis is focused on only one written genre: the research article. This genre will be defined in terms of its overall organization, general purpose including authors' aims and readers' expectations and linguistic choices made.

1.2.4 Research Articles

Much attention has been paid to the genre of research article. The reason for such deep interest in this genre is both the vast and growing number of published research articles and the need to provide help for postgraduate students and researchers, native and non-native alike, who need to produce texts appropriate for submission. On the basis of their texts, they aspire to be accepted to particular academic communities. Therefore many linguists consider the study of research articles vital for future knowledge-making. As Hyland writes in his *Academic Discourse,* "The research article remains the pre-eminent genre of the academy. Beginning life in the form of the letters published in *The Philosophical Transactions of the Royal Society* in the mid seventeenth century, the RA is now not only the principal site of disciplinary knowledge-making, but as Montgomery (1996) has 'it the master narrative of our time'." (Hyland, 2009:67)

Swales (1990) agrees with Ard (1983), who considers research articles to have developed from the letters through which scientists had communicated with each other. This viewpoint is also shared by Bazerman (1983), who examined articles published in *Transactions* from its founding in 1665 till 1800 in order to trace the internal development of scientific discourse and make a diachronic comparison of scientific articles. He claims that "the experimental report […] evolved through the needs, conceptions and creativity of the many authors who took it up" (Bazerman, 2000:59). Moreover, Bazerman sees the format of research articles as emerging from particular communication practices followed inside the scientific community. "By talking to each other in a specific format scientists were figuring out how to talk to each other and changed the format according to what they were figuring out." (quoted by Swales, 1990:112)

1.2.4.1 Definition of the genre of research articles

It is clear that the research article is one of the most prolific genres produced both by students and researchers. As English has become the lingua franca of the academy, research articles written in English are one of the major means of disseminating knowledge across the whole world. However, there is an uncertainty how to define them and whether it is possible to establish some stable set of features. Swales points out that even though the research article is "anything but a simple genre […] it is quite surprising to find that there have been very few attempts to define an appropriate procedural methodology for approaching texts of this kind" (Swales, 1990:128). Samraj expresses this uncertainty by making a reference to researchers who see that "this label tends to be used loosely" (Samraj, 2004:5). This suggestion is further supported by Johns, whose perception of genre knowledge as "abstract, schematic, enhanced by repeated, contextualized experiences with texts" (Johns, 1997: 21) leads to ambiguities, confusion and vagueness as it is much shaped by individualities and subjective memories. Further she writes that "some genres, particularly in pedagogical contexts, are loosely, and almost casually, named," (Ibid.) and she explicitly mentions research papers. According to Johns, it is "difficult for students to determine from the name what is required. The problem with defining […] is further exacerbated by the teaching of the research paper as a specific, fixed text type" (Johns, 1997:23).

Furthermore, what must not be forgotten is the role of the audience, we must bear in mind that particular genres are intended for particular readers. When producing an academic text, the author should take the presupposed level of knowledge of the readership into account. De Oliveira (2006) supports this claim by noting that the research article and the science popularization article are acknowledged as two distinct genres, differing foremost in their readership. However, this thesis does not investigate various audiences in greater detail as all the researched texts added to the corpus were found in academic journals and therefore are expected to be written for specific audience – scholars, students and fellow researchers – and it is supposed that this kind of journals does not attract interest of outsiders or lay public. Therefore the authors can rely on certain level of knowledge shared with their readership and do not have to adjust their linguistic choices in order to suit general public.

Many investigations have been carried out concerning the research article and its form. Among others, Halliday (1998, 2004), or Swales (1990, 2004) examined the language and organization of research articles and found considerable parallels. As mentioned above, each discipline, or academic community, perceives reality quite differently and it was observed that these various perceptions of the world have significant bearing on both the grammar and the style used (e.g.

active vs. passive voice; types of reporting verbs or types of dependent clauses). The conception of reality of a particular academic community influences its style of writing and thus contributes to knowledge making. This issue will be examined later in line with the specific grammatical and stylistic features found in the corpus designed for this thesis.

1.2.4.2 Structure of research articles

Research articles are organized according to a more or less fixed pattern. In his book *Genre Analysis,* Swales (1990) provided an innovative analysis of the textual structure of RAs. Deriving from the almost generally accepted IMRD pattern (Introduction-Method-Results-Discussion) for scientific articles established in the course of the twentieth century, Swales gathered studies that had examined the lexico-grammatical features occurring across the sections. These studies were focused for example on reporting statements, present and past tenses, passive voice or authorial comments. However, IMRD pattern has been not fully accepted and Swales himself admits that it is applicable to papers based on empirical research only, not to theoretical writing. In empirical RAs, findings are derived directly from observations or experiments. In comparison, theoretical RAs describe the development of a theory or compare different theories.

Many studies have examined the applicability of IMRD pattern across various academic disciplines. For example, Lin and Evans (2011) carried out a cross-discipline study examining the major structures of empirical articles based on a manual analysis of 433 recent articles from high-impact English-language journals in 39 disciplines in the fields of engineering, applied sciences, social sciences and the humanities. They found out that the IMRD pattern cannot be regarded as default, moreover, according to them, the most frequent pattern is ILM[RD]C, Introduction–Literature Review–Method–Results and Discussion merged into one section–Conclusion. Although Swales claimed that the IMRD framework is applicable to empirical research articles, according to Lin and Evans's research, only 53 out of 433 RAs were structured this way. Considering only the disciplines tackled by this thesis, the prevalent framework of linguistic articles is ILMRDC and biology articles are organized according to IMRD, IMRDC or IRDCM patterns. They further report that "no empirical RA was found in the field of history of art" (Lin and Evans, 2011:8). Thus, no major structural pattern can be traced.

Yang and Allison (2003) analysed 20 research articles in applied linguistics and the rhetorical choices that are made in the course of the closing sections (Results, Results and Discussion, Discussion, Conclusion and Pedagogic Implication). They found out that the Results section in empirical RAs appears obligatory. Following the Results section, in nineteen out of twenty cases, either Discussion or Conclusion occurs, making this a quasi-obligatory feature. Both sections

appear only in two articles. In line with many other studies (e.g. Holmes, 1997 or Posteguillo, 1999), Yang and Allison have also observed repeating cycles occurring in Discussion sections shifting the article from specific to general.

Swales investigated individual sections of the IMRD structure according to their aims and most of the studies following his ground-breaking work *Genre Analysis* also treat sections as separate items. In this thesis, the individual sections will be described briefly as well. As has been noted by many authors, "the notion of communicative purpose is central for analyses of both RA macrostructure and microstructure" (Yang and Allison, 2004:266). The most problematic section according to Swales is the Introduction. As the "opening paragraphs somehow present the writer with an unnerving wealth of options" (Swales, 1990:137), he, similarly to other linguists, decided to pay special attention to the Introductions of research articles and arranged his much-quoted CARS (*Create a Research Space*) model represented in Table 1 below. Each of the sections of RA is considered to include typical rhetorical practices, including reporting, quoting and adopting a stance on the claims of others. The model consists of three basic Moves, each including several Steps. A Move captures the rhetorical function of a part of the text, thus enabling "the categorization of chunks of texts in terms of their particular communicative intentions" (Yang and Allison, 2003:370). A Step is less general and more specific in following the rhetorical function and its realization.

Although Swales suggests that his model for Introduction sections is valid irrespective of academic field, much attention has been paid to the CARS model and its validity across various disciplines. Among others, Kanoksilapatham (2005) examined its application in biochemistry, Ozturk (2007) in applied linguistics, Samraj (2002) compared the structure of RA introductions from wildlife behaviour and conservation biology, Holmes (1997) analysed the validity of extended version of the CARS model in history, political science and sociology. Contrary to Swales's presupposition, these studies have indicated that there are significant differences across distinct or even related disciplines.

This thesis also compares the frequency of reporting structures and their distribution across the texts typical of soft and hard sciences. It is thus important to bear in mind the rhetorical functions of reporting structures and the reason for their usage in particular sections. Due to their aim and content, the Abstract and Methods sections do not typically include as many instances of reporting or citations as the Introduction and will be described here only briefly. More attention is paid to Introduction, Discussion and Results sections.

Abstract

Almost all research articles analysed in this thesis begin with the abstract. It is a particularly important part of the article as at this point, readers encounter the text for the first time and decide whether to continue reading (Hyland, 2003). The main rhetorical aim of this part is therefore to attract readers' attention and make them read the full text. For this reason it necessarily strives to show that it has something new to say. The abstract offers a much condensed account of the main problems and aspects of the research described. Even within these first lines of research articles, Hyland has observed a difference between the hard and soft sciences. "The hard knowledge abstracts tend to stress novelty and benefit, while writers in the social sciences largely draw on the notion of importance to promote their work." (Hyland, 2009: 70) This is further supported by stating that novelty is a pre-eminent concern of those scientific disciplines in which "innovation and progress are expected and practitioners look for new results to develop their own research." (Hyland, 2009: 71)

Introduction

Following the abstract, the article opens with the Introduction. In this section, the authors need to establish a space for their research, to prove its importance and also to show that they are acquainted with the state of knowledge concerning his topic. The Introduction involves a large number of intertextual references, quotations, paraphrases or summarisations of previous research because it is necessary to place the researched topic into the context of relevant existing literature. By this means, the authors not only show their knowledge and make the reader familiar with what is already known about this topic, but they also create space for their own contribution by emphasizing gaps, incompleteness, obscurities or controversies within the researched matters that they wish to fill in, complete or clarify.

Swales's CARS model is based on his examination of the Introductions to forty-eight articles from the field of natural and social sciences. He found out that the Introduction sections are organized more or less alike. Move 1 of CARS model, establishing a territory, aims to assert that the study is important and has a lot to contribute to the current state of knowledge. It typically includes three Steps, as provided in Table 1. Centrality claims aim to prove that the research is important and plays a significant role within a well-established research area. As soon as in this part, references to previous research are made. According to Swales (1990), Step 1 typically includes such phrases as for example: "there has been a wide interest in...", "many investigators have recently turned to...", "... has been studied by many authors" or "...researchers have become increasingly interested in...". Step 2, "making a topic generalization", involves general statements about knowledge, techniques

or phenomena. In Step 3, authors review items of previous research they find relevant. Therefore it is one of the parts in which it is necessary to mention other researchers working on the same topic. In Swales's words, the author needs to relate *what has been found* (or claimed) with *who has found it* (or claimed it). More precisely, the author needs to provide a *specification* (in varying degrees of detail) of previous findings, an *attribution* to the research workers who published those results, and a *stance* towards the findings themselves (Swales, 1990:148).

Move 1: Establishing a territory	
Step 1 and/or	Claiming centrality
Step 2 and/or	Making topic generalization(s)
Step 3	Reviewing items of previous research
Move 2: Establishing a niche	
Step 1A or	Counter-claiming
Step 1B or	Indicating a gap
Step 1C or	Question-raising
Step 1D	Continuing a tradition
Move 3: Occupying the niche	
Step 1A or	Outlining purposes
Step 1B	Announcing present research
Step 2	Announcing principal findings
Step 3	Indicating RA structure

Table 1: CARS model for article introductions (Swales, 1990:141)

This part of RAs is supposed to include the highest ratio of references and quotations, however, the question of validity of this presupposition for the corpus established for the purposes of this thesis will be discussed later.

After having established the territory, i.e. having explained the importance of their research, writers continue with Move 2: establishing a niche. In this Move, authors pave their way for their studies by pinpointing drawbacks, weaknesses or flaws in the current state of research and emphasizing problems that require solutions. With respect to the aims of this Move, it typically begins with an adversative connecting device (Swales explicitly mentions *however, yet, nevertheless, unfortunately* and *but*) which is logical considering the aim of Move 2.

Move 3, occupying the niche, is tightly connected to Move 2 because the author tries to "turn the niche established in Move 2 into the research space that justifies the present article" (Swales, 1990:159). This section opens either with the specification of the purposes, or with the description

of the present research. As has been observed by Swales, most of introductions end with this Step 1. If they do continue, there are two options. Either to announce their findings, or to announce the structure of the article instead.

Methods

Not much attention has been paid to the Methods section as it is generally regarded as relatively straightforward and unproblematic (Holmes, 1997). Its main aim is to describe procedures used in the course of the research and to assert their credibility (Skelton, 1994). However, particular procedures adopted in particular academic disciplines differ to a great extent, not only in accordance with the distinction between soft and hard sciences but also among individual academic fields and academic communities. Therefore the Methods section must inevitably also differ fundamentally according to particular scientific disciplines.

For example, Nwogu (1997), following Skelton (1994) analysed this section in medical research articles and observed the following regularities. At first, the author gives an account of the procedure used for data collection, then describes the experimental procedure adopted (if the research includes laboratory tests or other experimental work) and the final part of the Methods section consists of describing data-analysis procedure (if the research involves statistics or other quantitative methods). Martínez (2003) examined thirty RAs from the field of biology, concentrating on the Methods and Discussion sections. As she points out, the Methods section should provide the reader with facts and assure them that the research was conducted according to widely accepted norms and techniques.

However, the Methods section does not usually contain references to works of other researchers as well as direct quotations and thus no more elaborate analysis of this section is necessary.

Results and Discussion

The last two sections of the IMRD pattern for structuring RAs have been examined much less than the previous parts. In Swales's words, there is "much variation in the extent to which Results sections simply describe results and the extent to which Discussion section re-describe results" (Swales, 1990:170).

If the article includes both parts, final data are described but not yet interpreted in the Result section and the interpretation usually follows in the Discussion. According to Yang and Allison (2003), in Results section the findings are presented and the authors try to establish their place within the community and gain importance for their work. Hyland claims that in the Discussion section, "the current work is most vigorously 'sold' […and] previous research is treated as

background and introduced" with the intention to either support, compare or invigorate the newly conducted research (Hyland, 2009: 73). Together with other means of credibility, acknowledging other researchers' results and viewpoints appropriately helps writers to represent themselves as trustworthy members of a particular academic community.

Consistently with Brett (1994), Nwogu (1997) and Postegullio (1999), Yang and Allison also suppose that the Results section not only reports the findings but also comment on them. It is thus difficult to distinguish between the sections of Results and Discussion. Yang and Allison claim that these sections differ in the communicative purposes. In other words, their focus is shifted. Whilst the main communicative purpose of the Results section is to report results and the main focus is on the reporting, the purpose of the Discussion section is to comment on these results and the focus is thus shifted from reporting to commenting on the results. To make this clear, they offer a list of five Moves listed in Figure 3 that can be distinguished in the Results section. In the Discussion section, they have distinguished seven Moves, four of them overlapping with the Results and Moves 5-7 being optional.

Results	Discussion
1. Preparatory Information	1. Background Information
2. Reporting Results	2. Reporting Results
3. Commenting on Results	3. Summarising Results
4. Summarising Results	4. Commenting on Results
5. Evaluating the Study	5. Summarising the Study
	6. Evaluating the Study
	7. Deduction from the Research

Table 2: Moves according to Yang and Allison (2003)

These two final Moves are considered in a way similar to the Introduction. In the opening of RAs, authors establish their position in the context of the current state of knowledge and assert the importance of the object of their study. In this final part, they refer to existing literature and knowledge as well, this time trying to contextualize their results, observations or findings. As Skelton puts it, "this move validated the paper by claiming membership of the academic community, [and] it reflected move two, which was also designed to contextualize the study by reference to literature" (Skelton, 1994, 458). Overall, this sequence of Moves has been accepted, even though minor differences have been observed.

The above mentioned study of 30 biological articles carried out by Martínez defines the Discussion as an argumentative part of the article, its main communicative aim being to evaluate

and interpret the data obtained and also to persuade the audience that the research in question and the findings it has provided are relevant for the knowledge-making within the discipline of biology. The text proceeding from the Methods to Discussion shifts from "fact to abstraction [...], accompanied by increased dialogic intervention" (Martínez, 2003:107). This notion of dialogic intervention entails also increased number of citations and other references to previous research.

Conclusion

Although conclusion is not one of the main sections proposed by Swales, it is still considered one of the frequent parts. Yang and Allison (2003) have identified conclusion in thirteen out of twenty articles in applied linguistics. Its primary communicative purpose is to summarise the whole study, to evaluate it and to make deductions from the research. In contrast to Discussion, in which a commentary on specific results is provided, the Conclusion highlights overall results and seeks to evaluate them and to propose possible future research.

1.3 Intertextuality in research articles

By publishing RAs and other academic texts, researchers pursue several goals. First, they want simply to inform the public about their findings. Second, they need their results to be persuasive. This is closely connected to the notion of academic community and the aspiration of the authors to be accepted into this community. As suggested by Hyland (1999), academic knowledge is "now generally recognized to be a social accomplishment, the outcome of a cultural activity shaped by ideology and constituted by agreement between a writer and a potentially sceptical discourse community" (Hyland, 1999:341). All writers establish intertextual links to other writers in their discipline but they do it in different ways. By examining these various ways, disciplinary distinctions can be traced. For the following discussion the convention established by Thompson and Ye (1991) and accepted by a significant number of linguists, (among others by all whose works are mentioned in this thesis), was adopted. In keeping with this convention, the person who is citing is referred to as the "writer" and the cited person as the "author".

Considering citation from the diachronic point of view, it has gained an increasingly prominent role in constructing facts throughout the articles and become fundamental for the acceptance of claims. As Bazerman found out in his analysis of *Physical Review* articles, the number, function and type of references changed immensely during the last century. In the early years of research articles, referencing was used rather generally, and, "serving as a roll-call of previous work in the general area, references congregate at the beginning of the article, never to be raised in a significant way in

the course of the argument" (Bazerman, 2000:164). In the articles from 1910, Bazerman found in average only 1,5 references per article and only few of them were dated. In contrast, for the articles published in 1980 his graph shows more than 25 references per article. Not only have the references increased in number but they have also become more focused, pertinent, accurate and incorporated into the argument. In Bazerman's words, "common theory has become an extremely strong force in structuring articles and binding [them] to each other" (Bazerman, 2000:157).

As will be shown later, the average number of references has certainly increased even more since. Together with the growing number of references also the length of the Discussion section has increased, gradually replacing the Methods section and becoming the dominant base of persuasion. These features are generally regarded as interconnected. In other words, writers have intensified their efforts to embed new research in the literature. The contextualisation of research articles has increased. Nowadays, citations as responses to previous writing are "important constitutive features of research articles, contributing to how we identify and evaluate research writing in different disciplines" (Hyland, 2004:21).

As Martínez (2003:107) pointed out, doing science inevitably involves a degree of "manipulating objects in the real world, abstracting to generalise, classify and evaluate and relating to others in the discipline through first person use and citation." Reporting, quotations or making other reference to previous research is one of the means of integrating new findings and claims into a wider disciplinary framework. Thus, writers also gain relevance and assure their audience that they can be regarded as members of a particular academic community and as contributors to its endeavours. Their work is thereby embedded in the literature produced within the context of the academic community.

However, finding one's place in the academic community is not the only purpose of citation. As was discussed earlier in connection with the structure of RAs and communicative intentions of the individual sections, references to prior research occur predominantly in some parts, serving thus specific rhetorical intentions. The highest frequency of occurrence has been recognized in the Introduction and Discussion sections. To summarize the communicative strategies briefly, in the Introduction, the writer establishes a narrative context, creates a rhetorical gap for his or her own research and gains credibility through demonstrating the knowledge of previous work done within the field. This all is achieved by acknowledging previous research and pinpointing its weaknesses or flaws, in other words, by citing, summarising, paraphrasing or evaluating prior literature. Such "embedding of argument in networks of references not only suggests a cumulative and linear progression, but reminds us that statements are invariably a response to previous statements"

(Hyland, 1999:343).

1.3.1 Types of intertextuality

Since the concept of intertextuality is a broad phenomenon manifesting itself in various fields and disciplines, many writers, among them Hyland (1999) and Swales (1990) draw upon Norman Fairclough's systematic approach in order to make this concept efficient, comprehensible and easier to apply. Fairclough (1992) himself was influenced by intertextuality studies developed within the field of literary theory, particulary by the ideas of Mikhail Bakhtin. He distinguishes two basic types: manifest intertextuality and constitutive intertextuality (or interdiscursivity).

The latter type construes the configuration of discourse features and conventions, registers, or styles shaping the form of the text. In Fairlough's words, it refers to "the configuration of discourse conventions that go into production of the text" (Fairclough 1992:104). This type of intertextuality will not be of any interest in this thesis, although it is important to bear in mind that both types are related and cannot be separated clearly. In contrast, all references analysed here can be subsumed under the term manifest intertextuality. As Fairclough defines it, this type of intertextuality denotes cases where other texts are explicitly incorporated into the text in question in the form of quotes, paraphrases or citations and usually marked by quotation marks or reporting structures. In general, its function is to manifest the ideas of others in the particular discourse.

1.3.2 Citation

The term citation, as it is used in this thesis, in general denotes the attribution of propositional content to another author. The writer who chooses to report someone else's speech or thoughts must make several choices. Thompson (1996) proposes four basic steps that have to be followed: At first, it is necessary to choose whose voice to report, the second step of the referring procedure is the choice of how to present the content of the original message and it is closely related to the third step, the signal indicating that a language report follows. The last step is the attitude to or the evaluation of the original message by the present writer. As the choices made at each level of this four step procedure of reporting lead to different types of reporting, each step is described in a larger detail.

The choice of the voice opens a range of possibilities. The author of the original thought can be identical to the writer, or, specified other(s), unspecified other(s), community, unspecifiable other(s), or the source can be obscured for manipulative purposes. Self-citation as one of the self-mentioning structures will be discussed in detail in section 1.3.2.3. Specified other(s) simply refer to

other known speakers whilst if choosing unspecified other(s), the writer presents some information as a report but does not provide its source, although this is identifiable form the context. Community as understood by Thompson comprises folk quotes, families, friends, etc. and depends on shared knowledge. Usually, such instances do not include reporting signals and will be of no interest in this thesis. Reference to unspecifiable other(s) is rather controversial as only the reader's awareness determines if he or she recognises that the voice is not the writer's own. From all these options, only the self-reference and reference to specified other(s) is dealt with in this thesis.

The treatment of the message falls, according to Thompson, to five groups: quotation, echoing, paraphrasing, summarising or omitting. For the purpose of this thesis, only two distinct ways are distinguished. The first possibility for the writer is to use direct quotations, i.e. to use the exact words uttered or written by the author of the original idea. Quotations are particularly useful if they say exactly what the writer needs to say to support his or her own claim. Thompson (1996) further mentions that direct quotations may indicate a higher degree of credibility given to the original. On the other hand, direct quotations also may imply a particular degree of distance which the writer wishes to keep between himself and the proposition. In the case of direct quotation, the writer must make choices also in respect of the length of the quoted texts, ranging from a few words to extensive blocks of original writing (Swales, 1990). However, since the way of presenting the information is crucial for gaining credibility and acceptance, direct quotations do not occur so often as paraphrases or summaries.

The second possibility is to report the original idea using the author's own words, which is traditionally subsumed under the category of indirect speech (Thompson, 1996). Either this can be done by paraphrasing or summarising. The former simply means that the whole text or idea is paraphrased and roughly keeps the length of the original, while the latter option allows the writer to shorten the original text by pinpointing only the crucial ideas or ideas most fitting for the writer's research. Both fall under the category of reporting. These strategies represent especially suitable ways of achieving effective support for the new arguments as they enable the writer to emphasize and interpret what is being cited. Another option to use one's own words is to provide a generalization combining two or more sources.

Hyland (1999) examined a corpus of 80 RAs from ten leading journals in eight academic disciplines and investigated the contextual variants of quotations. His study comprised both soft and hard sciences: biology, physics, marketing, applied linguistics, philosophy, sociology, mechanical engineering and electronic engineering. Considering the presentation of cited work, he found no instances of direct quotations in hard sciences, in soft disciplines, they accounted only for 2% and

3% in philosophy and marketing respectively and 8% in both applied linguistics and sociology. The ratio of block quotes is obviously even lower. Summaries are highly represented in all disciplines accounting for 66% in electronic engineering (the lowest amount) or 89% in philosophy (the highest amount) and generalizations ranging from 8% in philosophy to 38% in biology.

1.3.2.1 Integral and Non-integral citation

Swales (1990) suggests a basic distinction between integral and non-integral forms of citation which has been further widely adopted by linguists working in the field of academic discourse, among others by Hyland (1999), Thompson (2005), Kwan (2006), Charles (2006) or Hewings et al. (2010). These two categories differ from each other mainly in the direction of focus. However, his model is easily applicable on reporting structures as well because it works with surface features of texts only. As Swales puts it,

> "[a]n integral citation is one in which the name of the researcher occurs in the actual citing sentence as some sentence-element; in a non-integral citation, the researcher occurs either in parenthesis or is referred to elsewhere by a superscript number or via some other device." (Swales, 1990:148)

Non-integral citations are considered the most common, in both hard and soft sciences (Hyland, 1999; Thompson, 2000; Thompson and Tribble, 2001; Hewings et al. 2010), from the disciplines studied in this thesis, RAs in biology display the highest ratio of non-integral citations, as much as 90%, physics 83% and applied linguistics 66% (Hyland, 1999; Thompson, 2000). The name of the author is included either in parentheses, usually occurring in the sentence-final position, or in the footnotes or endnotes referenced to by a number in superscript. Thompson (2005) mentions four possible purposes of non-integral citations: to attribute the proposition to its source, to provide reference to a text where the reader can find more information, to indicate the author of a particular theory, object or techniques, or to identify other articles alluded to in the current writing. Considering the function of non-integral citations, writers choose them if they wish to emphasize the proposition itself, for example the results of a particular research, whilst the author and his or her particular text or piece of research are not given much importance. According to Kwan (2006), non-integral citation forms with the name of the author and the date of publication in parentheses and with no reporting verb usually accompany content which is regarded as accepted knowledge.

The main focus of this analysis is on integral citations. It this type, the name of the author can stand in the position of a subject (followed by a reporting verb), can be an agent of a passive structure, it can be part of a possessive noun phrase (for example, *Swales' theory*) or stand in the adjunct position (*according to Swales*). The integral forms are more likely to present negotiable

information. However, this information, i.e. research findings, is only one of the possible research-related semantic features which can be exhibited in citations. The other features can be: research focus, research processes or participants in the research. In terms of focus, by being included in the sentence, the integral citations place the main emphasis on the author, or his or her particular piece of research, whereas the non-integral citation forms shift the focus on the cited proposition. Charles (2006) examined phraseological patterns used in citation structures in the disciplines of politics/international relations and material science, more specifically, she examined finite reporting clauses with *that*-clause component. Her data proved that the majority of integral citations occur with a human subject (amounting up to 96,7% in politics and 75,7% in materials).

citation	type		example
non-integral	source	the proposition is attributed to a source	Citation is central because it can provide justification for arguments (Gilbert, 1976)
	identification	identifies an agent within the sentence it refers to	...model has been developed to incorportate...(Potts, 1980)
	reference	includes the directive "see"; common in reference to procedures / proofs of lengthy arguments	DFID has changed its policy recently with regard to ELT (see DFID, 1998)
	origin	indicates the originator / creator of a concept / product	...software used was Wordsmith Tools (Scott, 1996).
integral	verb controlling	citation acts as an agent that controls a verb (either in active or passive voice)	Davis and Olson (1985) define a management...
	naming	citation is a noun phrase or a part of NP	Work by Samuel and East (1990) demonstrated that variety
	non-citation	the name of the author given without a year reference; common when the reference has been supplied earlier	The "classical" form of the disease, described by Marek, causes significant...

Table 3: Classification of citations according to Thompson and Tribble (2001)

Thompson and Tribble (2001) have further extended Swales' division on the basis of the examination of doctoral theses in the fields of agricultural botany and agricultural economics. They have proposed a further classification of non-integral and integral citations which is provided below in Table 3. They also examined the density of particular types of citations in different rhetorical sections of academic writing and found out that the highest number of citations can be found in the Introduction part (the prevalent types being: source, identification and verb-controlling citations) and in Discussion (where the same types have been employed), whereas the lowest density was

found in Methods (with reference, origin and naming types) and Results (source type). These findings correspond with the common presuppositions about rhetorical functions of the individual parts of research articles as explained in section 1.2.4.2 and different Moves and Steps involved.

According to Thompson and Tribble, the choice of the particular citation type and subtype is governed by the theme / rheme position. In addition to the functional sentence perspective, the choice also depends on how much prominence the writer wishes to give to the authors/researches involved, thus making decisions, for example, between non-integral identificational or integral verb-controlling citation types. These choices are, of course, also influenced by a particular academic community and its views on knowledge-making. For example, references to controlled experiments in hard sciences do not consider the human factor consequential.

1.3.2.2 Reporting and Non-reporting citation

Swales also sub-categorizes citaitons into reporting and non-reporting ones, distinguished on the basis of the presence of a reporting verb (*show, claim, suggest,* etc.) or its absence, respectively. Swales mentions that the repertoire of reporting verbs is quite large, he estimates the existence of 50 candidates. Hyland's analysis (1999) discovered over 400 different reporting verbs, although nearly half of them appeared only once. Such a high number of possible reporting verbs points to the fact that any verb referring to any action which has been involved in a research process in fact has the potential of being used as a reporting verb (Thompson and Ye, 1991). The semantic classification of reporting verbs will be discussed later in section 1.3.4, and the number of different verbs actually found in the analysed articles will be presented in section 3.4. Hyland further counted the ratio of reporting versus non-reporting structures and according to his findings, the ratio is roughly equal, accounting for approximately 40 per cent of reporting references across the examined disciplines. Philosophy and physics make exceptions, employing 67 and 27 per cent respectively.

1.3.2.3 Self-citation

Self-citation is commonly defined as a reference in which the writer and the author is the same person, or, in cases of multi-authored articles, it is sufficient if the texts have at least one author in common. This thesis draws on this definition of self-citation.

Many studies focused on reporting exclude self-citation completely, among others Hyland (1999). The reason why the phenomenon of self-citation is usually treated differently than that of other-citation is for many scholars the difference in its motivation. By contrast, Bonzi and Snyder (1991) examined 51 self-citing authors in natural science disciplines and found only few differences in motivation of self-citation and other-citation. This finding is supported by Aksnes (2003) who

analysed the scientific production of Norway by examining 45.000 publications, all from hard sciences. More authors claim that the motivation of self-citation does not in fact differ much from the motivation of other-citation. The comparison is put in a Table 4 provided by Hartley (2011) and shows that six out of eight reasons for citation are valid for self-citation as well as for other-citation.

In opposition to the authors who exclude self-citation form their research on citation, there has been an intense ongoing debate among scholars in the fields of bibliometrics and scientometrics aspiring on defining and interpreting self-citation, its role and purpose in the scientific communication and the influence it has on calculating impact factor. Many authors have focused on this issue, among others Bonzi and Snyder (1991), Aksens (2003), Hyland (2003), Kovačić and Mišak (2004), Glänzel, Thijs and Schlemmer (2004) or Hartley (2011).

O	S	Tell the readers where they can find the material being discussed
O	S	Provide evidence for the writer's claims
O	S	Draw the reader's attention to little-known and unknown work
O	S	Indicate to the reader the scholarship and the experience of the writer
O	S	Align the author with a particular school of thought
O	S	Show development of thought
O		Show the writer's respect for particular people
O		Mutual grooming—you cite them and they will cite you

Table 4: *Possible reasons for citing others (O) and self (S), (Hartley, 2011)*

Self-citation can surely be seen as part of a wider context of self-mention expressing writer's presence. Together with the use of first person pronouns, reference to the writers' own previous works is one of rhetorical strategies for presenting personal contribution to the topic of the RA. References to one's own work are one of the essential components of researchers' academic identities. On the other hand, scientometrics uses the number of citations of a particular text as a basis for calculating impact factor. The quality and productivity measures are thus based on citation counts and this fact leads to further problems in treating self-citations as they do not reveal much about the real scientific impact of the particular work. Although the validity of self-citation for impact factor is a serious problem, it is not the only controversial topic related to self-citation.

Some of the above named authors have mentioned that especially frequent self-citations are regarded as suspect (e.g. Bonzi and Snyder, 1991). Moreover, self-citation is sometimes condemned as means of artificially raising citation rates and thus strengthening the writer's position among other researchers from the same scientific community. (Glänzel, Thijs and Schlemmer, 2004). Another problem connected to the impact factor is pinpointed by Aksnes (2003). His findings have

shown that the majority (63%) of citations the particular article obtains within one year after its publication is formed by self-citations, whilst 15 years after the publication, only 9 per cent of all citations are self-citations. This is further acknowledged by Glänzel, Thijs and Schlemmer (2004) who investigated articles both from hard and soft sciences represented by 17 distinct disciplines. Their analysis has proved that the ratio of self-citations decreases from about 50 per cent in the year of publication to less than 20 per cent in the span of 10 years. This fact leads to the assumption that impact factor is highly influenced by self-citation as it is calculated "by dividing the number of current year citations to the source items published in that journal during the previous two years"[2].

Such high ratio of self-citation may seem to be in opposition to the widely accepted view that scientific writing has to be impersonal and overwhelmingly prefers passive forms to active sentences with the first person agent. Therefore the explicit self-mentions pose a great problem not only for students but also for experienced scholars. The convention of impersonality of scientific discourse is strong and supports the widely acknowledged presupposition that scientific research is empirical, objective and its results independent of human interaction. This is especially valid in the field of hard knowledge in which the avoidance of specific references to the writer and his personal viewpoints is downgraded to minimum. Soft disciplines have a different approach to knowledge-making and personal opinions play more significant role within this field. The focus is shifted from the scientific process to the author of the idea.

However, the academic discourse has been adopting a rather discrepant approach to self-mention recently. In fact, self-citations present a relatively high ratio of all citations used. Snyder (1998) examined patterns of self-citation in six disciplines comprising social science, humanities and physical science. In light of the above mentioned assumption about the impersonality of hard sciences, his results are surprising. He found out that in physical science, as much as 15 per cent of all citations were self-citations, in social science it was 6 per cent and in humanities, the self-citations accounted for only 3 per cent. Aksnes's (2003) analysis of Norwegian articles from hard sciences revealed that 36 per cent of all citations are represented by self-citations and that the number of self-citations increased in line with increasing number of all citations. Considering the individual disciplines, his study showed that the lowest ratio of self-citations, only 17 per cent is found in medicine whereas the highest is in astrophysics, 31 per cent. Biology, which is the second discipline representing hard sciences in this thesis, includes 23 per cent of self-citations. Kovačić and Mišak (2004) commented on self-citation in medical literature and pointed out that nearly one-fifth of all citations per year were author self-citations.

---

[2] accessed on 19th April 2012.

One of the reasons for self-citation is the need for the inclusion of relevant information. Especially in narrow-focused research, this usually includes the writer's previous work. According to Bonzi and Snyder (1991), it is often assumed that self-citation is used particularly when the current paper modifies, amplifies or corrects the previously reported results. With regard to highly specialized research topics it is assumed that a higher rate of self-citation may be caused by the fact that scholars tend to stick to one specific topic of research and continue to build upon their previous findings in their following research career.

Hyland (2003) explored the use of self-citation in a corpus of 240 RAs and 800 abstracts in eight disciplines and examined the patterns of self-mention, ranging from personal pronouns and determiners, both singular *(I, me, my)* and plural *(we, us, our),* to other forms of references to the authors *(this laboratory, the research group).* He found an imbalance between soft and hard sciences, because three quarters of all cases occurred in humanities and social science articles.

1.3.3 Construction of knowledge

As Hyland's (1999) analysis confirmed, the use of integral or non-integral forms also reflects the writer's decision to focus on either the author or the reported writing. The use of non-integral forms can also be required by the style of a particular journal and its convention to use endnote forms, though. The integral quotations or reporting tend to give "greater prominence to the cited author" (Hyland, 1999:346) and thus are, according to Hyland, more used in the humanities. He further claims that these disciplines employ a higher ratio of direct quotations and authors' names in the subject position than in the hard sciences.

This is closely connected to the different construction of knowledge used in the soft and hard disciplines. The knowledge-making in hard sciences is widely acknowledged to pursue in linear development, new findings being built upon the previous ones. This also entails that the individual findings are considered building blocks creating a wider picture of reality, helping to discover the truths lying in nature. Throughout this process, old claims are assimilated and, according to Hyland's interviews with expert members of scientific communities, readers are unlikely to come across quotations of such established names as Einstein or Oppenheimer since their findings belong to the assumed knowledge. Furthermore, the accomplishment of knowledge in hard sciences is perceived as consisting in correct application of appropriate scientific methods. The human factor plays only a minor role in this process. The researcher's role is only to deliver a description of phenomena existing in the reality independently on the writer. Hence, the researcher does not make any knowledge, only discovers, observes and describes the truth. As Hyland's analysis showed, hard

sciences consequently employ predominantly non-integral structures (see e.g. Bazerman, 2000; Hyland, 1999).

In contrast, humanities do not usually progress in any predictable direction. Writers often retrace the steps of other writers, moving in circles, reinterpret or revisit previous research adopting different viewpoints. Moreover, the topics examined by the humanities typically range over a broader academic field, using findings from other disciplines as background for the current research. Thus, roughly speaking, the writers have to cite more because they cannot be sure that their audience is familiar with the research done in the other fields (see e.g. Hyland, 1999).

Moreover, the fact that the writer chooses the reporting citation also comprises the choice of the reporting verb itself, its tense, voice, aspect and thematic position. Thus, these choices which the writer has to make carry both rhetorical and social meaning (Hyland, 1999). Swales (1990) sees this complex strategy as a powerful means of creating research space allowing the writer to adopt and manifest his or her position on the quoted or reported propositional content. For example, he explains the choice of a tense as related to abstract distance the writer wishes to make from the reported findings. Several studies were devoted to the usage of reporting verbs, for example Hunston (1993), Thompson and Ye (1991), or Thompson (1996).

1.3.4 Reporting verbs

Apart from being almost a necessary part of RA and serving all the functions named above, citations allow writers to adopt explicit personal stance towards the cited information. In other words, the choice of reporting structure helps writers to express their position in relation to the cited proposition they support or oppose. Apart from constructing a stance to the cited authors or propositions, citations also serve to construct a writer's self and, moreover, by naming scholars whose work he or she relates to, the writer establishes himself or herself as a professional member of a particular academic community. Moreover, citations provide the opportunity to promote the work of one's colleagues.

The cited propositional content is always reported with the interference of the writer, albeit minimal. For example, Shaw's (1992) study examines verb tenses. The results suggest that reporting verbs in past tense and active voice are highly suitable for providing detailed description, whereas perfect verbs in passive voice are the most appropriate to initiate new topics. Present perfect has also been found to serve as a signal that the writer is going to discuss the particular topic further. (Swales, 1999). The relation between temporal reference and the possible manipulation with the generality of the cited proposition is further acknowledged by Malcolm (1987). She has proposed

that generalisation is implied by the use of present tense, whereas specific experiments are more likely to be described by past tense and areas of inquiry are predominantly referred to in present perfect. Swales (1990) points out that 90 per cent of finite reporting claims are situated on the simple-present perfect-past scale, thus showing the distance the writer wishes to keep between himself and the reported message. This view is supported also by Hyland (2004) who approves Tadros's (1993) assumption that the choice of verb raises a possibility for writers to detach themselves from the information and, at the same time, this detachment predicts their declaration of their own point of view. Hunston (1993) claims that verb selection is crucial for ascribing presented information as accepted knowledge.

Much attention has been paid to different types of reporting verbs and their possible classifications. Verbs can be classified from several points of view. Considering the semantic domains of verbs proposed in *Longman Grammar of Spoken and Written English* (1999), most of reporting verbs fall into the categories of communication and mental verbs. The former group of verbs involve communication activities, the most common verbs are, for example, *discuss, explain, state, write, say, suggest,* whereas the latter group denotes cognitive processes, attitudes or desires, and perception. This category comprises both verbs denoting relatively dynamic cognitive activities as *assume, examine, find, decide, discover, study, consider*, and more stative mental verbs denoting cognitive states, for example *expect, think, believe, doubt, know, understand, remember.* As shown in the corpus research provided in *Longman Grammar*, the highest number of verbs fall into the category of activity verbs, however, only a handful of these can serve as reporting verbs, namely: *show, give* (when followed by a suitable noun to be regarded as reporting verb, for example *to give an explanation, give an example...*), *hold, provide, obtain* and *add.*

1.3.4.1 Denotative potential of reporting verbs

The majority of scholars researching this issue draw on the influential classification proposed by Thompson and Ye (1991). They distinguish two types of categorisation of reporting verbs: denotative and evaluative. Considering the denotative point of view, reporting verbs are classified on the basis of the type of the activity referred to into either cognitive or research act verbs. Thompson and Ye distinguish three categories of verbs, later adopted by Hyland (1999): textual verbs referring to processes based on verbal expressions (e.g. *state, write, point out, deny*), called by Hyland "discourse acts verbs"; mental verbs referring to mental processes (e.g. *believe, think, focus on, consider*), in Hyland's classification "cognitive acts verbs"; and research verbs referring to the mental or physical processes included in the research work, i.e. Hyland's "research acts verbs". Hyland also further distinguishes within the research category verbs occurring in statements of

findings (e.g. *discover, notice*) and those occurring in the statements of procedures (e.g. *analyse, explore*). Obviously, different disciplines prefer different groups of verbs according to their epistemological orientation. Thus, as observed by Hyland, for example hard sciences display an extensive use of research related verbs referring to procedures *(analyse, observe, explore, develop).* This observation fully matches the viewpoint of hard sciences which sees laboratory as the central place for generating knowledge, the place where all essential activities are undertaken. With respect to the relative complexity of his categorisation, his schema for both denotative and evaluative classification of verbs is provided in the next section about evaluation in Figure 2.

Another classification is provided in *Collins Cobuild Grammar Patterns* (1996), although it shares some similarities with Thompson and Ye's classification. Reporting verbs fall into four basic categories: the "ask" group of verbs concerned with different means of communication (*acknowledge, agree, argue, confirm discuss, emphasise, explain, remark, report, suggest, stress*); the "think" group of verbs concerned with mental processes (*believe, consider, determine, doubt, know, remember, underestimate, speculate*); the "discover" group of verbs concerned with coming to know (*analyse, calculate, deduce, establish, find out, pinpoint, prove, verify*) and the "show" group of verbs concerned with presenting an existing situation *(confirm, demonstrate, illustrate, indicate, reveal).* Obviously, the "ask" group corresponds to the textual verbs, the "think" group to the mental verbs, and the "discover" and "show" groups to research verbs.

As Charles (2006) points out, each of these groups predominantly focuses on different entities (the cited text itself, the author, the research process, etc.), which carries several implications. For example, in accordance with its focus on the cited text, the "ask" verbs tend to be used with present tense, whereas "discover" and "show" verbs refer to the conducted research and tend to be used with past tense. This is closely connected to the fact that different disciplines tend to employ different verbs. Thus, it is obvious that "ask" verbs are more likely to occur in social sciences whereas "discover" and "show" verbs in experimental hard disciplines.

1.3.4.2 Evaluative potential in reporting verbs

If writers decide to cite other authors, they need to clarify why they decided to do so and how the cited text relates to their research. This process implies adding evaluation of the cited author. Thus, it is obvious that the writer has to take several steps of choices beginning with what information to cite and ending with how to present it. This issue has been thoroughly investigated by Thompson and Ye (1991). As they pointed out, evaluation is a complex phenomenon. Firstly, it does not have to work only within the grammatical structure of a single sentence, but it may as well stretch across longer parts of texts. Secondly, the signals of evaluation also may be cumulative rather than occur at

one point, and thirdly, evaluation is strongly context dependent. Once the writer decides to adopt a stance towards the cited proposition, he or she thus ceases to be obliged to agree with its content and gains a space for arguing for or against this proposition. Although it has been said that evaluation is complex and may arise from the longer stretches of texts, this analysis focuses on the lexical choices constructing the clearest signals of evaluation, reporting verbs and evaluative adverbs used in integral citations.

For a careful investigation of evaluation, it is necessary to distinguish three basic but "largely separate factors: author's stance, writer's stance and writer's interpretation" (Thompson and Ye, 1991:372). Author's stance, that is his or her attitude towards the propositional content cited, can be presented as positive (true, correct) by using verbs like *accept, emphasize, note, point out*; negative (false, incorrect) using verbs like *attack, challenge, oppose, question*; or neutral using verbs like *examine, evaluate, focus on, quote, undertake.* Writer's stance distinguishes clearly the attitude of the writer to the proposition (Hyland, 2004). The information can be evaluated as either true or false, only by choosing particular reporting verb. For example, verbs as *acknowledge* or *point out* signal the writer's credit given to the cited information, whereas *fail, ignore* or *exaggerate* signal his or her disbelief in it. These types of representation are called factive or counter-factive, respectively. The third possibility is to express no clear attitude towards the presented information, using verbs like *believe, claim, generalise, propose*, and is classified as non-factive. At this point, it is important to mention that the counter-factive option is extremely rare in hard sciences and rarely chosen even in soft disciplines.

The third factor, writer's interpretation, is more diverse and heterogeneous. Thompson and Ye (1991) suggested four possible options which can be adopted by the writer. The first option is "author's discourse interpretation" which allows the writer to present how the cited content fits into the original text (e.g. *add, continue, mention, note, repeat*). The second possibility is "author's behaviour interpretation" in which the writer interprets the attitude of the author to the information (mostly verbs referring to textual processes (e.g. *admit, advocate, criticize, emphasize, insist*). Thirdly, a "status interpretation" serves to indicate what status will the reported information have within the present research (e.g. *account for, confirm, establish, prove*). The last option is "non-interpretation", which means that the writer presents the cited text objectively (e.g. *adopt, apply, observe, provide, see*). This category comprises a high ratio of research act verbs because evaluating someone else's research is more face-threatening than evaluating his or her text (Thompson and Ye, 1991).

Denotation

Research Acts   Cognition Acts   Discourse Acts

Findings   Procedures

Evaluation

Factive
(Writer acceptance)   Non-Factive
(No clear signal)   Counter-Factive
(Writer disagreement)

author positive   author neutral   author tentative   author critical
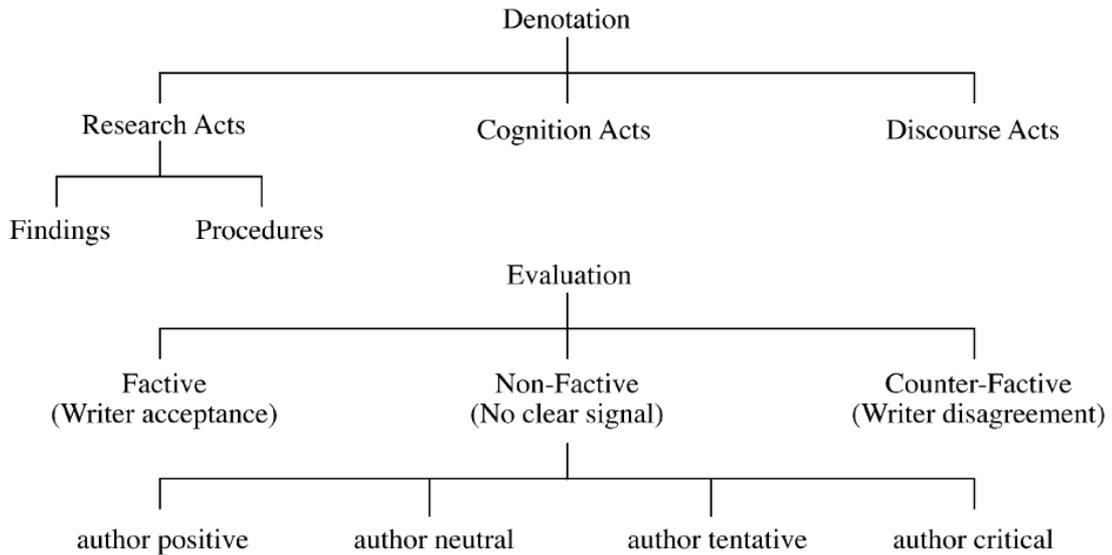
Figure 2: *Categories of reporting verbs (Hyland, 1999:350)*

Hyland (1999) adapts this taxonomy, simplifying the evaluative approach by leaving out the writer's interpretation. Instead, he has further elaborated the non-factive evaluation and distinguishes among verbs used within it. In the next step, a view to the source author is ascribed, he or she can be reported as positive (using verbs like *advocate, argue, hold, see*), neutral *(address, cite, comment...),* tentative (*allude to, believe, hypothesise, suggest*) or critical *(attack, condemn, object...).* His categorisation is shown in Figure 2.

Although the choice and the usage of attitudinal verbs has been in the centre of the linguistics interest, there are also other means of expressing evaluation. The writer can explicitly present his stance by using an evaluating adverbial which modifies the whole reporting structure. As Hunston (1993) points out, non-empirical articles tend to modify even non-interpretative verbs which obviously allow no modification in empirical articles as they express results of experiments and cannot be considered as correct, or incorrect. For example, writers in soft disciplines can reveal their attitude towards the propositional content by introducing it with a phrase *persuasively demonstrate*. Such differences prove that the meaning of some verbs may change according to their rhetorical functions in different disciplinary contexts (Tucker, 2003). As was already explained in section 1.3.3, the empirical disciplines suppose that correct observations of the real world inevitably results in correct findings. Therefore the only part of the research which can be in fact evaluated is the decision and planning of the experiment procedure employed. In contrast, in soft argumentative disciplines it is possible to evaluate the whole research because human intervention occurs at all its stages.

class	subject	attitude	examples
1 - Verbs of reporting	human	no information given	find, observe, report
2 – Verbs of subjective interpreting	human	information about the attitude of the subject of the verb	conclude, believe, suspect, interpret, suggest, characterise
3 – Verbs of objective interpreting	non-human	information about the attitude of the writer	show, indicate, demonstrate, establish, characterise, suggest
4 – Verbs of arguing	human	information about the attitude of the writer	note, point out
5 – Verbs of arguing	human	a potential difference of attitude between writer and subject of verb implied	claim, argue, admit, deny, insist, acknowledge, concur, declare
6 – Verbs of cognition	human	when gramatically positive – information about the attitude of the writer.	realise, be aware, understand
		when grammatically negative – a difference of attitude between writer and subject of verb implied	
7 – Verbs of reaction		a potential difference of attitude between writer and original researcher implied	find significant, be struck/impressed by, convince, persuade

Table 5: Classes of reporting verbs (Hunston, 1993:124)

The differences between the usage of attitudinal verbs in hard and soft sciences have also been examined by Hunston (1993). For this purpose, she has established her own classification of verbs, also partially derived from Thompson and Ye. However, she has adapted their framework to better suit her aims. Hunston examined the evaluation of propositions on the basis of discourse analysis of RAs from the disciplines of biochemistry, linguistics and history. The main focus of her research is on negotiation of controversial topics and conflicting claims of individual researchers. Her classification is thus particularly aimed at describing and distinguishing different attitudes presented in the texts and takes into account the potential of a reporting verb to imply also conflicting attitudes that may arise between the writer and the author of the cited proposition. Hunston's classification is provided in Table 5. Whereas verbs from classes 1-4 "simply imply a single judgement regarding the certainty of the following proposition, verbs from classes 5-7 […] involve a great deal more complex interaction of judgements" (Hunston, 1993:125). This presupposition leads to different usage of these verbs by soft and hard sciences. Classes 1-4 are preferred by more empirical articles and more reporting discourses, whilst the others are employed by more

argumentative and less empirical disciplines.

These verb classes have been adopted by other scholars, among others by Tucker (2003) who has examined evaluative strategies in the art-historical research articles. His findings are presented below in the next section.

1.3.4.3 Evaluating verbs in hard and soft disciplines

The denotative categories of reporting verbs reflect the traditional distinction between hard and soft sciences and their view of the world and construction of knowledge. Hyland (1999) himself pointed to some striking differences. For example, counter-factive structures have been found only in humanities and social sciences. Humanities further tend to use integral structures which allow to place the author in the subject position, are more likely to use direct quotes than hard disciplines and attribute a stance towards the cited proposition. In contrast, hard sciences employ a relatively high ratio of non-integral forms giving little attention to the authors of the cited contributions. By suppressing the agent, the human intervention or personal interests are also suppressed. Knowledge is thus represented as a result of the laboratory research, whereas soft domains see knowledge more likely as a result of the interpretive operations and verbal arguments (Hyland 1999). Regardless of these differences mentioned by Hyland, the prominent tendency among all disciplines is to indicate the position the writer has adopted to the cited information more indirectly. Soft sciences, for example, are more likely to evaluate the proposition by adding an adverbial component (e.g. *X argues here, correctly, that...*).

Tucker (2003), building on Hunston's categories of verbs, suggests that the discourse of art history differs to a great extent from other disciplines because, in contrast to other fields, in which "evaluation is always extrinsic to the construction of knowledge as such," in art history, evaluative language "may also participate directly in the construction of ideational meaning" (Tucker, 2003, 292). Thus, the propositional content itself has an essential evaluative function. Furthermore, previous interpretations referred to by the writer of the current article build an integral part of his or her argument and are corrected, doubted, impeached or supported. Prior arguments and interpretations thus often become the subject of the current analysis. According to Tucker, hard disciplines more likely regard the previous research as the scientific pretext of the current work.

Moreover, the discipline of art history works with a "broader range of human subjects" (Tucker, 2003:303), the text distinguishes among the cited author, the citing writer, arguments presented by artists themselves, arguments presented by other historical Figures, social or religious groups of people or schools of thought and, of course, also arguments presented by writers in other fields of

humanities (philosophy, history, anthropology and other related disciplines). Such nature of art history determines its prevalent use of verbs of subjective interpretation, as well as its preference of evaluative *as*-phrases accompanying mental or verbal process verbs.

2. Methods

2.1 Corpus analysis

Since its development in the 1960s, corpus analysis enables linguists to study language used in real life. Recently, corpus linguistics has been of increasing interest, one of its major advantages being its empirical approach allowing the researcher to discover phraseological patterns repeatedly occurring in natural languages. As pointed out by Flowerdew (2004), this view that some words tend to occur in the near context of each other inspired the main objections against corpus linguistics that have "emanated from the Chomskyan camp". However, in Flowerdew's point of view, these two approaches do not necessarily oppose each other as the corpus linguistics is concerned with what was actually said and the rationalist Chomskyan linguistics with what can be said. (Flowerdew, 2004:13).

Furthermore, corpus analysis has been criticized also for its incompleteness, for its partial and limited view on language. As observed by Hyland (2009), sociologists insist on the necessity of providing the context. Seeing the discourse as a process and writing as a social act, they raise the objection that corpus analysis describes texts as final products only, regardless of the social context of their creation. For example, Widdowson points out the importance of contextual connections and realities, shared perception and knowledge which make natural language meaningful. As he sums up, "reality does not travel with the text" (Widdowson, 1998:711). At this point, it is necessary to point out that this thesis is focused on a single academic genre which has been already described in large detail. Moreover, a special corpus built for this purpose only will be used, therefore the risk connected with the unawareness of context is reduced to minimum.

2.2 Creating a specialized corpus

This analysis uses corpus research as a tool for obtaining data and it is believed that no further justification of the choice of this method is needed. However, it is important to explain why it was necessary to create a special corpus instead of making use of one of the general corpora available. As Flowerdew puts it, "as general corpora have proved to be extremely useful for understanding how naturally-occurring language operates, then by the same token, specialized corpora can also prove to be of value in understanding academic and professional language." (Flowerdew, 2004:11) Under the term general corpus, Flowerdew and others understand such corpora which consist of

various texts, taken from both spoken and written language. The best known examples are the 100-million words BNC or the 425-million words COCA. Such large corpora were created to represent language as a whole and as universally as possible. They pose a bundle of texts "carefully balanced among the different types of texts for reception and production to reflect their importance in the culture, which means that there will be a limited representation of some genres" (Flowerdew, 2004:14). Moreover, many general corpora work on the basis of scanning texts samples only. For the purpose of this analysis, access to full texts is needed as thus a larger variety of data can be obtained. In addition to searching and counting of particular lexical items, the use of full texts allows for localizing specific items or lexical bundles in the text and further investigation of their meaning and function with respect to their context, position within the text and the overall organization of the article.

This issue is closely connected to Oakey's (2009) notion of isotextual version of corpus. Oakey examined fixed collocational patterns in eight academic disciplines comparing results from the isolexical and isotextual corpora. The former uses text samples of "an equal number of tokens in each discipline", whilst the latter approach works with "equal number of texts in each discipline" (Oakey, 2009: 140). As he found out, the isotextual type of research is better suited for comparative studies of patterns occurring in texts. These patterns serve different communicative functions and acts, thus it is necessary to keep the corpus balanced in terms of these communicative acts, not for the amount of language. As noted earlier, the highest occurrence of citations is expected to be in the Introduction and Discussion sections. Therefore each discipline should be represented by the same number of these sections, otherwise the analysis might "bias any quantitative results in favour of the discipline with the largest number of texts" (Oakey, 2009:150).

This requirement is another argument for creating a specialized corpus and not using the large corpora with text samples only, because the text samples show a particular part of texts only. Thus, the isotextual type of corpus has been created for this analysis, each discipline being represented by two articles, regardless of their word count. Thus, the total amount of language examined differs considerably, ranging from 133.065 words in linguistics and 101.010 in astronomy, to 85.238 in art history and only 53.847 in biology (word counts of the second sub-corpus, as explained below). The word counts for individual articles as well as the list of all articles are provided in Appendix C. The compilation of the corpus is described below in Section 2.2.1.

Moreover, the need of creating a specialized corpus also bears relation to Widdowson's objections levelled against corpus analysis which have been mentioned in the previous section. His requirement for the awareness of the context is thus met. By choosing the texts to include to the

corpus, the researchers ensure that they work only with texts they are familiar with. Or, to be more precise, they are familiar with the broader context of the texts, with the community these texts are aimed at, and with other social aspects determining the shape of the texts. In this respect, Flowerdew quotes Aston, saying: "It is much easier to *interpret* concordances or numerical data if you know exactly what texts a corpus consists of, since this allows a greater degree of top-down processing" (Aston, 2002, quoted in: Flowerdew, 2004:16).

2.2.1 Compilation of the specialized corpus

For all the above mentioned reasons, a special corpus was compiled to provide data suitable for this analysis. All the texts included in the corpus belong to the genre of academic research articles. The size of the corpus results from the length of the articles. Considering the frequency of the investigated phenomena, there was no need to create a large corpus. In total, seventy-two articles have been investigated, eighteen from each discipline. This corpus was divided into two sub-corpora, each consisting of nine articles from each discipline. Lists of articles from both sub-corpora are provided in Appendices B and C. The first sub-corpus was gone through manually within the preliminary step, whilst the second was converted to electronic text files and analysed using the AntConc programme. The texts have been downloaded from the academic journals accessible from Charles University Electronic Information Resources Portal. This choice was limited by the requirements concerning the origin of the journal, genre of the texts and date of publication.

Although all the articles are written in English, their authors are both native and non-native speakers of English. It was not found necessary to distinguish between native and non-native authors, furthermore, in some cases it was not possible to figure out what nationality the author was. In addition, some articles were written by two or more authors, one of them being native and the other non-native. Therefore, the focus was shifted to the institutions with which they cooperate. Internet search was conducted to find the home institution for each author, e.g. their home university or research society. On account of the above mentioned problems in identifying the origin of the authors, only one requirement considering this issue was set. Regardless of the nationality of individual writers, only journals published by institutions established and settled in the UK, USA or Australia have been used, or, journals the editors of which are British, American or Australian. For example *Journal of Pragmatics* is published by Elsevier branch based in Amsterdam, however, this publishing company is acknowledged by scholars all over the world and, moreover, its editor-in-chief is British. RAs have been selected more or less randomly from

particular issues of journals that suited these requirements.

The basic information about the selected journals form the second sub/corpus can be seen in Appendix A, which provides a short description of each journal, the name and/or nationality of the journal's editor, 5-year Impact factor[3], and information about the frequency of publication and the publisher. These pieces of information have been considered useful with the exception of the impact factor, which has been used only as one of the means of assuring that the quality of the particular journal was acknowledged by a wider academic community. However, this measure is applicable only within a single discipline, the impact factors cannot be compared across the fields. In consequence, no minimum threshold for accepting the journal has been set. This process was not applicable to art history as this is one of the soft disciplines for which the impact factor is not usually counted.

The rest of the information was considered useful for later examination of possible similarities and differences in reporting structures not only between hard and soft sciences but also within these two spheres and, presumably, also among the sub-disciplines of particular academic fields. It is obvious that the individual journals belonging to a single discipline, in fact differ in the extent of their academic focus. For example, the linguistic journals can be further divided into sub-fields such as cognitive linguistics, computational linguistics, psycholinguistics, sociolinguistics, pragmatics or general and applied linguistics. The art history journals cover aesthetics, art criticism, history of decorative arts, history of the classical art as well as contemporary art and new media and a feminist approach to art. The astronomical journals are focused on astronomy, astrophysics, cosmology, radiology, space research, astrobiology, radio astronomy, or positional and dynamical astronomy. The journals that belong to the field of biology are concerned with petrology, immunology, integrative biology, plant biology, behavioural biology and systematic biology. The journals were selected to cover as large number of sub-disciplines as possible with the intention to obtain general results.

According to the range specification explicitly drafted in the very beginning of this thesis, all articles examined have to be published between the years 2000 and 2008. With regard to the date of publication, the selected articles were equally distributed among these years. Each year is represented by the same number of articles. Considering the type of the articles, all fall into the genre of research article which has been described in section 1.2.4 at some length.

---

[3] Found on the ISI web of Knowledge, Journal Citation Reports, accessed through Charles University Electronic Information Resources Portal, , accessed: 27 April 2012.

2.3 Analysis of the corpus

2.3.1 Preliminary Step

In the first step of this analysis, nine articles were selected from each discipline, each of them representing a different year of publication ranging from 2000 to 2008. A list of articles included in the first sub-corpus is provided in Appendix B. In order to examine the issue of reporting structures properly, this sub-corpus of 36 articles has been gone through manually and all spotted instances of integral citations have been recorded in a table. The outcome of the preliminary step is a list of reporting verbs and nouns used in the four examined disciplines. These verbs and nouns were than used as search terms in the second step of this analysis, in the corpus search itself.

Among the references, two disputable features have been included. Although Hyland (1999) claims that self-citations differ considerably from other-citations not only in respect of motivation, and are less central to academic argument, self-citations have been after careful consideration included in this analysis. However, all instances have been clearly marked in the preliminary step.

The second problematic type of structure which was included in the analysis is noun phrases with the citation functioning as a pre-modifier, such instances as "Mendelian framework" or "Jaccard's coefficient". Hyland suggests that such structures are often accompanied by evaluative stance in humanities (providing examples from philosophy). In sciences, these structures are less common and, furthermore are used for different purposes, serving as shorthand references. In this analysis, no such structures were found in the articles from art history and only few instances occurred in linguistics. These can be considered well established concepts that need no further clarification, for example: "the Chomsky and Halle model", "Francis and Kučera past-tense-form frequency," which may not be established but was explained earlier in the text. On the other hand, far more examples have been found in the hard sciences. Majority of such structures in biology are names of tests, coefficients, functions, indexes, cycles, etc. To provide at least a few examples: "using Kruskall-Wallis test", or "the fractal index of Katz and George (1985)". In astronomy, the noun and possessive form structures are used for telescopes and other tools, profiles, lines, galaxies, numbers, forces, etc. and more than 40 instances have been found. For example: "the Lorentz force", or "the Riemann zeta function".

2.3.2 Corpus search

Following the preliminary step, the corpus research itself was conducted. Nine articles from each discipline were selected, different from those from the preliminary step, each of them representing

one year of publication. A list of the second sub-corpus is provided in Appendix C. Thus, the following part of Methods as well as the Results section of this thesis will describe only the data obtained from the second sub-corpus.

The articles were converted from .pdf to .txt files suitable for AntConc concordancer. This conversion included also several changes that had to be made in order to avoid bias in counting found instances. At first, the headers and footers including publication information were erased, as well as the Acknowledgements section, References (with the exception of those including endnotes) and Bibliography. These five parts include names and years and therefore have been considered risky for being interpreted as citations in the course of corpus search. Majority of the examined texts include charts, graphs, tables or pictures. All such non-textual elements have been erased, however, their descriptions have remained as some of them include citations. These textual units have been relocated in order to avoid breaking the surrounding paragraphs.

In cases of articles with footnotes, these have been moved to the end of the main text because in the converted .txt files some instances of footnotes broke apart the adjacent sentences. In contrast with References, both endnotes and footnotes have been left in the text file and thus also included to the word count. Formally, the .txt files with endnotes/footnotes and without them have been treated in the same way: when examining the density and distribution of citations, the positions of citations in the texts were recorded. In case of texts with endnotes/footnotes, the positions of citations correspond with the place where citations were inserted in the body of the text. In art history, endnotes usually comprise larger stretches of text, not only reference to the author and year of publication, however, such instances have been counted as a single citation being linked to by a single superscript number.

2.3.2.1 AntConc programme

Text files adapted in this way were loaded into AntConc, one of corpus analysis software available. Several types of searches were conducted. At this phase, all work was done using the concordancer, which is the central tool of the application.

At first, search terms were the reporting verbs found in the preliminary step. Strikingly, almost all verbs found in the first set of articles examined in the preliminary step were also found in the second set. This finding is important as it confirms that the verbs found were not used only exceptionally. AntConc programme enables the search term to be defined as a regular expression so it was possible to search for all conjugations at the same time, as well as for non-finite verbs and nouns derived from the reporting verbs in question. Thus, instances of integral citation have been

acquired. Next, the search term, also defined as a regular expression, was a four-digit number either preceded or preceded and followed by text, enclosed in parenthesis. This step has shown results for both integral and non-integral instances of citation. The last search conducted made use of the Word List tool. The generated word-list was searched through for names, thus covering the cases of non-citations.

The results obtained in this three-step search have been recorded in a table, each instance occupying one row. These occurrences have been sorted according to several aspects composing the individual columns of this table. The bipolar aspects have been: presence of a reporting verb, human vs. non-human agent, nominalized verb (e.g. *suggestion, claim*), name of the author, presence of the biographical data (date of publication, full name of the study, etc.), further information about the author, direct quotation, summary/paraphrase, further explanation of the quote/summary, evaluation (e.g. *correctly claim, its failures...*), passive vs. active voice, indexed reference pointing to footnotes or end-notes, *that* construction, further context provided. The data for the analysis of citation types (non-integral, integral, verb-controlling, naming, non-citation) and often used reporting verbs have thus been collected.

The third concern of this thesis is also the distribution of citations. For this step, the View File and Concordancer Search Term Plot tools were used. Here, all instances found in the previous searches had to be put together and found at once. Mostly, this did not pose any problem as vast majority of citations found included the year of publication. As a consequence, all instances of occurrence of a four-digit number ranging from 1400 to 2008 have been searched for. Number 1400 was chosen with regard to art historical articles focused on the art of old masters, number 2008 is, obviously, the last year available for possible citing. This search term was completed manually with examples of non-citation (citations without any year reference) found in the previous search. This process was successful for biology, linguistics and astronomy, with the exception of one article from *International Immunology* (2008).

This article contains endnotes and together with all articles from art history demanded a different approach. As already discussed, references in art history are accumulated in endnotes/footnotes. Thus, the only procedure which could allocate all instances of citations was to find all superscript numbers referring to endnotes/footnotes. This was, for a large part, done manually because the numbers do not always have the same format, they are usually following a comma, full stop, quotation marks or parentheses but this cannot be relied on.

AntConc produces visualisations of the relative position where the search terms can be found. (See concordance plots in Appendices D and E.) Each article has its own box, multiple lines

representing individual searched instances. Obviously, the boxes are equally big, regardless of the length of the articles so at first sight, the density of lines can be misleading. Longer articles seem to include citations accumulated much more densely, shorter articles allow the lines to spread more loosely. Therefore it has been considered useful to provide also the information about the word count for each article.

3. Results

In accordance with the process described in the Methods section, all data referred to from this point on have been obtained in the second step of the analysis, i.e. from the second sub-corpus consisting of thirty-six articles only. The numbers and types of reporting verbs and nouns found in the articles examined in the preliminary step will not be taken into consideration here.

3.1 Quantitative results

As already mentioned, the present thesis focuses on integral citations, which will be distinguished into categories proposed by Thompson and Tribble (2001) and further described. (For the categorization, see section 1.3.2.1.) The non-integral citations, however, contribute to the overall picture of referencing in the individual disciplines, and were therefore counted and their numbers are recorded – together with integral citations – in Tables 6–9 below. We distinguished also between other- and self-citations, since although instances of self-citation have been found in all four analysed disciplines, their number and type in individual academic fields differ to a great extent. The numbers of other citations are presented in Tables 6-9, those of self-citations in Tables 11, 13, 16, 18. Only references made to the authors' previous published research articles, and only such instances which include the names of the authors or titles of their works were counted. Thus, all occurrences of reference with the agent *I* or *we* or noun phrases with possessive pronouns *my* or *our* were excluded.

Journal,year	PE, 2000	BE, 2001	MBE, 2002	BE, 2003	SB, 2004	SB, 2005	DNA, 2006	ICB, 2007	IIM, 2008	total
Nr. of other-citations	**87**	**64**	**94**	**56**	**88**	**249**	**50**	**43**	**48**	**779**
Non-integral	77	50	87	47	82	235	49	32	48	707
Integral	10	14	7	9	6	14	1	11	0	72

***Table 6**: Type of other-citations in biology*

Journal, year	MN, 2000	AR, 2001	ABI, 2002	APH, 2003	MN 2004	MN, 2005	EA, 2006	AJ, 2007	AJS, 2008	total
Nr. of other-citations	**22**	**538**	**36**	**121**	**83**	**146**	**16**	**111**	**122**	**1195**
Non-integral	15	462	35	66	51	87	16	90	93	915
Integral	7	76	1	55	32	59	0	21	29	280

***Table 7**: Type of other-citations in astronomy*

At this step, no difference was made between citations which occur within the body of the text, and those in footnotes or endnotes. In section 3.3, the distribution of both other- and self-citations was captured using the function of concordance plot in AntConc software. The third section 3.4

discusses particular reporting verbs found.

Journal, year	JFD, 2000	CL, 2001	AL, 2002	JML, 2003	JP, 2004	JP, 2005	ESP, 2006	BL, 2007	JFD, 2008	total
Nr. of other-citations	**45**	**81**	**47**	**61**	**56**	**62**	**64**	**72**	**258**	**746**
Non-integral	20	9	31	54	19	40	32	65	213	483
Integral	25	72	16	7	37	22	32	7	45	263

Table 8: Type of other-citations in linguistics

Journal, year	Bur, 2000	LEO, 2001	MD, 2002	JWC, 2003	AB, 2004	Bur, 2005	JWC, 2006	WA, 2007	OA, 2008	total
Nr. of other-citations	**31**	**8**	**50**	**183**	**76**	**30**	**93**	**37**	**93**	**601**
Non-integral	17	1	19	66	42	25	71	18	50	309
Integral	14	7	31	117	34	5	22	19	43	292

Table 9: Type of other citations in art history

As can be seen from Figure 3, in all disciplines, the predominant type of citation is non-integral. Nevertheless, there appears to be a difference among hard and soft disciplines with respect to the mutual ratio of integral and non-integral citations. Indisputably, the prevalence of non-integral citations is most evident in in biology (see Table 6) with 707 instances of non-integral and only 72 instances of integral other-citation. In astronomy, the former type occurs in 915 cases and the latter in 280 (as is shown in Table 7). Both fields are considered hard scientific disciplines. The non-integral type of citations outnumbers the integral even in linguistics and art history; however, the prevalence of the non-integral type is not so marked in the soft disciplines. In linguistics, 263 instances of non-integral and 132 instances of integral citation have been found (see Table 8). In art history, the contrast is the smallest, non-integral citations occur in 309 cases, integral in 292 (see
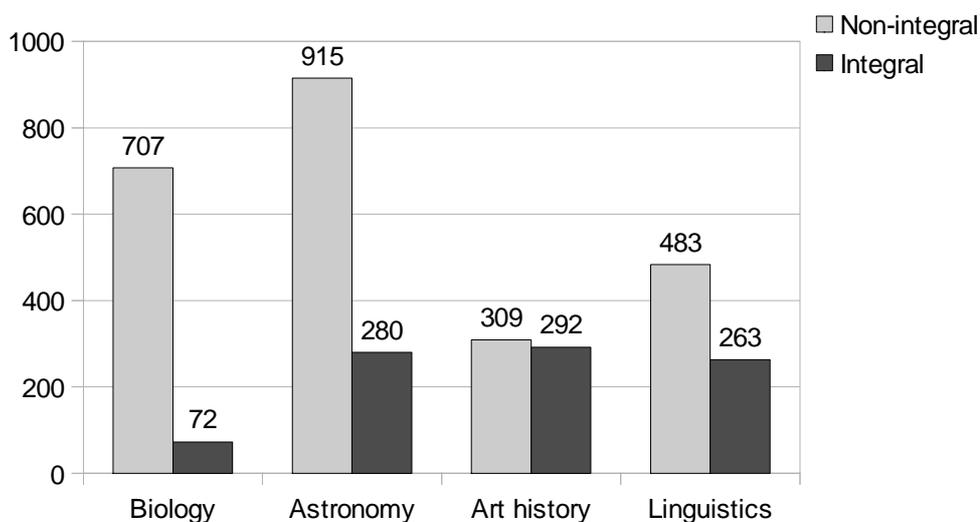
Figure 3: The representation of integral and non-integral citationsin the individual disciplines

Table 9). This discipline is specific in that all articles use the system of endnotes. Thus, in sentences which did not include any explicit reference, the superscript number pointing to endnotes, was treated as a non-integral citation. In sentences with integral citation forms, the endnote numbers were ignored and the citation was considered integral.

3.2 Integral citations

3.2.1 Integral citations in Biology

3.2.1.1 Types of other-citations

Journal, year	PE, 2000	BE, 2001	MBE, 2002	BE, 2003	SB, 2004	SB, 2005	DNA, 2006	ICB, 2007	IIM, 2008	total
Integral citations	**10**	**14**	**7**	**9**	**6**	**14**	**1**	**11**	**0**	**72**
Verb-controlling	4	11	3	6	5	7	1	3	0	40
Naming	6	2	2	3	1	4	0	1	0	19
Non-citation	0	1	2	0	0	3	0	7	0	13

***Table 10**: Types of integral citation in biology*

Biology articles show preference for the verb-controlling type of citations (40 out of s72 integral citations), in which the citation acts as an agent of the verb. From the functional sentence perspective point of view this type of citation gives more prominence to the author of the cited proposition. In the active verb-controlling citations (ex. 1), the citation itself is the theme of the sentence whilst the cited findings constitute the rhematic part. In the passive construction (ex. 2), the citation is, in fact, the new information, the rheme of the whole sentence. Usually, the active verb-controlling sentences ale followed by *that-* clauses, the passive constructions have *by* agents.

(1) For instance, Visscher and Seeley (1982) calculated that the circle encompassing 95% of their bees' foraging sites had a radius of 6 km. (BE, 2001)

(2) ... the method was employed by Lockhart et al. (1994) (SB, 2004)

The naming type, in which the citation itself is a noun phrase or a part of it, signifying a text rather than its authors, was employed in nineteen cases in biology. Usually it refers to models, diagrams, approaches, findings, or Figures which have been published by particular authors (ex. 3). However, naming citations occur also in adjuncts (*according to X; following X*) or simply as references to particular studies, followed by *in, from, of* (exx 4-5).

(3) … with distance between predator and prey, as in Ydenberg and Dill's (1986) model... (BE, 2003)

(4) According to During (1979) this strategy fits to a habitat which is present... (PE, 2000)

(5) Drawn by adding a second cost to Figure 1b from Ydenberg and Dill (1986).. (BE, 2003)

The least used type is non-citation, references providing the names of the authors but not the years. This type has been found in thirteen instances. Some of the non-citations are employed merely to avoid repetitiveness, others, however, serve different purposes. Either the cited author is well known (ex. 6), so the writer does not need to provide any further specification, or the citation refers to an established method, concept or practice (ex. 7).

(6) Darwin described natural selection as a process that "is daily... " (ICB, 2007)

(7) Bayesian analyses were performed using MrBayes 3.0b4... (SB, 2005)

3.2.1.2 Self-citation in biology

Journal,year	PE, 2000	BE, 2001	MBE, 2002	BE, 2003	SB, 2004	SB, 2005	DNA, 2006	ICB, 2007	IIM, 2008	total
Total nr. of citations	100	66	97	78	97	254	51	49	71	863
Nr. of self-citations	**13**	**2**	**3**	**22**	**9**	**5**	**1**	**6**	**23**	**84**
Non-integral	13	2	3	22	9	5	1	2	23	80
Integral	0	0	0	0	0	0	0	4	0	4

Table 11: Types of self-citation in biology

Similarly to other-citation, self-citation in biology shows a strong preference for non-integral type. In fact, unquestionable integral citations do not occur here at all. In one of nine biology articles, *Integrative and Comparative biology* (2007), four instances of self-citations have been found, that can be classified as borderline cases between the integral and non-integral type. These references are not integrated into the sentence itself but at the same time they form an expanded version of a non-integral citation. All are in identical passive constructions: "...*reviewed by* X (year)" (ex. 8) which reduces the importance of this finding.

(8) Many of the molecules involved in the generation, specification, and migration of these cranial neural-crest cells have been identified (reviewed by Helms et al. 2005). (ICB, 2007)

The numbers of self-citations in individual journals also differ considerably. At this point it is important to mention that the number of self-citations is not reliably dependent on the number of cooperating authors, although, as has been clarified earlier, in case of multi-authored articles a reference to a work of only one of them is sufficient to be regarded as self-citation. For example, as many as eight authors cooperated on the article published in *DNA Research* (2006) in which only one single instance of self-citation occurs, on the other hand, five authors cooperated on the article in *International Immunology* (2008) with 23 instances of self-citation.

3.2.2 Integral citations in Astronomy

3.2.2.1 Types of other-citation in astronomy

Journal, year	MN, 2000	AR, 2001	ABI, 2002	APH, 2003	MN 2004	MN, 2005	EA, 2006	AJ, 2007	AJS, 2008	total
Integral citations	**7**	**76**	**1**	**55**	**32**	**59**	**0**	**21**	**29**	**280**
Verb-controlling	3	55	1	39	16	20	0	18	13	165
Naming	4	19	0	16	14	29	0	3	14	99
Non-citation	0	2	0	0	2	10	0	0	2	16

Table 12: *Types of integral citation in astronomy*

In comparison with biology and linguistics, astronomy employs less non-citations. As results in Table 12 show, the prevalent type is, as well as in the other two disciplines, verb-controlling. This type is used in 165 instances which is more than 50 per cent of all cases of integral citations. The naming type has been found in 99 cases and the non-citation only in 16. Although these types and their use do not differ from biology and linguistics, it is worth pointing out that seven of the occurrences of non-citations refer to a well-established empirical relation (ex. 15) and two refer to models (ex. 16). In these cases, the name of the cited author is a pre-modifier further defining the head of the noun phrase. Only seven non-citations thus have a human agent controlling a reporting verb. This tendency is similar to that in biology, in contrast to linguistics, in which these structures normally occur.

(15) To predict a Tully-Fisher relation for our model we need to assign a maximum... (MN, 2005)

(16) J. B. Hyde & M. Bernardi (2007, unpublished) fit De Vaucouleurs models to SDSS images (AJS, 2008)

3.2.2.2 Self-citation in astronomy

Table 13 shows the results for astronomy articles. In comparison with both linguistics and biology, astronomy uses self-citation more extensively. However, corpus research has revealed some extremes as the number of self-citations ranges from only one instance (in two journals) to as many as 74 examples found in *AR 2001*. One of the possible reasons for such discrepancies may be also the diversity in the length of individual articles or variant number of collaborating authors. It must be noted here that the article from *AJS 2008* was excluded from analysis of self-citations because it provides a list of 80 collaborating authors and most of the text describes individual parts of research conducted by individual authors. Therefore its need of employing self-citation is far stronger and this article was thus not considered relevant for this examination.

Journal, year	MN, 2000	AR, 2001	ABI, 2002	APH, 2003	MN, 2004	MN, 2005	EA, 2006	AJ, 2007	AJS, 2008	total
Total nr. of citations	23	612	61	154	96	202	17	136	122	1423
Nr. of self-citations	**1**	**74**	**25**	**33**	**13**	**56**	**1**	**25**	**-**	**228**
Non-integral	1	63	24	22	4	10	0	22	-	146
Integral	0	11	1	11	9	46	1	3	-	82

Table 13: *Types of self-citation in astronomy*

The prevalent type of self-citations in astronomy articles is non-integral, amounting to 63 per cent across all nine analysed articles. This difference is not so significant as in biology articles and, taking into consideration also the types of other-citation, it can be assumed that academic writing in the field of astronomy combines citation practices of soft and hard disciplines. Considering the particular types of integral citations, the sum of naming type seems to outnumber the verb-controlling type (see Table 14). However, this is caused by a high ratio of the naming type in one article only, in *MN 2005* in which 33 of 46 integral citations are naming. The other articles show lower numbers of this type and therefore the use of the naming type cannot be assumed a general tendency of astronomy. The least used type is, similarly to other three analysed disciplines, the non-citation type which is represented by two instances (out of 88 instances of integral self-citations).

3.2.3 Integral citations in Linguistics

3.2.3.1 Types of other-citation in linguistics

Journal, year	JFD, 2000	CL, 2001	AL, 2002	JML, 2003	JP, 2004	JP, 2005	ESP, 2006	BL, 2007	JFD, 2008	total
Integral citations	**25**	**72**	**16**	**7**	**37**	**22**	**32**	**7**	**45**	**263**
Verb controlling	11	18	9	4	26	9	18	1	36	132
Naming	6	20	3	1	4	12	9	2	9	66
Non-citation	8	34	4	2	7	1	5	4	0	65

Table 15: *Types of integral citation in linguistics*

Linguistics uses both integral and non-integral citations, the integral being divided similarly to biology, as is shown in Table 15. The prevalent type is a verb-controlling citation, with 132 occurrences creating roughly 50 per cent of all integral citations. Its form and function is the same as in biology, however, the range of reporting verbs is broader, as is discussed in section 3.4. The majority of verb-controlling type are *that-* clauses. Usually, sentences are complex and show lexical and informational density (ex. 9), in comparison to biology, in which only a few instances of such complexity have been found.

(9) This concept has since been extended by Doerner and Nix (1986), who argued that the total displacement should be separated into... (ESP, 2006)

Considering the naming and non-citation type, almost the same numbers have been counted, 66 and 65 instances, respectively. The naming type, in which the subject is the text / result of the authors' research, not the authors themselves (ex. 10) is also a kind of means of achieving lexical and information density. Moreover, the naming type also contributes to depersonalization of the text by making the study or its results the subject of the clause. The non-citation type is used in longer paragraphs focused on a single author or research, therefore the repetition of dates is not necessary (ex. 11).

(10) The comment on historical sound change by Slobin (1977) also reflects the view that "forms which are late to be acquired by... (JFD, 2000)

(11) Second, Hyland examines research articles, while my data relate... (ESP, 2006)

3.2.3.2 Self-citation in linguistics

Journal, year	JFD, 2000	CL, 2001	AL, 2002	JML, 2003	JP, 2004	JP, 2005	ESP, 2006	BL, 2007	JFD, 2008	total
Total nr. of citations	61	86	53	80	60	76	64	94	262	836
Nr. of self-citations	**16**	**5**	**6**	**19**	**4**	**14**	**0**	**22**	**4**	**90**
Non-integral self citati	7	1	4	6	3	3	0	21	2	47
Integral self-citations	9	4	2	13	1	11	0	1	2	43

Table 16: *Types of self-citation in linguistics*

In general, the formal types of self-citation do not differ substantially from the other-citation, nor do the reporting verbs used. In linguistic articles, the total number of self-citations is almost equally divided between the integral and non-integral type, with the integral one represented by 43 hits found across all nine linguistic articles, and the non-integral one by 47 hits. The ratios of integral versus non-integral types of citations for individual articles are shown in Table 16. One article from *English for Specific Purposes* (2006) does not contain any instance of self-citation, either integral, or non-integral.

Journal, year	JFD, 2000	CL, 2001	AL, 2002	JML, 2003	JP, 2004	JP, 2005	ESP, 2006	BL, 2007	JFD, 2008	total
Integral self-citations	**9**	**4**	**2**	**13**	**1**	**11**	**0**	**1**	**2**	**43**
Verb controlling	5	0	1	7	1	4	0	0	1	19
Naming	4	4	1	4	0	7	0	1	1	22
Non-citation	0	0	0	2	0	0	0	0	0	2

Table 17: *Types of integral self-citation in linguistics*

As Table 17 shows, the integral self-citations in linguistics are almost equally divided between the naming and the verb-controlling type. The naming citations (ex 12), slightly prevail with 22 instances. The second option employed in 19 cases is the verb-controlling type of integral citation

(ex 13).

> (12) In work reported in Goldsmith and Reutter (1998), we have explored the usefulness of the present system for determining the linking elements used in German compounds... (CL, 2001)

> (13) Howell and Au-Yeung (1995a) investigated these factors in children who stutter... (JFD, 2000)

The rarely employed type of integral citation is non-citation (ex 14). This finding is not surprising given the general tendency of linguistics to use non-integral citation. The only example of self non-citation that has been found in *Journal of Memory and Language* (2003) occurs in context in which the year reference was supplied earlier. The writer thus wants to avoid repetitiveness.

> (14) Malt et al. argued that accounting for naming patterns requires reference not only to the understanding of stimulus properties by individual speakers of a language... (JML, 2003)

3.2.4 Integral citations in Art history

3.2.4.1 Types of other-citation in art history

Citation practices in art history differ greatly from those in the other three disciplines. All references are provided in endnotes, it is therefore difficult to distinguish among the types of integral citation as all these instances are, in fact, non-integral. However, two types of citations in art historical articles have been distinguished. The first type is purely non-integral, which means that the sentence itself contains no element which would reveal that it is a citation, except for the superscript number (ex. 17). The second type has been counted as integral, as it provides clear signal that it is a citation. Either the author's name, or the name of the text, is present, (scarcely both), usually followed by a direct quotation or paraphrase of the original text (ex. 18).

> (17) Its relatively small size would suggest that it was originally a private commission, and it may have been one of those canvases destined for the *portego* (the central front-to-back hall) on the *piano nobile* of a Venetian *palazzo*.[3] (BUR, 2000)

> (18) In Doyle's words, "Eva was an Abstract Expressionist painter, and I've always been an Abstract Expressionist sculptor."[6] (WA, 2007)

However, the citation format differs to such extent that it has been considered irrelevant to distinguish verb-controlling, naming and non-citation types. Firstly, the verb-controlling type occurs only in a few instances because the date is always provided in the endnotes. Moreover, in these few examples, the information about date is always a clause element integrated into the clause structure

(ex. 19), no instance of date in parentheses was found. Secondly, the naming type is rare for the same reason. The non-citation type is thus the prevalent type, however, it is not the same non-citation as in the other disciplines because the date is in fact provided, only at a different part of the text.

(19) By the time Rimmer published his Art Anatomy in the late 1870s... (MA, 2002)

3.2.4.2 Self-citation in art history

Art history tends to use much fewer self-citations of any type. Furthermore, in accordance with the practices adopted for other-citation, all instances occur in endnotes and are regularly non-integral. Usually, these endnote references begin with the directive *"see"* (ex. 20).

Journal, year	Bur, 2000	LEO, 2001	MD, 2002	JWC, 2003	AB, 2004	Bur, 2005	JWC, 2006	WA, 2007	OA, 2008	total
Total nr. of citations	31	8	53	189	77	32	98	37	108	633
Nr. of self-citations	**0**	**0**	**3**	**6**	**1**	**0**	**5**	**0**	**15**	**32**
Non-integral	0	0	3	6	1	0	4	0	15	31
Integral	0	0	0	0	0	0	1	0	0	1

Table 18*: Types of self-citation in art history*

(20) For information on the Peace Tower and other US collective works ... see Francis Frascina's
Art, Politics, and Dissent (Manchester University Press: Manchester, 1999). (OA, 2008)

It must be pointed out that self-citations are almost invariably non-integral, whilst other-citations occur in the integral form almost as frequently as in the non-integral one. In this respect, writers in art history differ greatly from scholars in linguistics, biology and astronomy, who handle self-citation identically to other-citation. The corpus research shows that self-citation in art history is restricted to endnotes, whereas other-citation occurs frequently in the body of the text as well.

3.2.5 Types of integral citation across the disciplines

Results for each discipline are displayed in Figure 4. Art history was completely omitted, for the above explained reasons. It is obvious that the predominant type of integral citation is the verb-controlling type, in all three disciplines. It is used in more than fifty per cent of integral citations.With regard to the above listed problems, these three types of citations have not been further distinguished among integral citations in art history, Table 18 thus shows only the distinction between non-integral and integral citations, which was made on the basis of the above stated condition.

The second most frequent is the naming type, in astronomy used in approximately 35 per cent, in biology in 26 per cent and in linguistics in 25 per cent of all instances of integral citations. The least

frequent type is non-citation.

The As Figure 4 shows, astronomy uses non-citations much less than biology (18 per cent) and linguistics (nearly 25 per cent), only in less than 6 per cent of all integral citations.
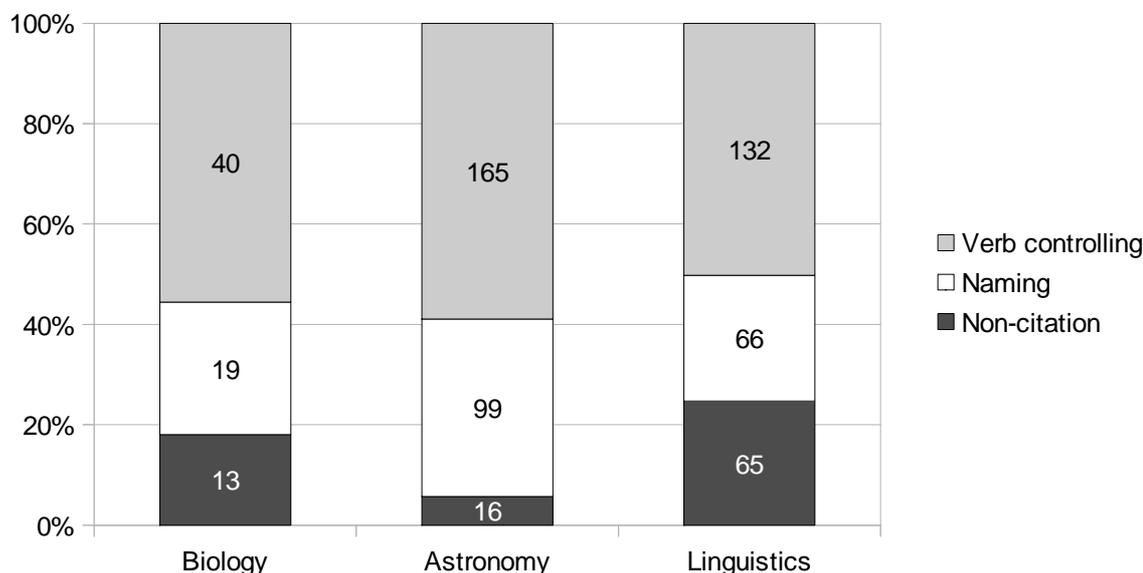

Figure 4: Types of integral citation in biology, astronomy and linguistics

3.3 Distribution of citations

For the examination of distribution of citations across the RAs, it is necessary to pay attention to the structure of these articles. A short description of structuring of the studied articles will be provided for each discipline. The typical IMRD (Introduction-Method-Results-Discussion) structure is described in section 1.2.4.2. Although Swales (1990) claimed that this structure is valid for all empirical research articles, many studies (e.g. Posteguillo, 1999; Yang and Allison, 2003; Lin and Evans, 2011) have shown that this structure varies considerably in different disciplines.

The distribution of citations is provided in visualised concordance plots acquired from AntConc. Only prototypical examples are shown in this section, all concordance plots can be found in Appendix D; the concordance plots of self-citations are provided in Appendix E.

3.3.1 Biology

#### 3.3.1.1 Structural organization of biology articles

In line with the findings of the above mentioned studies, biology shows the most regular organization of the research article. With the exception of one article from *Integrative and*

Comparative biology (2007)*,* all have a strict IMRD structure. In three articles, the Introduction section does not have any heading but it is evident that the text preceding Methods section aims to introduce the topic of the article. Also, it is worth noting that all but two articles explicitly name the second part Material(s) and Method(s), instead of Method(s) only. The only exceptional article (*ICB, 2007)* begins with an Introduction and continues with several sections comprising theoretical background with personal findings and observations. This article does not have any Discussion section.

3.3.1.2 Distribution of citation in Biology

Five out of nine biology articles show approximately the same distribution of other- and self-citations, the overwhelming majority being non-integral. As an illustrative example, the article from *Behavioral Ecology* (2001) will be examined. Figure 5 provides the visualised distribution of citations in this article. Obviously, the citations are distributed unevenly, with the highest density in the opening and slightly lower in the closing section.

Behavioral Ecology, 2001
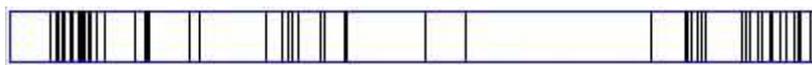
Nr. of citations: 66
Word count: 5917

Figure 5: Distribution of citations in biology (BE, 2001)

The Introduction is heavily loaded with citations, more instances are often gathered within a single sentence (ex. 21). In total twenty-four citations have been found in the Introduction with the word count of approximately 700 words.

> (21) Some evidence (reviewed in Schmid-Hempel, 1991; but see Visscher and Dukas, 1997) suggests that honey bees have a limited lifetime energy budget, equivalent to around 800 km of flying (Gould and Gould, 1988). Because flying is so energetic (Schmid-Hempel et al., 1985; Wolf and Schmid-Hempel, 1990; Wolf et al., 1989), by conserving their efforts... (BE, 2001)

The second highest density of citations is found in the Discussion section which includes 24 citations within its 1030 words. Similarly to Introduction, more citations often occur within one sentence (ex. 22). For comparison, the Methods section (in this case a description of various models) includes only twenty instances of citations and is over 7700 words long.

> (22) Seeley (1983) and Seeley and Visscher (1988) measured the time that scouts and recruits took to find a forage site. Surprisingly, recruits took significantly longer (60% longer, Seeley, 1983; 47% longer, Seeley and Visscher, 1988) than scouts. (BE, 2001)

A similar distribution of citations can be observed in the articles from *PE 2000*, whose Methods section is shorter, otherwise the density is in accordance with the other articles, *MBE 2002, BE*

2004 and *SB 2004* as is obvious from Figures in Appendix D. In the article from *DNA 2006,* the structure is slightly less evident, however, it can still be regarded as an example of the typical distribution in biology, as well as in the article from *IIM 2008*[4].

In contrast, two articles from *SB 2005* and *ICB 2007* use citations differently as can be seen in Figure 6. As already mentioned, *ICB 2007* is the only article whose overall design differs considerably from the IMRD pattern, as described above. In accordance with its structure, more citations can be found in the central part than in the beginning or in the end of the article. Moreover, the introductory part includes only four references to Charles Darwin (ex. 23) and two references to different authors. In the following text, the authors express their beliefs and findings in reference to several theoretical issues which, of course, demands mentioning previous research.

Integrative and Comparative Biology, 2007

Nr. of citations: 49

Word count: 2386

Figure 6: Distribution of citations in biology (ICB, 2007)

(23) In 1868, Charles Darwin introduced ''The Variation of Animals and Plants under Domestication'', a major work that expanded upon his initial thesis regarding the origin of species. In this work, Darwin's intention was to provide... (ICB, 2007)

Systematic Biology, 2005
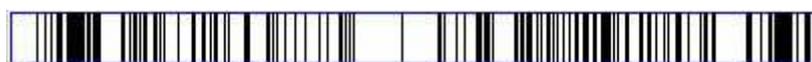
Nr. of citations: 254

Word count: 8760

Figure 7: Distribution of citations in biology (SB, 2005)

The second distinct article is from *SB 2005*, and contains an unusually high number of citations, with more citations occurring regularly within one sentence. These citations, as is obvious from Figure 7, are distributed throughout all parts of the article, with a slight gap in the narrow central part. This is caused by the nature of the Results section in which the author not only immediately compares his findings to previous ones (ex. 24), but also derives his calculations and charts from already calculated data (ex. 25), which he thus needs to cite. This all occurs in the Results section even though it is followed by Discussion.

(24) The unrooted arrangement of frog families shown in Figure 2 is very similar to those of Lynch (1973), Duellman and Trueb (1986), and Hay et al. (1995), and mainly differs from those of Maglia et al. (2001), Pugener et al. (2003), and Haas (2003) in the relative... (SB, 2005)

---

[4] This seems shifted at first sight but in fact, the last two pages of the text contain only endnotes, therefore there are no citations shown at the concordance plot. In this case, the distribution of the indexes referring to endnotes have been captured.

(25) The split between birds and mammals (diapsid versus synapsid reptiles) was set at 310+10 Mya (Benton, 1997). A minimum age of 338 Mya for the divergence between Lissamphibia and Amniota based on the aistopod fossil *†Lethiscus*, of Viséan age (Ruta et al., 2003).... (SB, 2005)

3.3.1.3 Distribution of self-citation in Biology

Results for biology are recorded in Figures in Appendix E. Similarly to other-citation, one representative article was chosen and Figure 8 shows the concordance plot of *Journal of Plant Ecology* (2000).

Plant Ecology, 2000

Nr. of self-citations: 14
Word count: 4752

Figure 8: Distribution of self-citations in biology (PE, 2000)

As this Figure shows, the distribution of self-citations is similar to that of other-citation. However, to obtain more precise results, the instances of self-citation have been counted manually. Nearly 32 per cent of self-citations occur in Introduction, 28 per cent in Methods/ Procedure/Terminology, 7 per cent in Results and as many as 23,5 per cent in Discussion or Conclusion. Self-citations in the Results section are not common because this section describes the results of the current research, however, some of the findings receive a support or explanation based on the current state of knowledge, which necessarily comprises also previous findings or claims made by the same authors.

3.3.2 Astronomy

3.3.2.1 Structural organization of astronomy articles

Articles from astronomical journals present the biggest problem. On one hand they are structured and the individual parts are clearly numbered and named. On the other hand, the only part that is invariably present in all articles is Introduction (and in all but one article this section also has a heading). Seven articles end with a Conclusion, the remaining two with Unsolved problems (*AR 2001*) and Results and discussion (*AJ 2007*). Falling between Introduction and Conclusion, several different unique parts can be found. They bear different headings and usually proceed from a description of the phenomena in question, of models and devices used and data collection, to Tables, calculations and diagrams. Very roughly, some similarity to Theoretical preliminaries – Methods and Data – Results may be observed, however, this cannot be relied on without detailed study and understanding of the particular issues.

3.3.2.2 Distribution of citations in astronomy articles

With regard to the above described uncertainty about the structural organization of articles in astronomy, it is also impossible to specify the exact locations of citations. In this discipline, the most reliable source of information about the distribution of citations are therefore the visualised concordance plots. As Figures in Appendix D shows, six of the astronomy articles are heavily loaded with citations, the number of instances exceeds much that of linguistics and biology (but to compare these two disciplines properly, the difference in the text lengths has to be kept in mind). Secondly, the distribution itself is very distinct from linguistics and biology, citations are not accumulated in the introductory and closing sections but they occur in all parts of the articles.

Monthly Notices of Royal Astronomical Society, 2005

Nr. of citations: 202

Word count: 14938

Figure 9: Distribution of citations in astronomy (MN, 2005)

However, there are three articles which do not employ citations so much. In two cases, *RA 2000* and *EA 2006,* the low number of citations is striking, the third article from *Astrobiology* (2002) includes apparently fewer citations than the others in astronomy, but not fewer than in linguistics and biology. The concordance plots of *RA 2000* and *EA 2006* are provided in Figures 10-11. The article from *EA 2006* deals with a new technique, discusses new use of an optical device. Almost the whole text illustrates and describes the design of this device, its layout and functions. An explanation of the exceptionally low number of citations may lie in the topic itself, authors writing about a new device do not need to refer back to previous devices as their technique does not have to be dependent on them. Of course, they need to demonstrate certain knowledge of the previous technical devices, which they do through the 17 references (ex. 32). From the formal point of view, this article does not include citations in the text but uses endnotes. In the list of references, there are 14 entries which are referred to 17 times.

Experimental Astronomy, 2006
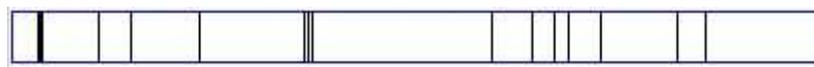
Nr. of citations: 17

Word count: 2935

Figure 10: Distribution of citations in astronomy (EA, 2006)

(32) The concept of two dimensional photon counting imaging is not new, as shown by the consistent use of, for example, multianode microchannel array (MAMA) cameras [12], intensified (EA, 2006)

The second distinct article from *MN 2000* includes only 23 instances of citation. Its Methods

section, which stretches to eight out of fifteen pages, is composed of short introductory texts followed by a number of equations, Figures, graphs and schemas. Almost no citations are found in this part (13 citations in a 6300 words long section). If there is some, it usually refers to problems and constraints which the preceding researchers met (ex. 33) and thus justifies the particular steps that the current authors decided to take. For comparison, the Introduction section includes ten instances of citation and is almost 650 words long.

Monthly Notices of Royal Astronomical Society, 2000

Nr. of citations: 23
Word count: 8715

Figure 11: Distribution of citations in astronomy (MN, 2000)

> (33) The range of the bar angle and axis ratio is also constrained by other observational constraints of the bar such as microlensing optical depth (Zhao & Mao 1996; Bissantz et al. 1997) and star counts (Stanek et al. 1994)... (RA, 2000)

3.3.2.3 Distribution of self-citation in astronomy

The situation in astronomy is recorded in visualisations of concordance plots in Figures in Appendix E, two illustrative examples are provided in Figure 12. The main difference from biology is the number of self-citations, as mentioned above. However, also the distribution is different, especially in the central part of the article.

Astrophysical Journal, 2003

Nr. of self-citations: 34
Word count: 11980

Figure 12: Distribution of self-citations in astronomy (AJ, 2003)

Whilst the Introduction section still includes many self-citations, the section Models / Measures / Procedures contains a much higher number of self-citations than any article in linguistics or biology. This strong tendency results from the narrow specialisation of various models, computations and methods which are substantial for research undertaken in this academic field. The fact that several different established models and formulas are needed for the new studies thus leads to a considerable demand of citations. Furthermore, this section is much longer than in the other disciplines and many of the analysed articles were written by more authors. There is only one article with one author, one article with two authors, the other articles were written by six or more authors, one by eighty.

57

3.3.3 Linguistics

3.3.3.1 Structural organization of Linguistic articles

All linguistic articles examined in this thesis begin with an Introduction; in 7 articles this is followed by Methods (slight variations in the naming of particular sections are ignored as long as the title clearly shows the rhetorical aim of this section) and Results. With the exception of *Computational linguistics* (2001), all articles end with either Discussion or Conclusion or both (one example). Two articles, in *CL 2001* and *JP 2005*, include several unique sections, however, these parts are considered irrelevant as they are devoted to descriptions of different theoretical concepts and models and are indicated with the asterisk. Linguistic articles are organized as follows:

JFD 2000: I-M-R-D

CL 2001:   I-Previous Research-Model- * - R- * -Remaining Issues

AL 2002:   I-M-R-Summary-C

JML 2003: I-Study (Methods, Procedure, Results)-D-C

JP 2004:   I-Aims-M- * -C

JP 2005:   I-Terminology-Concepts- * - C

ESP 2006: I-Data and Methods-Analysis and Findings-C-Pedagogical Implications

BL 2007:   I-Material and Methods-R-D

JFD 2008: I-Theoretical preliminaries-Aim, Predictions-M-R-D

3.3.3.2 Distribution of citations in Linguistcs

In general, two main tendencies can be seen in linguistics. The first one is similar to biology, with the highest density of citations in the Introduction and a bit lower in the Discussion section. This tendency is attested in five articles. Three of them: *CL 2001, AL 2002* and *JML 2003* are indisputable examples, and in the other two: *BL 2007* and *JFD 2008,* the tendency is not so straightforward, however, there is still a clear central part with almost no occurrences of citations. All concordance plots of this type are provided in Appendix D

Journal of Memory and Language, 2003

Nr. of citations: 80

Word count: 13232

Figure 13: Distribution of citations in linguistics (JML, 2003)

As a representative example of the first type, the article from *JML 2003* has been chosen. Its

concordance plot is provided in Figure 13. The Introduction section includes thirty-seven citations among more than 2050 words. Similarly to biology, more citations in a single sentence can be found (ex. 26), all providing background for the current research. Ten citations occur in the Study section, which is over 8880 words long and comprises Methods, Procedure and Results. However, eight from these ten occurrences are self-citations (ex. 27). For their current research, the authors use results obtained in their own previous studies. Thirty-one citations have been found in the General Discussion section the word count of which is 3050. In comparison to biology, Linguistic articles include much fewer citations given the length of the individual parts. It is also worth noting that in both disciplines, the number of citations in Introduction was roughly the same as in the Discussion section and the word count of Introduction is slightly lower than in Discussion.

(26) This possibility is compatible with suggestions that compelling structure exists in the world that "cries out to be named" (e.g., Berlin, 1992; Hunn, 1977), and that named categories are formed around clusters of correlated properties (e.g., Anderson, 1991; Jones & Smith, 1993; Rosch & Mervis, 1975; Smith & Heise, 1992). .. (JML, 2003)

(27) Our study is based on names for artifacts, in particular, the set of common containers studied by Malt et al. (1999). (…) We used the data collected by Malt et al. (1999) on names produced for a large set of common containers by... (JML, 2003)

The second tendency shows no such distinct preference for citing in the Introduction or Discussion sections, the citations are distributed across the whole body of the text. Three articles are considered to favour this tendency: *JFD 2000, JP 2005* and *ESP 2006.* The article from *Journal of Pragmatics* (2005) is, as mentioned above, generally different in organization. It does not, in fact, have any Methods or Results section and its structure follows individual theoretical issues underlying the main topic of spatial and temporal boundedness. Thus it is not surprising that the distribution of citations differ considerably from the typically structured articles, as can be seen in Figure 14.

Journal of Pragmatics, 2005
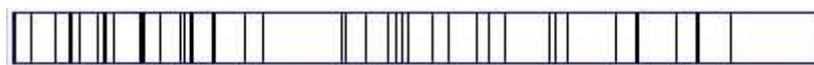
Nr. of citations: 76
Word count: 14441

Figure 14: Distribution of citations in linguistics (JP, 2005)

However, both *JFD 2000* and *ESP 2006* are organized according to the IMRD/C structure and it is therefore striking that their use of citations does not follow the general pattern, although a slight accumulation of citation can still be observed in the initial parts of both the articles. The organization of *ESP 2006* is not so strict and presents some minor irregularities. Following

Introduction and Methods, the Results section not only provides examples illustrating the findings, but the author distinguishes several categories of examples adding a theoretical background for each of them (ex. 28, 29). Thus also this part makes excessive use of citations in comparison to other already described Results sections.

(28) Following Jacoby (1987), I distinguish a category of research reports in which there is no specific reference point clearly identifiable. (ESP, 2006)

(29) Since the present research takes an approach based on patterns, I make use of the groups distinguished by Francis et al. (1996) for the V that pattern. (ESP, 2006)

Journal of Fluency Disorders, 2000

Nr. of citations: 61
Word count: 36085

Figure 15: Distribution of citations in linguistics (JFD, 2000)

Although the distribution of citations in the article from *JFD 2000* seems exceptional (see Figure 15), given its strict regular structure, a closer look at citations occurring in the central part reveals that all nine examples within the Results section are in fact only two alternating references to established procedures of statistical testing: "Tukey analysis" and "Friedman test" (ex. 30, 31), both falling within the category of non-citation. The authors thus do not provide any reference throughout the Results section, only describe how they acquired their results.

(30) Post hoc Tukey analysis showed that group 1 speakers used less... (JFD, 2000)

(31) Within each age group, a Friedman test was carried out on the ranking... (JFD, 2000)

3.3.3.3 Distribution of self-citation in linguistics

Two concordance plots representing both tendencies described in the previous section are provided in Figures 16-17, other visualisations can be found in Appendix E. Obviously, the most self-citations occur in Introduction and slightly less in Methods/Procedure/Terminology sections, the lowest number in Discussion/Conclusion. Most of self-citations used in Discussion or Conclusion refer to the previous articles on the same or similar topics and correct or acknowledge previous results.
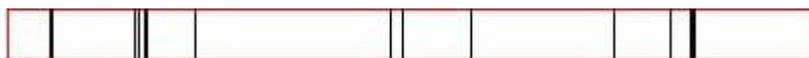
Journal of Memory and Language, 2003

Nr. of self-citations: 29
Word count: 13232

Figure 16: Distribution of self-citations in linguistics (JML, 2003)

Journal of Pragmatics, 2005

Nr. of self-citations: 16
Word count: 14441

Figure 17: Distribution of self-citations in linguistics (JP, 2005)

The distribution of self-citations does not differ substantially from that in biology. The only difference can be seen in the Discussion/Conclusion section which includes nearly as many instances of self-citations as the Methods section in biology articles, whereas the ratio in linguistics is different. The explanation for this could possibly lie in the nature of the discipline. Biology is one of hard sciences heavily based on empirical methods. As the discipline builds its knowledge on previous findings proceeding step by step, the researchers have to refer to those previous results more often. Scholars focused on highly specialized topics will of course cite their own work more often. In contrast to linguistics, self-citations in biology articles serve in the substantial majority of cases as a reference to the source of the basic presupposition taken into account in the current research. These sources are not further discussed, opposed or promoted, they are taken for granted. This issue is also connected to the fact that almost all self- and other-citations are non-integral, the writers thus only provide the source of a particular claim but do not add their own interpretation or evaluation to the findings.

3.3.4 Art history

3.3.4.1 Structural organization of art history articles

Art historical articles are completely different from the other three disciplines. In fact, they cannot be regarded as empirical articles and they do not display the organization typical of research articles in other disciplines. The art historical writing lacks any formal structure, it is not divided into any sections with headings. The texts begin with an introductory part, however, the other parts of the IMRD(C) structure are not suitable for research in this field as art history builds upon previous statements and discoveries and relies on historical facts, evidence and interpretations. There is no need for describing Methods. as these are almost invariably interpretative summarising and compiling historical evidence in the studies of old art; or interpretative observations in the case of contemporary art production or performance arts. These interpretations or pure descriptions of facts are then compared with previous studies which are constantly revalued and reappraised. Thus the discussion takes place through the whole body of the text. Usually, there is also no place for conclusion as research in the field of art history is not considered final, quite the opposite, it is prepared for further revaluation and reconsiderations. Therefore the distribution of citations in art history can be only roughly compared to the other disciplines, and only on the basis of visualised

concordance plots.

3.3.4.2 Distribution of citations in art history articles

Considering the distribution of citations, there is in principle no variation among the examined art historical articles. The citations occur throughout the whole text with almost equal density. As an illustrative example, the article from *Art Bulletin* has been chosen. Its concordance plot can be seen in Figure 18. As is described in the previous paragraph, there is no overt structure of art-historical articles, the texts are usually not divided into any sections. The authors refer to previous evidence, thoughts or beliefs and express agreement or disagreement with them. Thus, no part demands for heavier citing.

The Art Bulletin, 2004

Nr. of citations: 77

Word count: 12,611

Figure 18: Distribution of citations in art history (AB, 2004)

All articles have endnotes or footnotes and superscript numbers referring to them within the body of the text. In the case of endnotes, these were temporarily erased for the purpose of capturing the concordance plots in order to get a proper picture of the main text itself. It is also worth noting that both endnotes and footnotes differ substantially from those in the other three disciplines, as they provide not only the authors and the titles of their texts, but usually comprise much longer stretches of text discussing the issue in greater depth. Therefore one item in the list of references can mention more sources, usually contrasting (ex. 34):

> (34) 19. For the debate between photography as a personally expressive art versus one directed toward reproduction, see Hambourg (as in n. 9), 32-33. Although there were other points of view in this period-for example, Clarence White's emphasis on design as the key photographic principle, the argument between Stieglitz and Steichen, once close friends, was particularly well known in art circles (...) Steichen argued that "the great art in any period (...) and nail," and condemned the " 'art for art's sake' school" as "stillborn." Paul Strand, speaking for the Stieglitz camp, retorted snidely, "This indeed is a startlingly revolutionary... (AB, 2004),

or expand information given in the text (ex. 35):

> (35) 60. In the early 1920s, with his New York photographs, Sheeler experimented with (…) This was not a subtle editing of the image but a radical approach to picture making that nonetheless maintained the purity of the artistic vision. See Mora (as in n. 59), 86. (AB, 2004),

sometimes also adding personal opinion or often expressing gratefulness to colleagues (ex. 36):

> (36) 47 Pierre Cabanne, *Dialogues with Marcel Duchamp*, trans. Ron Padgett (Thames and

Hudson: London, 1971). Duchamp's discussion continues as follows: DUCHAMP: 'I don't know anything about it, I don't understand anything about politics, and I say it's really a stupid activity (…)" Duchamp's social irresponsibility was targeted by Joseph Beuys in his 1964 action (...). See David Hopkins, 'Re-thinking the "Duchamp Effect"', in Amelia Jones (ed.), *A Companion to Contemporary Art since 1945* (Blackwell: Oxford, 2006), pp. 145–63. My thanks to Alan Moore, Gregory Sholette and Blake Stimson for their helpful conversations on this topic during and after the CAA 2001 conference. (OAJ, 2008).

In some scarce cases, the endnotes do not serve the purpose of referring to other researchers at all, but only expand the main text (ex. 37). Such examples will not be examined and have not been counted as instances of citation, both in total numbers and in concordance plots.

(37) 7. For example, regarding film as a medium, intermedia aspects can be traced in film's early phase, insofar as film brought together discrete elements from literature, music, dance and theater. But the medium of film is not per se intermedia. (LEO, 2001)

As already said, the distribution of citations does not vary, however, there is a significant difference in their number. Two extremes are shown in Figures 19-20. The article from *JWC 2003* includes 184 instances of citation and the article from *LEO 2001* only 10. The former article, of course, is much longer. However, considering the precise length of the text itself (excluding extensive endnotes), the ratio is one citation for 811 words in *JWC* and for 3039 words in *LEO*. Such differences in art history are easy to explain. Articles focused on old masters employ citation more often than those dealing with contemporary art, new media and other "new" topics.

Leonardo, 2001

Nr. of citations: 8
Word count: 4812

Figure 19: *Distribution of citations in art history (LEO, 2001)*

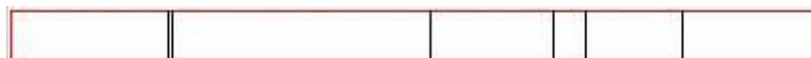Journal of the Warburg and Courtauld Institutes, 2003

Nr. of citations: 189
Word count: 28212

Figure 20: *Distribution of citations in art history (JWC, 2003)*

3.3.4.3 Distribution of self-citation in Art history

Journal of the Warburg and Courtauld Institutes, 2003

Nr. of self-citations: 6
Word count: 28212

Figure 21: *Distribution of self-citations in art history (JWC, 2003)*

As mentioned above, all instances of self-citation occur in endnotes. Thus, a search of index numbers referring to these particular endnotes was conducted and Figure 21 shows their location within the body of the text itself. The location of some instances (in *JWC 2003* and *JWC 2006* and in *OA 2008*) of self-citations remained in the endnote section, though. In these cases the particular self-citations were provided within a longer endnote which had in fact referred to different works by different authors. However, the writers needed to add reference to their own study as well and compare, support or oppose the other author mentioned within the same endnote.

3.4 Reporting verbs

3.4.1. Quantitative results

Reporting verbs used in individual disciplines have been counted, the results are provided in Table in Appendix F. In this section, sentences with nominalized subjects (for example *his suggestion, proposition...*) were omitted. In total, 139 different reporting verbs have been used, together found in 662 hits. As many as forty-six of these verbs were used only once in the whole corpus and twenty-two of them twice. This finding is closely connected with the question of stylistic diversity of academic texts. The discipline with the smallest number of distinct types of reporting verbs seems to be biology which employed only thirty-six types of verbs, however, the total number of all citations using reporting verbs in biology is much lower than in the other disciplines (63 tokens), as can be seen in Figure 22. On the other hand, the largest number of distinct types of verbs was found in art history which uses as many as seventy-nine types, however, the total number of integral citations including reporting verbs is much bigger in art history (392 tokens). In linguistics, sixty-eight distinct reporting verbs have been found and in astronomy sixty.

Mere numbers of different types of verbs do not say much about lexical diversity, though. For a more synoptic illustration of the ratio between tokens (the overall number of citations using reporting verbs) and distinct types of reporting verbs, see Figure 22. It can be seen that the total counts of reporting verb tokens vary to a great extent, ranging from 63 in biology to 392 in art history. In contrast, however, the numbers of verb types do not differ much, the scope is much narrower, from 36 types in biology to 79 in art history. From this point of view, it can not be assumed that biology uses a more limited variety of reporting verbs than the other disciplines. Moreover, the numbers for astronomy and linguistics are roughly equal, thus, no plausible conclusion about stylistic variety in academic writing in soft and hard sciences can be arrived at.
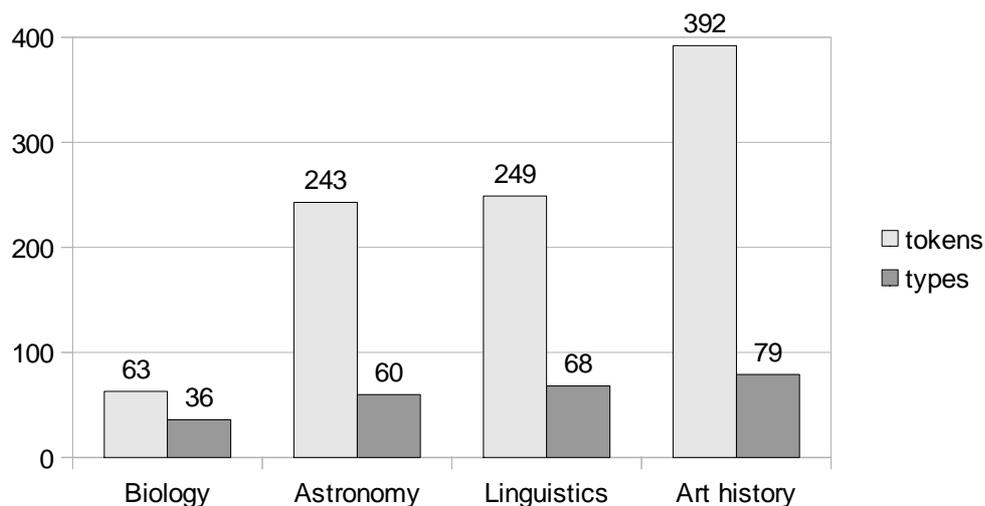
Figure 22*: The verbs used in citations: the overall number of citations using reporting verbs (tokens) and the distinct types of reporting verbs used*

3.4.2. Types of reporting verbs

All encountered verbs have been divided into three groups according to Hyland's (1999) classification of "discourse acts" verbs, "research acts" verbs and "cognitive acts" verbs (see section 1.3.4.1) and sorted in alphabetical order for better orientation. Numbers in these Tables show individual occurrences of each verb found across the four examined disciplines.

The most frequent verbs are *find* and *suggest,* both represented with more than 50 examples. Such high numbers, however, do not prove that these are the most frequently used verbs across all academic subjects. *Find* occurred 38 times in astronomy, but only 13 times in linguistics, 5 times in biology and not a single occurrence was found in art history. Quite the opposite, *suggest* was used mainly in art history (21 instances), in linguistics and astronomy 11 and 15 times, respectively, and only 5 instances were found in biology. Other verbs used more than 30 times are *argue, note, describe, discuss* and *show,* more than 20 times *point out, report, believe, state, propose.* From these, all 26 instances of *believe* occur in art history, similarly to *state,* which occurs 23 times in art history and only once in linguistics.

Considering the frequently used verbs in each of the four disciplines, great variation has been observed. In art history, the verb with the highest occurrence is *note* (30 examples found), followed by *believe* (26)*, argue* (24)*, state* (23) and *suggest* (21). According to the classification provided in Hyland (1999) and other studies summarised in section 1.3.4.1., *argue, state,* and *suggest,* together with other verbs employed almost exclusively in art history as *acknowledge, comment, explain, put*

forward, remark, say, and *write,* fall within the group of "discourse acts" verbs. (Ex. 38) On the other hand, verbs found frequently in other disciplines without any match in art history are *analyse, detect, find, present, report,* and *study,* which are, according to Hyland (1999) "research acts" verbs. The third group (ex. 39) of "cognitive acts" verbs is represented by *agree*, *believe*, *consider* and *think*, and with the exception of *consider,* all these verbs are used exclusively in art history.

(38) Duncan-Jones argues that Sidney's Urania is an allegorical Figure and identifies her as 'VenusUrania, a type of the Heavenly Beauty which inspires Heavenly Love ... (JWC, 2006)

(39) Venturi also believed the canvas to to be an early work.... (BUR, 2005)

The most frequently used verbs in astronomy are *find* (38 instances), *show* (24), *suggest* (15), *describe* (14) and *discuss* (14), these verbs are not used in astronomy only, however, especially *find* and *show* outnumbers the other disciplines. (Ex. 40) However, some verbs are found exclusively in astronomy: *detect* (7)*, obtain* (6)*, determine* (5) and *derive* (4)*.* (Ex. 41) As Table in Appendix F shows, the majority of verbs used in astronomy are "research acts verbs", the least used are "cognitive acts" verbs. For example, none of the above mentioned verbs *agree, believe, consider* and *think* occurs in astronomy.

(40) Hartigan et al. (1994) find a mechanical luminosity of 1.5 L©, or 6% of the... (AR, 2001)

(41) Murakawa et al. (2005) detected little (P10%) polarization in the southeast in... (AJ, 2007)

As has already been mentioned, the lowest number of reporting verbs was found in biology, therefore 28 verbs occurred only once, which is similar to the other disciplines (29 single occurrences in art history and linguistics and 34 in astronomy), in spite of their much larger number of reporting verbs in total. The most frequent verbs in biology are: *propose* (10 instances), *find* (5), *review* (5) and *suggest* (5), also other verbs employed in biology show no attitude towards the presented information (*show, present, measure, calculate*, etc.) and most of them belong to the category of "research acts" verbs (ex. 42). In contrast, *argue*, *note* and *discuss,* three of the most frequent verbs with more than 30 occurrences across all disciplines, do not occur in biology at all.

(42) For instance, Visscher and Seeley (1982) calculated that the circle encompassing 95% of their bees' foraging sites had a radius of 6 km... (BE, 2001)

Linguistics makes use of all of the most frequent verbs: *argue* (21), *suggest* (11), *describe* (9)*, propose* (9), *point out* (8), etc., all belong to the category of "discourse acts". (Ex. 43) The other frequently used are: *report* (17), *find* (13), *note* (7), *observe* (7) and others, belonging to the "research acts" verbs (ex. 44). Not many verbs of "cognitive acts" group have been found in linguistics, except for *consider* and *distinguish,* both with three occurrences and *assume* with two

occurrences. (Ex. 45) None of the more frequent verbs employed in all three remaining disciplines was absent from linguistics.

(43) In Syntactic Structures (Chomsky 1957, 51ff.), Chomsky famously argued that the goal of a linguistic theory that produces... (CL, 2001)

(44) Smit (1993) also noted that adults learning a second language also found... (JFD, 2000)

(45) Perkins, Kent, and Curlee (1991) consider that a disruption in pre-motor planning at the point of assigning phonetic (JFD, 2008)

These findings are illustrated in Figure 23, which shows the ratio of individual classes of reporting verbs used in the examined disciplines. This Figure shows results for distinct types of reporting verbs, as it has been found more transparent. For the total numbers of tokens, see Figure 24. Quantitative results are provided in Table in Appendix F as well.
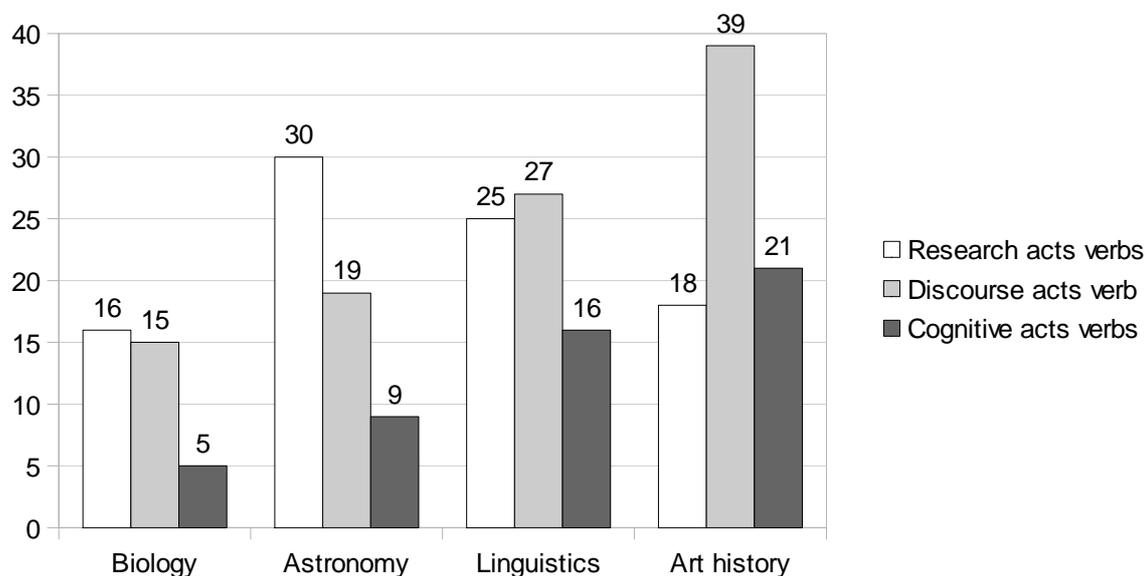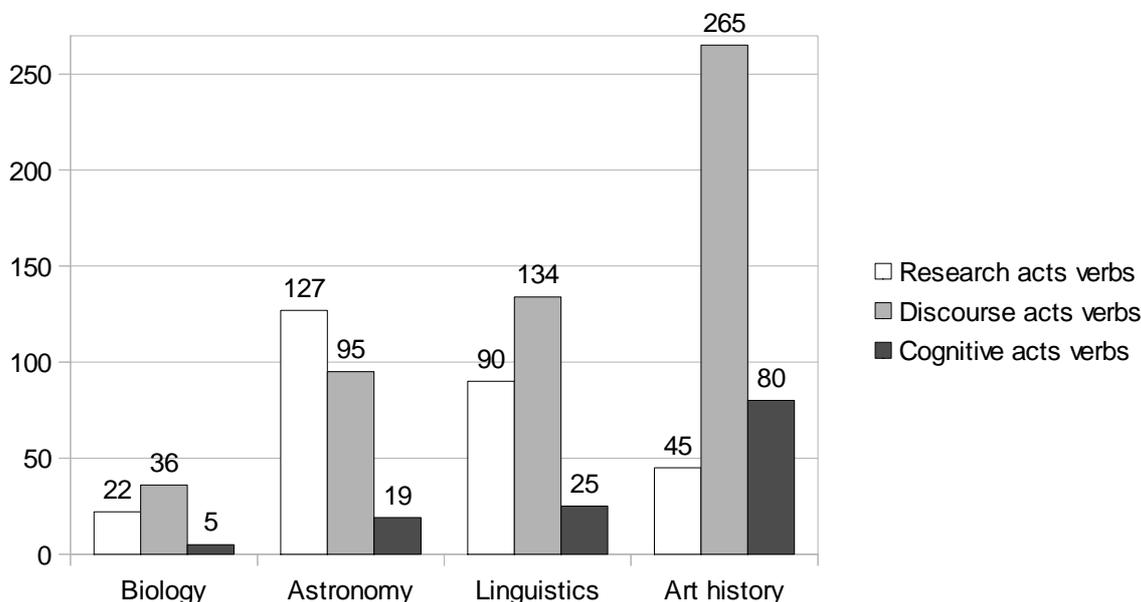

Figure 23: The representation of the individual classes of reporting verbs (Hyland, 1999) given in types

This Figure clearly illustrates differences between disciplines. In biology, the classes of "research" and "discourse" acts are employed roughly with the same frequency. However, if we look at the ratio of tokens, not types, verbs referring to "discourse acts" are used more often.

Within this category, the most frequent verbs are: *propose* (10), *review* (5), *suggest* (5) and *report* (4). With the exception of *review*, which is used in biology and astronomy only, all these verbs are heavily used in the other disciplines as well. Among the "research acts" verbs, only *find* (5), *measure* (2) and *show* (2) occurred more than once. Thirteen verbs occurred only once, the majority of them referring to research procedures. Only five verbs referring to "cognitive acts" were

found, each of them occurred only once. From these verbs, only *consider* is common also in the other disciplines, *imply* and *speculate* is shared with art history only, *follow* with astronomy and *support* is used exclusively in biology.

Figure 24: The representation of the individual classes of reporting verbs (Hyland, 1999) given in tokens

Astronomy displays a strong tendency for verbs referring to research acts, mainly to statements of findings, the most frequent verbs are: *find* (38 tokens), *show* (24)*, obtain* (7), *detect* (6) and *determine* (5). Less verbs refer to research procedures: *study* (6), *model* (6), *use* (5). Discourse acts verbs are also present in astronomy, *suggest* (15), *describe* (14), *discuss* (14), *note* (8), *give* (8 tokens; used with the following objects: *details, description, observations*), *present* (7) and *report* (6). Most of them are frequently used in all or at least in two of the other disciplines. The least employed type in astronomy is the "cognitive acts" class of verbs, in which only the verb *estimate* (7) was used more than twice, *recognize*, *infer*, *allow* all occurred twice.

In linguistics, most of the reporting verbs used refer to "discourse acts". The most frequent verbs are: *argue* (21), *report* (17), *suggest* (11), *describe* (9) and *propose* (9). As can be seen in Table in Appendix, many of the frequent "discourse acts" verbs in linguistics have been found in fairly high numbers in astronomy as well, with the exception of *claim* (8), which was not found in astronomy at all. Reporting verbs referring to "research acts" are in linguistics used slightly less often. Some of these verbs refer to statements of findings, for example *find* (13), *observe* (7) or *show* (6), however, there is no such strong tendency as in astronomy. This class is divided almost equally between statements of findings and statements of procedures, as for example: *carry out* (9, used with the

following objects: *analysis, test, research), develop* (7), *examine* (5) or *explore* (5). "Cognitive acts" verbs are used more than in both astronomy and biology, however, all of them were used only once, twice or thrice: *conclude, consider, distinguish* (3), *adopt, assume, hold, include* (2). As can be seen, none of them is shared with biology and only three of them (*assume, conclude, include*) were shared with astronomy.

Art history differs from the other three disciplines in its strong preference for verbs referring to "discourse" and "cognitive acts". Among the most frequent belong: *note* (30), *argue* (24), *state* (23), *suggest* (21), *describe* (18), *point out* (14), *discuss* (13) and *publish* (13). Art history stands out due to its preference for "cognitive acts" verbs as well. This type is not used as extensively as the "discourse" type, however, it is frequent in comparison with other disciplines. The most extraordinary is the usage of *believe* (26) which is one of the most frequent verbs in art history articles and, at the same time, is not used in any other discipline at all. Other frequent "cognitive acts" verbs are: *conclude* (9), *consider* (7), *assume* (6), *acknowledge* (5). With the exception of *acknowledge,* the remaining three verbs mentioned are shared with linguistics, *conclude* and *assume* are used in astronomy as well. "Research acts" verbs are the least common type in art history, however, some instances were found, for example: *identify* (8), *record* (7), *demonstrate* (5).

The following Tables summarise the findings about exclusively used verbs and verbs shared by more disciplines, which have been only briefly mentioned in the above described findings. Table 19 shows verbs used exclusively in one of the disciplines. Obviously, art history employs the biggest number of verbs that are not used in other disciplines, moreover, as has been already noted, some of these verbs occurred in a high number of tokens. On the other hand, seventeen of these thirty-one verbs were used only once. (For more precise results, see Appendix F.)

Art history	acknowledge, add, agree, believe, characterize, clarify, classify, connect, contribute, declare, dismiss, document, draw, expand, express, fail, formulate, inform, link, overlook, perceive, point to, predict, publish, recommend, record, relate, remember, reveal, say, share, think
Astronomy	compile, construct, derive, detect, determine, infer, list, modify, obtain, reproduce, set out, simulate, survey, switch off, take, term
Biology	cite, elaborate, evaluate, highlight, hypothesize, reintroduce, repress, support
Linguistics	adopt, approach, await, base on, carry out, collect, distinguish, extend, feel, hold, investigate, maintain, make a point, manipulate, offer, seem, select, talk, test, vary

Table 19: Verbs used exclusively in one discipline

Table 20 provides reporting verbs shared by two or more disciplines, verbs used in both soft

disciplines and both hard disciplines have been found, moreover, it was considered interesting to find also verbs shared by only one of the soft and both hard disciplines and vice versa. These results might support Hyland's (2009) theory of continuum between soft and hard disciplines described in section 1.2.2. Apart from verbs provided in Table 19 found in all disciplines, art history does not share a single "hard" verb with both astronomy and biology (which would not, at the same time, be used in linguistics), whereas linguistics shares five "hard" verbs. On the other hand, biology shares with both soft disciplines only one "soft" verb, *consider*, which appeared only once in biology articles, whilst astronomy shares thirteen "soft" verbs with both art history and linguistics.

Art history & linguistics	address, admit, aim, attempt, claim, compare, comment, conduct, confirm, develop, indicate, put forward, raise a question
Biology & astronomy	calculate, follow, map, model, perform, postulate, review
Art history & both hard	-
Linguistics & both hard	find, measure, present, report, use
Astronomy & both soft	argue, assume, conclude, discuss, emphasis, explore, give, identify, include, note, observe, provide, refer
Biology & both soft	consider
All disciplines	demonstrate, describe, employ, introduce, point out, propose, show, suggest

Table 20: Verbs shared by two or more disciplines

4. Conclusion

This thesis examined citation practices followed in research articles in four disciplines, astronomy, biology, art history and linguistics. These disciplines have been chosen in order to represent academic writing in hard and soft sciences, its similarities and differences. For this purpose, a specialized corpus of 72 research articles has been created; each discipline was represented by 18 articles. These articles were carefully chosen from academic journals written in English, published between the years 2000 and 2008, and accessible via the Charles University Electronic Information Resources Portal. An isotextual approach to creating the corpus has been adopted. This decision is based on the assumption that citations occur in particular communicative acts, therefore it is necessary to work with a corpus balanced in terms of these structures rather than in terms of the text length. Nine articles from each discipline were examined in the preliminary step, in which the most often used reporting structures were identified. A thorough corpus analysis of the remaining nine articles from each discipline has been conducted, derived from the results acquired in the preliminary step.

4.1 Integral and non-integral citation

The quantitative results show the ratio between integral and non-integral type of citation. All disciplines, hard and soft alike, prefer the non-integral type with the name of the author and the year of publication included either in parentheses, or in the footnotes or endnotes. The prevalence of non-integral citations is most evident in hard sciences, in particular in biology, in which the non-integral type constitutes as much as 91 per cent of all citations. In astronomy, it is 78 per cent. In soft domains, the prevalence of the non-integral type is not so marked. In linguistics, the non-integral type constitutes roughly 67 per cent and in art history roughly 51per cent of all citations.

These findings are in accordance with the results of previous research done in this field, which was summarised in section 1.3.2.1. In both hard and soft sciences, non-integral citations were considered predominant. Moreover, the particular numbers acquired in this thesis are in line with findings of Hyland (1999) and Thompson (2000) who reported the highest ratio of non-integral citations in biology (90%), and in linguistics, they found 66 per cent. Also the result for astronomy is similar to their results for physics (83%). In this type of citation, the writers choose to put the main emphasis on the proposition itself, the author or the particular text cited is not considered important and does not function as a sentence-element. The non-integral type of citation serves the

purpose of providing reference to a text where more information about the particular topic can be found; or simply of attributing the proposition, theory, technique etc. to its author (Thompson, 2005). Furthermore, this proposition is considered widely accepted knowledge which is neither evaluated, nor disputed.

The non-integral type of citing is well-suited for hard sciences working with experimental methods and research processes more or less independent of the human factor. The knowledge accomplished in hard sciences results mainly from correct application of appropriate methods, the personalities of the authors are suppressed as they only describe the already existing phenomena discovered. The knowledge-making in hard sciences pursues in linearly, the findings serve as building blocks upon which new ones can be based.

In contrast, soft domains move in circles and retrace the steps already taken. Old claims are usually not widely accepted as they result from subjective thoughts and conclusions, not from experimental human-independent procedures. Thus, they are constantly revisited, re-evaluated, reinterpreted and confronted with new ones, therefore it is crucial to attribute these claims to particular authors. Authors gain prominence in soft disciplines so it is not surprising that the integral type of citation, in which the name of the author is integrated into the sentence, is used more extensively than in the hard disciplines. Writing in humanities is also intended for a broader readership and often needs to provide information from different disciplines, making sure that the reader gets a sufficient background on the subject in question.

The quantitative findings of this thesis correspond with the processes of knowledge-making followed in the examined disciplines. Considering the hard versus soft distinction of academic disciplines as a continuum with two opposing poles, it can be concluded that biology is a typical hard discipline whereas art history is a typical soft one.

4.2 Types of integral citation

The integral type of citation has further been examined in order to identify its sub-types. At this point of analysis, serious problems were encountered in art history articles. All nine examined articles included endnote or footnote references, which are, in fact, non-integral. Thus, the distinction between integral and non-integral citations has been made according to the presence or absence a signal of citation (name of the author or publication). If no signal was given, the instance was considered non-integral. This process has disabled further possibility of distinguishing among the three types of integral citation because all instances included a reporting verb. Moreover, none

provided a year reference as this was always left to endnotes or footnotes. Citation in art history has thus been considered too different to fit the classification system applied.

In the three remaining disciplines, integral citations have been classified as verb-controlling, naming or non-citation, types proposed by Thompson and Tribble (2001). All disciplines show preference for the verb-controlling type with human agent and employ this type in more than fifty per cent of all integral citations. In biology, it is more than 55 per cent, in astronomy almost 59 per cent and in linguistics 50 per cent. These results suggest that all disciplines give particular prominence to the authors involved. Although the quantitative results for biology, astronomy and linguistics are similar, differences among the disciplines have been identified with respect to the range and type of reporting verbs and also to the complexity of the information presented. Linguistics tends to use complex reporting sentences providing extra information about the researchers, their studies or extending the cited proposition in a subordinated clause.

Following the verb-controlling type, the naming type is the second choice for all disciplines. Here, some differences in the ratio of the naming type have been observed. Astronomy uses the naming type more often than biology and linguistics, in 35 per cent of all integral citations. In biology and linguistics, this type amounts similarly to 25 and 26 per cent. Again, the disciplines differ slightly in the use of the naming type of citation. In biology, it usually refers to models, diagrams, approaches, or Figures which are attributed to particular authors. Less frequently, naming citations occur in adjuncts (e.g. *according to; following*) or refer to a place where the cited information can be found (*in, from, of...*). In astronomy, the use of the naming type is very similar to biology. However, in addition, some of the occurrences refer to well-established empirical procedures with the name of the author functioning as a pre-modifier of the head of the noun phrase. Considering the use of the naming citations, linguistics does not differ from biology and astronomy and employs the naming type frequently as it is also one of the means of achieving lexical density.

The least used type is non-citation. Astronomy employs this type only in scant six per cent, whereas in biology and linguistics it is eighteen and twenty-five per cent of integral citations, respectively. In biology, this result is rather surprising. Its overwhelming preference for the non-integral type suggests also overall preference for strict citation patterns, even in the integral types. Some of these instances refer to widely known authors as Darwin or Elton, some are examples of names of models, analyses and calculations, in which the citation functions as the pre-modifier of the noun. Only a few occurrences of non-citation do not belong to either of these groups. In one case, a verb-controlling citation is provided in the beginning of a paragraph focused on the results

of a particular experiment, within this paragraph, only non-citations occur for the purpose of avoiding repetitiveness. In the second case, reference is made to information acquired in a personal communication with the author, thus no year is provided. Avoiding repetitiveness is also one of the major purposes of employing non-citation forms in linguistics. Most of the instances of the non-citation type are used in longer paragraphs focused on a single author or research, therefore the repetition of dates is not necessary.

4.3 Distribution of citations

This issue is closely connected to the next Section of this thesis which focuses on the distribution of citations in research articles. At first, the structure of research articles in the individual disciplines had to be analysed. The strict IMRD structure has been found only in biology, with the exception of one article. The citations in biology tend to be distributed unevenly, with the highest density in the opening and slightly lower in the closing section. In the Introduction section, more instances of citation are often gathered within a single sentence. The second highest number of citations was found in Discussion. The results provided in visualised concordance plots show that the distribution of the citations is highly dependent on the structure, the only article with different distribution of citations is the one which is not structured according to the IMRD pattern.

Astronomy articles are organized differently, all of them begin with Introduction and seven end with Conclusion, however, the text falling between these two sections can be structured only roughly as proceeding from a description of the analysed phenomena, models and devices used, to data collection, Tables, calculations and diagrams. Astronomy does not tend to accumulate citations in the beginning and end of the article but all sections are heavily loaded with citations. Three articles include substantially fewer instances of citations, however, only one article shows the distributional pattern of citations similar to biology.

The predominant structure of linguistic articles is IMRD/C, with minor irregularities in some of the articles. Considering the distribution of citations, two tendencies have been observed. In five articles, the distribution is similar to that of biology, with citations occurring mainly in Introduction and Discussion. In the remaining articles, citations are distributed across the whole body of the text. However, these articles with "irregular distribution" show also some irregularities in structure, or the topic demands special rhetorical strategies.

The academic writing in art history is completely different from the other three disciplines. In fact, these articles cannot be considered empirical and they do not display any overall organization

comparable to the IMRD/C structure. The body of the text is not divided into any sections, it is similar to any narrative text. As the research in art history builds upon previous statements, historical facts, and interpretations, instead of empirical research, it works with compiling historical evidence or interpretative observations. These findings are constantly compared with other interpretations, which results in the need for citing throughout the whole text with almost equal density. This general tendency is followed almost invariably by all examined articles.

These findings about the distribution of citation are closely connected to the particular types of integral citations discussed above and are in accordance with the results of Charles (2006) and Thompson and Tribble (2001). They also examined the distribution of particular types of citations in different parts of academic writing. As their results show, each type of citation serves different purposes and is used in different rhetorical sections. The most frequent type is considered the verb-controlling citation which occurs in Introductions as well as in Discussions, whereas the naming type is typical of the Methods section, which generally includes lower number of any type of citations. The particular uses of these three types as described in the preceding paragraphs correspond with the aims of the research article sections.

The verb-controlling type gives prominence to the author of the cited proposition. Moreover, by using a reporting verb, it also allows the writer to adopt explicit personal stance towards the cited information, to ascribe particular stance to the author himself, or to manipulate with the generality of the cited proposition by employing different verb tenses. Thus, the verb-controlling type is well suited for resuming and summarising previous research in the field as well as for arguing and discussing the new findings with the previous ones. On the other hand, the naming type is used in the Methods section because it often refers to particular methods, calculations, procedures or models. As described above, this is true for all three examined disciplines. Furthermore, the naming type is used also as a reference to the place where additional information can be found. This is also useful in the Methods section in which the citations are employed mainly in order to justify particular procedures followed, or in order to provide the source of data used in the current research.

4.4 Self-citation

Apart from other-citation, self-citation has been analysed as well. The numbers of self-citations in individual disciplines and also in individual journals within one discipline differ considerably. As has been found out, the number of self-citations is not reliably dependent on the number of cooperating authors. The highest number of self-citations occurred in astronomy, however, the

differences are extreme as the number of self-citations ranges from one instance to as many as 74 instances. The use of self-citation in art history is significantly rare. Three of art history articles do not contain a single instance of self-citation. In accordance with the practices adopted for other-citation, all instances of self-citation occur in endnotes and are regularly non-integral.

Considering the types of integral self-citation, the ratios do not differ much from those of other-citation. These results are in line with the studies on self-citation by Bonzi and Snyder (1991), Aksnes (2003), and others, commented on in Section 1.3.2.3. In astronomy and linguistics, the naming type slightly prevails over the verb-controlling type, vast majority of the naming self-citations serve the purpose of referring to the place where additional information on the current topic can be found. This tendency is not surprising, given the fact that one of the major reasons for self-citation is the necessary inclusion of relevant information. Particularly in narrow-focused topics, this inevitably leads to the reference to the writer's previous work. The only discipline which handles self-citation differently from other-citation is art history. The corpus research shows that self-citation in art history is always restricted to endnotes or footnotes, whereas other-citation can be frequently found in main text as well. However, this tendency cannot be compared to other disciplines as art history makes use of references provided in endnotes much more extensively than the other ones, which use them scarcely (only in one astronomy and two biology articles).

The visualised concordance plots have been obtained for self-citation as well. These visualisations clearly show that self-citations are, similarly to other-citations, distributed unevenly. It is obvious that the highest ratio of self-citations can be found in the Introduction and Discussion/Conclusion sections. In contrast to other-citation, a considerable number of instances has been found also in the Methods section. This part does not typically include much other-citation as it is focused on the present study and procedures adopted by the present authors. The aim of the Methods section is to describe the research procedure and it is therefore likely that the authors who have been working on one topic for a longer period of time will refer to their previous methods. More self-citations have been found in descriptions of terminology used, probably for similar reasons: the writers use again the terminology they have already established in their previous studies.

4.5 Reporting verbs

The last part of this thesis focused on reporting verbs used in integral citations. At first, the total number of reporting verbs has been provided and the ratio between tokens and distinct types has

been counted. As these quantitative results show, the highest number of reporting verbs, both types and tokens, occurred in art history. Astronomy and linguistics show similar number and ratio of types and tokens. The lowest number of both types and tokens has been found in biology, which is not surprising as this discipline uses also the lowest number of integral citations. Quite the opposite, the results for biology show unexpectedly high number of distinct verb types employed.

Considering particular types of reporting verbs, the classification proposed by Hyland (1999) has been adopted. According to this classification, reporting verbs can be distinguished into three categories with respect to the activity referred to: research acts verbs (further distinguishable between statements of results and statements of procedures), discourse acts verbs and cognitive acts verbs. Some verbs were not straightforwardly classifiable, so attention has been paid to particular occurrences in context.

Biology employs both research and discourse acts verbs with roughly the same frequency. Very slightly prevailing are references to research acts, either statements of findings or statements of procedures. This is in accordance with the experimental nature of biology. Closely connected to statements of findings, verbs referring to discourse acts occur in biology as well. The type which is used the least is cognitive acts verbs, which is not surprising given the already described process of knowledge-making in biology which suppresses the personality of the author. Thus, there is also no place left for the description of mental processes. Reference is established to various procedures, experiments and results, not to thoughts and beliefs of the researcher. The acquired results which are later used as building blocks for other researchers have to be considered final and correct, not open to further discussion. Therefore it is likely that results referred to by reporting verbs of cognitive acts are treated as suggestions or considerations demanding further examination or verification.

In comparison, astronomy tends to use research acts verbs more often than the other types, the statements of findings prevail with *find* and *show*, which are the most frequently used reporting verbs in this discipline. Discourse acts verbs are also used much in astronomy, from this category, *describe, discuss, suggest* rank also among the verbs with the highest number of occurrences. The predominance of verbs referring to research acts is in line with the nature of astronomy and its notion of construction of knowledge. Similarly to the procedures in biology described above, astronomy also uses empirical methods, calculations, observations and experiments and builds its knowledge on the results thus acquired. The above stated reasons for low numbers of cognitive acts verbs are valid for astronomy as well.

In linguistics, the discourse acts verbs prevail, closely followed by research acts verbs. Among

the verbs referring to discourse acts, the most frequent verbs comprise *argue, suggest* and *report*. Apart from this distinction from the hard disciplines, linguistics employs more cognitive acts verbs, however, none of the most frequent verbs belong to this category. Linguistics also shares several verbs with both hard disciplines, for example *find, measure, report*. These verbs are well-suited for hard disciplines as they not only refer to research processes, but also do not provide any information about the writer's or author's attitude towards the cited proposition. These verbs therefore do not allow for any evaluation as is typical for verbs favoured by hard disciplines.

Art history differs from the other three disciplines in its strong preference for verbs referring to discourse and cognitive acts. The most frequent discourse acts verbs comprise *believe* and *think*, the discourse acts verbs are represented by *note, state* or *suggest*. This tendency obviously points to the discursive character of art history writing, its lack of experimental methods and research practices. The verb *believe* is worth pointing out as it is the second most frequent verb in this discipline and, at the same time, is used exclusively by this discipline.

This issue of shared and exclusive verbs leads to interesting results. The highest number of reporting verbs which have not been found in the other disciplines occurs in art history. This supports the suggestion that art history differs to a great extent from the remaining disciplines, no matter if soft or hard. Not only is the structure of RAs completely different, but also the endnotes style of citation and the whole conception of knowledge-making and research practices differ greatly. Art history shares verbs with linguistics, the majority being discourse acts verbs. Not a single verb is shared by art history, biology and astronomy, which suggests that none of the verbs typical of hard disciplines is suitable for art history. On the other hand, linguistics surprisingly shares several verbs with both hard disciplines, as mentioned above. Also astronomy shares some verbs with both soft domains, for example *argue,* which has an evaluative potential as it may suggest the difference of attitude between the author of the cited proposition and the writer.

A few reporting verbs occurred in all four examined disciplines, with the exception of *employ*, all eight verbs refer to discourse acts. This category is therefore considered the most universal. Of course, the range of verbs falling into this category is broad, but it can still be concluded that both soft and hard disciplines tend to refer to the statements of others. This category also enables a particular level of manipulation with the cited proposition as the information is inevitably influenced by the reporting verb used. For example, *suggest*, a verb of subjective interpreting, would carry different connotations than *demonstrate,* a verb of objective interpreting. Both verbs have been found in all four disciplines.

Assuming that soft disciplines prefer different types of reporting verbs, dependent on different

notion of the construction of knowledge, it can be concluded that art history is a "softer" field than linguistics and biology is "harder" than astronomy.

4.6 Soft and hard disciplines

As pointed out at several steps of this analysis, Hyland's notion of a continuum between soft and hard domains has been adopted. Many studied features of citation practices in biology, astronomy, linguistics and art history suggest that the differences between the imaginary opposing soft and hard poles are not as significant as the differences observed among individual disciplines.

Various results confirmed biology as a typical hard science, its citation practices corresponding to the widely accepted notion of knowledge-making followed in hard disciplines. The preference for non-integral citations is perfectly in line with the construction of knowledge proceeding linearly by building upon previous findings. Non-integral citations allow for no evaluation, manipulation or adoption of personal stance. Moreover, the cited proposition does not include any signal of including a citation. Sometimes it is difficult to distinguish which part of a complex sentence is the cited one. This type is thus used for citing widely accepted knowledge that can be taken for granted. Considering the integral type of citations, the class of reporting verbs employed also shows that biology pursues practices of hard sciences, preferring exact methods independent of the human factor.

The results for astronomy do not attest the "hardness" of this discipline so obviously. Not only does astronomy share many reporting verbs with both linguistics and art history, but it also uses a much higher number of integral citations. The case of linguistics is similar, considering the non-integral and integral citations. However, the choice of reporting verbs suggests that the human factor plays a more significant role in linguistics and, moreover, that not all statements are considered accepted knowledge. Linguistics mixes citation practices followed by hard and soft disciplines, it uses both the strict non-integral form and the integral citation with the potential of evaluating or adopting a stance towards the cited information.

Art history should be placed on the "soft" end of Hyland's continuum. As several results from different viewpoints show, academic writing in art history differs significantly from the other three disciplines. Citations are provided in endnotes and thus do not intrude into the main text itself. On the other hand, the endnotes usually provide longer pieces of information complementing the information given in the main text. Writers in art history aim at greater stylistic diversity using a wide range of reporting verbs, mostly referring to cognitive acts. This fact shows the prominence

given to the author of the cited propositional content. The thoughts, interpretations, observations and other claims are constantly revisited and evaluated, not instantly regarded as accepted truths. In general, articles in art history are not divided into any sections and thus resemble narrations of stories rather than reports of research results.

5. Resumé

Tato diplomová práce zkoumá způsoby, jimiž se ve vědeckých článcích odkazuje dosavadní výzkum. Práce porovnává typy citací, které se užívají ve čtyřech různých akademických disciplínách, astronomii, biologii, dějinách umění a lingvistice. Ověřuje hypotézu, že způsoby citování jsou závislé na procesu poznávání, který se v humanitních a přírodních vědách liší. Výchozím materiálem pro tento výzkum je specializovaný korpus sestavený z vědeckých článků publikovaných v odborných časopisech mezi lety 2000 a 2008. Každá disciplína je zde zastoupena osmnácti články, po dvou z každého roku, korpus tedy čítá 72 článků různé délky. Tento přístup je založen na předpokladu, že citace se typicky objevují v určitých mluvních aktech, tedy v určitých částech vědeckých článků. Proto bylo nutné sestavit korpus, v němž budou jednotlivé disciplíny zastoupeny stejným počtem textů bez ohledu na počet slov (Oakey, 2009).

Korpusová analýza byla rozdělena do dvou kroků. Výsledkem prvního kroku, v němž bylo pročteno devět článků z každé disciplíny, je seznam sloves uvozujících citace. Druhým krokem byla důkladná korpusová analýza zbylých devíti článků z každé disciplíny. Slovesa získaná v prvním kroku sloužila jako termíny pro vyhledávání, ukázalo se, že naprostá většina sloves nalezených v první polovině korpusu se užívá i v druhé.

5. 1 Typy citací

První část této práce se věnuje užívání dvou základních typů, integrovaných a neintegrovaných citací (viz Swales, 1990). Získaná data ukazují, že neintegrované citace převládají jak v přírodních, tak v humanitních vědách, ovšem poměr mezi oběma typy se liší. Nejmarkantnější je tendence používat neintegrované citace v biologii, kde tento typ tvoří 91 procent všech nalezených referencí, v astronomii je zastoupen 78 procenty. V humanitních vědách není preference neintegrovaných citací natolik silná, v lingvistice tvoří zhruba 67 procent a v dějinách umění jen 51 procent všech citací. Tyto údaje jsou v souladu s předchozím výzkumem užívání citací v akademických textech. Například Hyland (1999) či Thompson (2000) uvádějí 90 procent neintegrovaných citací v biologii a 66 procent v lingvistice.

V neintegrovaných citacích není jméno autora a rok začleněno do věty, je uvedeno v závorce nebo v poznámkovém aparátu. Tento typ klade důraz na obsah samotného citovaného sdělení, nikoli na autora či text. Slouží buď jako odkaz na studii, kde může čtenář nalézt více informací o daném tématu, nebo jen připisuje dané sdělení, teorie či metody jejich autorům (Thompson, 2005). Takto citovaná informace je pak považována za obecně přijatý fakt, který již není třeba dále přezkoumávat

či hodnotit. Z tohoto hlediska jsou neintegrované citace velmi vhodné pro přírodní vědy, které pracují s experimentálními metodami a zakládají se na výzkumných procesech na osobě vědce nezávislých. Přírodní vědy postupují lineárně, výsledky předchozích výzkumů slouží jako stavební kameny pro studie následující.

Naproti tomu humanitní vědy takto lineárně nepostupují, vrací se po vlastních krocích a neustále přehodnocují a prozkoumávají výsledky, kterých již dosáhl předchozí výzkum. Ty nejsou většinou považovány za definitivní, jelikož pramení ze subjektivních myšlenek, interpretací a závěrů. Vzhledem k tomuto procesu je zde kladen větší důraz na osobu autora, proto humanitní vědy častěji užívají integrovaných citací, ve kterých je citace větným členem.

V dalším kroku analýzy byly integrované citace rozděleny na tři podskupiny (viz Thompson and Tribble, 2001): „verb-controlling", „naming", a „non-citation". Do této části práce mohly být zahrnuty jen tři disciplíny, jelikož články z dějin umění používají zcela odlišný formát citování. Ve všech zkoumaných článcích jsou citace v poznámkách, tedy ze své podstaty neintegrované. Přesto zde však lze sledovat dvě základní tendence. Zaprvé, samotný text neobsahuje žádný signál, že jde o citaci, pouze odkaz na položku v poznámkách, tento způsob byl započítán jako neintegrovaná citace. Druhou možností, použitou zhruba v polovině případů, je začlenění jména autora či publikace do věty, v tomto případě byly citace považovány za integrované. Vzhledem k tomuto způsobu rozdělení by další rozlišování citací na tři podskupiny nebylo relevantní.

Všechny tři zbývající disciplíny používají především typ „verb-controlling", tedy citace, ve kterých je jméno citovaného konatelem slovesného děje. V biologii je tento typ zastoupen 55 procenty, v astronomii 59 procenty a v lingvistice 50 procenty všech integrovaných citací. Jednotlivé disciplíny se ale liší v několika aspektech, zejména v rozmanitosti a typu sloves uvozujících citace nebo v komplexnosti citovaného textu.

Druhým nejčastěji užívaným typem je „naming" citace, ve které je jméno autora součástí jmenné fráze, tedy může fungovat jako reference na zdroj, ve kterém lze nalézt více informací, nebo slouží jako premodifikátor podstatného jména. V astronomii je tento typ užíván nejčastěji, tvoří 35 procent všech integrovaných citací, v biologii a lingvistice je 25 a 26 procent. Malé rozdíly byly zjištěny v použití tohoto typu v disciplínách. V biologii a astronomii jsou „naming" citace nejčastěji užívány pro modely, diagramy, grafy či postupy výzkumu. Méně často zde figurují jména autorů v adjunktech (*according to; following*) nebo odkazují na zdroj sdělení (*in, from, of*).

Nejméně užívaným typem je „non-citation", tedy jméno autora bez reference na rok vydání. V astronomii byl tento typ nalezen jen v necelých 6 procentech, zatímco v biologii a lingvistice

tvoří 18 a 25 procent. Výsledek je překvapující zejména pro biologii, vzhledem k naprosté převaze neintegrovaného typu citací lze předpokládat, že i v rámci integrovaného typu bude biologie preferovat přesné citační vzorce. Všechny disciplíny tento typ používají v případech, kdy je celý odstavec například věnován určitému experimentu. Po prvním uvedení celé citace se jménem autora a rokem vydání už je dále opakováno pouze jméno. Takových příkladů bylo nalezeno nejvíce v lingvistice.

5.2 Rozmístění citací

Další část této práce je zaměřena na umístění citací ve vědeckých článcích. Aby bylo možné analyzovat, ve kterých sekcích článků se citace používají, musela být nejprve popsána struktura zkoumaných článků. Striktně dodržované členění IMRD (Swales, 1990), tedy úvod, metoda, výsledky, diskuse, bylo zjištěno pouze u biologie, s výjimkou jednoho článku. Citace jsou zde umístěny nerovnoměrně, nejvíce se objevují v úvodu, kde lze nalézt i několik citací v jediné větě. O něco méně četné jsou citace v diskusi, nejméně se objevují ve výsledcích a metodě. Výzkum ukazuje, že umístění citací je zcela závislé na členění textů.

V astronomii je situace odlišná. Všechny články sice začínají úvodem a sedm z devíti končí závěrem, ale text mezi těmito dvěma sekcemi je členěn různě, u každého článku jinak. Jen zhruba text postupuje od popisu analyzovaného problému, přes modely a popis použitých přístrojů, k tabulkám, výpočtům a diagramům. V astronomii se citace hojně vyskytují ve všech částech článků, nikoli jen v úvodu a závěru. Tento vzor je platný pro všechny analyzované články kromě jednoho, ve kterém jsou citace rozmístěny podobně jako v biologii.

Struktura článků z lingvistiky většinou odpovídá členění IMRD/C (s volbou mezi diskusí a závěrem). Ukazují se zde dvě hlavní tendence. První je podobná vzorci z biologie a následuje ji pět článků. Ve zbývajících čtyřech jsou citace rozmístěny ve všech částech textů. Ve struktuře těchto článků se objevují nepravidelnosti, nebo se texty věnují tématům, které vyžadují odlišné rétorické postupy.

Dějiny umění se naprosto liší od ostatních oborů, tyto články však nemohou být považovány za empirické a samotný text není členěn na žádné sekce. Výzkum na poli dějin umění staví na jednotlivých výrocích, historických faktech a subjektivních interpretacích, nikoli na experimentálně získaných datech. Studie jsou neustále porovnávány s odlišnými interpretacemi, což vyžaduje citování po celé délce textu. Tato tendence se ukazuje platnou pro všechny zkoumané články.

Rozmístění citací v jednotlivých sekcích odborných článků úzce souvisí s typy integrálních citací

a výsledky této práce odpovídají předchozím studiím v této oblasti. Charles (2006) a Thompson and Tribble (2001) například tvrdí, že každý z typů je vhodný pro odlišné rétorické cíle. Zatímco citace, ve kterých je jméno autora konatelem slovesného děje, přiznávají autorovi větší důležitost. Užití slovesa uvozujícího citaci navíc umožňuje zaujmout osobní postoj jak k obsahu či správnosti citovanému sdělení, tak vyjádřit postoj samotného autora. Volba slovesného času pak ovlivňuje, zda bude sdělení vnímáno jako nové, vhodné k přezkoumání, či obecně přijaté. Typ „verb-controlling" je tedy vhodný jak pro shrnutí předchozího výzkumu v dané oblasti, tak pro diskusi a konfrontaci nových poznatků s předchozími. Tento typ je nejčetnější a užívá se především v úvodu a diskusi.

V sekci o metodě se naopak objevuje typicky „naming" citace, protože často odkazují na určité dříve následované metody, kalkulace, postupy či modely. Dále uvádějí, kde je možné se o daném problému dočíst více, proto mohou v rámci sekce o metodě sloužit jako obhájení volby příslušných postupů či odkaz na zdroj dat, která jsou pro daný výzkum použita. Jak již bylo řečeno, toto se potvrzuje ve všech třech disciplínách, ve kterých byly jednotlivé typy integrálních citacích zkoumány.

5.3 Citace vlastních studií

Kromě citací jiných autorů byly zkoumány i citace vlastních studií. Počet těchto citací se výrazně liší, nejen mezi jednotlivými obory, ale i mezi jednotlivými články. Nejvíce příkladů bylo nalezeno v astronomii, i v rámci této disciplíny jsou však velké rozdíly, v jednom článku byl nalezen pouze jeden příklad, zatímco v jiném 74. Naopak nejméně citují vlastní studie autoři z oboru dějin umění. Ve třech článcích nebyl nalezen ani jeden případ, i v ostatních článcích byly citace vlastních textů spíše výjimečné.

Co se týče typu a užití, citace vlastních výzkumů se v biologii, astronomii a lingvistice neliší od citací ostatních autorů. V dějinách umění jsou citace sebe sama vždy omezeny jen na poznámky, zatímco citace ostatních se objevují i v samotném textu. Tuto tendenci však nelze srovnat s dalšími třemi disciplínami, které, kromě jednoho článku z astronomie a dvou z biologie, nepoužívají poznámkový aparát, pouze seznam literatury. Distribuce citací vlastních studií se také příliš neliší od citací ostatních, největší koncentraci nalezneme v úvodu a diskuzi / závěru. Jediný rozdíl byl zjištěn v metodické sekci, kde se citace vlastních děl objevují relativně často. Autoři zde především odkazují na postupy, které v dané problematice ověřili dříve, či terminologii, kterou zavedli.

5.4 Slovesa uvozující citace

Poslední část práce ze zabývá slovesy, která byla nalezena v integrálních citacích. Jak ukazují

kvantitativní výsledky, v článcích z dějin umění se objevuje největší množství sloves, a to jak odlišných lemmat, tak jejich jednotlivých výskytů. V astronomii a lingvistice bylo nalezeno zhruba stejné množství sloves, nejméně obsahují články z biologie, které však obsahují i mnohem menší počet integrálních citací celkem. Zároveň ale biologie užívá nadprůměrné množství různých typů sloves, vezmeme-li v úvahu poměr lemma/výskyt.

Slovesa jsou dále rozdělena dle klasifikace (Hyland, 1999) na tři kategorie podle aktivity, kterou popisují. „Research acts verbs" odkazují na procesy spojené se samotným výzkumem, dále je lze dělit na slovesa popisující výsledky a slovesa popisující postupy. Druhou kategorií jsou „discourse acts verbs" odkazující na řečové akty a poslední „cognitive acts verbs" popisující myšlenkové procesy.

V biologii se objevují zhruba ve stejné míře slovesa řečových aktů a slovesa popisující výzkumný postup či jeho výsledky, která velmi mírně převažují. Nejméně užívaná jsou slovesa odkazující na myšlenkové procesy. Tyto výsledky nejsou překvapující, úzce souvisí s experimentální povahou výzkumu v biologii, který staví osobu autora i jeho subjektivní myšlenkové procesy do pozadí. Výsledky, kterých bylo dosaženo, jsou považovány za konečné a správné, bez nutnosti dalšího přezkoumávání. Sdělení, která by byla uvozena slovesem popisujícím myšlenkový proces, by byla považována za návrhy, jenž je nutné dále potvrdit.

V astronomii byl nalezen největší počet „research acts verbs" zejména pak sloves popisujících výsledky, nejčetnější jsou *find* a *show*. Slovesa řečových aktů jsou zde též častá, především pak *describe, discuss* a *suggest*. Preference pro tyto typy souvisí s procesem poznávání v astronomii, podobně jako v biologii.

V lingvistice převládají slovesa řečových aktů, těsně následovaná slovesy odkazujícími na výzkumné procesy. Nejčetněji užívaná slovesa této disciplíny jsou *argue, suggest* a *report*. Lingvistika dále užívá většího množství sloves pro myšlenkové procesy než obě přírodní vědy. Na druhou stranu však užívá i mnoho sloves typických pro přírodní vědy, jako například *find, measure, report,* která jsou čistě objektivní a nevyjadřují žádný postoj k citovanému sdělení.

Dějiny umění se i z tohoto hlediska významně liší od zbylých třech disciplín svou preferencí pro slovesa řečových aktů a myšlenkových procesů. Mezi nejčastěji užívaná slovesa patří *believe* (které je zde druhé nejčetnější a zároveň se neobjevuje v žádném jiném oboru) a *think*, z řečových aktů pak *note* (první nejčetnější)*, state,* či *suggest*. Tato tendence jasně ukazuje diskurzivní charakter psaní v oboru dějin umění, stejně jako proces poznávání, který nepracuje s experimentálními metodami.

Kvantitativní údaje o četnosti jednotlivých sloves ukazují, která slovesa se objevují výlučně v jedné disciplíně a která jsou typická pro přírodní či humanitní vědy. Například největší množství sloves, která nenalezneme v žádném jiném oboru, užívají dějiny umění, což je dalším v řadě výsledků dokazujících odlišnost tohoto oboru. Články z dějin umění se liší ve své struktuře, ve způsobu citování v poznámkovém aparátu a konečně i v celkovém pojetí procesu poznávání. Dějiny umění sdílí některá slovesa s lingvistikou, nebylo však nalezeno žádné sloveso typické pro přírodní vědy (takové, které se objevuje jak v biologii, tak v astronomii, nikoli však v lingvistice), které by se objevilo v dějinách umění. Naproti tomu lingvistika užívá několik sloves nalezených v obou přírodních vědách. Astronomie naopak užívá některá slovesa typická pro vědy humanitní (nalezena v dějinách umění a lingvistice, nikoli v biologii), mezi nejčastější patří *argue*.

Několik sloves je sdíleno všemi čtyřmi disciplínami, většinou jde o slovesa odkazující na řečové akty. Tato kategorie se tedy ukazuje jako nejuniverzálnější, zahrnuje také široké spektrum sloves, od sloves subjektivní interpretace jako s*uggest,* až po slovesa objektivní, například *demonstrate*.

5.5 Závěr

Výsledky všech částí této práce odpovídají procesům poznávání v humanitních a přírodních vědách. Ukazuje se, že rozdíly mezi přírodními a humanitními vědami nejsou tak markantní jako rozdíly mezi jednotlivými disciplínami. Různé výsledky dokládají, že biologii můžeme považovat za typicky přírodní vědu. Preference neintegrovaných citací, stejně jako výběr sloves uvozujících citace a striktní členění vědeckých článků odpovídají výše popsanému procesu poznávání v přírodních vědách. Výsledky získané pro astronomii nejsou tak jednoznačné. Nejen, že užívá velké množství sloves, která sdílí s oběma humanitními vědami, ale obsahuje i větší počet integrovaných citací. Poměr integrovaných a neintegrovaných citací je v astronomii podobný jako v lingvistice. Slovesa uvozující citace užívaná v lingvistice odpovídají předpokladu, že lidský faktor zde hraje větší roli než v přírodních vědách, objevuje se zde také více sloves umožňující subjektivní interpretaci. Všechny části této práce dokládají, že dějiny umění se od ostatních oborů liší v mnoha ohledech. Citace jsou v poznámkovém aparátu a nezasahují tak vůbec do samotného textu. Volba sloves je rozmanitá, většina z nalezených sloves popisuje myšlenkové procesy a staví tak do popředí osobu autora citovaného sdělení.

Uvažujeme-li tedy o kategoriích přírodních a humanitních věd jako o kontinuu, nikoli o dvou protichůdných pólech (viz Hyland, 2009), rozlišení na integrované a neintegrované citace, stejně jako slovesa tyto citace uvozující, dokládají, že biologie se ukazuje jako typická přírodní věda, dějiny umění jako typicky humanitní, lingvistika a astronomie se nachází mezi nimi.

References

Aksnes, D. W. (2003) 'A macro study of self-citation'. *Scientometrics* Vol. 56(2), 235-246.

Bazerman, C. (2000) *Shaping Written Knowledge: The Genre and Activity of the Experimental Article in Science.* Available on-line from http://wac.colostate.edu/books/ bazerman_shaping/ 2000 (accessed 13 January 2012).

Berliner, D. C. (2002) 'Educational Research: The Hardest Science of All'. *Educational Researcher,* Vol.31 (8), 18-20.

Biber, D., et al. (1999) *Longman Grammar of spoken and written English.* Harlow: Pearson Education Limited.

Biber, D. (2006) *University Language, a corpus-based study of spoken and written registers.* Philadelphia: John Benjamins Publishing Company.

Bizzell, P.. (1992) *Academic discourse and critical consciousness.* Pittsburgh: University of Pittsburgh Press.

Bonzi, S. and H. W. Snyder (1991) 'Motivations for citation: A comparison of self-citation and citation to others'. *Scientometrics,* Vol. 21 (2), 245-254.

Brett, P. (1994) 'A genre analysis of the results section of sociology articles'. *English for Specific Purposes,* Vol. 13 (1), 47-59.

Bruce, I. (2009) 'Results sections in sociology and organic chemistry articles: A genre analysis'. *English for Specific Purposes,* 28, 105-124.

Calsamiglia, H. and L. C. Ferrero (2003) 'Role and Position of Scientific Voices: Reported Speech in the Media'. *Discourse Studies* 5 (2), 2003, 147-173.

Chandler, D. (1997) 'An Introduction to Genre Theory'. Available on-line from http://www.aber.ac.uk/media/Documents/intgenre/intgenre.html (accessed 23 January 2012).

Charles, M. (2006) 'Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines'. *English for Specific Purposes,* 25, 310-331.

Crystal, D. and D. Davy (1969) 'The Language of Science' in *Investigating English Style*, London: Longman.

Dušková, L. a kol. (2003), *Mluvnice současné angličtiny na pozadí češtiny.* Praha: Academia.

Flowerdew, L. (2004) 'The argument for using English specialized corpora to understand academic and professional language', in **Connor, U. and T. A. Upton** (eds.) *Discourse in the Professions: Perspectives from corpus linguistics,* pp. 11 – 33. Philadelphia: John Benjamins Publishing Company.

Francis, G., S. Hunston and E. Manning (eds.) (1996) *Collins Cobuild Grammar Patterns 1: Verbs. HarperCollins Publishers Ltd., 1996.* Available on-line from https://arts-ccr-002.bham.ac.uk/ccr/patgram/ (accessed 29 May 2012).

Glänzel, W., B. Thijs and B. Schlemmer (2004) 'A bibliometric approach to the role of author self-citations in scientific communication'. *Scientometrics.* Vol. 59 (1), 63-77.

Halliday, M. A. K. (1978) *Language as Social Semiotic.* London: Edward Arnold.

Halliday, M. A. K (1988) 'On the Language of Physical Science' in: **Webster, J. J.** (ed.) (2004) *The Language of Science,* pp. 140 – 158. London: Continuum.

Hartley, J. (2011) 'To cite or not to cite: author self-citations and the impact factor'. *Scientometrics*, DOI: 10.1007/s11192-011-0568-6

Hewings, A., T. Lillis and D. Vladimirou (2010) 'Who's citing whose writings? A corpus based study of citations as interpersonal resource in English medium national and English medium international journals'. *Journal of English for Academic Purposes* 9, 102-115.

Hewings, A., T. Lillis and D. Vladimirou (2010) 'The geolinguistics of English as an academic lingua franca: citation practices across English-medium national and English-medium international journals'. *International Journal of Applied Linguistics* Vol. 20 (1), 111-135.

Holmes, R. (1997) 'Genre Analysis, and the Social Sciences: An Investigation of the Structure of Research Article Discussion Sections in Three Disciplines'. *English for Specific Purposes,* Vol. 16 (4), 321-337.

Hyland, K. (1999) 'Academic Attribution: Citation and the Construction of Disciplinary Knowledge'. *Applied Linguistics* 20 (3), 341-367.

Hyland, K. (2003) 'Self-citation and self-reference: Credibility and promotion in academic publication'. *Journal of the American Society for Information Science and Technology* 54 (3), 251-259.

Hyland, K. (2004) *Disciplinary Discourses: Social Interactions in Academic Writing.* The University of Michigan Press.

Hyland, K. (2009) *Academic Discourse: English in a Global Context.* London: Continuum International Publishing Group.

Huddleston, R. and G. K. Pullum (2002), *Cambridge Grammar of the English Language.* Cambridge: Cambridge University Press.

Hunston, S. (1993) 'Professional Conflict Disagreement in Academic Discourse', in **M. Baker, et al.**, (eds.) *Text and Technology: In Honour of John Sinclair*, pp. 115-134. Philadelphia/Amsterdam: Johns Benjamins Publishing Company.

Johns, A. M. et al. (2006) 'Crossing the boundaries of genre studies: Commentaries by experts'.

Journal of Second Language Writing 15, 234-249.

Kanoksilapatham, B. (2005) 'Rhetorical structure of biochemistry research articles'. *English for Specific Purposes* 24, 269-292.

Kovačić, N. and A. Mišak (2004) 'Author self-citation in medical literature'. *CMAJ,* 170 (13), 1929-1930.

Kwan, B. (2006) 'The schematic structure of literature reviews in doctoral theses of applied linguistics'. *English for Specific Purposes* 25, 30–55.

Lin, L. and S. Evans (2011) 'Structural patterns in empirical research articles: A cross-disciplinary study'. *English for Specific Purposes,* Vol. 31 (3), 150 – 160.

Malcolm, L. (1987) 'What rules govern tense usage in scientific articles?'. *English for Specific Purposes,* 6, 31-44.

Martín-Martín, P..(2006) 'Scientific Writing: A Universal or a Culture-Specific Type of Discourse?'. *Revista de Lenguas para Fines Específicos,* Nos 11y 12 (2005 y 2006), 191-204.

Martínez, I. A. (2003) 'Aspects of theme in the method and discussion sections of biology journal articles in English'. *Journal of English for Academic Purposes*, 2, 103-123.

Messer-Davidow, E. (1992) 'Book Review: Academic Tribes and Territories: Intellectual Enquiry and the Cultures of Disciplines by Tony Becher'. *Signs*, Vol. 17 (3), 676-688.

Myers, G. (2003) 'Discourse studies of Scientific Popularization: Questioning the Boundaries'. *Discourse Studies* 5 (2), 265-79.

Nwogu, K. N. (1997) 'The Medical Research Paper: Structure and Functions'. *English for Specific Purposes,* Vol. 16 (2), 119-138. (1997)

Oakey, D. (2009) 'Fixed Collocational Patterns in Isolexical and Isotextual Versions of a Corpus'. Baker, P. (ed) *Contemporary Corpus Linguistics.* London: Continuum. 140 – 158.

Oztruk, I. (2007) 'The textual organization of research article introductions in applied linguistics: Variability within a single discipline'. *English for Specific Purposes,* 26, 25-38.

Posteguillo, S. (1999) 'The Schematic Structure of Computer Science Research Articles'. *English for Specific Purposes,* Vol. 18 (2), 139-160.

Quirk, R. et al. (1985), *A Comprehensive Grammar of the English Language*. London: Longman.

Samraj, B. (2002) 'Introductions in research articles: variation across disciplines'. *English for Specific Purposes,* 21, 1-17.

Samraj, B. (2004) 'Discourse features of the student-produced academic research paper: variations across disciplinary discourses'. *Journal of English for Academic Purposes*, Vol. 3, 5-22.

Shaw, P. (1992). 'Reasons for the Correlation of Voice, Tense, and Sentence Function in Reporting Verbs'. *Applied Linguistics* 13 (3), 302-319.

Skelton, J. R. (1994) 'Analysis of the structure of original research papers: an aid to writing original papers for publication'. *British Journal of General Practice,* 44, 455-459.

Skelton, J. R. and S. J. L. Edwards (2000) 'The function of the discussion section in academic medical writing'. *BMJ,* Vol 320, 1269-1270.

Snyder, H. (1998) 'Patterns of self-citation across disciplines (1980-1989)'. *Journal of Information Science*, Vol. 24 (6), 431-435.

Storer, N. W. (1967) 'The Hard Sciences and the Soft: Some Sociological Observations'. *Bulletin of the Medical Library Association*, 55 (January 1967), 75-84. Available on-line from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC198502/ (accessed 15 February 2012)

Swales, J. M. (1990) *Genre Analysis. English in academic and research settings.* Cambridge: Cambrigdge University Press.

Thompson, G. and Y. Ye. (1991) 'Evaluation in the Reporting Verbs Used in Academic papers'. *Applied Linguistics,* Vol. 12 (4), 365-382.

Thompson, G. (1996) 'Voices in the Text: Discourse Perspectives on Language Reports'. *Applied Linguistics,* Vol. 17 (4), 501-530.

Thompson, G. and C. Tribble (2001) 'Looking at Citations: Using Corpora in English for Academic Purposes'. *Language Learning and Technology,* Vol.5 (3), 91-105.

Tucker, P. (2003) 'Evaluation in the art-historical research article'. *Journal of English for Academic Purposes,* 2, 291-312.

Widdowson, H. G.. (1979) 'English for science and Technology', in *Explorations in Applied Linguistics*, pp. 21 – 65. Oxford: Oxford University Press**.**

Widdowson, H. G.. (1998) 'Context, community and authentic language.'. *TESOL Quarterly* 32 (4): 705-716**.**

Yang, R. and D. Allison (2003) Research articles in applied linguistics: Moving from results to conclusions'. *English for Specific Purposes,* 22, 365-385.

Yang, R. and D. Allison (2004) Research articles in applied linguistics: structures from a functional perspective'. *English for Specific Purposes,* 23, 264-279.