

**UNIVERSITY GENRES IN DISCIPLINARY DOMAINS:
SOCIAL SCIENCES AND HUMANITIES AND BASIC SCIENCES
AND ENGINEERING (*)**

(Gêneros acadêmicos nas diferentes disciplinas: ciências sociais e
humanidades e ciências básicas e engenharia)

Giovanni PARODI
(Pontificia Universidad Católica de Valparaíso – Chile)

***ABSTRACT:** The issue of disciplinarity is becoming increasingly salient in discourse studies. Questions of how differences in the structures of intellectual fields and curricula help shape educational experiences and outcomes are the focus of studies across a variety of disciplines using a range of approaches. One way to access the specialized written genres employed by academia is to begin from the tenet that all materials read by students during their university training reveal relevant data about disciplinary genres. This article presents research that focuses on the collection, construction, and description of an academic corpus based on texts collected in four disciplinary domains of knowledge: Industrial Chemistry, Construction Engineering, Social Work, and Psychology. A review of the concepts of genre and academic discourse is presented. This is followed by a description of the procedures of collecting and organizing the Academic Corpus PUCV-2006, which comprises almost 60 million words. In addition, a preliminary genre typology of the 491 text corpus is provided.*

***KEY-WORDS:** Academic genres; written discourse; university literacy.*

***RESUMO:** O tema da disciplinaridade está se tornando cada vez mais relevante nos estudos do discurso. Uma maneira de conhecer os gêneros escritos específicos utilizados na universidade, é ter acesso ao material que os estudantes lêem durante sua formação universitária. Este artigo apresenta uma investigação que tem como foco a coleta, construção e descrição de um corpus acadêmico baseado em textos de quatro áreas: Química Industrial, Engenharia da Civil, Serviço Social e Psicologia. Apresenta-se uma revisão*

* Funded by Research Project FONDECYT N° 1060440.

dos conceitos e gêneros do discurso acadêmico. Segue-se uma descrição dos procedimentos de coleta e organização do corpus acadêmico PUCV-2006, que compreende um total de 60 milhões de palavras e uma tipologia dos gêneros a partir dos 491 textos do corpus. PALAVRAS-CHAVE: Gêneros acadêmicos; discurso escrito; letramento universitário.

Introduction

Empirical research from several linguistic approaches has documented the relevance of analysis based on corpus as a way of describing linguistic and discourse variations through the disciplines and through prototypical genres (Biber, 1988, 1994, 2005, 2006; Biber, Connor & Upton, 2007; Martin & Veel, 1998; Wignell, 1998; Williams, 1998; Swales, 1990, 2004; Flowerdew, 2002; Parodi 2005, 2006a, b, 2007a, b). This focus on diversity has begun to describe and explain the common and divergent construction of specialized knowledge within discourse communities. At the same time, there is a lack of research and available data, based on corpus linguistics principles, to fill the gap that exists between a general approach and a more specific one based on naturally occurring language use. This specific approach should be one that is situated, data-driven, dynamic, and pedagogical (Herrington & Moran, 2005; Thaiss & Zawacki, 2006; Beaufort, 2007; Bazerman, 2008).

Our research has an ultimate aim to favor understanding of the written discourse of academia and professional workplace by collecting and studying the written texts that university students read, and which provide them with knowledge particular to their chosen discipline. We examine assigned student readings in four university programs. More specifically, this article describes a research project currently being carried out at the Pontificia Universidad Católica de Valparaíso, Chile. This project involves the collection, construction, and description of a corpus of written texts belonging to four disciplinary knowledge domains: Social Work, Psychology, Construction Engineering, and Industrial Chemistry. The first part of this article presents some theoretical background that frames the research. The second part establishes the parameters of the constitution of the corpus, and undertakes a general description of the nine genres that have been found in these four disciplinary domains during the five-year university course of study.

1. Theoretical assumptions

The approach taken towards discourse is decidedly interdisciplinary and of a psycho-sociolinguistic nature (Parodi, 2005, 2006a, 2007c). Hence, the texts chosen are linguistic units immersed in a cognitive and social context, that is, whose function is determined cognitively and contextually. From this perspective, texts are linguistic units with meanings constructed by producers/speakers and readers/hearers in particular contexts and with defined purposes, with prior knowledge constructed from human cognition in specific social contexts. In other words, the texts are conceived of as meaning processes and products of cognition and context.

Corpora of natural, annotated texts have had a significant impact on linguistic analyses over the previous two or three decades. In particular, research into the English language, as well as certain European and Asian languages, has revealed that linguistic studies based on large corpora of digital texts do not always corroborate the researchers' initial intuitions. Genre descriptions must be based on sufficient text samples of naturally occurring language use to ensure that the regularities and patterns observed reveal actual characteristics of the genres under study. The use of computer-supported corpora as well as the availability of computer programs that help in dealing with them has boosted linguistic research in a way that was previously unpredictable.

Empirical research from varied linguistic approaches has documented the relevance of analysis based on corpus as a way of describing linguistic and discourse variations through the disciplines and through genres (Biber, 1988, 1994, 2005, 2006; Biber, Connor & Upton, 2007; Martin & Veel, 1998; Wignell, 1998, 2007; Williams, 1998; Swales, 1990, 2004; Flowerdew, 2002; Parodi 2005, 2006b, 2007a, b). Also, genre research has capitalized enormously from the data gathered from large corpora of complete texts in digital format sometimes available on line, thus surpassing or overcoming the study of fragmented texts or exemplary documents. A corpus approach, in other words, enables large-scale evidence-based research into language use and helps discover with reliable methods discourse variability across disciplines, registers and, of course, genres.

The increasing importance of genre variation across disciplines as an explanatory factor for diverse knowledge construction within discourse

communities has been recognized over the past years. The perception that there is no core disciplinary discourse per se and that it is better to talk about disciplinary discourses in the plural (Hyland, 2000) is becoming more accepted among researchers (Bhatia, 2004; Parodi, 2009).

The idea behind a corpus is that it represents a language variety in some domain. This makes the approach ideal for researching the prototypical and identifying features, for example, of academic and professional genres. In other words, priority is given to what frequently and typically occurs based on descriptions of the commonest uses and also based on the comparison of one genre type to others, or of genres on varying degrees of specialization, or of genres on different language modes (e.g. written or spoken). Essentially, genres are sociocognitive constructions that help group texts together and represent how writers/speakers and readers/listeners communicate and use language to interact in recurring situations. One way of characterizing genres is in terms of the communicative purposes that the writer/speaker conveys through language: e.g., to report, to describe, or to argue. Also, the participants may be specified or described in terms of the degree of specialization, the number of people included, the relation between them: e.g., specialist to specialist, a group of scientists to a large audience, or from a teacher to his or her students. Discourse organization in terms of rhetorical and linguistic recurring features also applies (e.g., narration, exposition, or argumentation).

1.1. Specialized discourse

Academic discourse (AD) and professional discourse (PD) are analyzed as part of specialized discourse (SD). SD includes a varied set of discourse genres, but with certain prototypical features. It is precisely this idea of heterogeneity of texts within a scale of gradation that Parodi (2005) applies when approaching the notion of SD. According to this notion, SD must necessarily be understood as a continuum in which texts are aligned along a diversified gradient that runs from a high degree to a low degree of specialization. Thus, SD could be conceived of as a supercategory of AD and PD.

Parodi (2005) defines SD by using a series of characterizing co-occurring linguistic features. Many researchers also agree that there are a set of

features that identify SD and many of them consider that specialized lexicon is highly important (Cabré, 1993; Burdach, 2000; Cabré, Doménech, Morel & Rodríguez, 2001; Ciapuscio, 2003; Cabré & Gómez, 2006). Academic and professional discourses are made operational through a set of texts that can be organized along a *continuum* in which the texts are linked together, from general school discourse to university academic discourse to professional discourse in a workplace environment. This is presented graphically in Figure 1.

Figure 1 illustrates a conception of discourse in academic and professional fields along a continuum that follows a process of permanent updating and multiple interactions. SD, in part, comes from AD and, in turn, is linked to and interacts with PD. This distribution of specialized knowledge organization is mainly proposed from a student's perspective, i.e. one in which the discourse continuum is traced from a learner who faces the process of instruction. In other words, this is not a researcher's or university professor's point of view because interactions would be different. For example, if research articles are considered, it is clear they overlap in academic and professional life, given that they are discourse constructions employed in both fields.

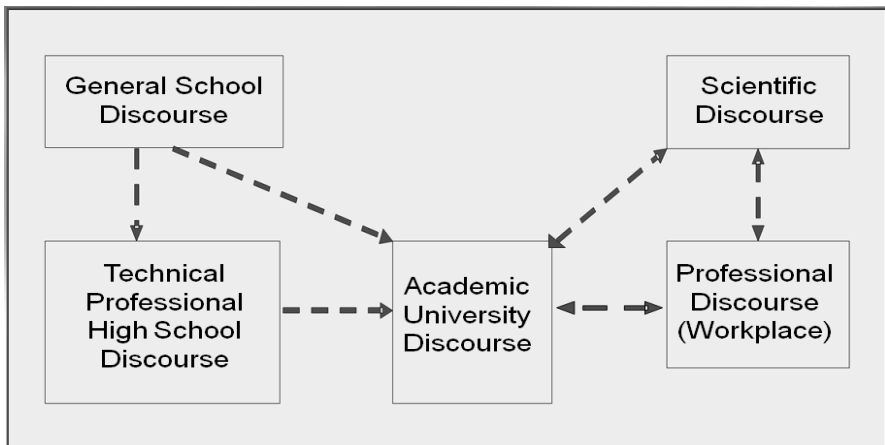


Figure 1: Continuum of texts in academic and professional fields.

1.1.1. *Academic discourse*

As Flowerdew (2002) suggests, there has been little systematic research into exactly what AD is. When one undertakes the study of this discourse genre the following questions need to be addressed: a) Are there any existing criteria that accurately define this genre? If so b) what type of criteria are they? Below three approaches to AD will be considered: 1) a functional communicative approach, 2) a contextual approach, and 3) a textual approach.

First, in functional communicative terms, AD is characterized by the predominance of description with, in some varieties, more persuasive and didactic purposes. Furthermore, it is a discourse genre that carries with it credibility and prestige because of the writer's authority of the subject matter. Second, from the contextual criteria, AD is that which is used in academic contexts or for academic purposes (Kennedy, 2001; Flowerdew, 2002; Dudley-Evans & St. John, 2006). However, it is evident that academic purposes are varied and not always easy to determine, which makes the criteria rather complex. This is due to the fact that AD does not have clear limits and may be confused or assimilated into other genres in nearby fields, such as technical-scientific, professional, pedagogic, or institutional ones (López, 2002; Flowerdew, 2004). Hyland (2000) argues that identifying the interactions and the participants involved in the interactions is indispensable. In other words, analyzing the texts as social practices is critical. This approach includes an analysis of the mediums in which these texts circulate and are used (Gunnarsson, 1997); so, AD is considered a manifestation of a specific community (Valle, 1997). Third, since AD is oriented towards the transmission of knowledge, generally through definitions, classifications, and explanations (Wignell, 1998), writers of AD use linguistic features that ensure clarity and conciseness. This manifests itself in an economy of words, an absence of empty adjectives, and the elimination of redundancy and repetition. It also has a more rigid and controlled syntax, and a higher proportion of nominalizations, than does non-academic discourse (Ciapuscio, 1992; Halliday, 1993; Lang, 1997; Gotti, 2003; Charaudeau, 2004; Parodi & Venegas, 2004; Cademártori, Parodi & Venegas, 2006).

Non-linguistic aspects frequently found in this genre, such as chemical formulas, physics equations, virtual recreations, mathematical repre-

sentations, and symbols, must also be considered. In addition, items such as graphs, tables, figures, diagrams, and other graphic representations are found in this discourse. In view of the above considerations, Lemke (1998) suggests that AD is a hybrid semiotic system. This has also been studied as multimodal texts (Kress & van Leeuwen, 2001).

1.1.2. Professional discourse

Characterizing professional discourse (PD) faces similar problems as these two terms tend to overlap, e.g. when the term PD is used in a general sense that includes AD, and vice versa. The current investigation will clearly define these two areas of discourse based on the environment in which the texts are collected. That is to say, PD will be that which is collected in contexts of professional use and circulation, while AD will be that which is collected in contexts of academic activities normally faced by university students. Nevertheless, there is an inevitable area of overlap or intersection between these two discourse genres. Therefore, our general research objectives include identifying and describing those texts that are used in both domains, and which form a nexus between the academic and professional domains.

Bazerman and Paradis (1991a) suggest that the structure of PD is founded on a textual dynamics that gives form to a profession. Bazerman and Paradis (1991b) review a series of related articles describing the way in which professional communities organize themselves based on their own relevant texts. So, PD is formed by those texts which bring together specific knowledge of the world and specific meanings that emerge from the purposes of the participants and the communicative interactions (Berkenkotter, Huckin & Ackerman, 1991; Doheny-Farina, 1991; Bathia, 1993, 2004; Christie & Martin, 1997).

Macrostructure and the superstructure have not been extensively examined by any analysis of PD. However, studies can be found that apply these categories to administrative language (MAP, 1995). López (2002) applies some rhetorical microstructures to the analysis of a text about economic policy, but there are no systematic studies with respect to how these rhetorical microstructures are distributed in each professional field. With regard to linguistic traits, the lexical level is the area that has received the

greatest attention in academic studies, especially terminological analyses associated with particular professions (Ciapuscio, 2003). At the morphological level, an issue that has been extensively investigated is the role that nominalizations perform (Chafe, 1982, 1985; Biber, 1986; Ciapuscio, 1992; Halliday, 1993; Lang, 1997; Parodi & Venegas, 2004; García, Hall & Marín, 2005; Cademártori, Parodi & Venegas, 2006).

2. The research: describing the Academic Corpus PUCV-2006

This research is the preparation and the first result of a descriptive-comparative study of linguistic-textual features, beginning with the texts that were read in the academic areas of Basic Sciences and Engineering, as well as the Social Sciences and Humanities. This was accomplished by collecting and examining an academic corpus and a professional corpus using a methodology based on corpus linguistics principles (Sinclair, 1991; Leech, 1991; Stubbs, 1996, 2006; Tognini-Bonelli, 2001; Teubert, 2005; Parodi, 2006b, 2007a, c). The academic corpus was collected from four undergraduate university programs offered by Pontificia Universidad Católica de Valparaíso, Chile: Industrial Chemistry, Construction Engineering, Social Work and Psychology. The professional texts were collected in the professional settings in which graduates of those university programs work and the criteria used reflected the workplace environment.

Concentrating on the university material, this article describes the corpus collection processes and quantitative and qualitative analysis of the PUCV-2006 Academic Corpus of Spanish (for the Professional Corpus see Parodi, 2008).

2.1. Constitution of Corpus PUCV-2006

As stated above, the aim is to collect as much of required reading and reference material for the respective university programs as possible. The specific methodology of research is divided into different stages according to the status and focus of each corpus. Table 1 summarizes the general steps followed to collect and process the corpus.

Nine steps followed to collect and process the Academic Corpus PUCV-2006
Step 1: Construction of a database with the complete curricula of the four university programs (including the syllabi of all required courses)
Step 2: Construction of a database with obligatory bibliographic references of all required courses
Step 3: Collection of complementary materials that all professors provide through prepared note files and photocopied materials
Step 4: Preparing a survey for all the professors of each of the four university programs, which included a request for the complementary materials mentioned above.
Step 5: Searching the Internet to find those titles already available in digital format, thus minimizing time spent on digitalization
Step 6: Collecting the texts from the corresponding libraries and professors
Step 7: Photocopying of each text in order to maintain a database in paper format
Step 8: Training a team of people to scan and compile all texts
Step 9: Processing all plain texts (*.txt) through tagger and parser El Grial and uploading of all texts in the online web site www.elgrial.cl

Table 1: Steps used to conduct the corpus collection and computer processing.

By following these nine steps we were ensured the creation of a database that accurately reflect the written texts to which subjects were exposed to during their university careers. The steps outlined in Table 1 correspond to general procedural activities that help construct an on-line tagged corpus available at www.elgrial.cl. These texts would be analyzed and described in detail in the following section of the article; also, from these texts the final discourse genre classification will emerge.

3. Results and discussion

3.1. *The PUCV-2006 Academic Corpus*

In this section, we define the PUCV-2006 Academic Corpus in quantitative terms as distributed among the academic disciplines and the four university programs. Also, a first classification of the texts that comprise the total PUCV-2006 Academic Corpus is given. This description defines nine text types identified and quantifies their overall occurrence in the total corpus.

In Figure 2, the proportional distribution of the total number of texts (491) of the PUCV-2006 Academic Corpus is presented.

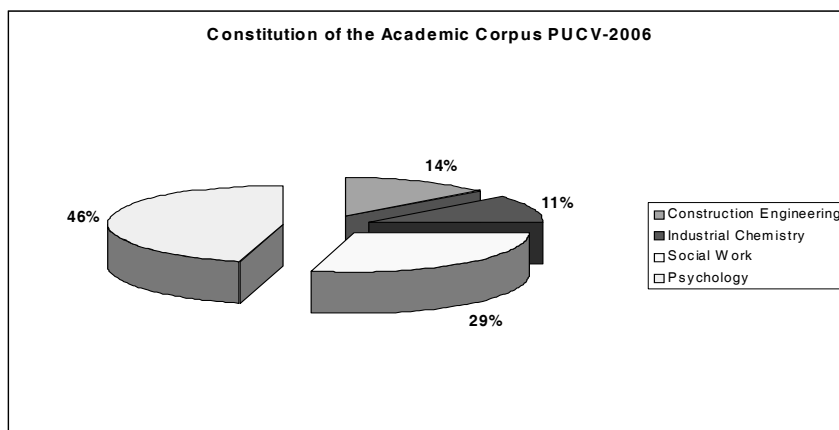


Figure 2: Constitution of the Academic Corpus PUCV-2006: number of texts.

There is a progressive increase and a substantial difference between the quantity of texts in the fields of Basic Sciences and Engineering and Social and Human Sciences, as well as a considerable difference between the specific university programs themselves. A preliminary interpretation might lead one to believe that students in the Social Sciences and Humanities are exposed to much more than students in Basic Sciences and Engineering. Psychology students would read up to four times the number of texts than Industrial Chemistry students. However, an actual word count decreases the disparity, even though Psychology students would still be exposed to more than twice the number of words as Industrial Chemistry students, as revealed by the data in Table 2.

Table 2 reveals students in Psychology and Social Work (30% and 39%) would be asked to read (in terms of number of words) more than two times as much as students in Industrial Chemistry and Construction Engineering (15% and 16%). This same comparison in terms of books is doubled, that is, it is almost four times. Therefore, as already pointed out by Parodi (2007c), there is a growing and progressive tendency based on the number of texts and the number of words, the university career, and the disciplinary domain to which they belong.

	Number of Texts	%	Number of Words	%
Psychology	227	46	22,163,379	39
Social Work	142	29	16,343,175	30
Construction Engineering	69	14	8,813,663	15
Industrial Chemistry	53	11	9,304,407	16
Total	491	100%	56,624,624	100%

Table 2: PUCV-2006Academic Corpus: Number of texts and words.

There is no other report of a written academic corpus available in the Spanish language of such dimension that is so representative and so thematically focused. A corpus of such size, close to 60 million words, in digital format, organized by subject matter and genres, becomes a fundamental tool for cutting edge research in Spanish. Adding the PUCV-2006 Academic Corpus to those pre-existing corpora at www.elgrial.cl website increases the amount of available diversified material for corpus research (more than 120.000 million words) (Parodi, 2007b).

3.2. The emerging genres

A first attempt at a more in-depth analysis of the written material collected and a classification as to the genre types follows, using the communicative-functional and textual-discursive linguistic taxonomy, as proposed by Parodi, Venegas, Ibáñez and Gutiérrez (2008). Below, in Table 3, nine genres are identified along with figures of their frequency of occurrence.

The organization of the information in Table 3 follows the alphabetical order for the names applied to each genre. Simple, everyday names in Spanish were selected; names of easy accessibility and transparency in their usage for native academic speakers of Spanish. Definitions for each of the nine genres are presented in Table 4. Some of the variables involved in

these definitions are: purpose, participants, contents, organization discourse mode, and formats.

A quite heterogeneous panorama with clear concentrations emerges from the data presented. Two genres are by far the most frequent: Disciplinary Texts (DT), with 270, and Textbooks (TB), with 126. This provides an overall initial situation that combines, on the one hand, disciplinary knowledge as presented in subject matter books concentrating high thematic specialized knowledge in each domain (DT), sometimes with a high degree of discourse complexity; and Textbooks, on the other hand, which, although oriented towards disciplinary knowledge, have a didactic and more disseminating character. TB generally use more educational resources, such as graphs, tables, diagrams, etc., in a more systematic manner and incorporate exercises and other practical applications in order for readers to access, develop and test their knowledge.

Other genres are less common. For example, Didactic Guideline (DG), although they are the third largest genre numerically, are only about a fifth as large as the two largest. Perhaps even more surprising is the low representation of Research Article (RA), which appears only about one eighth the rate of the textbooks and the disciplinary texts. It is important to note that this discourse genre of transmitting specialized knowledge would have been expected to occupy a more prominent position, particu-

Genres in the PUCV-2006 Academic Corpus	N° of texts
Lecture	1
Didactic Guideline	40
Dictionary	2
Disciplinary Text	270
Regulation	13
Report	13
Research Article	23
Test	2
Textbook	126
Total	491

Table 3: Distribution by genres.

Genres	Definitions
Dictionary	Discourse genre whose main macropurpose is to record concept definitions or procedures on a specific discipline or topic. The ideal circulation context is the pedagogical field, and the relationship between participants may be between expert writer and lay or semi-lay reader. Preferably, the descriptive mode of discourse organization is used, and multimodal resources are employed.
Didactic Guideline	Discourse Genre whose main communicative macropurpose is to instruct about a specific disciplinary topic and/or procedures. It ideally circulates within the pedagogical field, and the relationship between participants is between expert writer and lay or semi-lay reader. Preferably, an argumentative mode of discourse organization is used; and, occasionally, multimodal resources are employed.
Disciplinary Text	Discourse genre whose communicative macropurpose is to persuade on a viewpoint of one or more topics regarding a specific discipline. It ideally circulates within scientific field, and the relationship between participants is between expert writer and expert reader. Preferably, an argumentative mode of discourse organization is used. Multimodal resources are employed.
Lecture	Discourse genre whose main macropurpose is to persuade an audience within a relationship established between expert writer and expert or semi-expert reader in a scientific field. Preferably, the argumentative mode of discourse organization is employed, and multimodal resources are used.
Regulation	Discourse genre whose main communicative macropurpose is to rule behaviors and/or procedures. It circulates within a universal field, and the relationship between participants is between expert writer and expert or semi-lay reader. It is usually monomodal, and it is mainly written using a descriptive mode of discourse organization.
Report	Discourse genre whose main macropurpose is to record situations, procedures and/or problems. It ideally circulates within working field, and the relationship between participants is between expert writer and expert reader. It is usually monomodal and is written using descriptive mode of discourse organization.
Research Article	Discourse genre whose main communicative macropurpose is to persuade on a given viewpoint, which is taken based on a theoretical research or on results obtained from an empirical study. It ideally circulates within scientific field, and the relationship between participants is between expert writer and expert reader. Preferably, an argumentative mode of discourse organization is used, and it is supported by multimodal resources.
Test	Discourse genre whose main communicative macropurpose is to record psychological features of an individual. It circulates within working field, and the relationship between participants is between expert writer and lay reader. It may be multimodal, and it is preferably written using a descriptive mode of discourse organization.
Textbook	Discourse genre whose main communicative macropurpose is to instruct on concepts and/or procedures of a specialized topic. It ideally circulates within pedagogical field, and the relationship between participants is between expert writer and lay or semi-lay reader. Preferably, a descriptive mode of discourse organization is used, and multimodal resources are employed.

Table 4: Nine academic genres in the PUCV-2006 Academic Corpus of Spanish.

larly in Basic Sciences and Engineering. So, the occurrence of only 23 RA out of a total of 491 texts in the corpus indicates that it is not a common genre in undergraduate training. The genres identified in this corpus appear to be clearly concentrated in TB and DT. These findings reveal the two points of the continuum of genres, from general to specialized, as mentioned earlier. They do not, however, represent two extreme points. Both the TB and the DT are oriented towards greater specialization, but with a clear tendency towards mainstream dissemination of information. DG,

which represents the extreme point of the generalized-specialized continuum, does not appear in sufficient number to be significant in the overall corpus.

As expected, the textbook, irrespective of the discipline it is associated with, serves a clear common didactic purpose across academic settings. These kinds of texts disseminate discipline-based knowledge and are seen, in Hyland's words (2000), as "repositories of codified knowledge", which through some rhetorical structures may grant access to the most specialized professional communication. At the same time, these two text types interact with the audience in a writer-reader relationship that is appropriate to the educational and disseminating context; i.e. the writer acts as the specialist and the reader as the non-initiated student approaching a new knowledge and trying to become part of the discourse community.

Figure 3 compares the frequency of occurrence of each genre in the Social Sciences and Humanities. The resulting figure reveals a more in-depth analysis of the findings displayed in Table 3.

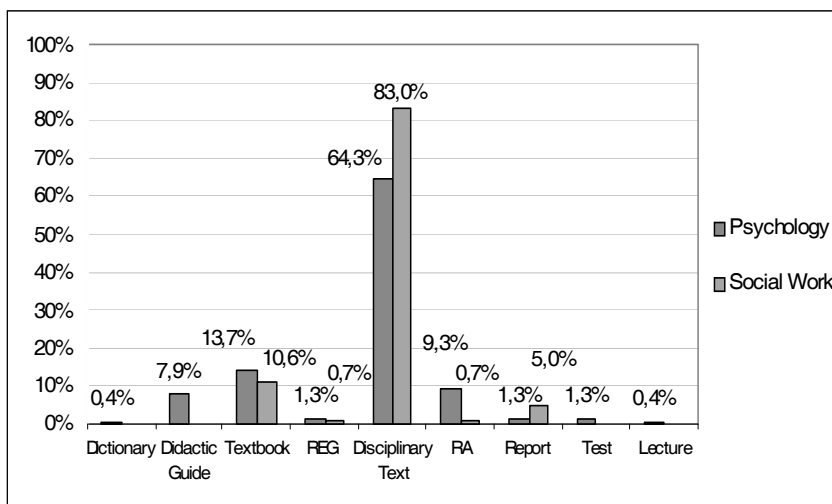


Figure 3: Genres in Social Sciences & Humanities (PUCV-2006 Academic Corpus).

As is obvious from these figures, only five common genres are detected in both careers. It is indeed noteworthy that the area with the most genre types is Psychology (9), not only in Social Sciences and Humanities, but in the total corpus including Basic Sciences and Engineering (see next Figure 4). Social Work presented only five of these genres (with an important concentration in two of them). DT and TB were the highest frequency genre types detected in both university areas of study, revealing themselves as the most common instruments of reading material students are asked to use while attending five-year university programs. This distribution clearly reflects the kind of written texts through which students access discipline-specific knowledge. These texts are those which help students acquire professional expertise and become part of the academic community to which they will eventually belong. In the case of Psychology, seven genres show relatively low occurrence: Research Article (9,3%), Regulation (1,3%), Lecture (0,4%), Didactic Guideline (7,9%), Report (1,3%), Test (1,3%), and Dictionary (0,4%). Although they are part of the kind of readings students engage in during university training, they contribute in only a minority way to student's text exposure.

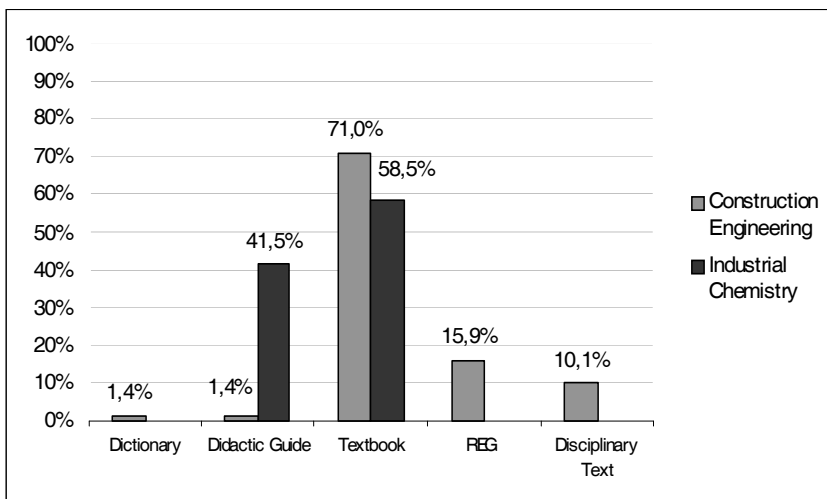


Figure 4: Genres in Basic Sciences & Engineering (PUCV-2006 Academic Corpus).

Similarly, we now compare the percentages of occurrence of text types in Basic Sciences and Engineering. Figure 4 shows the findings expressed in percentages.

This quantitative analysis reveals interesting differences of internal genre variability not found when comparing Social Sciences and Humanities. Not only are there fewer genre types in Industrial Chemistry and Construction Engineering (as shown in Table 2), but there is less genre diversity, as seen in Figure 4.

Only five text types were identified in Construction Engineering, with TB dominating the distribution. There are four other genres that are part of the collected corpus: two are more closely related to the professional workplace (Regulation and DT), and two more typical of academic environments (DG and Dictionary). It is worth noting that in Industrial Chemistry only two genres were collected: DG and TB. This was unanticipated. There is a clear reader-oriented focus, recognizing the dialogic dimension of disciplinary instruction and directing readers to some action and understanding of the truths and facts under study. These two genres represent important academic tools that open pathways to knowledge to novice students.

This variety of reading materials depicts the most common academic writing students must encounter in their daily university discourse activities. The primary genre identified in this research is one oriented to disseminating knowledge. While this genre uses a disciplinary prose, also combines instructional devices such as examples, diagrams, and problem solving exercises. TB is highly employed in Construction Engineering and Industrial Chemistry (71% and 58,5%), while DT appears most frequently in Social Work and Psychology (83% and 64,3%).

Figures 3 and 4 illustrate that the comparisons favored genres that disseminate knowledge (TB and DG), as well as more highly specialized genres (DT). Taken together, these figures reveal an important pattern of situating and distributing academic genres in these four university fields of study. This provides evidence that the social, cognitive and discourse interactions of members of these academic communities help shape their interactions, with the written material. It also yields information about the way a university organizes its academic curricula.

3.3. *Academic discourse: An emerging mixed discourse*

As mentioned at the beginning of the article specialized discourses within disciplines tend to show a relatively compact and homogenous unit and the identification of nine discourse genres through four disciplinary domains emphasizes that academic discourse shows differences in the discourse construction of knowledge in each discipline.

Figure 3 illustrates four types of genre, based upon the classification proposed by Parodi (2008 applying two of the different criteria proposed: “ideal context of circulation” and “communicative macropurpose” serving the context for the texts for texts of that genre From this point of view, it would be possible to distinguish more specialized contexts, of restricted circulation, as well as others that are more ample and general. Regarding macropurpose, this could be considered the best criterion for a general objective fulfilled by each genre. However, in our opinion, one unique criterion is not enough to capture the deep essence of a genre. From these two criteria, it is possible to operate scientific, labour, universal and academic genres, as is shown in the following figure.

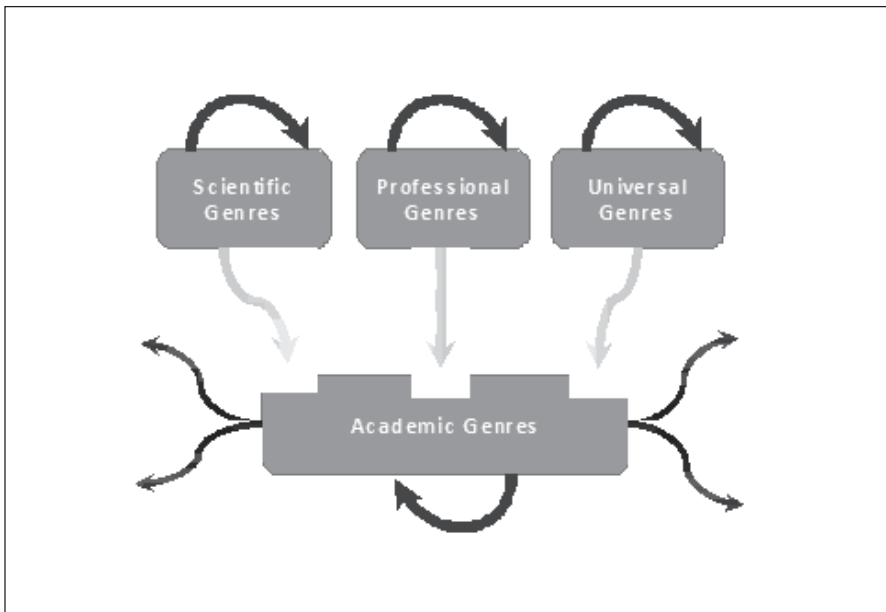


Figure 3: Interactions among genres.

Rather than focusing on the identification processes of these groups of genres, what Figure 3 aims to capture is the idea that, in each particular discourse community, there circulate genres that have been produced within that context in order to satisfy their emerging communicative demands. This idea is the result of the two abovementioned criteria. It is important to emphasize that the set of genres named “academic” due to the fact of having – in this case – been collected in university settings, have not always been produced in this same particular context nor do they possess original communicative macropurposes of these discourse communities. The unidirectional arrows, coming from the groups of scientific, labour and universal genres, show that, from these genres, it is possible to supply discourse tools for the constitution of academic genres. All this means that, although within the academic context specific genres may be produced, this context also is nurtured by other genres that have not originally been created to satisfy specific demands of this university setting. Undoubtedly, it is also feasible that genres elaborated within university academic settings circulate within scientific, labour and universal contexts. Nevertheless, it is possible that the communicative macropurposes may be very similar. In fact, in Figure 3, the arrows emerging from the academic genres illustrate this fact.

The above infers the constitution of a type of discourse that is very heterogeneous in its nature, in the sense that many of the genres that circulate within the academic context could certainly not be originated from this setting. This might not be a surprise or constitute an important scientific event in itself. However, in our opinion, it does unveil theoretical and applied research niches of high interest from linguistic, cognitive, social, and educational perspectives. The empirical findings indicate, for example, that in disciplines such as PSI, nine diverse genres are detected in the academic training of future professionals, whilst in ICH only two genres are identified. Thus, it becomes highly relevant to know in detail whether these genres are academic and if –for example– their original communicative macropurposes point towards the satisfaction of discourse exchanges or come from and has been originally created for other contexts and in order to fulfill other macro-purposes. For example, the Disciplinary Text and the Test are not genres originated in the academic context nor do they have initial purposes of the pedagogic kind. The educational and psycholinguistic processing implications are, at least, some of the issues that can be initially observed.

Following the characterization proposal of Parodi (2008), Figure 3 illustrates the manner in which the scientific, labour and universal genres are distributed and how they are imbedded in the **Academic Discourse** and thus come to belong to the group of academic genres. At the same time, in this figure the proper academic genres are consigned (Didactic Guideline, Textbook and Dictionary), i.e., this means that in their origin they emerge from a pedagogic context.

As observed in Figure 3, there are six genres which, according to this research data, form an integral part of the academic discourse: RA, LECT, DT, REP, TEST and REG. Nevertheless, according to our characterization, they have been originally created within other contexts and, in order to comply with other communicative purposes, i.e., not directly for those pertaining to academic instructing tasks. As may be inferred from this analysis, the academic discourse identified in this research from a set of nine written genres is clearly a **mixed discourse**. By this, we wish to draw the attention to this exceptional discourse constitution and prompt a first data to emerge in respect of the heterogeneous nature of academic discourse. By unveiling this mixed nature, certainly not in its origin but in its realization configuration, we emphasize the diversity of the academic discourse in university settings.

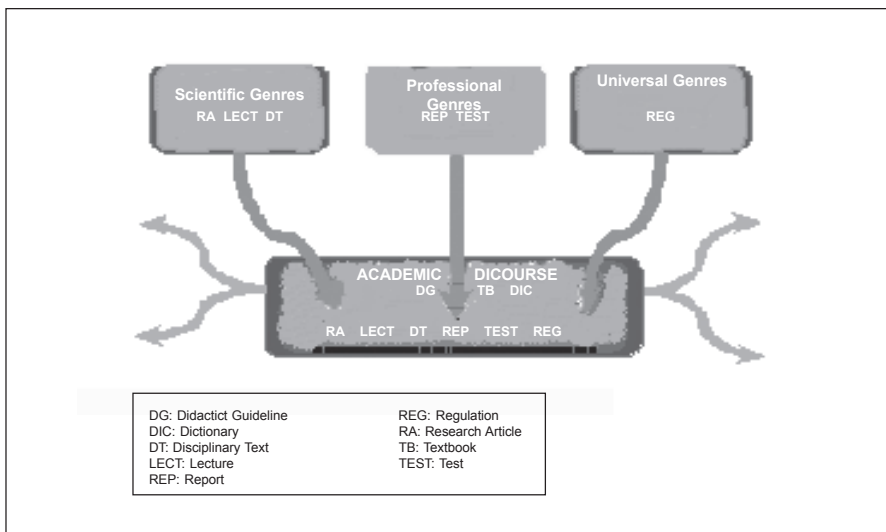


Figure 4: Genres constituting the Academic Discourse.

4. Summary and conclusions

The general results showed a genre classification of nine academic types. Some expected cross-discipline similarities are detected, but most interesting are the inter-disciplinary variations, where, for example, Industrial Chemistry and Psychology are at the extreme poles of the continuum. In general, more disseminating-reader-oriented genres were found in the fields of Basic Sciences and Engineering, with a particular high frequency of Didactic Guideline and Textbook (especially in Industrial Chemistry). Social Sciences and Humanities showed a richer variety of genres, but with a mayor concentration in disciplinary-specific perspectives with less emphasis on didactic resources (important occurrence of Disciplinary Text).

With regard to the quantitative and qualitative analysis of the academic corpus PUCV-2006, the data presented in the areas of the Social Sciences and Humanities and in the Basic Sciences and Engineering reveal differences both in the number and variety of written genres. It is evident that Psychology and Social Work tend to employ a greater quantity of texts with a relatively more extensive range (at least in the number of words) during the course of their university programs, when compared to Industrial Chemistry and Construction Engineering, which use a smaller number of texts and more limited range of texts in terms of the number of words.

The current study identified an interesting heterogeneity of genres, which would confirm the validity of the idea of a continuum of genres between poles of higher discipline-specific complexity on one end and a focus on teaching on the other.

Empirical findings have helped to design a characterization of the academic discourse, which emerges as a highly heterogeneous discourse wherein genres coming from other contexts are embedded, which were not originally created for those settings. It is also clear that very probably, the original communicative purposes of these genres (scientific, labour, universal) do not correspond with those of their present circulating context. Due to all this, we have named this group of diverse genres as **mixed discourse**.

We believe the corpus data collected is reliable in sketching a first approach to the study of genres across academic disciplines, in order to

report and describe the actual genres that are read by students at four university careers. All of this has followed ecological and situated principles in particular contexts and in one particular tertiary academic institution in Chile.

In short, the reported findings in this study constitute a preliminary approximation to the distribution of discourse genres through four disciplines in university domains, from a “corpus-based” focus. Undoubtedly, the analysis from this perspective, supplies benefits which a case study could not render in the same manner. The revealing differences between the areas of knowledge under study cause discipline to emerge as a vital factor in forming the way of constructing knowledge by means of written discourse genres. This means that this study proves that, belonging to one discipline or another, in university education, implies a divergent discourse access for the students. This causes the nature of the discipline to determine very particular associated pedagogic mechanisms and protocols of the prototypic analytical procedures that originate a certain way of constructing science.

From the information presented in this article, we see that through corpus linguistics it is possible to provide rich, accurate descriptions of language use in institutionalized contexts; also, the findings have helped gained insights into the ways that discourse practices vary across disciplines. Thus, we are beginning to realize and understand that the texts employed as reading material in one academic field (*hard sciences*) are not the same than in others (*soft sciences*). The data obtained from corpus-based research of written specialized discourse can provide valuable contributions in the areas of disciplinary reading and writing processes at university level and of specialized material development. The use of corpora and computers allows researchers to explore complex issues related to disciplinary written genres from ecological perspectives including large amounts of texts, and all this collected information can contribute to the knowledge-base used to shape better access and paths to disciplinary discourse communities and to help university readers engage in specific genres as part of their academic and professional lives.

Recebido em julho de 2008
Aprovado em fevereiro de 2009
E-mail: gparodi@ucv.cl

REFERENCES

- BAZERMAN, Ch. (Ed.) (2008). *Handbook of research on writing. History, society, school, individual, text*. New York: Erlbaum.
- BAZERMAN, Ch. & PARADIS, J. (1991a). Introduction. In Ch. BAZERMAN & J. PARADIS (Eds.), *Textual dynamics of the professions* (pp. 3-12). Wisconsin: The University of Wisconsin.
- _____. (1991b). *Textual dynamics of the professions*. Wisconsin: The University of Wisconsin.
- BEAUFORT, A. (2007). *College writing and beyond*. UTAH: Utah State University Press.
- BERKENKOTTER, C., HUCKIN, T. & ACKERMAN, J. (1991). Social context and socially constructed texts: The initiation of a graduate student into a writing research community. In C. BAZERMAN & J. PARADIS (Eds.), *Textual dynamics of the professions* (pp. 191-215). Wisconsin: The University of Wisconsin.
- BHATIA, V. (1993). *Analysing genre: language use in professional settings*. London: Longman.
- _____. (2004). *Worlds of written discourse. A genre-based view*. London: Continuum.
- BIBER, D. (1986). Spoken and written textual dimensions in English: Resolving the contradictory findings. *Language*, 63, 384-414.
- _____. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- _____. (1994). Using register-diversified corpora for general language studies. In S. ARMSTRONG (Ed.), *Using large corpora* (pp. 180-201). Cambridge: The MIT Press.
- _____. (2005). Paquetes léxicos en textos de estudio universitario: Variación entre disciplinas académicas. *Revista Signos. Estudios de Lingüística*, 38, 19-30.
- _____. (2006). *University Language. A corpus-based of spoken and written registers*. Amsterdam: Benjamins.
- BIBER, D., CONNOR, U., & UPTON, T. (2007). *Discourse on the move. Using corpus analysis to describe discourse structure*. Amsterdam: Benjamins.
- BURDACH, A. (2000). El léxico científico técnico: Un recurso publicitario persuasivo. *Onomazein*, 6, 189-208.
- CABRÉ, T. (1993). *La terminología. Teoría, metodología, aplicaciones*. Barcelona: Antártica/Empuréis.

- CABRÉ, T. & GÓMEZ, J. (2006). *La enseñanza de los lenguajes de especialidad. La simulación global*. Madrid: Gredos.
- CABRÉ, M., DOMÉNECH, M., MOREL, J., RODRÍGUEZ, C. (2001). Las características del conocimiento especializado y la relación con el conocimiento general. In M. CABRÉ; J. FELIÚ (Eds.), *La terminología técnica y científica* (pp. 173-186). Barcelona: Instituto Universitario de Lingüística Aplicada.
- CADEMÁRTORI, Y., PARODI, G. & VENEGAS, R. (2006). El discurso escrito y especializado: caracterización y funciones de las nominalizaciones en los manuales técnicos. *Literatura y Lingüística*, 10, 243-265.
- CHAFE, W. (1982). Integration and involvement in speaking, writing and oral literature. In D. TANNEN (Ed.), *Spoken and written language: Exploring orality and literacy* (pp. 35-53). Norwood, N.J.: Ablex.
- _____. (1985). Linguistic differences produced by differences between speaking and writing. In D. OLSON; N. TORRENCE & A. HIDYARD (Eds.), *Literature, language and learning: The nature and consequences of reading and writing*. Cambridge: Cambridge University Press.
- CHARAUDEAU, P. (2004). La problemática de los géneros. De la situación a la construcción textual. *Revista Signos. Estudios de Lingüística*, 37, 23-39.
- CHRISTIE, F. & MARTIN, J. (Eds.) (1997). *Genre and institutions. Social Processes in the workplace and the school*. London: Continuum.
- CIAPUSCIO, G. (1992). Impersonalidad y desagentivación en la divulgación científica. *Lingüística española actual*, 14, 183-205.
- _____. (2003). *Textos especializados y terminología*. Barcelona: Instituto Universitario de Lingüística Aplicada.
- DOHENY-FARINA, S. (1991). Creating a text/creating a company: The role of a text in the rise and decline of a new organization. In C. BAZERMAN & J. PARADIS (Eds.), *Textual dynamics of the professions* (pp. 306-335). Wisconsin: The University of Wisconsin.
- DUDLEY-EVANS, T. & ST. JOHN, M. (2006). *Developments in English for Academic Purposes. A multidisciplinary approach*. Cambridge: Cambridge University Press.
- FLOWERDEW, J. (Ed.). (2002). *Academic discourse*. Cambridge: Cambridge University Press.
- FLOWERDEW, L. (2004). The argument for using English specialized corpora to understand academic and professional language. In U. CONNOR & T. UPTON (Eds.), *Discourse in the professions. Perspectives from corpus linguistics* (pp. 11-33). Amsterdam: Benjamins.

- GARCÍA, M., HALL, B. & MARÍN, M. (2005). Ambigüedad, abstracción y polifonía del discurso académico: Interpretación de las nominalizaciones. *Revista Signos. Estudios de Lingüística*, 38, 49-60.
- GOTTI, M. (2003). *Specialized discourse. Linguistic features and changing conventions*. Bern: Lang.
- GUNNARSSON, B. (1997). On the sociohistorical construction of scientific discourse. In B. GUNNARSSON, P. LINELL & B. NORDBERG (Eds.), *The construction of professional discourse* (pp. 99-126). Essex: Longman.
- HALLIDAY, M. (1993). On language and physical science. In M. HALLIDAY & J. MARTIN (Eds.), *Writing science. Literacy and discursive power* (pp. 54-68). Pittsburgh: University of Pittsburgh Press.
- HERRINGTON, A. & MORAN, C. (2005). *Genre. Across the curriculum*. UTAH: Utah State University Press.
- HYLAND, K. (2000). *Disciplinary discourses: Social interactions in academic writing*. London: Longman.
- KENNEDY, J. (2001). Language use, language planning and EAP. In J. FLOWERDEW & I. PEACKOC (Eds.), *Research perspectives on English for Academic Purposes* (pp. 67-92). Cambridge: Cambridge University Press.
- KRESS, G. & VAN LEEUWEN, T. (2001). *Multimodal discourse*. London: Arnold.
- LANG, M. (1997). *Formación de palabras en español*. Madrid: Cátedra.
- LEECH, G. (1991). The state of the art in corpus linguistics. In K. ALJMER & B. ALTENBERG (Eds.), *English Corpus Linguistics. Studies in honor of Jan Svartvik* (pp. 8-29). London: Longman.
- LEMKE, J. (1998). Multiplying meaning. Visual and verbal semiotics in scientific text. In J. MARTIN & R. VEEL (Eds.), *Reading science. Critical and functional perspectives on discourses of science* (pp. 132-149).. London: Routledge.
- LÓPEZ, C. (2002). Aproximación al análisis de los discursos profesionales. *Revista Signos*, 35, 195-215.
- MARTIN, J. & VEEL, R. (Eds.) (1998). *Reading science. Critical and functional perspectives on discourses of science*. London: Routledge.
- Ministerio de Administraciones Públicas (MAP) (1995). *Manual de documentos administrativos*. Madrid: MAP.
- PARODI, G. (Ed.) (2005). *Discurso especializado e instituciones formadoras*. Valparaíso: EUV.
- PARODI, G. (2006a). Reading-writing connections: Discourse-oriented research. *Reading & Writing Interdisciplinary Journal*, (20): 225-250.

- _____. (2006b). Discurso especializado y lengua escrita: Foco y variación. *Estudios Filológicos*, 52, 165-204.
- PARODI, G. (Ed.) (2007a). *Working with Spanish Corpora*. London: Continuum.
- PARODI, G. (2007b). El Grial: Interfaz computacional para anotación e interrogación de corpus en español. In G. PARODI (Ed.), *Lingüística de Corpus y discursos especializados: Puntos de mira* (pp. 31-52). Valparaíso: EUV.
- _____. (2007c). El discurso especializado escrito en el ámbito universitario y profesional: Constitución de un corpus de estudio. *Revista Signos. Estudios de Lingüística*, 63, 147-178.
- _____. (2009). *Lingüística de Corpus*. Frankfurt: Iberoamericana.
- PARODI, G. (Ed.) (2008). *Géneros académicos y géneros profesionales: Accesos discursivos para saber y hacer*. Valparaíso: EUV.
- PARODI, G. & VENEGAS, R. (2004). BUCÓLICO: Aplicación computacional para el análisis de textos. Hacia un análisis de rasgos de la informatividad. *Lingüística y Literatura*, 26, 223-251.
- PARODI, G., VENEGAS, R., IBÁÑEZ, R. & GUTIÉRREZ, R.M. (2008). Los géneros discursivos del Corpus PUCV-2006 del español: Criterios y definiciones. In G. PARODI (Ed.), *Géneros académicos y géneros profesionales: accesos discursivos para saber y hacer*. Valparaíso: EUV.
- SINCLAIR, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- STUBBS, M. (1996). *Text and Corpus Analysis*. Oxford: Blackwell.
- _____. (2006). Corpus analysis: The state of the art and three types of unanswered questions. In S. HUNSTON & G. THOMPSON (Eds.), *System and corpus: Exploring connections* (pp. 15-36). London: Equinox.
- SWALES, J. (1990). *Genre analysis. English in academic and research settings*. Cambridge: Cambridge University Press.
- _____. (2004). *Research genres. Explorations and applications*. Cambridge: Cambridge University Press.
- TEUBERT, W. (2005). My version of corpus linguistics. *International Journal of Corpus Linguistics*, 10, 1-13.
- THAISS, C. & ZAWACKI, T. (2006). *Engaged writers dynamic disciplines*. Portsmouth: Boyton/Cook.
- TOGNINI-BONELLI, E. (2001). *Corpus linguistics at work*. Amsterdam: Benjamins.
- VALLE, E. (1997). A scientific community and its texts: A historical discourse study. In B. GUNNARSSON; P. LINELL & B. NORDBERG (Eds.), *The construction of professional discourse* (pp. 76-98). Essex: Longman.

- WIGNELL, P. (1998). Technicality and abstraction in social science. In J. MARTIN & R. VEEL (Eds.), *Reading science. Critical and functional perspectives on discourses of science* (pp. 297-326). London: Routledge.
- _____. (2007). *On the discourse of social science*. Australia: Charles Darwin University Press.
- WILLIAMS, I. (1998). Collocational networks: Interlocking patterns of lexis in a corpus of plant biology research articles. *International Journal of Corpus Linguistics*, 3, 151-171.