CHAPTER 2

Describing the domain of academic journal writing

2.1 Introduction

A great deal of research on academic writing acknowledges that different disciplines use language in different ways. This is evidenced by the abundance of research that compares disciplines (as in many of the studies mentioned in Chapter 2), and by the many "how to" books like Zeiger's (1999) *Essentials of Writing Biomedical Research Papers* or Robinson, Stoller, Costanza-Robinson, and Jones's (2008) volume *Write like a Chemist* that focus on offering detailed descriptions within a single discipline. This type of research operates on the belief that each discipline follows its own discourse conventions and patterns.

However, the summary of literature focusing specifically on research articles published in academic journals (Chapter 1) has shown that very little research accounts for variation in the *types* of journal articles published both within and across disciplines. As a consequence, almost no research has documented linguistic variation that occurs due to differences in article type.[1] In using the term 'article type,' I put forth the claim that research articles can be described according to situational characteristics that go beyond the principal features of being written by professionals and experts within a field, published in professional journals, with an informational purpose of relating the results of research. For example, we can also characterize research articles based on research methodologies and ask questions that link these specific characteristics of articles with linguistic patterns, such as 'do empirical studies which report on qualitative research use the same linguistic features to the same extent as studies which report on quantitative research?' By explicitly considering the types of articles that are produced within and across

1. This is not to say that this body of research does not recognize different registers within an academic discipline. For example, *Write like a Chemist* considers the written registers of journal articles, conference abstracts, posters, and research proposals. Rather, these research studies and how-to volumes typically do not consider differences in the *types* of journal articles, or variation within journal articles.

disciplines, linguistic analyses can both describe important patterns of variation that have been unidentified in the research to date, as well as offer more comprehensive explanations of variation that go beyond simple accounts of disciplinary differences.

However, in order to analyze linguistic variation across article types, the first step is to carry out an analysis of the target domain, to identify the range of possible article types and describe the non-linguistic (i.e., situational) features that distinguish different types of research articles. The basic premise here is that article types differ in terms of certain characteristics, and these differences allow us to consider sub-registers within the texts published in academic journals. Because the goal of this study is to break away from the traditional and somewhat monolithic term 'research article,' the purpose at this stage of the research is to survey academic journals in many disciplines in order to document the full range of publication types (including non-empirical research) that occur in these journals, along with the basic characteristics that define these article types. The idea is to situate the often-studied 'research article' within the larger context of academic journal publishing, as well as identify systematic categories within 'research articles'. To do so, a careful consideration of what features identify a particular article as belonging to a certain type is also needed, as the ultimate goal is to represent these article types in a corpus. Thus, the first objective of this chapter (Section 2.2) is to establish a taxonomy that identifies the range of possible text types (registers), and enables the categorization of academic journal articles into different registers according to situational characteristics (such as the type of research that is being reported on).

This target domain description serves as the foundation for corpus design, helping to ensure that a corpus is representative. Two types of corpus representativeness are important relative to corpus design and the interpretation of corpus-based findings. External or situational representativeness is the most commonly addressed type. The second type is internal or linguistic representativeness (McEnery et al. 2006; Biber 1993), and refers to the stability of the linguistic findings from the corpus: whether the corpus represents the range of linguistic variability in the target domain (Biber 1993: 243). Biber (1993) argues that internal/linguistic reliability only occurs if external representativeness has been achieved; thus, this type of situational analysis that describes the target register and informs corpus design is a much-needed step in attaining corpus representativeness – both external/situational and internal/linguistic representativeness.[2]

After the range of publication types have been identified and described in a way that allows for the categorization of an individual article within the larger

2. It is important to note that achieving external/situational representativess does not necessarily mean that a corpus is internally/linguistically representative.

framework (i.e., to identify the register that a particular text belongs to), disciplines can be described in terms of their publication patterns for these registers. In turn, such a description can inform the design of a corpus that can be used to investigate disciplinary and register influences on linguistic variation. Thus, the second objective (Section 2.3) is to identify how the use of these various article types varies across disciplines. This analysis can then inform the design and construction of the corpus that will serve as the foundation for the research.

2.2 Surveying the domain of disciplinary journal writing

In order to develop a taxonomy of research article types, a survey of journals in 11 varied disciplines was undertaken to develop descriptions of article types and identify the use of these article types across a range of disciplines. As mentioned above, the motivation for this type of taxonomy is two-fold. First, if different types of articles published in academic journals are to be considered distinct registers, then there must be aspects of the situational characteristics of these articles that differ, and a taxonomy allows us to systematically describe these differences. The second motivation is methodological: if the goal of the research is to empirically investigate linguistic variation across these registers, then a taxonomy based on non-linguistic features of these texts is needed in order to design and build a corpus to represent a particular register. That is, corpus builders need a reliable method for categorizing texts into register categories so that the corpus is a suitable representation of that register. In this section I describe the process used to develop a taxonomy for defining registers published in academic journals.

2.2.1 Procedures

The first step in the development of the taxonomy was to develop a list of possible article types based on my own background knowledge as a reader of academic journals and as a trained researcher. Along with this list, a preliminary or draft inventory of distinguishing characteristics that an article of a particular type might exhibit (such as presence or absence of observed data, type of data, and so on) was compiled. These lists served as starting points for the analysis in that they allowed texts to be grouped into broad categories that could then be refined based on the second stage of taxonomy development: an inductive survey of journals in a range of disciplines.

The inductive survey involved reading carefully but widely in four disciplines (chemistry, economics, sociology, and philosophy), categorizing as many articles as possible in four to seven journals in each discipline. Journals were identified based on their placement in topic groupings in the periodical section of the

library, and were selected from those published in 2001 or later. Journal descriptions enabled the sample to be restricted to peer-reviewed journals, and to ensure the inclusion of both 'general' and more specialized journals within each field. The 'general' journals were included systematically because they covered a wide range of topics within a field. Other journals examined during the survey represented publications focused on a range of more specific sub-topics or areas. The journal description also helped to identify journals which might publish different types of articles. For example, most journal descriptions included a statement of the types of research that they accept for publication.

These selection guidelines were chosen based on a desire to describe the current state of respectable academic research, and to capture as many different types of writing as possible. For example, in sociology, the journals *Theory and Society* and *Qualitative Sociology* were reviewed because their titles and journal descriptors indicated the presence of theoretical and qualitative research articles respectively. In economics, the journal *Quarterly Journal of Economics* was included in the survey based on its description as covering "all aspects of the field," while *Computational Economics* was selected based on its focus on a subfield not often mentioned in other journal descriptions, and the statement in its Aims and Scopes that it publishes three specific types of articles: state of the art reports, brief reports, and critical reviews.

All journals examined in each discipline are listed in Appendix A. For this initial survey, every article in one issue of each journal was examined (however, "editorials" written by the editor of the journal, book announcements, or obituaries were not included). The following aspects/portions of the text served as the basis for the analysis:

1. the title
2. the abstract, if present
3. a goal/purpose statement, typically located in the first few pages of the article
4. internal headings
5. descriptions of data and/or procedures, if present
6. textual aspects like the presence/absence of data tables and figures, formulas
7. labels assigned to the text by the journal itself (e.g., *article, commentary, book review, note,* etc.)
8. key words throughout the article that described the study as dealing with data and methods, such as *experiment, treatment, survey, interviews, case study, observational,* and so on.

Based on the reading at this stage, the taxonomy categories and operational definitions were revised to include additional categories and descriptors. The second stage involved reading in 2–4 journals in additional disciplines (geology, physics,

applied linguistics, psychology, political science, pediatric medicine, and general and civil engineering). I applied the revised framework to articles in those journals in order to evaluate the applicability of the taxonomy to disciplines not included in the initial survey. The journals examined in this stage are also listed in the Appendix A. The next section contains the revised general taxonomy for article types along with the operational definitions, as well as an illustration of how the taxonomy was applied to categorizing texts.

2.2.2 A taxonomy of academic journal registers

The taxonomy of published journal writing developed for this study has three 'meta' article types: empirical research reports, theoretical articles, and evaluative documents. Empirical research reports are those that analyze some type of observed data. The term 'empirical' is used here in a traditional sense, so that it encompasses both quantitative data and qualitative data. However, as I will discuss a bit later, a major question in categorizing texts are distinct disciplinary conceptualizations of what constitutes 'observed data.' Theoretical articles are those that discuss matters of theory within the discipline, and a key feature of these is that they do not analyze any empirical data. Rather, they focus on explicating and extending key premises in the discipline. Evaluative texts are those whose primary purpose is to offer critique or summary of the state of the field, a particular article, book or product. Each of these three types contains subsets of texts within them, and each subtype is described in terms of its distinguishing characteristics (or operational definition) in Table 2.1 below.

Empirical research reports include quantitative, qualitative, and mixed methods research. Quantitative research includes experimental/quasi-experimental research as well as observational research. Qualitative research analyzes any observed data which is not quantitative in nature and encompasses a variety of research methods and data types, such as ethnographies, case studies, focus groups, interviews, and field observations. Mixed methods research is a term reserved for studies which use a nearly equal focus on quantitative and qualitative data analysis, and only occurs in disciplines where both quantitative and qualitative research is already conducted.

Theoretical articles can also be one of three subtypes. Some articles focus on explicating details of a famous scholar's work and ideas on a theory, and this constitutes what I term 'author interpretation' articles. Other theoretical articles are logic-based and rely on formal statements of logic to parse through theoretical constructs. All other theoretical articles are grouped under a 'general theoretical' article subtype.

Evaluative texts can be of three main types. The first type, which is fairly common, is the book or product review. These texts are typically shorter than full

articles, and focus on summarizing a book or product while incorporating a degree of evaluation/critique. The 'commentary/forum' evaluative texts focus on critique or put forward an argument on an issue in the field, and they often take the form of a forum where scholars with opposing viewpoints each present a critique/argument. The third type is a review article which synthesizes current research in a particular area of the field without presenting new data analysis. In this type of article the focus is on the synthesis and not necessarily on evaluation.

The operational definitions presented in Table 2.1 were formed based on prior knowledge and on the analysis of the texts in the journals listed in Appendix A. In order to illustrate how this worked, several examples are provided below. As mentioned above in the methodology section, many pieces of evidence in each text were examined, from the title to internal headings, to the presence of tables or formulas in the text, to purpose statements, and so on.

The first piece of evidence examined was how a text was labeled (if labeled by the journal) in the table of contents. For those labeled "commentaries," "notes," and "book reviews" (labels that clearly fit within the evaluative type of article), several additional features from the operational definitions were confirmed before the texts were labeled with the corresponding sub-type of evaluative texts. For example, book reviews are a very straight forward type of text to determine. Not only are they always labeled 'book review', but across disciplines, they typically begin with a citation to the book or product being reviewed (rather than a creative title), they are usually 2–3 pages long, and have a focus on summary that is illustrated by the fact that the text usually contains a lot of markers like "in Chapter 4" and "the focus of the second part of X's book is….".

For texts labeled as "article," the analysis was a bit more complex. Often the title of the text gave a first indication of the type of article. For example, the article "Wittegenstein on Metaphysical/Everyday Use" (Baker 2002) follows a typical pattern for a theoretical article title that is an author interpretation. That is, the title follows the pattern "X on Y." In addition, the statement of the paper's aim or goal explicitly identifies the purpose of the text as an interpretation of Wittegenstein's claims: the author states that he "shall make a case for a very different reading of this remark" (Baker 2002: 289). A further look through the article reveals a heavy focus on the author's ideas. All of this evidence confirms that Baker's article should be labeled 'theoretical: author interpretation.'

Empirical articles were identified by the presence of observed data, and a data mention could appear in the abstract or in a section labeled "Data" or something similar. Take, for example, the article "The long arm of the law: Effects of labeling on employment" (Davis & Tanner 2003). Here, the title indicates that it is probably a quantitative analysis (the use of "effects of X on Y"). Within the abstract, the authors state that they will use "the National Longitudinal Survey of Youth, a large and nationally representative sample, to examine …" (Davis & Tanner 2003: 385).

Table 2.1. Operational definitions for the text taxonomy

	Article Type	Content Operational Definition	Textual/Genre Features Operational Definition
Empirical *Analyzes observed data*	*Quantitative: Experimental*	– analyzes numerically-based data – object of study is manipulated in some way, either physically or through a 'treatment' – includes quasi-experimental – usually includes comparison of groups	– generally includes a 'experimental' or 'procedure' section – usually includes tables, figures illustrating quantitative analysis – key words: *experiment, control, group, laboratory setting, treatment, procedure*
	Quantitative: Observational	– analyzes numerically-based data – object of study is not submitted to any type of treatment or manipulation – data comes from a variety of sources depending on discipline, e.g., survey or demographic data, test scores, observations from nature	– usually includes a description of the data – usually includes tables, figures illustrating quantitative analysis – key words: *survey, demographic, measure*
	Qualitative	– observational in nature – does not analyze quantitative data – typically does not manipulate/apply treatment conditions to get data	– usually includes description of data – only limited statement of 'procedure' or methodology – key words: *ethnography, longitudinal, interviews, focus groups*
	Mixed Methods	– uses both quantitative and qualitative methods with an equal focus	– has features of both qualitative and quantitative research reports – key words: *qualitative/quantitative or mixed methods*

Table 2.1. (Continued) Operational definitions for the text taxonomy

(*Continued*)

	Article Type	Content Operational Definition	Textual/Genre Features Operational Definition
Theoretical *No observed data is analyzed. Advances theory within field.*	*General*	– discusses/advances a theoretical aspect of the field	– lack of distinguishing features of other types (e.g., not labeled 'review,' no data description or methods section)
	Author Interpretation	– comprehensive and in-depth description/explication of one author's ideas/theories on a particular issue	– title usually includes "X on Y" – internal headings often include author's name, or first sentences if no headings
	Logic-based	– uses formulas to advance logic, but no data	– includes a progression of formulas within text of analysis
Evaluative *Offers critique of state of field, issue, article, book, or product*	*Commentary/ Forum*	– Presents critique/evaluation of state of field, issue within the field, or a particular article – Focus is on critique with less summary	– typically labeled 'commentary', 'discussion note' – fewer references than a theoretical article – begins with statement introducing article to be critiqued – initial critiques have descriptive, clear title, e.g., "A critique of X" – response typically titled "A Reply to X"
	Synthesis/ Review	– focus is on synthesizing what is known in the field or recent research on a particular area	– often termed 'review articles'
	Book/Product Review	– offers summary and evaluative comments regarding a book or product – focus is on summary, while critique is there but often backgrounded	– typically shorter (2–3 pages) than articles – title usually a citation to reviewed book – usually at end of journal issue – labeled "Book Review" by journal/title – typically have an internal structure of summary marked by adverbial phrases like "In Chapter 2..." or "Ch. 2 deals with..."

Also indicative that this text is an empirical article is the internal headings "Research Questions and Methods" and "Measures" (Davis & Tanner 2003: 391). The presence of a large table showing descriptive statistics for their measures further confirms that it is a quantitative study. Because the data is described as a survey, this study is labeled 'observational'.

The examples that I have presented above represent fairly straightforward applications of the taxonomy. However, in practice, applying the taxonomy is much more complicated. In what follows, two of the most problematic issues that I encountered in applying the taxonomy are discussed: (1) that the lines between certain pairs of registers are not as transparent as others, and (2) that some types of articles bridge registers.

2.2.3 Some issues in applying a taxonomy of research articles

The first issue in applying this taxonomy is that without familiarity in a discipline, it can be difficult to distinguish between certain registers. At the heart of this issue is what is considered 'observed data' in a discipline. One of the most difficult distinctions to make is between qualitative research and theoretical research. For example, a study reporting on an ethnography in which focus groups, observations, and interviews were conducted is clearly qualitative research, in part based on the fact that the article calls itself 'ethnographic' (a widely-acknowledged qualitative methodology), as well as meta-language in the article that labels and describes data. However, it is not as clear (at least to an outsider of the field) whether a political science article that provides an analysis based on legislative decisions and court records is also qualitative (i.e., empirical). An article such as the latter typically does not have a section in which the data is systematically described, but it is also not purely theoretical. While I have considered an article such as this qualitative research in my taxonomy, an important consideration is the perspective from inside the discipline. Taking into account what members of the discipline consider data is key to understanding the discipline and the writing that takes place in the discipline.

Related to this issue is the fact that as an outsider of these disciplines, it can be difficult to distinguish between data types. For example, in chemistry and physics, despite careful reading, it was often difficult to determine if research was based on experimental or observational data. Thus, for registers and disciplines represented in a corpus of journal registers, discussions with disciplinary informants are important in writing operational definitions that can be applied on a more discipline-specific level.

The second main issue is that some articles do not clearly fit into any one category. For example, some articles make a theoretical or methodological argument, and then present a brief data analysis to support that argument. The data analysis is not the focus of the article, but rather is used for illustrative purposes. While the article may have some characteristics of an empirical article, such as quantitative

data, tables and figures, and so on, they less often contain key sections such as a description of procedures. This type of hybrid article is not accounted for in the current taxonomy (and is also not included in the corpus designed for this study).

Furthermore, a type of article that I am not distinguishing in this analysis is that of the brief report. Brief reports carry the same features as empirical articles, but are labeled by journals as 'brief reports' or something similar. One journal described brief reports as reports of research that are much abbreviated, or reports of research still in progress. Interestingly, primarily the 'hard' disciplines tended to publish these (e.g., chemistry, pediatrics, physics, and engineering).

In sum, the major consideration for most articles is the presence or absence of data, and if present, the nature of that data and the methods through which the data was analyzed. In the next section, I present the results of my analysis of eleven disciplines across a range of academic areas, and discuss the variation that I observed within.

2.3 Journal registers in the disciplines

While carrying out the taxonomy development reported on in Section 2.2, I also analyzed the extent to which each discipline surveyed publishes each article type. This analysis is presented in Table 2.2, where ++ indicates that this type occurs frequently in journals in that discipline, + indicates that it occurs on a regular basis, +– indicates that a few examples of that type of article were found but that they occurred rarely, and – indicates that no articles of this type were observed in the discipline. It should be noted here that these estimations are just that – estimates based on a survey of a limited number of journals in these disciplines. For example, although a register may be labeled with a – in Table 2.2, it is possible that the register would occur in other journals or sub-disciplines not considered in this survey.

Several interesting trends emerge from this analysis. First, the evaluative types of articles, particularly book reviews, show a decreasing trend as disciplines move (in traditional terms) from soft to hard disciplines. Likewise, commentary/forum articles are nearly non-existent along that same parameter. A second trend is that theoretical articles are more frequent in philosophy and political science. Empirical research is generally not present in philosophy. As we move into disciplines which fall more in the middle between hard and soft disciplines (such as sociology and economics), quantitative observational studies become more frequent, and as we move further into the 'hard' realm, experimental research studies become more common. In addition, all theoretical article types decrease as the disciplines approach the 'hard' side of the continuum. Physics is an exception to this rule, however, because of the sub-discipline of theoretical physics, which one of the journals surveyed in this task represented.

Table 2.2. Types of articles by discipline

Discipline	Empirical				Theoretical			Evaluative	
	Quantitative		Qualitative	Mixed	General	Author Interp.	Logic-based	Commentary, Forum	Book/ Product Review
	Exp.	Obs.							
*Chemistry†		++	–	–	+–	–	–	–	–
Physics†		++	–	–	+	–	++	–	–
Medicine (Pediatrics)	+	++	+–	–	-	–	–	+–	+–
Geology	+	++	–	–	+–	–	–	–	+–
Engineering (General & Civil)†		++	–	–	+–	–	–	–	+–
*Economics	+	++	–	–	+	+	+	+–	+
*Sociology	+–	++	++	+–	+	–	–	+–	+
Psychology	++	++	–	–	+	–	–	+	+
Applied Linguistics	+	++	++	+	+	–	–	+	+
Political Science	–	++	++	–	+	+	+–	+	+
*Philosophy	+–	–	–	–	++	++	+	+–	+

Key: ++ frequently occurs, + occurs with regularity, +– occurs rarely, – not found
*Discipline investigated in more detail as part of initial taxonomy formation
† A field in which as a non-specialist, I could not reliably distinguish between experimental and observational research. However, my belief is that these are primarily experimental research designs, particularly in engineering and chemistry.

The case of physics brings up an important point about variation across sub-disciplines and in variation across journals. For several of the disciplines, one sub-discipline is theoretically-oriented. Thus, journals within that sub-discipline publish only theoretical (and sometimes evaluative) articles. While some variation occurred across sub-disciplines in all the disciplines, psychology was perhaps the most varied, and this is most likely because psychology is a very broad and diverse discipline with a wide range of sub-disciplines.

In addition, this analysis reveals much more variation due to the effect of journal than expected. For example, in each of the disciplines, relatively few journals published evaluative texts, particularly forums. Some journals did not publish book reviews, while one journal only published reviews. In particular, I noted that most journals publish primarily either theoretical work or empirical work (there are a few exceptions to this), even if their descriptions state that they publish both. The journals that do publish both types tend to publish primarily empirical work, and a theoretical article may appear once in a while.

While there are some journals that are more general in nature, other journals can be associated with a particular sub-discipline, and thus publish more of a certain type of article. Some journals have distinct article types that were not represented elsewhere. For example, the journal *Philosophy and Phenomenological Research* contained a section of articles labeled "Book Symposium" in which an author of a book writes an article introducing the book, which is then followed by several reviews of the book by other scholars, and concluded with a response from the reviewed book's author.

Two major points can be summarized from this analysis. First, there is wide within-discipline variation that often follows along sub-disciplinary lines, and which is reflected in journals that are aimed at those specific sub-disciplinary areas of inquiry. Thus, when selecting disciplines to represent registers, care will need to be taken to select disciplines in which sampling from a wide range of journals is possible in order to reach a desired number of texts for inclusion in the corpus.

Second, and perhaps most notably, there are no disciplines in which all journal registers were found. That is, disciplines typically published a small number of registers with greater frequency, rather than a broad range of registers. The implications for these two trends are discussed in Section 2.4, as well as a description of how the results of this survey and taxonomy development have been applied to the design of the corpus that serves as the foundation for the analyses contained in this book.

2.4 Implications for corpus design

The premise behind creating a general taxonomy of published journal article types was to identify potential disciplinary differences in the types of journal articles which are published, and thus aid in the selection of the disciplines and journal registers to be represented in the corpus. This analysis has revealed that each discipline publishes a variety of texts, but that most disciplines do not publish *each* type of text. Therefore, two principles for corpus design are apparent. First, in order to investigate across-discipline variation, two disciplines should represent each type of article included in the corpus. Second, in order to investigate the possibility of within-discipline variation due to register differences, disciplines should be represented by at least two article registers whenever possible.

The first way in which the survey and taxonomy development (Section 2.3) is applied to corpus design is in the selection of journal *registers* to be included in the present study. While my original conceptualization for this project had been to include both research reports and evaluative texts, the findings displayed in Table 2.2 above illustrate that these evaluative texts are published much less frequently than research reports. In addition, I found that commentaries and forums, in which academics engage in more interactive scholarly discussion, were much

less frequent than even book reviews (the second type of evaluative texts). In fact, the publication of forums and commentaries is highly dependent upon journal – that is, only a few journals ever publish this type of text. Furthermore, evaluative texts as identified by this taxonomy are largely absent from the natural sciences (e.g., chemistry, physics, geology, etc., see Table 2.2). A further complicating factor is that these evaluative texts are not usually published in each issue or even volume of the journal. It would be difficult to sample these evaluative texts in a manner consistent with the sampling of theoretical or empirical articles, which occur most frequently in all disciplines. Therefore, evaluative registers have been excluded from the corpus design for this study.

The corpus for the present study has been limited to the primary article types of empirical and theoretical research. Within empirical research, the analysis shows that both quantitative and qualitative research is common, with great variation across disciplines. That is, research is exclusively quantitative in the hard sciences, while social sciences often publish both types of research. Mixed methods research is less common overall (Table 2.2) and is limited to disciplines which publish both quantitative and qualitative research. Thus, within empirical registers, I have chosen to represent quantitative and qualitative research in the corpus. Within theoretical research, the two specific types of theoretical research (logic-based and author interpretation) are generally less frequently published. Based on the decision to sample from physics and philosophy (described next) to represent theoretical articles, these specific types of theoretical articles have been collapsed into one general 'theoretical' register.

The second way in which the taxonomy development and survey has informed the corpus design is in the selection of disciplines to be studied in the project. In fact, the selection of disciplines has been influenced by several factors: (a) quantitative trends in the types of research published in disciplines, as reported in Table 2.2, (b) the desire to represent each of the selected registers (theoretical, quantitative, and qualitative research) with at least two disciplines, (c) the desire to represent disciplines by more than one register whenever possible, (d) the desire to include disciplines from a range of academic areas, and (e) the benefits of including disciplines that have been analyzed in previous linguistic research. Consequently, six disciplines were selected.

To represent theoretical articles, physics and philosophy showed a frequent use of this register, as well as representing disciplines clearly situated on opposite ends of the 'hard' and 'soft' continuum. Quantitative research reports are also readily available in physics. Two social science disciplines, political science and applied linguistics, exhibited strong publication rates for both quantitative and qualitative research, and thus have been selected for inclusion in the corpus. Finally, two disciplines, biology and history, which were not included in the survey summarized in Section 2.3, were selected for inclusion in the corpus based on their frequent presence in

research on disciplinary variation and because they are disciplines that are characterized by their quantitative and qualitative research methods respectively. Table 2.3 summarizes the registers and disciplines selected for the present study.

Table 2.3. Disciplines and registers represented in the corpus

Discipline	Theoretical	Qualitative	Quantitative
Philosophy	✓		
History		✓	
Political Science		✓	✓
Applied Linguistics		✓	✓
Biology			✓
Physics	✓		✓

To the extent possible, two registers were chosen to represent each discipline. However, for philosophy, history, and biology, it was not possible to sample more than one journal register because the discipline relied primarily upon one type of article. Although differences surely exist within these disciplines, making those highly fine-grained distinctions would have been unreliable within the scope of the present project.

The six disciplines that have been identified in this chapter represent a range of fields along the 'hard' and 'soft' parameter of disciplinary variation (see Becher, 1994), and are capable of representing the three journal registers selected for the study. As a result, the corpus design specified here allows me to investigate linguistic variation within and across both discipline and register. More specifically, however, several comparisons for describing variation are possible:

1. comparisons across all disciplines and registers,
2. comparisons across discipline for a single register type (e.g., quantitative research reports in political science, applied linguistics, biology, and physics), and
3. comparisons across register type within a discipline (e.g., theoretical versus quantitative research in physics).

In order to build a corpus of these disciplines and registers and enable these comparisons, the taxonomy presented above in Table 2.1 was refined on a disciplinary basis in consultation with expert informants from the discipline. The process undertaken to revise the taxonomy and the resulting operational definitions that were used to construct the corpus are detailed in Chapter 3, along with the general analytical methods employed in the study.