Susan Conrad

Beyond Grammar Description: Applying Corpus Analysis to Disciplinary Education

Abstract Corpus-based studies of grammar have greatly increased our understanding of language use. As a field, however, corpus linguistics has been less successful in moving beyond description to substantive impacts. Many corpus studies claim important implications for education, but outside of second language teaching and translation, the results are rarely applied. In contrast, this paper describes a project designed to advance engineering education in the United States. The project has conducted several kinds of corpus-based grammar analyses of student and practitioner writing, and then applied the findings to materials that improve the preparation of students to write as professional engineers. Additional corpus analyses are used to analyze the impact of the materials on student writing. This paper traces the process used in the project and discusses its successes and challenges, encouraging other corpus linguists to apply their skills to diverse disciplines.

Keywords Corpus-based research applications, English corpus linguistics, engineering writing, corpus-based grammar teaching

1 Introduction

In recent decades, corpus-based analyses have contributed greatly to our understanding of English. Reference grammars produced since the late 1990s have differed greatly from traditional grammars that focused on accurate structure. For example, Biber, Johansson, Leech, Conrad and Finegan (1999) present over 300 analyses of variation in grammatical features' use, and McCarthy and Carter (2006) have chapters addressing spoken language and grammar, and utterances and discourse. A new generation of English as a second language (ESL) grammar textbooks also includes information about frequencies of features, patterns of lexis and grammar, and common learner errors. Most notably, Cambridge

University Press uses its Cambridge English Corpus seal on back covers of textbooks such as the recent *Grammar and Beyond* series (e.g., Reppen 2012), assuring readers that "you can be fully confident the language taught is useful, natural and fully up-to-date." Other publishers also offer corpus-based textbooks, such as Pearson Education's *Real Grammar* (Conrad & Biber 2009), which identifies itself as "a corpus-based grammar of English" that supplements traditional textbook information.

Other language-related fields have also been influenced by corpus linguistics work. In translation, for example, corpus-based studies have made it possible for the field to move from comparing single originals and their translations to examining – among other things – language patterns in translations more generally and translation-related shifts that occur regardless of the languages involved (see review in Bernardini 2015). Corpus techniques have been used in concrete applications in translation, not only advancing machine translation (e.g., Koehn 2005) but also providing a lexical and syntactic perspective for evaluating the quality of translations (Freire 2009).

Unfortunately, however, within education, corpus-based work has had little influence beyond language-centered fields such as translation and second language teaching. This is particularly surprising since there is ample evidence that almost all students – even native speakers – are challenged by the use of language as they enter a new discipline (see review in Wingate 2015). The findings of corpus-based analyses seem likely to be helpful for training in many disciplines, but impacts have been limited. Some corpus analyses that have included many disciplines are designed to be descriptive, not to have a direct application (e.g., Biber 2006). Other disciplinary work does have the potential for a direct application. For example, with a combination of corpus-based and experimental techniques in a study of German court decisions, Hansen, Dirksen, Küchler, Kunz, and Neumann (2006) found that reading comprehension was enhanced when the decisions were rephrased with simpler syntactic structures. They suggest their findings be used to teach law students. Few such implications become applications, however.

In this chapter, I urge corpus linguists to strive to have more impact – that is, to move beyond descriptive work into its application. I provide an example of a project that has used corpus analysis to examine an educational problem in the United States, to make teaching materials to address the problem, and to assess the effectiveness of the materials. The example demonstrates that, collaborating with disciplinary experts, corpus linguists can clarify and address student needs with great success.

In the next section I introduce the project, which focuses on civil engineering. I then present three corpus-based grammar analyses, illustrating different kinds of analyses that are useful in the project. Next, I exemplify how the analysis

results are applied in the development of teaching materials and briefly describe the additional corpus analyses that assess the outcomes from the new materials. The final section reflects on the project, highlighting characteristics that have made it successful and that are still challenging.

2 Civil Engineering and Corpus Linguistics

Most people come in contact with civil engineering every day through use of infrastructure such as roads, bridges, tunnels, water systems, buildings, and retaining walls. However, with the exception of engineers themselves, few people realize the important role communication plays in civil engineering. Studies within the industry have found that communication is the single most important factor in the success of infrastructure projects (Thomas, Tucker, & Kelly 1998) and poor communication has contributed to costly legal battles, structural failures, injuries, and deaths (Banset & Parsons 1989, Parfitt 2008, Parfitt & Parfitt 2007). Since large infrastructure projects are expensive and paid out of public tax funds, effective communication by engineers is also a financial concern for society. From a business perspective, too, writing is important; most firms' only product is written documents, and easy-to-understand writing is critical to clients' satisfaction and timely work.

There is a clear need, then, for civil engineering students to develop strong writing skills. In fact, this need has been discussed for decades, but employers and new graduates of engineering programs continue to express dissatisfaction with the preparation they receive (Berthouex 1996; Sageev & Romanowski 2001; Donnell, Aller, Alley & Kedrowicz 2011). The only studies of writing in engineering practice use surveys, small case studies, and anecdotal text evidence, and they rarely mention civil engineering (e.g., see Tenopir & King 2004, Winsor 2003, Sales 2006). Numerous textbooks for technical writing exist, but they have no empirical basis, and some studies have found they neglect the needs of engineering students (Wolfe 2009).

When I learned about the need to improve writing instruction within civil engineering, I immediately saw the usefulness of corpus linguistics to address this problem. With funding from the U.S. National Science Foundation and collaborators at three universities and in the local engineering community, I undertook a corpus-based project to investigate the gap between practitioner and student writing, clarify student needs, and develop materials to address the needs.

2.1 The Civil Engineering Writing Project

Figure 1 provides a schematic of the overall process in the Civil Engineering Writing Project.

The first phase, begun in 2009, compiled a corpus of 400 student papers from four universities and 400 workplace documents from 50 firms and agencies, covering ten registers (e.g. e-mails, technical memoranda, reports, plan sheet notes; see further Conrad, Pfeiffer & Szymoniak 2012). We then analyzed the corpus to investigate differences between student and practitioner writing. With the input of engineering practitioners in industry, we identified the most serious student writing weaknesses. In phase 2 of the project, currently underway, we develop teaching materials that address those writing weaknesses. In the intervention step, the materials are used in existing civil engineering courses. Students' papers from these courses – the post-intervention papers – are then analyzed and comparisons made with the pre-intervention papers, to assess the impact of the materials.

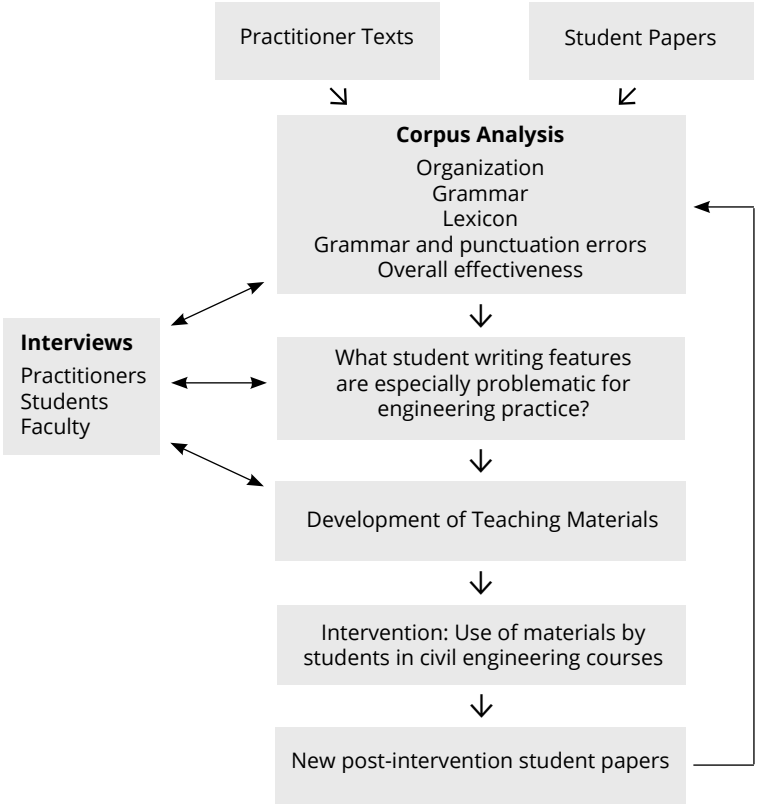

Figure 1: Overview of the Civil Engineering Writing Project process.

Three characteristics of the project might be surprising to readers more familiar with descriptive projects rather than teaching interventions. First, although the corpus has grown to over 1500 texts, the analyses typically focus on small subcorpora. The situational characteristics of many registers differ greatly (e.g., the content, communicative purpose, and audience of a student lab report are very different from a practitioner design report) and an overall description of the linguistic variation – though interesting to linguists – is not especially helpful for designing teaching materials.

A second notable characteristic of the project is the interplay of the corpus analysis with interview data. Corpus projects often consult disciplinary experts for corpus design issues or to understand disciplinary conventions, but this project relies even more heavily on input from practitioners and students. An especially useful step has been sharing the results of corpus analysis with interviewees. Student reactions help us to understand the "why" behind their writing choices, something no corpus analysis can reveal. Practitioner explanations allow us to understand which student writing problems are the most important to address and which changes in student writing are most effective. Practitioners also contribute to the teaching materials, commenting on drafts and checking that all information – even if it is simplified for a beginning-level course – is consistent with engineering practice. The examples in the next sections share some specific contributions from interviews, based on interviews with 22 students and 16 practitioners. (Faculty are also interviewed but are not the focus of this paper.)

The third characteristic concerns the diversity of the universities who participate in the project. Compiling a corpus from multiple universities is more time-consuming than focusing on just one, but for this project it was crucial for identifying weaknesses shared by different student populations and investigating the impact of the materials with diverse groups. The project is based at Portland State University in the northwestern U.S. and includes three other universities: the California State Polytechnic University at Pomona, Howard University in Washington, D.C., and Lawrence Technological University in the Midwest. All offer an accredited Bachelor's degree in civil engineering and seek to train students to become effective practitioners, but they differ in size, geographic region, entrance requirements, and typical student academic and ethnic backgrounds.

3 Grammar Analyses

This section summarizes three of the grammar-related analyses from the first phase of the project, which revealed differences in student and practitioner writing and also challenged many claims about engineering writing. I highlight just a few of the most important aspects of the analyses; further details about the

methods and results can be found in other publications about the project, especially Conrad (2015, 2017, and 2018).

3.1 Passives and Impersonal Style

It is widely claimed that engineers overuse passive voice and make texts too impersonal. For example, Gwiasda berates the high frequency of passive voice in student writing as "the perfect vehicle for documents that record material of no intended consequence to anyone at all" (Gwiasda 1984: 150). Sales (2006: 18) describes practicing engineers as "consciously avoiding any use of the personal pronouns" in order to be more objective. There is no systematic evidence to support these claims, but previous corpus-based investigations of academic prose (e.g., Biber 1988) have found engineering to use a higher frequency of passives than most academic texts.

For an analysis of passives and impersonal style features in the civil engineering texts, I used a sub-corpus chosen so that practitioner and student writing was as similar as possible and represented a typical workplace writing task – reports written to clients, addressing real situations (Table 1). This is a task typically given to students in their fourth (final) year of the degree. For a comparison with professional academic texts, I also included 50 research articles.

Table 1: Texts used in the passive voice and impersonal style analysis.

Category	Number of texts	Sources	Words
Practitioner Reports	60	10 firms	201,700
Student Reports (for clients)	60	9 courses	207,700
Journal Research Articles	50	10 journals	270,900

The analysis used a technique well established in corpus linguistics – Multidimensional (MD) analysis, as introduced by Biber (1988). MD analysis uses a factor analysis to calculate the co-occurrence patterns of linguistic features in texts. Groups of features that tend to occur together in texts are identified statistically; no a priori assumptions are made about which features should be grouped together. The factors are interpreted in terms of their communicative functions as dimensions of register variation. In the study of 23 registers of spoken and written English conducted by Biber (1988), one factor had four kinds of passive structures – agentless passives, passives with *by* prepositional phrases, past participial clauses, and past participial noun postmodifiers (Table 2). In addition, two kinds of connecting words loaded onto the same factor: linking adverbials and multi-functional subordinators. This dimension was characterized as

Table 2: Features on the Impersonal Style dimension.

Language Feature	Example	Factor loading
linking adverbials	*therefore, however, in conclusion*	.48
passive verbs, agentless	The bridge *was built* in 1923.	.43
past participial clauses	*Designed by a local engineer*, the bridge won an international award.	.42
passive verbs with *by* phrases	The bridge *was designed* by a local engineer.	.41
past participial noun postmodifiers	The recommendations *included in this report* cover …	.40
adverbial subordinators with multiple functions	*since, while, whereas, such that*	.39

Impersonal Style, reflecting the high frequency of passives and lack of human agents. The connectors were found to overtly structure the logical relationships in the often dense, technical texts. I applied this dimension for the analysis of the engineering texts.

I used the standard procedures for the MD analysis as outlined in Conrad and Biber (2001). I grammatically "tagged" the files with the Biber tagger and checked and corrected features with another program. Grammatical features in the engineering registers were counted and standardized to the findings of Biber's (1988) analysis so that comparisons could be made with a range of English discourse. In Figure 2, which displays the results of the analysis, 0 represents the mean for the 23 registers in Biber's analysis, and each positive or negative unit represents a standard deviation.

As Figure 2 shows, the results of the analysis are generally consistent with claims that engineering writing is highly impersonal; relative to a wide range of English discourse, the three registers of engineering all have a markedly high mean score on the Impersonal Style dimension. Their use of impersonal features is, for example, far higher than conversation, fiction, and popular nonfiction (magazines and books for a non-specialist audience). However, when the engineering registers are compared among themselves, the differences are important. An analysis of variance found a statistically significant difference among the three engineering registers ($F(2, 167) = 19.89$, $p < .0001$, $\eta^2 = .19$), with the student papers and journal articles using more impersonal style features than the practitioner papers. Post-hoc Scheffe pairwise comparisons found a statistically significant difference between the practitioner reports and student reports, and between the practitioner reports and journal articles, but not between the student reports and journal articles. In other words, in the frequency of impersonal style features, the student reports resemble academic journal articles more than the practitioner reports they are meant to imitate.

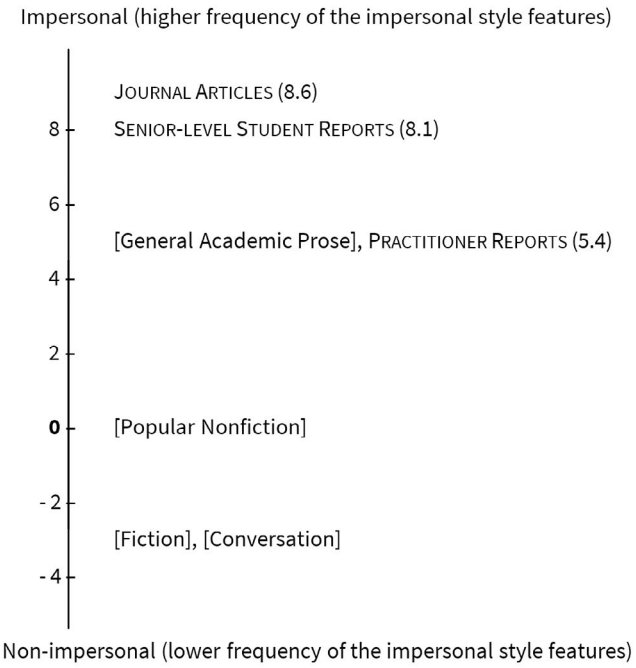

Figure 2: Mean scores for three civil engineering registers on the Impersonal Style dimension. *Note:* General academic prose, popular nonfiction, fiction, and conversation are from Biber (1988) for comparison.

Several of the important characteristics from the Impersonal Style analysis are exemplified in this excerpt from a practitioner report:

(1) On August 15 and 19, 2003, we drilled five exploratory borings with a portable drill rig using solid stem auger techniques. These borings were drilled to provide data for retaining wall and signal pole foundation design. The boreholes were drilled to depths ranging from ±2 to 6 m.

Surprisingly, the paragraph begins with a human agent and active voice (*we drilled*). Although not as common as passives, these structures appeared regularly in practitioner texts with a variety of verbs (*we observed..., the subject team conducted..., ABC Engineering recommends..., we anticipate...*). In interviews, practitioners commented that occasional overt statements of responsibility were important; they not only made it "easy for readers to read fast" but they were important to "manage liability in a field where you are hired for subjective judgments." Contrary to the claims in the literature about engineers seeking to sound objective, these practitioners emphasized making subjective judgments based on

observed data. They discussed the need to be explicit about responsibility for observations and judgments. They especially emphasized being explicit about recommendations because recommendations from a licensed engineer have a legal status; they must be followed unless they are changed by another licensed engineer.

The second and third sentences in example 1 use passives. They illustrate three functions that commonly occur with passive voice. First, they allow objects, processes, or concepts to be the grammatical subject and thus a consistent topic of discourse (here: *these borings, the boreholes*). Second, the passive constructions conform to the principles of information structure and end weight (Biber et al. 1999). That is, the subject noun phrases in the passives refer back to the topic established in the previous sentence (*borings*), and the information after the verb (*to provide data for..., to depths ranging...*) is new information that is longer than the subject noun phrase. Only the first of these three functions is typically mentioned in technical writing materials even though conforming to typical information structure and end weight can be crucial for making technical information easy to read.

An additional important characteristic that accounted for fewer passives in practitioner writing was the more frequent use of inanimate subjects with active voice verbs. Objects, processes, and documents often do things in these engineering texts – for example, *this document reports the analysis...* and *our analysis assumes a factor of safety of....*

The journal articles and student papers used passives more consistently. Passives were regularly used for the kind of actions practitioners expressed in active voice, such as recommendations and observations:

(2) a. It is recommended that these new equations and charts should be included in the revision of the AASHTO Bike Guideline. (journal article)
b. Due to the design of the intersection, initially it was thought that cyclists would merge to the right lane and be forced to compete with merging freeway traffic, but it was observed that most cyclists merged safely into the left car lane well before reaching the intersection. (student report)

Since recommendations in journal articles do not entail any legal meaning, the lack of explicit responsibility and use of the hedge *should be* do not have a critical impact, as they might in a practitioner report. The writing in (2b), however, is meant to imitate a practitioner report. Instead, its passives leave the reader wondering who is responsible for this work: what mysterious group was hypothesizing about how cyclists will merge to the right lane? And was it that group or another who observed the cyclists merging safely? The difference from

practitioner reports is striking, but in interviews, most students said they had learned that technical writing should not use personal pronouns or refer to people. They commented on "...the technical writing thing of don't use *I* or *we* or *us*" and stated "You need to use objective language." Some writers clearly thought the absence of human agents automatically created objective meaning; they used expressions such as *it was believed...* or *it was felt...*, but – even in passive voice – beliefs and feelings are not appropriate evidence for engineering.

When students were shown examples like (2b) in interviews, many also commented that they used such sentences because they were long or looked "fancy." This desire to look fancy also contributed to a high frequency of linking adverbials and subordinators in texts. Unfortunately, the fancy sentences were also often ineffective; in (3), for instance, the important conclusion – the recommendation to use bike lanes and bioswales – is minimized by being in a subordinate clause:

(3) ... *Moreover*, SW Elm is fully paved with standard asphalt (highly impermeable) and relies fully on gutters to carry off rainwater. *Thus*, water overflow can occur on the site during heavy rain seasons, *while* having permeable pavements and bioswales could solve this issue.

The analysis of the impersonal style features added to our understanding of student and practitioner writing in notable ways. It countered the image of all engineering writing being like academic writing; in fact, workplace writing incorporates more human agency because explicit responsibility and unambiguous content is valued. It provided systematic evidence for claims that passives are often useful in writing that focuses on objects, but it also highlighted passives' usefulness for conforming to typical end weight and information structure. It also revealed the student's weakness for "fancy" sentences, which is taken up in the next analysis.

3.2 Sentence Structure

Another widespread belief about engineering writing is that sentences are needlessly long and complicated. An online website for career and education information for a professional engineering society, for example, quotes a technical writing consultant with 25 years of experience: "I have met very few engineers who are comfortable with using simple language, organizing documents for the readers' benefit, keeping sentences and paragraphs short, and getting to the point" (Crawford 2012: 2).

One approach for investigating sentence complexity in corpus-based studies is to use automatic counts of complexity features, but in pilot work we found

that some student texts had such numerous sentence structure and punctuation errors, it was difficult to automatically identify clause structure. For this example, then, I illustrate a different kind of analytical technique that is useful in the project – coding a sample by hand.

For the sentence structure analysis, we sampled sentences in the texts in Table 3. Originally interested in development as students progressed in their major, we included third-year student lab reports, the most common type of third-year writing students do. For fourth-year students and practitioners, we included reports and technical memoranda – two registers that are common in the workplace and final-year courses. It turned out that preliminary analyses found no difference in the two student groups, so they were combined in the analysis reported here.

Table 3: Texts used in the sentence structure analysis.

Category	Number of texts	Sources
Practitioner reports and technical memoranda	86	10 firms + 1 public agency
Student reports and technical memoranda (senior level)	78	9 courses
Student laboratory reports (junior level)	122	4 courses

For the analysis, I made a simple distinction between sentences that were "complicated" or non-complicated, defining complicated as having dependent or embedded clauses. The more detailed categories typical of linguistic studies, such as finite versus nonfinite dependent clauses or postnominal versus adverbial clauses, were more specific than needed for the general comparison of sentence complexity we sought and too detailed for the engineers to understand quickly.

I followed a standard procedure of multiple samples, often used in corpus-based studies that require hand-coding of data (Biber, Conrad & Reppen 1998: 91–93). Specifically, for each of the writer groups, I analyzed three random samples of 100 sentences. The proportions of complicated sentences was within 5 % for each sample, so I took them as representative of the group. The complete sample was thus 600 sentences.

A chi-square test found a statistically significant difference between the frequency of complicated sentences in the practitioner and student writing (χ^2 = 51.3, df = 1, p < .001, φ= .293) with the students using more complicated sentences. Over half of the student sentences had complex or embedded structures, while only about a quarter of the practitioner sentences did (Figure 3).

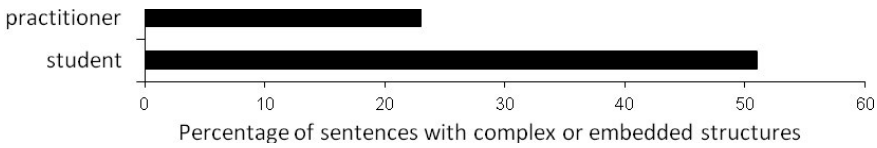

Figure 3: Use of complicated sentence structure by students and practitioners.

Practitioner writing had more sentences expressing a single idea, as in the following examples:

(4) a. The rainfall depth was obtained from the City of Granson, County of Wilson. For the 25-year storm event, 24-hr rainfall depth is 4.0 inches for the site.
b. The lower portion of the embankment, below ±El. 475 to 480 and near Harmony Creek, is graded at approximately 1½(h):1(v).

Sentences like (4b) look long to students and might contribute to student beliefs about "fancy" sentences. However, linguists can easily see that the length comes from phrasal complexity, especially long noun phrases and prepositional phrases that make information very precise (see further discussion in Conrad 2015: 325–6). The clause structure remains simple. Commenting on the frequency of simple sentence structures, practitioners again noted the need to make information as easy as possible for clients to follow. They commented, for example, "Clients want to be able to read fast or skim," and "Simple sentences are more concise. And they are less likely to be ambiguous or be misinterpreted."

Student sentences, on the other hand, tended to have more complexity on the clausal level, as illustrated in this sentence from a transportation report, which has multiple clausal constituents and one subordinate clause embedded within another subordinate clause:

(5) [This particular modeling detail does not seem [to greatly affect the output of the simulation] [because [although it appears unrealistic], it does not affect the flow of traffic greatly and only seems [to occur on occasion]]].

Such student sentences are, at best, hard to follow. Sometimes they even became so complicated that their literal meaning was inaccurate. In interviews, however, students expressed no concern for making texts easy to read and unambiguous. Instead, when students were asked to comment on complicated sentences, typical explanations for choosing them were:

"It looks better if it's longer. I think it's that simple."
"Make it fancy."

"I kind of felt like I had to sound professional and smart. I mean, you want to sound really knowledgeable about things, and it seems like the easiest way to do that is to be wordy."

Overall, this analysis was useful because it provided systematic evidence that it is students – not practitioners – who write with complicated sentence structures. The interviews made clear that practitioners valued the simpler clause structure for their ease of reading and the complex phrases for the specificity of information. The analysis provided evidence that students' writing and their beliefs about writing were the opposite of practitioners'.

3.3 Errors in Grammar and Punctuation

Initially, I did not plan to include error analysis in the project because grammatical choices and their impacts, not basic accuracy, seemed most important for writing. However, it was soon obvious that errors had a large impact on students' writing effectiveness. Furthermore, several civil engineering faculty firmly believed that it was only ESL papers that had a high frequency of errors when I suspected errors were more widespread. I therefore added an error analysis to the project.

The analysis investigated the extent to which writers conformed to standard written English grammar and punctuation. It followed procedures for hand-coding errors as in traditional learner corpus studies. Because the coding of errors is time-consuming, the analysis covered a subset of the papers in Table 3 (above), using 45 texts each from the practitioners, senior-level students, and junior-level students. The senior-level and junior-level papers were counted separately since the frequency of errors varied greatly.

Errors were categorized into five major categories (Table 4) by trained research assistants. The errors typical of ESL students provided a rough means of assessing whether ESL-type errors dominated the analysis. Native speakers of English also make these kinds of errors, but they tend to be more common in ESL texts.

Table 4: Error categories in the error analysis.

Error Category	Description
1. Verb errors	Tense, aspect, formation of infinitives and other verb forms, any verb errors other than S-V agreement
2. Sentence structure	Any structure errors that make sentence ungrammatical in English, includes relative clause or participle clause errors
3. Punctuation	Commas, semi-colons, sentence-final punctuation, and other punctuation
4. Spelling and typos	Errors related to spelling or typing
5. Articles, prepositions and other errors typical of ESL learners	Errors with articles, prepositions, plurals, subject-verb agreement and pronoun-antecedent agreement

Errors in each category and total errors were counted per text and normed per 1,000 words. Figure 4 displays the median error frequencies across the groups: just over 2 for practitioners, about 13 for senior-level papers, and almost 16 for junior-level papers. On a double-spaced, printed page, these frequencies mean about one error on every other page for practitioner documents, about three per page for senior-level papers, and about five per page for junior-level lab reports. A Kruskal-Wallis one-way analysis of variance test found a significant difference in the three groups' error rates overall ($H(2) = 60.855$, $p < .001$). There was a statistically significant difference between the practitioner writing and senior-level writing ($p < .001$, $r = 0.67$) and between the practitioner writing and junior-level writing ($p < .001$, $r = 0.75$), but not between the senior-level and junior-level writing.

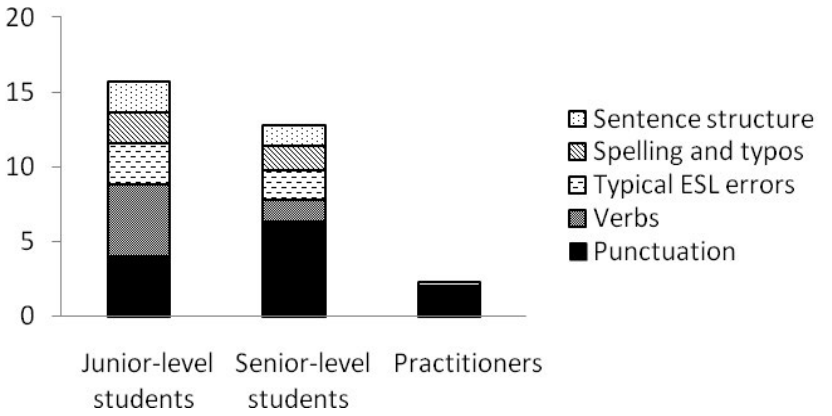

Figure 4: Median error rates in student and practitioner writing.

Although a few student papers were almost error-free, the median rates show that many student papers had enough errors to be distracting and damaging to the writer's credibility. The errors were also more widespread than ESL students would account for, especially since the senior-level papers were written in groups and interviewees commonly reported that native English speakers edited ESL writers' contributions.

The student and practitioner texts also differed in the types of errors they included and their impacts on comprehensibility. In the practitioner documents, punctuation accounted for the vast majority of errors, as Figure 4 shows. The majority of these errors involved isolated comma errors that did not interfere with meaning. Student errors, on the other hand, covered all categories. Some errors were just odd, such as unusual punctuation choices (example (6a)), perhaps related to the desire to "make it fancy." Some errors made sentences literally nonsensical, such as the dangling modifier in example (6b). The most serious usually involved sentence structure errors and made the main idea difficult to discern, as (6c) exemplifies.

(6) a. The map displays the geologic conditions; with the basalt layers in darker colors.
b. As a civil engineer, the strength of concrete is highly affected by the curing time.
c. But the brittleness of each coupon varied with coupon #3 having little necking and being the most brittle of the three coupons, coupon #13 had more necking than #3 but less than #7 and thus concluding it had moderate ductility of the three coupons.

When discussing errors, practitioners' most common comment had to do with engineering being a detail-oriented profession. They were concerned about errors inadvertently changing meaning and also making the firm look unprofessional. One interviewee summed up a credibility problem for the writer: "Errors convey carelessness. Who wants a careless engineer?" Some mentioned that they were shocked by the level of errors in some job applications they received and that those applications went straight into the trash.

All the students said they proofread their papers at least once, but many reported spending little time because they perceived errors to have little influence on their grade. This perception was consistent with a review of lab reports that received grades of 90% or above; they included papers with some of the lowest and highest error rates. Many students also reported that, even when they did proofread thoroughly, they had little confidence in their ability to recognize and correct errors.

This analysis provided evidence to counter the faculty impression that errors are a problem only for ESL students. They are a serious problem for many students. They also constitute a serious matter for the practice of engineering. Errors can undermine the credibility of a new graduate applying for a job, a practicing engineer, or the professional reputation of a firm.

4 Applying the Corpus Research to Improve Teaching

The results of the analyses were used to develop the new teaching materials. These materials are free-standing units that cover genre expectations, grammatical and lexical choices, and grammar and mechanics errors. This section uses examples of the materials related to the grammar analyses described above. More details can be found at the Civil Engineering Writing Project website, www.cewriting.org, and in Conrad, Kitch, Smith, Lamb & Pfeiffer (2016).

4.1 Features of the New Teaching Materials

Each unit is drafted by applied linguistics and engineering faculty and is then reviewed by at least two practitioners, who check that advice is consistent with workplace practice. Here I highlight four features that set the materials apart from typical technical writing instruction, made possible by the combination of the corpus analysis and interview data.

First, the units provide information about the patterns of language features that differ between student and practitioner writing and, with practitioner quotes, tell why the language features matter within civil engineering practice. The opening of a sentence structure unit illustrates these features (see appendix). Students see a figure comparing the percentage of simple sentences in student and practitioner reports. The findings are described for the students, and the target for revising is explicit (use more sentences that express one idea). The importance of simple sentences for engineering practice is reemphasized by comments from practitioners.

Each unit also contains numerous examples of practitioner writing. For many students, this is a first experience seeing sentences from practitioner documents. We choose examples that illustrate the most important corpus findings. We also provide explanations that use simple terms to direct students' attention to linguistic features. Figure 5 provides an example from the unit about simple sentence structure.

Effective Simple Sentence Structure	
Examples	Explanation
1. **The existing bridge is** a 9-span timber trestle bridge with a concrete deck. **It is** 217 feet long and 30 feet wide. **The posted speed is** 25 mph. *(Report)*	Each sentence has one main idea. It has a subject (in purple) and a verb phrase (in red). The verb is close to its subject.

Figure 5: Opening of a section exemplifying and explaining practitioner writing.

Many units also contain "Myth buster" boxes. These boxes present information that directly counters the misconceptions that students expressed in interviews and that underlie ineffective writing choices. For example, the unit about passive voice counters the idea that passive voice automatically expresses objectivity (Figure 6). It addresses the fact that engineering requires judgment and ties it to the use of human agents with active voice. It goes on to urge students to strive for accurate meaning in verbs, rather than relying on passives such as "it was felt that..." since "feeling" is not adequate evidence in any voice.

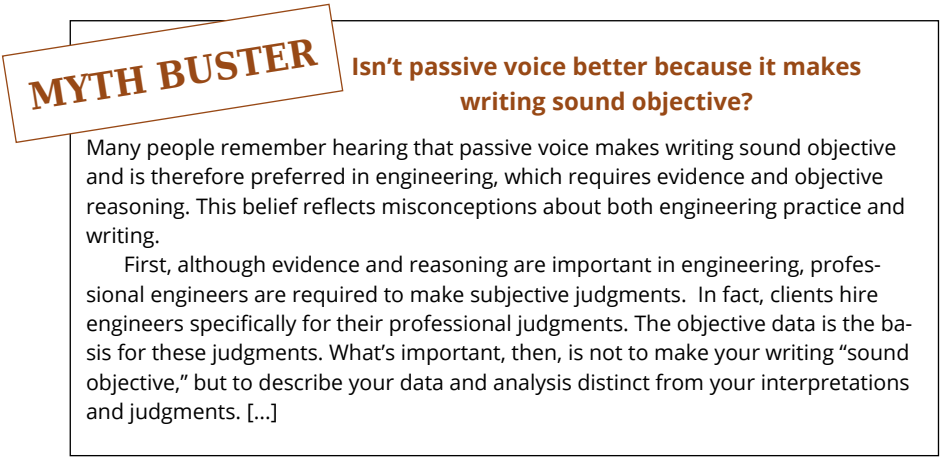

Figure 6: Example of a "myth buster" box from the passive voice unit.

The units also cover specific revision techniques and provide practice activities for them. This kind of practice is not unusual in writing materials, but using the corpus allows us to include real student sentences, and give students realistic revising practice that addresses common problems. The unit on passive voice, for example, includes tips on using inanimate subjects with active verbs (Figure 7). The units that address grammar and mechanics address the most common

errors, some of which – like the overuse of semi-colons – would not have been recognized without the corpus analysis.

Technique 4: Use an inanimate subject + active voice verb.	
Original Sentence Needing Revision	Revision
1. [*Note: preceding paragraph describes the basis for the liquefaction analysis]* A potential for liquefaction in the loose sand between 15 and 30 feet was indicated. *(Report)*	1. The results of the analysis indicate a potential for liquefaction in the loose sand between 15 and 30 feet.
Explanation. The original of example 1 has a long subject before the verb. The revision uses a shorter, inanimate subject + active verb (*results indicate*) for easier reading. The revision also now follows expected information structure in two ways: it explicitly moves from data analysis to the engineers' interpretation of it (see Unit 4, Part 1) and it follows known-new information sequencing (see Unit 4, Part 2).	

Figure 7: Example revision technique for reducing overuse of passive voice.

4.2 Assessing the Effectiveness of the Materials

After the materials are used in courses in civil engineering departments, students write papers that are compared to pre-intervention student papers. Currently, we have results from four universities, three levels (first-, third- and fourth-year courses), and 16 different courses. The materials have been implemented in a variety of conditions. Class size has ranged from 12 to 80 students. The amount of class time versus homework time for the materials has varied from a writing workshop day in class to no class time at all. Some courses had writing teaching assistants; most did not. Although this variability can make assessment more challenging, we want the materials to be piloted in realistic conditions.

The same techniques used for analyzing differences in practitioner and student writing are used to analyze the change in student papers. This includes the techniques described above, plus a separate analysis of passive main verb effectiveness, word choices, and genre organization (further information can be found in Conrad, Kitch, Pfeiffer, Smith, & Tocco, 2015). In addition, the assessment includes a holistic evaluation of effectiveness by a practitioner since changes in linguistic forms do not always amount to an improvement in overall effectiveness. The results are summarized in Table 5, with the grammar features described in this paper in the top half of the table, and other features in the bottom half. As the summary in the table shows, the results have been consistently positive.

Table 5: Summary of post-intervention results (16 courses).

Language feature	Change in student writing
Passive Voice	Statistically significant reduction in frequency of passive voice Active voice used appropriately for responsibility
Sentence Structure	Statistically significant reduction in complicated sentences No complicated sentences with inaccurate meaning
Grammar and Punctuation Errors	Statistically significant decrease in targeted errors Decrease in errors that interfere with meaning
Word Choices	Statistically significant reduction in vague or inaccurate words
Genre Analysis (organization)	Statistically significant increase in effectiveness of content sequencing, inclusion of expected content, and decrease in extraneous content
Evaluation by Practitioner	Statistically significant increase in overall effectiveness rating

We also ask students for their reflections and suggestions after they use the materials. Their reflections show that the materials can impact attitudes and beliefs that underlie some of the ineffective features of student writing. Typical comments have included the following:

"The information that made the biggest impression on me was that engineering writing is different from literature writing and can cost me a job."
"The thing that impressed me most today was how poor my grammer [sic] and editing skills are."
"I think the biggest challenge for me in writing for CE will be to ignore the temptation to sound fancy and smart."

The only consistent suggestion we have received is to include more examples even though the units are already longer than we planned for easy incorporation into courses.

Of course, the positive results of the assessment do not mean every post-intervention student paper is strong. In fact, it occasionally appears that a student did not look at an assigned unit at all. Certain individuals, for example, never stop overusing complex sentences, and we hope to investigate this individual variation more in the future.

5 Conclusion

The evidence from the Civil Engineering Writing project suggests that corpus-based grammar description can indeed be applied to have positive impacts in disciplinary education. To conclude, I reflect on some of the most important factors for the success of the project and others that continue to be our biggest challenges.

One characteristic that contributes to the success of the project is the highly specialized nature of the corpus. Even if the corpus focused on all engineering rather than only civil engineering, it would be impossible to identify student weaknesses as specifically because work contexts could vary so greatly. It is even more important that we were able to compile a corpus to represent the kind of workplace writing students hope to do after graduation, not just academic writing. Compiling a corpus of workplace texts is easier in civil engineering than many fields because the documentation of any publicly funded project is open to the public; in many other fields, issues of confidentiality would likely make corpus compilation more difficult.

Civil engineering is also well suited to a corpus-based project because the field is data-oriented. Engineers expect to see data analysis, especially quantitative data, as a basis for decision-making. Even if they do not understand all the linguistic details of an analysis, they generally appreciate the quantitative evidence in conjunction with explanations of language functions. Other fields in science, technology, engineering and mathematics are likely to be equally appreciative partners in a corpus project, but some other fields might consider the quantitative analysis less valuable.

Success has also depended on having access to helpful disciplinary experts. Numerous practitioners have been generous with their time, both in teaching me about civil engineering generally and in answering numerous writing- and language-related questions. They are aware of how important writing skills are in their profession, and many struggled in their own first attempts to write in industry. Without their input, we simply could not target workplace writing skills as we have.

Project success is also dependent on civil engineering faculty, who help develop the materials and try them in their courses. Many faculty have contributed, but this continues to be one of the most challenging aspects of the project. Most faculty have no training in teaching writing, nor do they have any metalanguage for explaining language choices. Even those who are enthusiastic about using the materials in courses admit it takes some time to be comfortable with them and to feel prepared to answer the kinds of questions students typically ask. Many faculty also find it challenging to add anything more to their already full syllabi. A number of faculty are resistant to using the materials at all. A

shortcoming of the project is that I did not plan faculty training seminars, which would likely increase enthusiasm for using the materials.

Finally, another continuing challenge in the project concerns teaching linguistic phenomena to an audience that generally has little language training and little metalanguage for referring to language. In materials, it is often difficult to be accurate about linguistic phenomena, but also easy enough for the audience to understand. Even referring to sentence structure is difficult because terms like phrases, clauses, and subordination are not known. Effective descriptions often require multiple rounds of drafts, feedback, and revisions. I also find it a satisfying challenge, however, because people untrained in linguistics learn to recognize how to manipulate language in more effective ways and even how to explain effective choices to each other.

All of these factors – and others – make an applied, corpus-based project challenging. Nonetheless, I have found any aggravations well worth seeing the improvements in student writing. Corpus-based descriptions provide a basis for work that other approaches cannot match. I urge other corpus grammarians to consider the wider audiences who might benefit from the applications of their work and to start working with them. Otherwise, though corpus linguistics will continue to be known within linguistics and language studies, it will not help to solve problems in other disciplines, where corpus analysis can make such a valuable contribution.

Acknowledgements

Partial support for this project is provided by the U.S. National Science Foundation (awards DUE-0837776 and DUE-1323259). I am also grateful to the civil engineering practitioners, students, and faculty who make the project possible.

Appendix – Example opening of a unit about sentence structure

Civil Engineering Writing Project – Language Unit 3
EFFECTIVE SENTENCES: SIMPLE SENTENCE STRUCTURES

What do you need to know about effective writing in civil engineering practice?

Experienced engineering practitioners use **simple sentence structure** in most of their writing. Simple sentence structure is effective because it conveys one main idea. Simple sentence structure makes comprehension easier for readers especially when sentences have complex, precise technical information.

Students use fewer simple sentences than practitioners do (Figure 1). In other words, students use complicated sentences more often. Students' sentence structure is more similar to academic journal articles than practitioner documents. In addition, students' complicated sentences often make content ambiguous or inaccurate. Revising sentence structure can therefore be an important step towards effective writing.

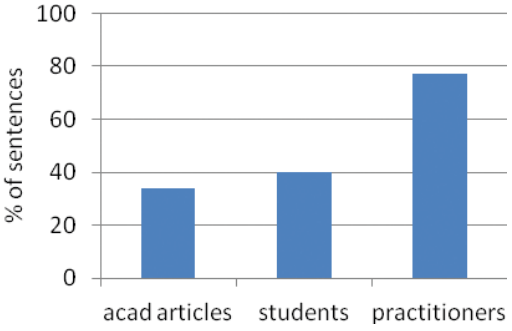

Figure 1: Percentage of sentences with **simple sentence structure** in student reports, practitioner reports, and academic journal articles

What experienced engineering practitioners say

"Clients want to be able to read fast or skim."

"Simpler sentences are more concise. And they are less likely to be ambiguous or be misinterpreted."

References

Banset, Elizabeth and Gerald Parsons. 1989. Communications failure in Hyatt Regency disaster. *Journal of Professional Issues in Engineering* 115: 273–288.

Bernardini, Silvia. 2015. Translation. In Douglas Biber and Randi Reppen (eds.), *The Cambridge handbook of English corpus linguistics*, 515–536. Cambridge: Cambridge University Press.

Berthouex, P.M. 1996. Honing the writing skills of engineers. *Journal of Professional Issues in Engineering Education and Practice* 122: 107–110.

Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, Douglas. 2006. *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.

Biber, Douglas, Susan Conrad and Randi Reppen. 1998. *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan. 1999. *The Longman grammar of spoken and written English*. Harlow, UK: Pearson Education.

Conrad, Susan. 2015. Register variation. In Douglas Biber and Randi Reppen (eds.). *The Cambridge handbook of English corpus linguistics*, 309-329. Cambridge: Cambridge University Press.

Conrad, Susan. 2017. A comparison of practitioner and student writing in civil engineering. *Journal of Engineering Education* 106: 191–217.

Conrad, Susan. 2018. The use of passives and impersonal style in civil engineering writing. *Journal of Business and Technical Communication* 32: 38–76.

Conrad, Susan and Douglas Biber (eds.). 2001. *Variation in English: Multi-dimensional studies*. Harlow, UK: Pearson Education.

Conrad, Susan and Douglas Biber. 2009. *Real grammar: A corpus-based approach to English*. White Plains, New York: Pearson Education.

Conrad, Susan, William Kitch, Timothy Pfeiffer, Tori Smith and John Tocco. 2015. Students writing for professional practice: Collaboration among faculty, practitioners, and writing specialists. *Proceedings of the 2015 American Society for Engineering Education Annual Conference and Exposition*. http://dx.doi.org/ doi:10.18260/p.24769.

Conrad, Susan, William Kitch, Tori Smith, Kenneth Lamb and Timothy Pfeiffer. 2016. Faculty-practitioner collaboration for improving civil engineering students' writing skills. *Proceedings of the 2016 American Society for Engineering Education Annual Conference and Exposition*. http://dx.doi.org/10.18260/p.26892.

Conrad, Susan, Timothy Pfeiffer and Thomas Szymoniak. 2012. Preparing students for writing in civil engineering practice. *Proceedings of the 2012 American Society for Engineering Education Annual Conference and Exposition*. https://peer.asee.org/21817.

Crawford, Mark. 2012. How engineers can improve technical writing. https://www.asme.org/career-education/articles/business-writing/how-engineers-can-improve-technical-writing.

Donnell, Jeffrey, Betsy Aller, Michael Alley and April Kedrowicz. 2011. Why industry says that engineering graduates have poor communication skills: What the literature says. *Proceedings of the 2011 American Society for Engineering Education Conference and Exposition*. https://peer.asee.org/18809

Freire, Evandro. 2009. A corpus-based approach within Juliane House's model for translation quality assessment. *The ESPecialist* 30: 193–211.

Gwiasda, Karl. 1984. Of classrooms and contexts: Teaching engineers to write wrong. *IEEE Transactions on Education* E-27: 148–150.

Hansen, Sandra, Ralph Dirksen, Martin Küchler, Kerstin Kunz and Stella Neumann. 2006. Comprehensible legal texts – utopia or a question of wording? On processing rephrased German court decisions. *Hermes – Journal of Language and Communication Studies* 36: 15–40.

Koehn, Philipp. 2005. EuroParl: A parallel corpus for statistical machine translation. *Proceedings of the Tenth Machine Translation Summit*, 79–86. http://mt-archive.info/MTS-2005-Koehn.pdf.

McCarthy, Michael and Ronald Carter. 2006. *Cambridge grammar of English*. Cambridge: Cambridge University Press.

Parfitt, M. Kevin. 2008. Building failures: Avoiding the mistakes of yesterday leads to the successes of today. *Journal of Architectural Engineering* 14: 31.

Parfitt, M. Kevin and Elizabeth Parfitt. 2007, January. Failures education: The key to better engineering design. *Structure Magazine*, 10–12. http://www.structuremag.org/wp-content/uploads/2014/09/p10-12C-EdIssues-FailuresEducation-Jan071.pdf.

Reppen, Randi. 2012. *Grammar and beyond 2*. New York: Cambridge University Press.

Sageev, Pneena and Carol Romanowski. 2001. A message from recent engineering graduates in the workplace: Results of a survey on technical communication skills. *Journal of Engineering Education* 90: 685–693.

Sales, Hazel. 2006. *Professional communication practices in engineering*. Basingstoke, UK: Palgrave Macmillan.

Tenopir, Carol and Donald King. 2004. *Communication patterns of engineers*. Hoboken, New Jersey: John Wiley/Institute of Electrical and Electronic Engineers.

Thomas, Stephen, Richard Tucker and William Kelly. 1998. Critical communication variables. *Journal of Construction Engineering and Management* 124: 58–66.

Wingate, Ursula. 2015. *Academic literacy and student diversity*. Bristol, UK: Multilingual Matters.

Winsor, Dorothy. 2003. *Writing power: Communication in an engineering center*. Albany, New York: State University of New York Press.

Wolfe, Joanna. 2009. How technical communication textbooks fail engineering students. *Technical Communication Quarterly* 18: 351-375.