

Discourse Versus Corpus Analysis in Materials Design: A Case Study

Michael Guest
University of Miyazaki

Reference Data:

Guest, M. (2013). Discourse versus corpus analysis in materials design: A case study. In N. Sonda & A. Krause (Eds.), *JALT2012 Conference Proceedings*. Tokyo: JALT.

In this paper I argue that a discourse analysis approach may be preferable when applying corpus data to the development of teaching materials. This argument is based on a combination of my research and subsequent classroom experience. Having initially taken a corpus analysis (CA) approach to nursing English speech data collected from 3 Asia-based hospitals, and applying some of the findings directly to nursing English classroom materials, I noted that student responses to these initial corpus-based materials were not positive. Subsequently, the materials were drastically revised using a discourse-analysis (DA) approach, that is, by first establishing and utilizing wider categories of discourse to provide meaningful frames for the content applied to the materials. The resultant student response to the revised materials was much more positive. This experience suggests that when applying authentic data for the purpose of developing EFL classroom materials, a DA-based approach may be more effective.

本論文は、教材開発にコーパスデータを応用した場合に、ディスコース分析が適当か否かについて論じたものである。論文著者の研究およびその後の授業体験に基づいて議論を進める。看護英語領域データとして、アジア圏の3箇所から収集した会話データをコーパス分析し、そこから得た知見をそのまま看護英語教育への応用を試みた。しかし、コーパスを基本とした教材に学生は肯定的な反応を見せなかったため、ディスコース分析アプローチを用いて大幅に教材を変更したところ、学生のモチベーション並びに成績の両方で劇的な改善が見られた。

CORPUS DATA has long been used as raw material for classroom applications. In fact, most current materials make some mention of being informed by recent corpus data. The difficulty for many teachers, however, lies in making procedural and methodological decisions as to exactly what data from a corpus should be applied, how to apply it so as to make it most relevant to learners, and to what degree, if any, the raw data should be adjusted for pedagogical purposes. How corpus data might best be applied to materials has been a topic of considerable controversy since corpus-based approaches were first proposed.

This paper outlines some problems encountered by two researchers (a colleague and I) who attempted to apply corpus data directly to classroom materials using a corpus analysis (CA)-based approach and our subsequent attempts to revise those materials using a discourse analysis (DA)-based approach when the former method failed. This showed that a DA-based approach is more suitable when applying narrow instances of corpus data to students studying English in a specialized field, in this case nursing.



We collected numerous samples of real-time spoken nursing English from hospitals and clinics in Singapore, the Philippines, and Malaysia over the course of 2010 and 2011. The data was collected as handwritten notes and was selective rather than exhaustive or comprehensive, meaning that greater attention was paid to particular interactions and speech events (see Guest & Nambu, 2011a; 2011b).

The immediate, short-term goal of the data collection was to analyze the data and concordance the results with existing nursing and general corpora. The long-term goal was to apply the results to developing nursing English materials for Japanese university nursing students at my university. This, it was felt, would help provide Japanese nursing students who wish to train, study, or practice abroad with a coursebook of realistic and useful nursing English.

Particular attention was paid to authentic instances of turn-taking, strategic competence, and pragmatics (particularly illocutions), all areas in which both existing nursing English textbooks in Japan and student performance were felt by us to be deficient. However, we encountered problems after collecting this data and applying it to classroom materials, which persuaded me to adopt a different theoretical approach when undertaking revisions.

Applying Corpus Data to Materials Development

Corpus-Based Versus Corpus-Driven

Although corpus data is regularly applied to teaching materials, there still exists controversy as to what degree materials should be based upon raw corpus data, as opposed to data merely informing those materials, a dichotomy usually known as corpus-based vs. corpus-driven. We might describe the dichotomy as follows, with the corpus-driven features on the left and corpus-based features on the right:

- big picture vs. details,
- macro vs. micro,
- holistic vs. discrete points, and
- top-down vs. bottom-up.

When applying corpus data, materials makers have a choice of applying the raw data directly to materials or reformulating the data in order to make it suit a particular classroom need or syllabus. The former method implies a CA-based approach while the latter would be corpus-driven, allowing for an analysis of wider discourse patterns. After the speech data described earlier was collected, we opted for a CA-based procedural model for materials development. The process we utilized was as follows:

1. pattern recognition and identification,
2. generation of frequency data, and
3. utilization of concordance.

This CA process, since it is more quantitatively focused, is typically procedural and mechanical. Once speech patterns are identified, the researcher generates frequency data, which usually involves some type of concordancing, either with the new corpus or with existing corpora or with both. The most frequent items become the primary teaching targets employed in the materials.

From Corpus to Discourse

The fundamental differences between a CA corpus-based approach and a DA approach might be summed up as follows, with the characteristics of CA noted on the left and those of DA on the right:

- use of representative samples vs. integrity of whole text,
- quantitative vs. qualitative, and
- language per se vs. contents expressed by language.

Over the past 10 years the CA-DA dichotomy has been bridged somewhat. There have been numerous articles published, both pedagogical and research-based, that outline approaches for incorporating raw corpus data into teaching and materials development. At the forefront are two practical guidebooks suggesting methods of applying corpus data to classroom materials, the first by Hunston (2002) and the second by O'Keefe, McCarthy, and Carter (2007). Prior to this, Carter and McCarthy (1995) originally attempted to link corpus data with the classroom by advocating a methodological refocus, from the old presentation-practice-production (PPP) paradigm to one that emphasizes illustration-interaction-induction (I-I-I). Johns (1997) advocated data-driven learning (DDL), using students as language detectives, such that they discover forms by themselves from corpus examples and apply them as they see fit, thus bypassing the role of the teacher as analyst.

Critiques of a Corpus-Based Approach

The corpus-based approach, championed primarily by Sinclair (1997), who was less concerned with practical applications to teaching, argues for the inclusion of real examples only “based on descriptive data” (p. 30), and has had its critics, most notably Widdowson (2000), who claimed that “Corpus linguistics describes text (as product), not discourse (as process). Discourse significance cannot be read off from the data” (p. 9). He added that, “[CA] cannot account for the complex interplay of linguistic and contextual factors whereby discourse is enacted.” In other words, Widdowson believed that any utilization in the classroom already implied that authenticity had already been compromised, the classroom already being an artificial construct.

Widdowson’s objections have found kindred voices. Martin (1999) said that “Analyzing a lot of text from a corpus at one time forces the analyst to ‘lose contact’ with the text” (p. 50).

McEnery, Xiao, and Tono (2006) argued that, “Corpora rarely provide explanations for observations. Explanations must be developed using other methodological tools, such as intuition” (p. 121), while Hunston (2002) felt that, “Corpus-based approaches obscure . . . the character of each text as text . . . [and] the role of the text producer and society of which they are a part” (p. 110).

Each of these arguments implies that when developing raw data into materials there must be some type of further analysis or reinterpretation that includes wider context in order to make any teaching or learning using corpus-based materials viable. This is precisely where the wider scope of DA comes into play as an option for materials writers. After all, as McCarthy (1998) famously wrote, “Discourse drives grammar” (p. 78). A similar point was made by Kennedy (1998), who claimed that, “What we say and how we say it is influenced by who we are talking to and where the interaction is taking place” (p. 174).

Thus it is clear that wider analytical tools that focus upon discourse are needed in order to make materials more relevant to learners and further implies that frequency is not the sole arbiter of importance or relevance to learners. As Kennedy (1998) argued, “Frequency should be only one of the criteria used to influence instruction” (p. 290), a position echoed by Renouf (1987), who noted that, “. . . raw frequency should be adjusted for use in a syllabus (p. 168).”

Raw Data Examples

As stated earlier, we initially opted for a more CA-based approach. This meant identifying and isolating significant samples of turn-taking strategies, strategic competence, and illocutions, and concordancing these, using the concordancing software Antconc 3.2.4 (Anthony, 2011) with either the British National Corpus (BNC, 2011), or the University of Michigan’s nursing English corpus, in order to determine relative frequency. More

frequent forms were given priority in terms of being incorporated into the teaching materials. Once these most frequent forms were isolated and identified, models were developed by the researchers as dialogues and sample texts to be used in the coursebook. Salient among the samples collected and applied to the new materials were those showing that illocutions were commonly generated in response to treatment suggestions, especially as refusals or denials, such as:

Nurse 1: So we can change to a smaller drip?

Nurse 2: Platelet count still elevated so . . .

Strategic competence was also widely noted, with repair being most often initiated by repetition of a keyword alone followed by some confirmation or elaboration, such as:

Nurse 1: Panadol still required.

Nurse 2: Still?

Nurse 1: But now at 150 cc.

Also widely noted through concordancing, and subsequently applied to the initial coursebook materials, was the disproportionate usage of certain high register and genre specific terms (such as, *presented with, manifestations, acute [adj./n] discomfort, after consultation with, [noun]+intake*) in comparison to general usage. These chunks of authentic nursing discourse collected from the field became the foundation for the new classroom materials.

Student Responses to Initial Materials

Noticing a rather doubtful and hesitant response by the nursing students to the new materials, I decided to interview selected students regarding how they felt about the new materials. Although these interviews were informal, I asked students to be candid regarding their opinions because I hoped to improve the materials. I explained to the students that the samples were based on

authentic nursing English noted in several Asian hospitals and clinics, but neither the authenticity nor normative quality of the material appeared to be motivationally persuasive.

Among the most common negative responses were those to the effect that the focus of these new materials was too narrow and detailed. The argument was presented by students that although they recognized these forms as authentic and valid, they did not view them as being learning priorities, given that they were 1st- and 2nd-year students who had not yet developed basic competency in the language. They did not understand how to fit the various data samples into the bigger framework of nursing English. This indicated a scaffolding problem. To these students an ESP nursing focus seemed murky and indistinct.

Connected to this was a second commonly stated claim, specifically that the forms being taught seemed artificially applied to create the materials. Students were alert to the fact that forms had been isolated and dialogues built around them. The surrounding texts were created to give the forms context but they appeared to the students forced and awkward.

In other words, there appeared to be a disjunction between the rather narrow scope of the CA-based observations and the more general goal of producing ESP learning materials. Sensing that the students' negative responses were valid, I decided to take a different approach to both selecting the data and creating the materials.

Revision of Materials Based on Discourse Analysis

In order to provide a more understandable framework for our students I decided to approach the data from a wider, discourse-based perspective. Rather than focusing on specific forms, I instead analyzed the original data for instances of more general discourse frames.

Among the research questions asked when gathering this data were: Who typically talks to whom about what? What are the most common speech events? What are the goals/purposes of these speech events? How is the discourse typically managed? What, if any, are the identifying features of this discourse? Among the most significant answers were the following:

- About 80% of nurse speech is nurse–nurse.
- Four speech events (roll call, handover, instructional, medicine administration) comprised over 75% of all nurse speech.
- Three of these four speech events use highly abbreviated forms (ellipsis) as a default, and this is typical of workplace speech grammar.
- About 80% of nurse speech is directly related to a written document, such as a nursing admission form or patient chart.

These insights helped provide a framework for revising the teaching materials. The process of materials development using this discourse-based approach was reformulated as

- speech events: participants, goal or purpose;
- text management: openings and closings, turn-taking, and register; and
- discrete features: semantic and pragmatic.

In other words, the discourse was now framed as a hierarchy in which wider discourse features, such as the main speech events carried out by nurses and their relevant features, such as the participants and goals or purposes of the speech event, were identified first. This was followed by noting features of the management of these speech events, such as openings and closings and turn-taking strategies. Only once these frameworks became clear and identifiable were discrete features of semantics and pragmatics (including samples of strategic competence) finally introduced. This represents a near reversal of a typical CA-based procedure.

When applied to classroom materials, this meant that broader frameworks of discourse which could be more easily grasped by students were presented first, allowing the discrete features (such as strategic competence and illocutions) to be placed within that discourse more naturally.

Two Classroom Examples

Two examples of these revised materials are presented below in abbreviated form. The first focuses upon reporting patient data from nurse to nurse. Three task questions are asked, followed by a full sentential report of the data:

- How would you report the following patient information in a hand-over?
- How would you write this information as chart notes?
- How would you give the same information to the patient or the patient's family?

Patient data: In room 603 there is a 17-year-old female patient. *Her FBC count is 0.8. We did a monospot trace which was positive. Her increased intake of greens has reduced her blood platelet count to 135,000. This morning she had a fever of 38.7 degrees. She is currently receiving amiodarone intravenously at a rate of 150 mg per day using a 14 point drop. She also had an x-ray taken which just came back and indicated Eagle Syndrome.

The second is a sample handover nurse–nurse patient report interview, based upon data contained in an authentic clinic admission form. This is used as a writing and speech information gap exercise reflecting the type of abbreviated discourse noted in roll call speech events. Students are instructed to develop their own invented data, fill in the empty patient admission form, and then report it to a partner in a similar manner:

Next, John Smith bed 14, Male, aged 55.

- Any allergies? What? What kind of reaction?
- Any medical problems or infectious diseases?
- Any past surgeries or hospitalizations? When? Why? Where?
- Has he ever had a blood transfusion?
- Any current medications? What? How often? When was the last dose?
- Any assistive devices?
- Any contractures or deformities?
- Is he able to walk? Feed and drink by himself? Dress and groom? Use the toilet and bathe? Turn in bed?
- Is there any fall risk?
- What's his skin color? Temperature? Condition?

With this new discourse-based focus in mind, the revised course syllabus for nursing students was amended and rewritten (see Figure 1).

Nursing English 2: Workplace English
<ul style="list-style-type: none"> • Admission forms: Asking for, reporting and writing data • Roll call & handover: Using authentic documents • Instructions (preceptor–preceptee + trainer–trainee) • Procedural and administrative English • Nurse–patient and nurse–patient's family discourse • Nurse–doctor discourse

Figure 1. Revised Course Syllabus for Nursing Students

Student Responses to the Revised Materials

I immediately noticed a much more positive response to the introduction of these new materials both in terms of task engagement and quality of evaluation projects based upon these materials. This more positive response was confirmed after again conducting an informal survey among selected students. The following are among the more commonly cited reasons for the more positive response:

1. Students claimed that they were better motivated by having access to more general knowledge about how nursing workplace discourse is managed. They could see how the content was practical and applicable to day-to-day experience. This seemed to better suit their expectations as to what should be studied and practiced in an ESP course.
2. Students claimed to better understand the discrete strategic, semantic, and pragmatic items introduced in the revised materials since they could now be meaningfully placed within a larger, more concrete discourse picture frame.

There are also indications, based on subsequent test and task results, that these discrete items were being better internalized as a result of the revised approach. Student-generated role-plays incorporating this material was of a higher standard than when based upon the initial materials. New items were more appropriately located in the student discourse. Results of discrete-item paper tests measuring student ability to understand and apply the usage of these items also improved considerably.

Conclusions

Two conclusions, one general and one procedural, were reached as a result of this research-to-materials-development process. First, I advise caution in using the direct application of corpus

data into classroom materials (the corpus-based approach), particularly if the students are still at an intermediate or pre-intermediate level. Students who still lack productive competency in overall English skills might understand items procured from corpus data in an abstract sense but lack the holistic ability to connect them to a bigger picture. The result is likely to be, at best, a slightly raised awareness of a few isolated discrete items rather than a general improvement in English skills. At worst, the lack of wider perspective means that motivation and performance both decrease.

The second (procedural) suggestion is that when making ESP materials it is best to establish discourse categories first. The students for whom the materials were intended were nursing students, and the course ESP-based. Students in such courses need general frameworks in order to locate language in such a way that it becomes meaningful and practical for them. A DA, corpus-driven approach to interpreting the raw data, resulting in meaningful cognitive scaffolds such as participants and common speech events prior to the application of discrete language samples, is better suited to this type of student and course than a purely CA-based approach. For linguistics students, or those otherwise particularly interested in unraveling the nuances of linguistic minutiae, a CA-based approach may be welcomed by the learners, but caution should be applied in the case of ESP students. I would also suggest that the materials developer should establish which speech events and which discourse features therein will be most relevant to learners. Noting that set ESP speech events such as roll call, handover, instruction sessions, and medicine administration dominate workplace nursing English allowed both the learners and the materials developer to establish real-world references that were meaningful to learners. By locating discourse features such as strategic competence or turn-taking within such speech events, students could understand their roles and functions more fully. As Willis (1994) said,

It is the learner who has to make sense of the insights derived from input, and learners can only do this by considering new evidence about the language in the light of their current model of the language. This argues against presenting them with pre-packaged structures and implies that they should be encouraged to process text for themselves... to reach conclusions which make sense in terms of their own systems. (p. 56)

Therefore micro-features of the type noted in the corpus should be inserted into discourse samples and tasks only at the final stage, as a flavoring. This is because just as grammatical rules are often better internalized only when the wider discourse frame is clear, so too do students appear to better understand discrete language points when these are placed naturally within clear frameworks and not artificially forced into dialogues and other texts. The top-down approach that I advocate appears to better ensure that discrete points may be internalized than does a bottom-up approach, in which the materials developers start by creating text around the discrete items they hope will be learned.

Using the more methodologically sound DA-based, corpus-driven approach in proceeding from corpus data to materials development appears from our experience to be more sensitive to the cognitive frames and scaffolds employed by ESP learners. This sense of cognitive congruence and suitability in turn helps to motivate students to better internalize the materials.

Bio Data

Michael Guest is Associate Professor of English in the Faculty of Medicine at the University of Miyazaki. He is a regular monthly columnist on ELT affairs in the *Daily Yomiuri* newspaper and an active blogger on similar issues at eltnews.com <mikeguest59@yahoo.ca>

References

- Anthony, L. (2011). Antconc 3.2.4. [Computer software]. Retrieved from <http://www.antlab.sci.waseda.ac.jp/software.html>
- British National Corpus. (2011). Retrieved from <http://www.natcorp.ox.ac.uk/>
- Carter, R., & McCarthy, M. (1995). Grammar and the spoken language. *Applied Linguistics*, 16, 141-158.
- Guest, M., & Nambu, M. (2011a). Framing nursing discourse for English for specific purposes materials' development. *Journal of Medical English Education* 10(3), 78-83.
- Guest, M., & Nambu, M. (2011b). Strategic competence and ESP materials development: A pilot study in authentic nursing discourse. *Research and Practice in ESP*, 10, 47-57.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Johns, T. (1997). Contexts: The background, development and trialling of a concordance-based CALL program. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 100-115). London: Longman.
- Kennedy, G. (1998). *An introduction to corpus linguistics*. London: Longman.
- McCarthy, M. (1998). *Spoken language and applied linguistics*. Cambridge: Cambridge University Press.
- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. New York: Routledge.
- Martin, J. (1999). Grace: The logogenesis of freedom. *Discourse Studies*, 1, 29-36.
- O'Keefe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom*. Cambridge: Cambridge University Press.
- Renouf, A. (1987). Moving on. In J. Sinclair (Ed.), *Looking up: An account of the COBUILD project in lexical computing* (pp. 167-178). London: HarperCollins.
- Sinclair, J. (1997). Corpus evidence in language description. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 27-39). London: Longman.
- Widdowson, H. (2000). The limitations of linguistics applied. *Applied Linguistics*, 21, 3-25.
- Willis, D. (1994). A lexical approach. In M. Bygate, A. Tonkyn, & E. Williams (Eds.), *Grammar and the language teacher* (pp. 56-66). Hemel Hempstead, UK: Prentice Hall International.