

Corpus-based Genre Analysis for Language Assessment

| | |
|------------------------------|---|
| 著者名(英) | Siwon Park |
| journal or publication title | 神田外語大学紀要 |
| volume | 22 |
| page range | 275-294 |
| year | 2010-03 |
| URL | http://id.nii.ac.jp/1092/00000624/ |

CORPUS-BASED GENRE ANALYSIS FOR LANGUAGE ASSESSMENT

Siwon Park

INTRODUCTION

Use of corpora and corpus-based genre analysis in language assessment is still an underdeveloped area. The potential of such use has been recognized and discussed in literature on corpus linguistics (e.g., Hunsten, 2002); yet, actual attempts to apply corpus analysis to practical language assessment are hardly seen (for an exception, see Biber et al., 2004). Therefore, the discussion of this paper on the topic will be around exploring the potential of corpus-based genre analysis for language assessment, particularly focusing on its potential role in addressing specific problems in language assessment. I will first consider two approaches to genre – Swales and Biber’s – with a focus on how they have attempted to explain variation in text. I will connect the first discussion with a supposition that variation in text (i.e., genre) may require different language knowledge and skills for language processing and performance. That is, the variation in genre may be responsible for the variation in language performance. Finally, I will consider the potential roles of corpus-based genre analysis in dealing with problems often identified in language assessment, particularly on task-based performance assessment.

GENRES, TEXT TYPES, AND REGISTER

A common understanding of genre is as a particular class of speech events which are considered by the speech community as being of the same type (e.g., prayers,

sermons, letters, novels, etc.). Swales (1990) defines a genre as something that “comprises a class of communicative events, the members of which share some set of communicative purposes” (p. 58). Fourteen years later, Swales (2004) revisits his definition of such and reformulates it as a set of metaphors with multifaceted perspectives in defense of his definition of genre in terms of communicative purposes (p. 68). Knapp and Watkins (2005) view a genre in terms of processes, such as describing and arguing, rather than products (i.e., text types).

Discussing the importance of classification of texts into different genres and calling for the need for language teachers and researchers to better understand the exact characteristics of the language they deal with, Lee (2001) argues that the reason one would want to deal with a specific genre is to be able to define the scope of any generalizations they would want to make (p. 37). I believe the same argument can be applied to the use of genre-based corpora in language assessment. That is, we are rarely concerned with learner language performance in the testing context in focus. Rather, we hope to make generalizations of performance on a task to another and performance in one context to that in another (Brown et al., 2002). For this goal to be achieved, characteristics of tasks and test contents must be clearly identified so that comparisons can be made across different tasks as well as contexts. Clearly defined genre-based corpora will help us understand to what extent we could generalize our observation from one instance of task/context to another within a genre (or domain) or across different genres.

With respect to corpora and corpus linguistics, many corpus-based studies are based implicitly and explicitly on the notions of genre, registers, and/or text types. Without understanding the characteristics of the target corpus (or its sub-sections), arguments based on the findings from a corpus analysis can never be justifiable. Therefore, classification of genre/text types enables more systematic investigations

of the linguistic features of language; by means of comparison and contrast. Also, in corpus linguistics, sub-corpus labeling/classification is important, as it serves for the users to select appropriate text types for their intended uses (e.g., EAP, ESP, discourse analysis, lexicogrammatical and collocational studies, etc.).

The way we distinguish between genres and text types is using internal and external criteria: genre is defined based on the external (conventional) criteria (e.g., audience, purpose, etc.) and text types on the internal criteria (e.g., linguistic characteristics). Another commonly notable term is register. Register concerns stylistic variety, that is, variation in a person's speech or writing. It also refers to a speech variety used by a particular group of people – e.g., sharing the same occupation

TWO APPROACHES TO GENRE ANALYSIS

Two approaches are notable to genre analysis: Swales' ethnographic approach and Biber's corpus-based statistical approach. This section briefly reviews the two approaches.

Ethnographic approach (Swales) To Swales, genre analysis concerns more with the specific context of a professionally or socially recognized discourse type. Therefore, genre analysis is an attempt to find patterns of conventional formulations accounted for by processes of use and production. The formulations of such are not necessarily based on grammatical or cohesive features. Rather, purposes shared by a discourse community serve the basis of individual genre (Swales, 1990, 2004).

Corpus-based approach (Biber) Biber (1988, and others later) has investigated the typology of text types using multi-feature, multi-dimensional analyses. To Biber, text-types exist on a continuum, and their differences can be explained by an analysis of internal linguistic properties of texts using corpora that are grammatically tagged,

as opposed to external social and rhetorical features. The premise of this approach is that formal (linguistic) differences reflect functional differences; hence, can be interpreted functionally. For instance, Biber (1988, 1989) identified 67 linguistic features from LOB and LLC based on 481 text samples. Text typology was identified based on common syntactic and lexical (not functional) features of text samples (using factor analyses). Also, text dimensions were identified, and functional interpretations were made accordingly, following that typological classification. Three notions of function (of forms) were identified:

1. Roles of forms in discourse (e.g., ‘passives’ being used to rearrange information structure of a sentence, etc.)
2. Situational or processing constraints (e.g., ‘hedges’ to reflect the difficulty of more precise lexical expression under real-time language production)
3. Situational or social distinction index (e.g., ‘jargon’ as an index of a particular group membership)

Decision criteria for textual dimensions used were functional interpretations and statistical reliability and significance.

Biber and Finegan (1991, p. 213) address two issues regarding genre and text types in their study: to what extent genre categories are linguistically homogeneous and also if linguistic heterogeneity entails that a genre category is invalid – i.e., if not all genres are equally homogeneous. They report that the genres of English can effectively be categorized linguistically, and there are large (and statistical) differences among them (Biber, 1988). Moreover, some genres include distinguishable sub-genres (e.g., newspaper with different sub-genres) and have much wider (functional and/or developmental) variation than others (e.g., science

fiction).

Differences of the two approaches Differences between Swales and Biber's approaches come mainly from how variation in texts is considered (and possibly measured). Biber in his register-analysis approach to genre has attempted to measure variation in texts by the (co)occurrence of grammatical features while Swales has considered such variation due to differences in conventionally recognized instances of language in a specific discourse community. Furthermore, Biber's approach claims that a set of co-occurrence features has similar functions throughout the language. To Biber (1989), "genre distinctions do not adequately represent the underlying text types of English...; linguistically distinct texts within a genre represent different text types; linguistically similar texts from different genres represent a single text type" (p. 6). Swales, however, assumes that grammatical forms have different functions for different discourse settings, and that within a genre a rhetorical move may be realized differently depending on what linguistic features can be adapted according to the practices of the discourse community. Swales (1990, p. 42) therefore distinguishes between genre as a conventionally recognized instance of language in a discourse community and register as the 'language of' a certain field, such as science or journalism. However, to Biber (1995), register and genre cannot be separated. Particularly, he views register as "the general cover term associated with all aspects of variation in use" (p. 9).

GENRE AND LANGUAGE KNOWLEDGE AND SKILLS

Researchers have repeatedly proposed the possibility that genre/register may be responsible for differential language performance, and the effect may be significant to be even treated as a source of test bias (Alderson, 2000; Bygate, 1999; Douglas & Selinker, 1993; Shohamy, 1984; Shohamy & Inbar, 1991; Shohamy 1997). For

instance, Alderson (2000), in his discussion on variables that affect the nature of reading, points out that when deciding topics or text contents in reading assessment, the test designer needs to be aware that “variation in text content might be expected to lead to different test results” (p. 63). Bygate (1999) found that EFL learners’ performance on different tasks – narrative and argumentative – were significantly different on their production of grammatical patterns. Douglas and Selinker (1993) have revealed that a different combination of contextual features in mathematics and chemistry versions of a speaking test produced not only different scores, but also different rhetorical structures of test-takers’ responses. Shohamy (1997) argues that “in a number of studies of oral testing tasks, it was shown that different testing tasks have an effect on test takers’ scores... thus test takers’ scores on oral tests varied depending on the elicitation method, discourse styles and genres” (p. 341). For instance, Shohamy and Inbar (1991) found that test-takers’ scores were significantly different on a listening test with three discourse types – interview, lecture, and a news broadcast, even after having controlled for the topic factor of the genres. According to Shohamy (1994), test task genre is responsible not only for language production, but also for processing and strategies used by test-takers. Therefore, distinct genres of texts or tasks may require different language knowledge and skills in language processing (i.e., comprehension) and performance. As such, understanding of text/task characteristics may be crucial to promote our understanding of task demands and sources of differential performance.

Corpus analysis has the potential of being a research means to help analyze and understand such relationship between distinct genres and the language knowledge and skills that they may require. That is, corpus linguistics could help examine if and how distinct genre of texts or tasks require different language knowledge and skills in language processing and performance. Although rare, there are signs of such an

application in language assessment (e.g., Ball, 2001; Brown et al., 2005; Biber et al., 2004; Yamada, 2005).

Regarding the potential use of genre-based approach to analyzing the characteristics of spoken and written texts, Brindley (1998) states:

The genre-based approach is derived from systemic-functional linguistic theory (Halliday, 1985) and is concerned with exploring the way in which language is used to realize a variety of social purposes. The genre-based approach to text analysis involves specifying the discourse structure and particular linguistic features of text types which are relevant to the target population (e.g., argument or reporting a process in the case of secondary school learners). It thus offers testable hypotheses concerning the specific linguistic features which need to be controlled in order for certain tasks to be carried out; and these features can, in turn, serve as criteria for determining task achievement (p. 129)

The problem of variation The development of assessment materials requires on-going judgment about language use, to decide on the linguistic contents that should be represented in these materials (Biber et al., 2004). Such decision has often been made based on the test developers' intuition, field experts' judgments, or, at best, discourse analysis using small language samples. Unfortunately, however, such bases of decision making are often found incorrect (Biber & Conrad, 2001; Paltridge, 2001). Nonetheless, understanding the linguistic contents, in particular, their variations within and across genres, is crucial in accumulating content validity arguments and generalizations of the score interpretation; genre analysis using corpora may be found useful in that regard.

In promoting our understanding on the variation in genres and registers, corpus

based research has been proven useful. For instance, Bondi (2001) examined ways in which small corpora can be used to study language variation across genres within given discourse areas. Her study shows how corpus linguistics can be used to study variations of language features across genres in a small corpus such as functional units and lexico-grammatical patterns. Biber (1988, 1992, 1993, 1995), Biber and Finegan (1991, 1997), Biber, Conrad, and Reppen (1994, 1998), and Greenbaum and Nelson (1995) are examples of such research. Studies conducted by Biber and others all used the multi-dimensional approach to analyzing the linguistic co-occurrence patterns associated with register variation in data-based quantitative terms. Greenbaum, Nelson, and Weitzman (1996), using a manual analysis of a modest-sized corpus, revealed that ‘complement’ clauses are the most frequent subordinate clauses, with that-clauses preferred over clauses with ‘zero’ complementizers in formal language modes. In spoken texts, the majority of complement clauses are ‘finite’, while, in writing, non-finite complements are the most frequent. Greenbaum and Nelson (1995; also in Biber, 1988) found that for syntactic complexity, “within each mode (or genre), there is considerable variation, and this variation tends to be greater than that between speech and writing as such.” Swales (2001) claims that previous studies on academic genre are based on too small language samples and that academic speech is much more variable in structure, function, and style than academic writing.

Variation has also been noticed in speech acts. Deutschmann (2003) found that the age and social class differences were important factors affecting the use of the apology form. She also found that the number of participants present during a conversation affects the total apology rates; more participants led to more frequent use of the form. This finding was further supported by the apology rates in the twelve genres, which could be positively correlated to the number of conversational

partners present during the conversations in those genres. In addition, as mentioned earlier, our intuition about the typicality of text characteristics is often found to be wrong. For instance, our intuition of core vocabulary within a specific genre may not always be correct, and the definition of core vocabulary may not be held consistent across different genres (Lee, 2001). Rowley-Jolivet (1999) has shown that our stereotyped understanding about the discourse of conference speech tends to be more conversational than what we had thought before. To make the issue more complicated, the variation of such characteristics discussed above is not stable and changes over time.

Needs for target discourse/context analysis The issues of variation underscore the needs of target discourse/tasks/context analysis in performance assessment in general and task-based assessment in particular. Brown et al. (2002) inform that “task-based performance assessment has implications for (a) the intended educational contexts in which such assessment is used, (b) test methods appropriate for such assessment, and (c) performance evaluation methods and interpretive criteria related to such assessment” (p. 9). Notice the term context used in their elaboration of the implication of task-based assessment. Typically, the context is to be identified based on the needs using needs assessment. However, defining a context has not been successful in practical assessment, in part due to the lack of data sources against which we could use to validate the characteristics of contexts. Moreover, in (task-based) performance assessment, the terms, real world tasks, and associated situational and interactional characteristics where communication plays a central role have taken the priority stance to highlight its significance. Yet, under such claim, one must assume *a priori* a proper means of defining what he/she means by real, situational, and interactional characteristics, not by intuition or by needs subjectively identified, but based on actual use/production data collected within the

context where the interpretation of performance is to be related and from the target population who the test-takers will interact with. As mentioned earlier, the intuitive definitions and operationalization of these terms may in fact be found not correct and hence pose a threat to test validation. Needs assessment is not immune to such a possibility, as needs are often identified through felt needs, rather than the actual needs passed unnoticed for some reasons.

Corpora properly designed, sampled, and systematically structured from a clearly defined target context (e.g., the academic English used in American colleges and universities) can provide a means to help clarify the problems mentioned above (e.g., Biber et al., 2004; Brown et al., 2005). Biber et al.'s (2004) goal of their corpus study was to suggest a data-based source to help deal with such problems. In their statement of the motivation of the project, Biber et al. (2004) claim:

To better understand the nature of the tasks that incoming international students encounter in the university, and ultimately to help students develop the language skills required for those tasks, we need a comprehensive linguistic description of the range of university spoken and written registers that are predominant in the university context (p. 2).

They go on to argue that under the current popularity of the task-based approach to language teaching and assessment, teachers and researchers face difficulty identifying and analyzing the language demands of the college-university setting. They claim such identification and analysis can be done using a large scale corpus and corpus analysis. This is in close line with what Lee (2001) has identified as the major purpose of text classifications into different genres. As mentioned earlier, he argues that genre is important since it provides us with a means to understand exactly

CORPUS-BASED GENRE ANALYSIS FOR LANGUAGE ASSESSMENT

what kind of language we are to examine or describe. That is, corpora and corpus-based genre analysis enable us to define the scope of generalizations we would want to make of the language sample we are dealing with.

The problem of test validation and generalizability The potential definitional problem mentioned earlier also connects with an issue of generalization of the interpretations about students' abilities from performance on one task to performance on a set of related tasks (Brown et al., 2002; p. 11). Researchers in language assessment have often faced difficulty with validation (especially of content/context) and generalization of inferences observed from test performances to non-testing ones. More specifically, researchers have suffered from identifying to what extent the textual materials or test tasks are representative of the linguistic characteristics of the target registers. In addition, the recent popularity of performance assessment has observed a methodological challenge of defining or identifying the real life tasks. Also, needs have been expressed frequently for a methodology to properly examine the characteristics of tasks and to identify the sources of task demand.

POTENTIALS OF CORPUS-BASED GENRE ANALYSIS FOR LANGUAGE ASSESSMENT

The use of corpora in language teaching and language assessment has become popularized due to their accessibility, cost effectiveness, and the wide range of tools available for searching corpora. Used properly, corpora can reveal many aspects of language use quickly and accurately, while helping reduce researchers' labor. With respect to the application of corpus linguistics to language assessment, Barker (2005) claims that corpora can be used to develop and validate language tests. As a future possibility, she also considers that corpus-informed testing can be used to compare test performances with that in less formal interactions. There are signs of

her suggestions being realized in language assessment and examples of such studies include test validation studies of Cambridge First Certificate in English (FCE) using small corpora (Galzaczi, 2003; Lu, 2003), a study on the quality of test-takers' discourse on two different TOEFL task types at different levels of proficiency (Brown et al., 2005), and a study to identify linguistic characteristics of spoken and written academic registers (Biber et al., 2004). The potential of corpus-based genre analysis for language assessment can be examined from two perspectives – task-related and product-related uses of corpora for language assessment.

Investigation of characteristics of tasks/texts/contexts The characteristics of test texts/tasks/contexts can be examined *a priori* using corpus-based genre analysis. Douglas (2000) suggests that a corpus of field specific discourse is a potentially useful tool for the investigation of target language use, both written and spoken (p. 271). He further suggests that corpora “can add detail to our understanding of the linguistic properties of specific purpose speech and writing (p. 273). Citing Alderson (1996), Douglas goes on to say that corpora of written texts may be used for selection in language tests and corpora could inform test developers of content, syntax, and level of vocabulary for their test specifications. Brindley (1998) argues that the genre-based approach to text analysis can “offer testable hypotheses concerning the specific linguistic features which need to be controlled in order for certain tasks to be carried out; and these features can, in turn, serve as criteria for determining task achievement” (p. 129). To Brindley (e.g., 2000, 2002, and Wigglesworth, 2000), achieving parallel tasks have been a major concern in his outcome-based assessment; therefore, a research means that allows systematic investigation of variations in task condition and characteristics must suggest great merits. In that regard, the genre-based corpus analysis can help specify the discourse structure and identify particular linguistic features of text types most relevant to the target population. That is, such

CORPUS-BASED GENRE ANALYSIS FOR LANGUAGE ASSESSMENT

an analysis can be used as a means for text grading and classification. That will help researchers (Brindley, 1998) develop sound assessment criteria as well, e.g., rating scaling. The genre-based corpora and corpus analysis can help obtain sufficiently valid information about the linguistic features of the texts and how texts with such features are processed by test-takers. Using that information, one may be able to manipulate the processing/performance demands on the text/task.

Investigation of content coverage and representitiveness Another use of corpus-based genre analysis in language assessment is to investigate content coverage or representativeness of test tasks (Alderson, 1996), which eventually concerns content validation. A corpus of a context defined clearly in genre/register terms can be constructed and used for such type of investigation. A good example of such investigation is Biber et al. (2004). In an attempt to evaluate the representativeness of English as a second/foreign language materials and assessment instruments, they realized a lack of a data source (i.e., a representative corpus) for the research purpose. They therefore undertook a project to construct the TOEFL 2000 Spoken and Written Academic Language (T2K-SWAL) Corpus and used the corpus to develop diagnostic tools to indicate whether the language used in T2K listening and reading comprehension tasks is representative of real-life language use. In Europe, UCLES has been actively involved in the area of corpus application to language assessment (Ball, 2001). Essentially, the same approach as the TOEFL team was used by UCLES, but it was initiated much earlier. UCLES has however found somewhat different uses of corpora in corpus-informed research for language assessment. Their main focus of corpus research has been to develop learner corpora as well as native corpora, and the use of them 1) to develop examination materials and 2) to standardize test tasks especially using learner corpora collected from previous test administrations of candidates with different backgrounds. The second

use of corpus linguistics for research in UCLES is a topic for the next section.

Examination of language performance product The genre-based approach to language production data analysis is a more systematic method to investigate the characteristics of language production data in an *a posteriori* manner. Such approach is to answer if there is any systematic (i.e., causal) relationship between variation in genre and variation in language performance. Also, just like UCLES practices, linguistic characteristics of learner performance on presumably parallel tasks can be compared across different test administrations of them. In fact, the issue of examining to what extent the tasks are parallel concerns the validity and reliability of the assessment.

Furthermore, using corpus-based genre analysis, a comparison can be made between test-takers' performance on a test task and the real life task that are captured in the corpus like T2K-SWAL. The information drawn from this approach will suggest to what extent a test-taker's performance in a testing task could be generalizable to non-testing (natural) tasks/contexts. There are signs that the language (task) characteristics we assume to be real are found different from those in the real language use data, corpora. Especially, the assumed validity of language functions (e.g., in speech acts) engendered in test tasks may not be authentic/real when we consider the actual instances revealed by corpus analysis, an issue of which is directly concerned with test validation. Notice that Brown et al. (2002), the most extensive research in task-based performance assessment, recognize as the first limitation of their research method that "they [we] did not investigate the relationship between outcomes on their [our] task-based tests and real-world language use (i.e., outside the testing context)..." (p. 119).

CONCLUSION

In constructing tests, it is important to include texts and tasks that mirror as closely as possible those that test-takers have been exposed to or are to meet in their future target contexts (Weir, 2005). Unless the context is clearly definable, it may not be easy to achieve such a goal. Even areas such as EAP and ESP, where the contexts and needs are relatively distinct and hence identifiable, have been found difficult to evaluate different groups of candidates. Therefore, it is not surprising to see Brindely and his colleagues have been suffering from achieving equivalence of different test tasks in their outcome-based assessment in Australia. In his discussion of context validity, Weir (2005) strongly emphasizes the importance of making test tasks as close to the real life ones as possible, by saying that:

Achieving such realism in tests for general English students may not be so easy, but the emphasis must still be on giving the interlocutors as realistic, and as needs-based a purpose as possible. Full authenticity of task may not be achievable, but we need to make our tests as valid as possible if we are to measure anything of value. The more we compromise the more difficult it will be to make meaningful statements about what candidates can or cannot do on the basis of test results [all italics added] (p. 61).

Through this paper, I have discussed how corpora and corpus-based genre analysis can be used to achieve such realism in language assessment. As such, Biber et al.'s (2004) study was suggested as the example that had come the closest to achieving such a goal. One major drawback of the corpus-based approach to test construction and validation may be about the nature of extensive and intensive labor that has to be put in the construction of a corpus. This is noticeable as well by the number

of authors, nine, involved in Biber's project. However, as more institutions (e.g., Kanda Learner Corpus of Spoken English) are willing to invest their resources to build corpora for assessment purposes, the labor intensive nature of corpus building would not be considered a major challenge anymore. Besides, for the purpose of test validation, the size of the corpus may not matter much. Rather, the quality in terms of the representiveness of target contexts and population may be an aspect that requires more attention.

REFERENCES

- Alderson, J. C. (1996). Do corpora have a role in language assessment? In J. Thomas and M. Short (Eds.), *Using corpora for language research* (pp. 248-259). London: Longman.
- Alderson, J. C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.
- Ball, F. (2001). Using corpora in language testing. *Research Notes*, 6, 6-8, Retrieved from: http://www.cambridgeesol.org/rs_notes/
- Barker, F. (2005). What insights can corpora bring to language testing? (Lecture handout) Retrieved from: <http://www.ling.lancs.ac.uk/groups/crile/docs/crile%20lectures/barker0105.doc>
- Berber, S. T. (1996). *EFL writing assessment and corpus linguistics: The 'sound right' factor*. Applications of corpus linguistics seminar, Aston University, 19th April 1996.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, D. (1989). A typology of English texts. *Linguistics*, 27, 3-43.
- Biber, D. (1990). Methodological issues regarding corpus-based analyses of linguistic variation. *Literary and Linguistic Computing*, 5, 257-269.

CORPUS-BASED GENRE ANALYSIS
FOR LANGUAGE ASSESSMENT

- Biber, D. (1995). *Dimensions of register variation: a cross linguistic comparison*. Cambridge: Cambridge University Press.
- Biber, D., & Conrad, S. (2001). Quantitative corpus-based research: Much more than bean counting. *TESOL Quarterly*, 35(2), 331-336.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Biber, D., Conrad, S., Reppen, R., Byrd, P., Helt, M., Clark, V., Cortes, V., Csomay, E., Urzua, A. (2004). *Representing language use in the university: Analysis of the TOEFL 2000 spoken and written academic language corpus*. (TOEFL Monograph No. MS-25). Princeton, NJ: ETS.
- Biber, D., & Finegan, E. (1989). Drift and the evolution of English style: a history of three genres. *Language*, 65, 487-517.
- Biber, D., & Finegan, E. (1991). On the exploitation of computerized corpora in variation studies. In K. Aijmer and B. Altenbert (Eds.), *English corpus linguistics* (pp. 204-220). New York: Longman.
- Biber, D., & Finegan, E. (1994). *Sociolinguistic perspectives on register*. New York: Oxford University Press.
- Bondi, M. (2001). Small corpora and language variation: reflexivity across genres. In M. Ghadessy, A. Henry and R. L. Roseberry (Eds.), *Small corpus studies and ELT: Theory and practice* (pp. 136-174). Amsterdam: John Benjamins.
- Brindely, G. (1998). Describing language development? Rating scales and SLA. In L. F. Bachman and D. Cohen (Eds.). *Interfaces between second language acquisition and language testing research* (pp. 112-140). Cambridge University Press.
- Brown, J., D., Hudson, T., Norris, J., & Bonk, W. (2002). *An investigation of second language task-based performance assessments*. (Technical Report #24).

Honolulu: University of Hawai'i, Second Language Teaching & Curriculum Center.

- Brown, A., Iwashita, N., & McNamara, T. (2005). *An examination of rater orientations and test-taker performance on English-for-Academic-Purposes speaking tasks*. (TOEFL Monograph No. MS-29). Princeton, NJ: ETS.
- Conrad, S. (1996). Investigating academic texts with corpus-based techniques: an example from biology. *Linguistics and Education*, 8, 299-326.
- Deutschmann, M. (2003). *Apologising in British English*. Unpublished PhD Thesis. University of Umeå, Sweden.
- Douglas, D. (2000). *Assessing languages for specific purposes*. Cambridge: Cambridge University Press.
- Douglas, D., & Selinker, L. (1993). Performance on a general versus a field-specific test of speaking proficiency by international teaching assistants. In D. Douglas and C. Chappelle (Eds.), *A new decade of language testing research* (pp. 234-254). Alexandria, VA: TESOL.
- Galaczi, E. D. (2003). Interaction in Paired Speaking Tests: The case of the First Certificate in English. *Research Notes*, 14, 19-23. Retrieved from: http://www.cambridgeesol.org/rs_notes/
- Greenbaum, S., & Nelson, G. (1995). Clause relationships in spoken and written English. *Functions of Language*, 2(1), 1-21.
- Greenbaum, S., Nelson, G., & Weitzman, M. (1996). Complement Clauses in English. In J. Thomas and M. Short (Eds.), *Using Corpora for Language Research* (pp. 76-91). London: Longman.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Kennedy, G.. (1998). *An introduction to corpus linguistics*. New York: Longman.

CORPUS-BASED GENRE ANALYSIS
FOR LANGUAGE ASSESSMENT

- Knapp, P., & Watkins, M. (2005). *Genre, text, grammar: Technologies for teaching and assessing writing*. Sydney: University of South Wales Press Ltd.
- Lee, David, Y. W. (2001). Genres, registers, text types, domains, and styles: clarifying the concepts and navigating a path through the BNC jungle. *Language Learning & Technology*, 5, 37-72.
- Liu, D. (2003). The most frequently used spoken American English idioms: a corpus analysis and its implications. *TESOL Quarterly*, 37, 671-700.
- Lu, Y. (2003). Insights into the FCE Speaking Test, Research Notes, 11, 15-19. Retrieved from: http://www.cambridgeesol.org/rs_notes/
- McCarthy, M. (1998) *Spoken language and applied linguistics*. Cambridge: Cambridge University Press.
- Paltridge, B. (2001). Linguistic research and EAP pedagogy. In J. Flowerdew and M. Peacock (Eds.), *Research perspective on English for academic purposes* (pp. 55-70). Cambridge: Cambridge University Press.
- Reppen, R., & Simpson, R. (2002). Corpus linguistics. In N. Schmitt (Ed.), *An introduction to applied linguistics*. London: Arnold.
- Shohamy, E. (1984). Does the testing method make a difference? The case of reading comprehension. *Language Testing*, 1(2), 147-180.
- Shohamy, E., & Inbar, O. (1991). Construct validation of listening comprehension tests: The effect of text and question type. *Language Testing*, 8(1), 23-40.
- Shohamy, E. (1997). Testing methods, testing consequences: are they ethical? are they fair? *Language Testing*, 14(3), 340-349.
- Simpson, R., & Mendis, D. (2003). A corpus-based study of idioms in academic speech. *TESOL Quarterly*, 37, 419-441.
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Swales, J. M. (2004). *Research genre: Exploration and application*. Cambridge: Cambridge University Press.

Yamada, M. (2005). Task proficiency and L1 private speech. *IRAL*, 43, 81-108.

Weir, C. J. (2005). *Language testing and validation: an evidence-based approach*. New York, NY: Palgrave MacMillan.