Exploring the Potential for Corpus-Based Research in Speech-Language Pathology

Alison Ferguson, Hugh Craig, and Elizabeth Spencer
University of Newcastle, Australia

1. Introduction

Individual speakers vary their use of language from moment to moment and from situation to situation, and the cognitive and social sources of such variation have been subject to considerable research (Pishwa, 2006). However, observed in groups of speakers are also patterns of language variation that are related to different levels of proficiency, that is, associated with language development through childhood (McLaughlin, 2006); second language acquisition and learning (Van Patten & Williams, 2007); changes over the lifespan associated with healthy aging (Brownell & Yves, 1993) as well as with age-related pathology such as dementia (Bayles & Tomoeda, 2007); and acquired pathologies, for example, brain impairment from stroke or traumatic brain injury (Murray & Clarke, 2006). The study of these patterns of language variation in groups of speakers needs to take into account a range of biopsychosocial factors that contribute to language variation (Holmes, 2001; Milroy & Gordon, 2003), including health status, cognitive ability, education, and socioeconomic group or 'class' (Labov & Auger, 1993).

1.1. Variation and Speech-Language Pathology

The study of these patterns of language variation is important to speech-language pathologists in providing a reference base of 'normal' or 'typical' language use to assist with assessment and diagnosis, as well as providing a framework for intervention (Ball, Perkins, Muller, & Howard, 2008). In the field of speech-language pathology, research into normal variation has been conducted within the experimental research paradigm, through the use of closely matched normal control groups (Irwin, Pannbacker, & Lass, 2008), and carefully selected normal subjects' performance on assessment tools (McCauley & Swisher, 1984). The increasing application of the qualitative research paradigm in speech-language pathology has promoted the use of authentic language sampling, yet such research continues to make use of matched controls in seeking to draw a comparison with typical language use (Simmons-Mackie & Damico, 2003). While these methodologies serve the purposes of particular research projects and assessment tools, they provide a picture of normal language use that has questionable validity, since the intrinsic requirements for control of factors thought to influence language variation mean that the full range of normal variation is effectively screened from view.

One area where a greater understanding of normal language use is needed is the identification of Specific Language Impairment in children. Here the identification of pathology must be made in the context of developing language. Another area is in the differential diagnosis of language pathology associated with neurological damage (e.g., stroke, traumatic brain injury), particularly when either the pathology is mild (sometimes described as 'higher level' language problems) or for language users who had high levels of linguistic proficiency prior to injury. Standardized tests are not adequate for these purposes, since they are designed to control for the effects of education, and so have low ceilings. Increased sensitivity of assessment is important for intervention, since, for example, in the case of the onset of dementia, early detection can now prompt promising pharmacological preventative (or at least slowing) treatments.

Chapman and Ulatowska (1994) have argued for discourse analysis to play an important role in differential diagnosis. While such data do not replace the pivotal role of medical case history in diagnosis, in cases where medical data is lacking or ambiguous, then the patterns of discourse

 Selected Proceedings of the 2008 HCSNet Workshop on Designing the Australian National Corpus, ed. Michael Haugh et al., 30-36. Somerville, MA: Cascadilla Proceedings Project.

associated with the different acquired language impairments provide an important source of information. From their review of the research literature, Chapman and Ulatowska compared the patterns of linguistic pathology observed in aphasia, 'normal old elderly', Alzheimer's dementia, progressive aphasia, right-hemisphere damage, and closed head-injury, identifying linguistic features that are common across disabilities (and hence of limited use in differential diagnosis) and those features that are distinctively associated with particular disabilities (and so more use to differential diagnosis) - see Table 1. However, to date, such discourse analyses have relied on manual analyses, and lack normal reference data for comparative purposes.

1.2. Variation and Corpus Linguistics

In the field of linguistics, however, the developments in corpus-based research have provided a complementary methodology to study language variation (Svartvik, 1992). As corpus-based research has gained currency, the term 'corpus' is being adopted in the field, but sometimes refers simply to the body of data collected for the purpose of the research, rather than specifically to corpus-based methodology. In this paper we will confine our discussion of examples to research that uses large, authentic, data sets that are suitable for computational analysis and which were collected in order to provide a source of reference for later interrogation for the purposes of other research. The selection of samples that make up the corpus is often wide-ranging, yet principled, that is, collected and selected according to particular criteria, with those criteria generally designed to increase representativeness, for example, across particular registers/genres, and spoken/written modes. While the actual sizes of different corpora vary considerably, they are generally larger than those obtained in experimental research. In speech-language pathology, corpus-based research relies predominantly on archival databases. Corpus linguistics is also used in critical discourse studies, for example, looking at the shading of meaning reflecting differing cultural ideologies (see Adolphs, 2006, ch. 6 for an introduction), and is also used in corpus stylistics (Craig, 2004). While both stylistics and critical discourse approaches have made some impact in the field of speech-language pathology, the use of corpus-based methodologies in relation to these foci has yet to emerge in the field.

Because of the size of corpora, the methods of linguistic analysis typically involve computerized software-generated analyses (Adolphs, 2006; Lebart, Salem, & Berry, 1998), rather than more traditional pen and paper analyses. The combination of these two things – that is, a corpus analyzed by computer – has meant that some prefer the term 'electronic text analysis' to 'corpus linguistics'. Many linguistic corpora have made use of 'tagging' procedures (including manual coding of all items), which are then subject to computerized collation and sorting depending on the particular research focus (Adolphs, 2006; Smith, Hoffman, & Rayson, 2008).

More recently, developments in computerized analyses have allowed for analyses that do not require 'tagging', and instead make use of principles of text recognition and probabilistic statistical procedures (Hinneburg, Mannila, Kaislaniemi, Nevalainen, & Raumolin-Brunberg, 2007) (e.g., principal components analysis, discriminant analysis) to locate patterns of linguistic features that are then subjected to further analysis and interpretation. Biber and colleagues (Biber, Reppen, & Conrad, 2002) describe this kind of analysis as 'multi-dimensional analysis', in which computational analysis identifies dimensions with underlying functions that typify particular registers and modes, which they then describe as 'models'. While corpus-based research provides a less well-controlled methodology than that obtainable through experimental methods, it has the advantage of increased statistical power (though large data sets) and increased validity (through large and wide scope of sampling from authentic contexts).

Corpus linguistics has been applied in fields that share some common issues with speech-language pathology, for example, teaching English for academic or special purposes (Flowerdew, 2005), and teaching English as a second language (Braun, 2005). These fields also need to establish a reference point for what is the 'normal' or target level of performance, and need to identify where language users are divergent from that reference standard, and have an interest in using the development and frequency data to guide teaching and interventions. These fields also discuss using corpora as a resource for the language users to access directly as a model for improving their own proficiency. This potential use has not been discussed in the field of speech-language pathology. Braun (2005) notes that the uptake in

practice in these fields is not as widespread as might appear from the published literature, and suggests that this may reflect some inadequacies in the corpora, and the need to develop ways to use corpora in teaching.

Table 1.
Differential Diagnostic Patterns in Discourse (based on Chapman & Ulatowska, 1994)

Linguistic patterns	Normal elderly	Aphasia (non-fluent)	Aphasia (fluent)	Progressive Aphasia	Cognitive comm. disability (dementia)	Cognitive comm. disability (Right CVA)	Cognitive comm. disability (Closed head injury)
General							
'Empty'	X		X	X	X	X	X
Tangential	X		X	X	X	X	X
Verbose	X		X		X	X	
Disfluent		X		X	X		
Errors							
Anomia	X	X	X	X	X	X	X
Reference	X	X	X	X	X	X	
Paraphasia		X	X	X	X		X
Paragrammatism			X	X	X		
Agrammatism		X					
Jargon			X		X		
Repetition		X	X	X	X		
Perseveration		X	X	X	X		

Note. Comm. = communication.

2. Current Corpus-Based Research in Speech-Language Pathology

Corpus-based research in linguistics has influenced both research and clinical work in speech-language pathology, although the term 'corpus-based' is not commonly used in the discipline. One example of this is the use of the Brown Corpus of American English (Francis & Kucera, 1982) in developing stimuli for assessments of naming in aphasia. Because word-finding difficulty ('anomia') in acquired adult aphasia is known to be affected by word frequency, Francis and Kucera's word frequency data were used to develop lists of high and low frequency words within the test batteries for the Psycholinguistic Assessments of Language Processing in Aphasia – PALPA (Kay, Lesser, & Coltheart, 1992) . Speech-language pathologists are aware of the significance of word frequency in ease of retrieval and speed of linguistic processing, and control for this factor in their assessment and treatment protocols, but may be less aware of the research that underpins such frequency information. This lack of awareness of the data source is potentially problematic, since it means that researchers may not be able to evaluate the suitability of particular usage data to the clinical task at hand. For example, the Brown Corpus is based on American English, and is primarily written data, which raises questions regarding its validity for Australian English speakers.

Some research in speech-language pathology is harnessing the computational approach associated with the methodology of corpus-based research to address some of the long-standing issues in speech-language pathology research and clinical methodology. For example, Tomasello and Stahl (2004) address the issue of the size of language samples (which traditionally in speech-language pathology was influenced by Lee's DSS approach of 50 utterances, and Crystal's rule of thumb guideline (Crystal, 1992; Crystal, Fletcher, & Garman, 1976) of 100 utterances in child language sampling). Tomasello and Stahl's approach suggested that much larger sampling is required, particularly when the linguistic feature under examination occurs rarely.

In the field of speech-language pathology, corpus-based research is dominated by the Child Language Data Exchange System (CHILDES) database, which has in more recent years been

incorporated within a larger data base known as TalkBank. TalkBank is an interdisciplinary research database developed and directed by Professor Brian MacWhinney at Carnegie Mellon University. The goal of TalkBank is to foster fundamental research in the study of human and animal communication. The TalkBank database grew out of the initial CHILDES, which was developed in the mid-1980s initially as a research archive of child language data (MacWhinney, 1996, 2000). The CHILDES system has enabled the development of international standards for transcription and analysis (Codes for the Human Analysis of Transcripts – CHAT, and Computerized Language Analysis – CLAN). TalkBank now also includes PhonBank and, in 2007, further funding was received from NIH to develop AphasiaBank – see http://talkbank.org for details.

From information available on the TalkBank site, the corpora included in CHILDES totalled over 180 million characters at 1996 (MacWhinney, 1996), with about half the corpora in English (over 27 corpora of typically developing children and 9 corpora of children with language impairment in the mid-1990s) and by 2008 there were over 3,000 publications based on the use of CHILDES.

For example, in the area of research into Specific Language Impairment (SLI), Aguilar-Mediavilla, Sanz-Torrent, and Serra-Raventos (2007) compared their longitudinal data from five children with SLI with data they collected for the purposes of the study from five age-matched controls, and with data from one of the corpora within the CHILDES database (the Serra-Sole corpus). Their analysis of the data was primarily statistical, using the Mann-Whitney U test for testing differences between groups, as well as linear regression for exploring relationships between variables, for example, between phonological and morphosyntactic errors.

Bastiaanse and Bol (2001) were able to mine the database corpora to conduct a study with eight participants in three groups - adults with aphasia, children with SLI, and normal children. The data for the adults with aphasia were drawn from previous research by Bastiaanse, and the data for the children were from a database that is from Bol's previous research and which is now available through CHILDES. The samples were 300 words in length. Statistical analyses were conducted to determine correlation (Pearsons *r*) between type token ratio, mean length of utterance, and a 'finiteness index' (number of finite verbs, divided by the number of clauses containing a verb).

What is striking from these examples is the use of very small corpora (for example, data drawn from less than 10 speakers). Also, it is common for the corpora to be drawn from the researchers' own work, subsequently made available to others through CHILDES. For much of the research in the field, the main use of CHILDES is in the use of the transcription (CHAT) and analysis software (CLAN), which is freely available through TalkBank.

The AphasiaBank database currently includes 18 corpora, but only four of these are of substantial size (that is, over 100 samples). While some of these corpora have been discussed in the research literature prior to their entry to AphasiaBank, to our knowledge there are as yet no published studies that have accessed these data for corpus-based research. While the current corpora in AphasiaBank have been provided by researchers from their own projects, a prescribed protocol has been established for the prospective data collection for AphasiaBank. The development of this protocol was designed to accommodate the strong prevailing experimental research paradigm which is concerned with comparability, and hence control, of sampling conditions. The protocol involves only monologic discourse (e.g., describing illness; describing a picture; describing a set of procedures) and test data. The lack of conversational data and the structured elicitation methods raise questions as to the scope of potential uses of the corpus. Another problem for the corpus will be the lack of data from speakers who do not have a communication disability, since 'normal' performance on such structured tasks may differ from that in natural interactions.

3. Issues for Speech-Language Pathology in Corpus-Based Research

Many of the issues for speech-language pathology in relation to corpus-based research are common to the field of linguistics in general, as well as shared by much research in the field of applied linguistics (as previously discussed).

One key theoretical issue is the question of how adequately this type of research can deal with sociocultural context. Flowerdew (2005) comments that the samples that make up the English for Special Purposes Corpus are taken out of their context of speaking, writing, and use, and may well only

involve a segment of the entire text, so that it may not be possible to identify genre-specific features or to explore intertextuality. Some of these problems can no doubt be overcome through careful corpus design, or by selecting samples within a corpus and focusing for a given trial on a particular text type.

A more fundamental theoretical challenge is the extent to which corpus research can be used to establish theories of grammar, or normative notions as to typical grammatical use. Newmeyer (2003, p.696) puts this case strongly when he says "There is no way that one can draw conclusions about the grammar of an individual from usage facts about communities, particularly communities from which the individual receives no speech input." While he recognizes that language users are often 'ungrammatical' and that instances of this occur in corpora, his argument is that these sorts of examples of language use cannot be taken as the foundation for claims about underlying 'grammar'. This is an opposite view to that of Halliday (1992, p.66), who argues that the "...relative frequency in the corpus is the same thing as probability in the grammar....frequency in the corpus is the *instantiation* (note, not realisation) of probability in the grammar." Given the doubts Flowerdew and Newmeyer raise, it would be prudent for speech-language pathology researchers to complement corpus-based research with other approaches if they are seeking a comprehensive determination of 'typical' or 'normal' reference standards with which to compare the language used by individuals with communication disability.

In terms of methodology in corpus-based research, speech-language pathology is in accord with other disciplines. In particular, the size of corpora can be considered crucial, particularly since it is generally accepted that some of the concerns in relation to sample selection previously described are addressed through large sample sizes. For example, Gierut and Dale (2007, p. 429-430) found that "...any type of lexical corpus (child, adult, expressive/spoken, receptive/written) may be appropriate for use in the study of children's language acquisition, but only if it is comprised of a substantial number of entries."

The methods used in corpus-based research for the preparation of data to enable manual and automated searching and sorting are highly specialized, and require the development of knowledge and experience in systems of mark-up and annotation, for example, Standard Generalized Mark-up Language (SGML), and Extensible Mark-up Language (XML), both of which are part of the Text Encoding Initiative (TEI) system for the scholarly exchange of marked-up electronic text. It is also useful to develop a familiarity with parsers (e.g., Charniak parser, Part of Speech parser – POS), and concordance software such as WordSmith (Smith et al., 2008). Both the size of corpora and the technology involved in analysis can prevent ease of access (Braun, 2005).

4. Future Directions for Corpus-Based Research in Speech-Language Pathology

Corpus-based research provides the field of speech-language pathology with a methodology that can be usefully applied across the main theoretical perspectives that inform research in this area. Both psycholinguistic and sociolinguistic research can make use of large authentic data sets.

One of the immediately available directions for speech pathology is to maximize the use of current corpus-based research. For example, the work on children's development of literacy as discussed by Biber et al. (2002) provides empirically well-grounded guidance for developmental hierarchies in children's control of linguistic devices associated with particular spoken and written genres. Another example is in the work of Roland, Dick, and Elman (2007) which explicitly draws attention to the usefulness of the comprehensive set of frequencies of particular linguistic forms that are common points of focus for research in both language development and acquired language impairment: cleft sentences, relative clauses (subject relative, object relative, passive relative, infinitive subject relative, infinitive object relative, infinitive passive relative), verb subcategorisation (e.g., intransitive, passive), and word order (SVOO, SVO, SV etc). There is a need to increase awareness and accessibility of this information to both speech-language pathology researchers and clinicians. It would, however, take some transformation of the information, for example, development of reference tables, to make it accessible for clinical use.

Current plans within TalkBank (as previously discussed) are aiming to develop a larger and more authentic corpus of pathological language use in adults. As previously discussed, the inclusion of more natural interaction within this data set would improve the authenticity of the corpus. Also, the use of existing corpora of adult language as a reference set when investigating communication disability in

adults would provide an alternative, or complementary, methodology to the current practice of obtaining matched control data. For example, the following corpora include spoken Australian English: Australian International Corpus of English (ICE - http://www.ucl.ac.uk/english-usage/ice/), English Language Interview Corpus as a Second Language Application (ELISA - www.corpora4learning.net), and ICAME Collection of English Language Corpora (www.hit.uib.no/icame/cd).

A further direction is to explore the potential of computational analysis for differential diagnostic purposes. Such research needs to evaluate this methodology's capabilities in relation to the identification of key differentiating features of discourse. There may be potential for this methodology to be clinically applied by enabling comparison of individuals with a reference corpus. Our research team (Ferguson, Craig, and Spencer) is currently seeking to explore this direction. Such research also holds the possibility to develop applications to gauge recovery and response to therapy.

5. Conclusion

Corpus-based research in the field of speech-language pathology is currently making use of small specialized corpora collected for specific research projects which are being collected within large archival databases. Existing corpora are also being used to inform the development of assessment methods and selection of assessment stimuli, and to provide data sets for comparison with clinical populations. In this paper, we have suggested that existing large-scale corpora could inform future psycholinguistic and sociolinguistic research in speech-language pathology but that there are currently barriers of awareness and technology to be overcome. We have also suggested that further research is needed to explore the applications of corpus-based computational analyses to diagnostics and outcome measurement.

6. Acknowledgements

The support provided for the preparation of this paper through a Fellowship from the University of Newcastle to the first author, is gratefully acknowledged.

References

Adolphs, Svenja. (2006). *Introducing electronic text analysis*. London: Routledge.

Aguilar-Mediavilla, Eva, Monica Sanz-Torrent, & Miguel Serra-Raventos. (2007). Influence of phonology on morpho-syntax in Romance languages in children with Specific Language Impairment (SLI). *International Journal of Language & Communication Disorders, 42*(3), 325-247.

Ball, Martin J., Michael R. Perkins, Nicole Muller, & Sara Howard. (Eds.). (2008). *The handbook of clinical linguistics*. Malden, MA: Blackwell Publishing.

Bastiaanse, Roelien, & Gerard Bol. (2001). Verb inflection and verb diversity in three populations: Agrammatic speakers, normally developing children, and children with Specific Language Impairment (SLI). *Brain and Language, 77*(3), 274-282.

Bayles, Kathryn, & Cheryl Tomoeda. (2007). *Cognitive-communication disorders of dementia*. San Diego, CA: Plural Publishing.

Biber, Douglas, Randi Reppen, & Susan Conrad, S. (2002). Developing linguistic literacy: Perspectives from corpus linguistics and multi-dimensional analysis. *Journal of Child Language, 29*(2), 458-462.

Braun, Sabine. (2005). From pedagogically relevant corpora to authentic language learning contexts. *ReCALL: Journal of the European Association for Computer Assisted Language Learning, 17*(1), 47-64.

Brownell, Hiram H., & Joanette Yves. (Eds.). (1993). *Narrative discourse in neurologically impaired and normal aging adults*. San Diego, CA: Singular.

Chapman, Sandra B., & Hanna K. Ulatowska. (1994). Differential diagnosis in aphasia. In Roberta Chapey (Ed.), *Language intervention strategies in adult aphasia* (pp. 121-131). Baltimore: Williams & Wilkins.

Craig, Hugh. (2004). Stylistic analysis and authorship studies. In Susan Schreibman, Ray Siemens & John Unsworth (Eds.), *A companion to digital humanities* (pp. 273-288). Malden, MA: Blackwell.

Crystal, David. (1992). *Profiling linguistic disability*. London: Whurr.

Crystal, David, Paul Fletcher, & Michael Garman. (1976). *Grammatical analysis of language disability*. London: Whurr.

Flowerdew, Lynne. (2005). An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: Countering criticisms against corpus-based methdologies. *English for Specific Purposes, 24*, 321-332.

Francis, W. Nelson, & Henry Kucera. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.

Gierut, Judith A., & Rachel A. Dale. (2007). Comparability of lexical corpora: Word frequency in phonological generalization. *Clinical Linguistics & Phonetics, 21*(6), 423-433.

Halliday, Michael A. K. (1992). Language as system and language as instance: The corpus as a theoretical construct. In J. Svartvik (Ed.), *Directions in corpus linguistics: Proceedings of Nobel Symposium 82, Stockholm, 4 - 8 August, 1991* (pp. 61-77). Berlin: Mouton de Gruyter.

Hinneburg, Alexander, Heikki Mannila, Samuli Kaislaniemi, Terttu Nevalainen, & Helena Raumolin-Brunberg. (2007). How to handle small samples: Bootstrap and Bayesian methods in the analysis of linguistic change. *Literary and Linguistic Computing, 22*(2), 137-150.

Holmes, Janet. (2001). *An introduction to sociolinguistics* (2nd ed.). Harlow, England: Longman.

Irwin, David L., Mary Pannbacker, & Norman J. Lass. (2008). *Clinical research methods in speech-language pathology and audiology*. San Diego, CA: Plural.

Kay, Janice, Ruth Lesser, & Max Coltheart. (1992). *Psycholinguistic assessments of language processing in aphasia [Kit]*. Sussex: Lawrence Erlbaum.

Labov, William, & Julie Auger. (1993). The effect of normal aging on discourse: A sociolinguistic approach. In Hiriam H. Brownell & Yves Joanette (Eds.), *Narrative discourse in neurologically impaired and normal aging adults* (pp. 115-133). San Diego, CA: Singular.

Lebart, Ludovic, André Salem, & Lisette Berry. (1998). *Exploring textual data*. Dordrecht: Kluwer Academic Publishers.

MacWhinney, Brian. (1996). The CHILDES system. *American Journal of Speech-Language Pathology, 5*, 5-14.

MacWhinney, Brian. (2000). *The CHILDES project: Tools for analyzing talk*. Mahwah, NJ: Lawrence Erlbaum.

McCauley, Rebecca J., & Linda Swisher. (1984). Use and misuse of norm-referenced tests in clinical assessment: A hypothetical case. *Journal of Speech & Hearing Disorders, 49*, 338-348.

McLaughlin, Scott. (2006). *Introduction to language development* (2nd ed.). Clifton Park, NY: Thomson Delmar Learning.

Milroy, Lesley, & Matthew Gordon. (2003). *Sociolinguistics: Method and interpretation*. Malden, MA: Blackwell Publishing.

Murray, Laura L., & Heather M. Clarke. (2006). *Neurogenic disorders of language: Theory driven clinical practice*. Clifton Park, New York: Thomson Delmar Learning.

Newmeyer, Frederick J. (2003). Grammar is grammar and usage is usage. *Language, 79*(4), 682-707.

Pishwa, Hanna. (Ed.). (2006). *Language and memory: Aspects of knowledge representation*. Berlin: Mouton de Gruyter.

Roland, Douglas, Frederic Dick, & Jeffrey L. Elman. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language, 57*(3), 348-379.

Simmons-Mackie, Nina, & Jack S. Damico. (2003). Contributions of qualitative research to the knowledge base of normal communication. *American Journal of Speech-Language Pathology, 12*(2), 144-154.

Smith, Nicholas, Sebastian Hoffman, & Paul Rayson. (2008). Corpus tools and methods, today and tomorrow: Incorporating linguists' manual annotations. *Literary and Linguistic Computing, 23*(2), 163-180.

Svartvik, Jan. (Ed.). (1992). *Directions in corpus linguistics: Proceedings of Nobel Symposium 82, Stockholm, 4 - 8 August, 1991*. Berlin: Mouton de Gruyter.

Tomasello, Michael, & Daniel Stahl. (2004). Sampling children's spontaneous speech: How much is enough? *Journal of Child Language, 31*(1), 101-121.

Van Patten, Bill, & Jessica Williams. (Eds.). (2007). *Theories in second language acquisition: An introduction*. Mahwah, NJ: Lawrence Erlbaum.

Selected Proceedings of the 2008 HCSNet Workshop on Designing the Australian National Corpus: Mustering Languages

edited by Michael Haugh, Kate Burridge, Jean Mulder, and Pam Peters

Cascadilla Proceedings Project Somerville, MA 2009

Copyright information

ISBN 978-1-57473-435-5 library binding

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, e-mail: sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Ferguson, Alison, Hugh Craig, and Elizabeth Spencer. 2009. Exploring the Potential for Corpus-Based Research in Speech-Language Pathology. In *Selected Proceedings of the 2008 HCSNet Workshop on Designing the Australian National Corpus*, ed. Michael Haugh et al., 30-36. Somerville, MA: Cascadilla Proceedings Project.

or:

Ferguson, Alison, Hugh Craig, and Elizabeth Spencer. 2009. Exploring the Potential for Corpus-Based Research in Speech-Language Pathology. In *Selected Proceedings of the 2008 HCSNet Workshop on Designing the Australian National Corpus*, ed. Michael Haugh et al., 30-36. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2285.