

A CORPUS APPROACH TO ECOLOGICAL DISCOURSE ANALYSIS AND L2 WRITING

PEDAGOGY

by

Robert Poole

---

A Dissertation Submitted to the Faculty of

SECOND LANGUAGE ACQUISITION AND TEACHING

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2015

THE UNIVERSITY OF ARIZONA  
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Robert Poole, titled A Corpus Approach to Ecological Discourse Analysis and L2 Writing Pedagogy and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

\_\_\_\_\_  
Jonathon Reinhardt Date: 04/10/2015

\_\_\_\_\_  
Christine Tardy Date: 04/10/2015

\_\_\_\_\_  
Linda Waugh Date: 04/10/2015

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

\_\_\_\_\_  
Dissertation Director: Jonathon Reinhardt Date: 04/10/2015

## STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of the requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that an accurate acknowledgement of the source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Robert Poole

**TABLE OF CONTENTS**

LIST OF FIGURES.....	5
LIST OF TABLES .....	6
ABBREVIATIONS AND ACRONYMS.....	7
ABSTRACT.....	8
AN INTRODUCTION TO THE ROSEMONT MINE DEBATE, ECOLINGUISTICS, & CORPUS LINGUISTICS .....	10
A CORPUS-BASED ECOLOGICAL DISCOURSE ANALYSIS OF THE ROSEMONT COPPER MINE DEBATE.....	42
INTEGRATING GIS AND CORPUS LINGUISTICS FOR THE ANALYSIS OF ENVIRONMENTAL DISCOURSE .....	76
A CORPUS-AIDED APPROACH FOR THE TEACHING AND LEARNING OF RHETORIC IN AN UNDERGRADUATE COMPOSITION COURSE FOR L2 WRITERS.....	104
CONCLUSION .....	133
APPENDIX .....	145
REFERENCES .....	160

**LIST OF FIGURES**

FIGURE 1: THE PROPOSED MINE SITE	84
FIGURE 2: THE ROSEMONT MINE SITE	84
FIGURE 3: GEOGRAPHICAL PLACE MENTIONS IN THE RMT CORPUS (RAW FREQUENCY)	94
FIGURE 4: PLACE NAME MENTIONS IN RMT AND RCC	96
FIGURE 5: PLACE NAME MENTIONS IN THE STATE OF ARIZONA	96
FIGURE 6: REFERENCES TO MONEY, PAY, AND BUSINESS	98
FIGURE 7: ECOLOGY-INFORMED REFERENCES	99

## LIST OF TABLES

TABLE 1: CORPORA OF RCC AND RMT	53
TABLE 2: KEYWORDS IN THE ROSEMONT MINE TRUTH CORPUS	58
TABLE 3: RHETORICAL CONSTELLATION OF UNCERTAINTY/DOUBT IN ROSEMONT MINE TRUTH CORPUS	60
TABLE 4: RHETORICAL CONSTELLATION OF AESTHETICS/STEWARDSHIP IN ROSEMONT MINE TRUTH CORPUS	62
TABLE 5: A SAMPLE RMT TEXT	64
TABLE 6: KEYWORDS IN THE ROSEMONT COPPER COMPANY CORPUS	66
TABLE 7: RHETORICAL CONSTELLATION OF CERTAINTY/AUTHORITY IN ROSEMONT COPPER COMPANY	68
TABLE 8: RHETORICAL CONSTELLATION OF DOMINION/ECONOMICS IN ROSEMONT COPPER COMPANY	70
TABLE 9: CORPORA OF RCC AND RMT	87
TABLE 10: SAMPLE DATA	89
TABLE 11: ROSEMONT MINE TRUTH HIGH FREQUENCY/SIGNIFICANT SEMANTIC TAGS	90
TABLE 12: ROSEMONT COPPER COMPANY HIGH FREQUENCY/SIGNIFICANT SEMANTIC TAGS	93
TABLE 13: THE TEXTS	115
TABLE 14: STUDENT HANDOUT #1 (ABRIDGED)	120
TABLE 15: STUDENT HANDOUT #2 (ABRIDGED)	121
TABLE 16: THE KEYWORDS ACTIVITY HANDOUT	125
TABLE 17: SHORT ESSAY ASSIGNMENT	126
TABLE 18: SHORT ESSAY EXCERPT #1	127
TABLE 19: SHORT ESSAY EXCERPT #2	127
TABLE 20: SURVEY QUESTIONS	129
TABLE 21: POTENTIAL DESIGN FEATURES OF THE ECO-CORPUS	138

**ABBREVIATIONS AND ACRONYMS**

Arizona Corps of Engineers	ACOE
Corpus linguistics	CL
Cognitive Linguistics	CogL
Critical Discourse Analysis	CDA
English for Academic Purposes	EAP
English for Specific Purposes	ESP
Environmental Protection Agency	EPA
Geographical Information Systems	GIS
Rosemont Copper Company	RCC
Rosemont Mine Truth	RMT
Save the Scenic Santa Ritas	SSSR
Second Language	L2
Second Language Acquisition	SLA
Systemic Functional Linguistics	SFL

## ABSTRACT

This three-article dissertation emerges from interests in corpus linguistics (CL), corpus-based discourse analysis, and corpus-informed pedagogy for second language (L2) writing classrooms. A brief summary of each article follows:

Article #1: Using the localized, place-based discourse of the Rosemont Copper Mine debate of southern Arizona, the first article produces a corpus-based discourse analysis of texts from the primary interest groups involved in the mine proposal. The ecolinguistic analysis details linguistic patterns within the interest groups' texts and discusses how these grammatical and semantic features form rhetorical constellations, i.e. patterns of linguistic features performing a shared rhetorical purpose, within the debate. Findings show that the industry group produces rhetoric of authority, certainty, and dominion through deployment of particular constellations of lexicogrammatical features while the linguistic elements in the environmental advocacy texts construe uncertainty, doubt, aesthetic value, and environmental stewardship.

Article #2: The second article details an integration of geographical information system (GIS) and CL techniques with an ecolinguistics-informed analytical framework for the analysis of the same contentious environmental debate from southern Arizona. The application of GIS and CL procedures enabled the mapping of place name mentions present within two interest group corpora as well as the frequency of particular semantic tags and semantic tag sets that co-occur with specific places prominent in the debate. The findings and the GIS visualizations exhibit how different interest groups refer to and represent geographical places within their discourse and how these references to places index ideological positions towards the environment.

Article #3: The final article details a study in which twenty-one international students in the second course of an undergraduate writing program sequence at a U.S. university studied the local debate regarding the Rosemont Copper Mine. The participants analyzed texts from two primary interest groups, a local, environmental group and an international mining company, and participated in a series of corpus-aided activities using corpus data derived from texts from the opposing groups. The contrastive analyses made possible through the study of texts and corpus data from the two sharply distinct groups enabled students to notice, analyze, and discuss the meaningful and purposeful variation in word choice and rhetorical strategies present in the texts, the data, and the debate. The article provides a model for how corpus data can be integrated into writing classrooms for advancing students' abilities to analyze language and increase rhetorical awareness.

The introductory chapter provides an overview of the Rosemont Copper Mine debate, ecolinguistics, and corpus linguistics. This opening chapter is followed by three articles (corresponding to Chapters 2, 3, and 4), and the dissertation concludes with a discussion of implications of the findings and potential for future research.

## **AN INTRODUCTION TO THE ROSEMONT MINE DEBATE, ECOLINGUISTICS, & CORPUS LINGUISTICS**

From Keystone Pipeline and its origin in Alberta, Canada and its 2,000 mile traverse of farmlands and rivers to the Gulf of Mexico, to the Pebble Beach Mine Project of Bristol Bay, Alaska, on to hydraulic fracturing across the Marcellus Shale formation, and to the proposed Rosemont Copper Mine in the Santa Rita Mountains of southern Arizona, the exploitation and degradation of the environment continues essentially unheeded despite ample and overwhelming evidence of climate change, species loss, irrecoverable ice cap melting, and a host of other human-made environmental tragedies. Each of the aforementioned projects will undeniably alter and perhaps make ruin to ecosystems for numerous species across North America from jaguars and ocelots in the Santa Ritas, sockeye salmon in Bristol Bay, and numerous animals across the path of the Keystone Pipeline. These massive encroachments make significant local impacts on water resources, clean air, recreational space, and local ecosystems across the United States. However, the so-called “development” marches on and although concerned citizens are increasingly unifying and pressing for sustainability, the reality is we have done little of substance to actually produce the kind of change necessary for the long-term viability of the human species on the planet. While Earth may indeed persevere, it is increasingly evident that humanity’s long-term viability is in jeopardy. However, to prolong our stay and perhaps even slow the rate of environmental degradation, hope and opportunity for change does remain if more people from more walks of life bring their knowledge and expertise to bear on a problem that touches us all. As linguist M.A.K. Halliday (1990/2001) asserted, this crisis is not one simply for the chemists, biologists, and scientists but that all within the academy, and especially those within applied linguistics, have a role to play, for this problem is too grand and too urgent

to be engaged by only a few. The environmental crisis is truly an “All hands on deck!” scenario as the consequences of limited action are truly severe. This crisis will not be successfully addressed by fractured departments, self-preserving research institutions, and disparate government agencies pursuing distinct agendas with little collaboration and cooperation. Instead, it calls for increased trans-disciplinary collaboration that expands and extends existing research capacities through successful and comprehensive integration and elaboration of approaches and theories from various research disciplines to enable “new forms of activity” (p. 176) to engage the crisis. And certainly, within this space, applied linguists have an important role to play for the language of the environment must also be subject to greater inspection.

This dissertation is a humble attempt to integrate various approaches and frameworks in the investigation of environmental discourse to hopefully produce new insights into the patterning and production of features common in environmental debates and the various, often flawed and incongruent, conceptualizations produced by our language and our discursive production of environment and place. This project follows the challenge to applied linguists offered by Halliday (1990/2001) and his call for a more active role from applied linguists in addressing the environmental crisis and Mülhäusler and Harre’s assertion “that ecological and environmental studies need to take a linguistic turn” (1999, p. 1) for this most urgent of crises “is at root a discursive phenomenon” (p. 3). With particular focus on the discourse of the Rosemont Copper Mine Proposal of southern Arizona, this work produces a series of applications: a corpus-based ecological discourse analysis of grammatical and semantic features in interest group texts, a study of place and its importance in environmental discourse through the integration of GIS and CL , and an illustration of a corpus approach for the teaching and learning of L2 writing.

To frame this work, first, the ongoing debate surrounding the Rosemont Copper Mine proposal will be introduced followed by an overview of corpus linguistics and ecology/ecolinguistics as well as a brief discussion of the potential for integrating corpus linguistics and ecology.

### **The Rosemont Mine Proposal**

In 2005, Augusta Resources, a Canadian-based mining corporation, bought the rights to the Rosemont Ranch area in the Santa Rita Mountains approximately 20 miles south of Tucson, Arizona and months later submitted an official mining plan as part of the National Environmental Protection Act (NEPA) project protocol for the construction of a massive open pit copper mine on their acquired lands. The initial proposal was revised and supplemented in 2007 in a second submission which attempted to more fully and comprehensively detail the company's plan for the mine and outline their various mitigation schemes and efforts devised to limit environmental impact in the area. However, the proposed copper mine site is situated on the eastern boundary of the Coronado National Forest and the western edge of the Las Cienegas National Conservation Area, a region well-known for its abundant birdlife, countless recreation trails, rich ranching history, deep cultural importance to native peoples, and the approximately forty local wineries which flourish on the Sonoita Plain. For these reasons, local opposition towards the mine has generally been strong, and in 2008, the first organized opposition to the mining plan materialized as the Las Cienegas Conservation Group made public their objections to the proposal. Soon after, other groups formed with Save the Scenic Santa Ritas Organization (SSSR) emerging as the most active and influential; SSSR is also the primary litigant in the ongoing legal battles between the mine developers and the environmental groups. Thus, SSSR and its affiliate, Rosemont Mine Truth (RMT), the organization leading the public relations

campaign against Augusta Resources and their subsidiary, Rosemont Copper Company, and whose board and membership generally reflect the SSSR's, are the primary opposition voices. While the mine debate has captivated southern Arizona, the project first received national attention in 2013 from both the New York Times and the Wall Street Journal and in 2014 was featured in publications such as National Geographic and the Los Angeles Times.

It seemed for several years the primary strategy of SSSR and RMT was to delay the construction of the mine through legal maneuvers and repeated calls for additional environmental impact studies in hopes that the delays would produce a financial burden too significant for Augusta Resources and RCC to sustain. Indeed, in the final months of 2013, this strategy was appearing to have success as RCC and Augusta's stock prices fell to nearly one U.S. dollar and the loss of nearly \$57 million in company value following the SSSR posting of a confidential letter written by the EPA suggesting permitting be denied (Perry, 2013). Further, the company was forced to seek an additional \$26 million dollar loan from London-Based creditors, Red Kite International (Perry, 2013). Despite these issues, in the spring of 2014, Augusta and RCC was hopeful for an impending decision and the conclusion of the review process but instead suffered again from major setbacks in their efforts to secure Environmental Protection Agency (EPA) and Army Corps of Engineers (ACOE) approval. First, the U.S. Forest Service announced it would not meet an April 30, 2014 deadline for responding to the hundreds of objections filed against the agency's Final Environmental Impact Statement (FEIS) in the public comment and objection's phase of the process. Additionally, and perhaps most damaging, ACOE informed Augusta and RCC that the company's application for a Section 404 Clean Water Act permit still failed to meet regulatory standards. These setbacks pushed potential approval dates many months beyond Augusta and RCC's desired approval dates of the

summer of 2014. Further, with new images surfacing of a male jaguar and ocelot near the proposed mine site, the potential for new studies to confirm the existence of the endangered species in the region and the effect the mine would have on these animals seemed to open a very real possibility of additional delays for the mine's approval. Many activists in Tucson and the surrounding areas increasingly believed the possible delays would further cripple Augusta's finances and were exceedingly optimistic that they had finally won in their battle to protect the Santa Rita Mountains and the rich flora and fauna of the region.

These initial hopes for victory though were dashed when in the summer of 2014 a much larger and more financially secure international mining operation seized control of Augusta Resources, RCC, and all rights to the Rosemont Ranch and proposed mining area through a corporate takeover. The company, Hudbay Minerals Inc., announced quite forcefully in the first days following the acquisition that they had the financial means to continue to push the approval process through whatever litigation and reviews would be necessary while asserting they possessed both the patience and financial capability to persevere through any additional reviews. Most concerning for local environmental activists, Hudbay stated, "Our objective, in addition to building Rosemont, is to give Pat (Merrin) free reign to start to accumulate property positions, hopefully within spitting distance of the industrial complex we'll build at Rosemont." Suddenly, the battle to preserve the Rosemont Ranch area expanded to areas throughout the Santa Ritas, the Sonoita Plain, and the Patagonia Mountain region. As of March 2015, Hudbay has continued the approval process but with little progress to date. Indeed, their stated interests in developing a much larger mining complex has prompted many to argue that the approval process must begin anew as recently issued statements seemingly fall beyond the original plans.

This brief summary conveys the contested nature of the ongoing debate in southern Arizona. The identified key parties from Augusta Resources, RCC, RCT, SSSR, and Hudbay have advanced aggressive public relations measures to convince the citizens of Sonoita, Patagonia, Tucson, Pima County, and southern Arizona of the veracity and legitimacy of their positions of support or opposition. These persuasive measures may be encountered and experienced through scores of press releases, blog posts, brochures, videos on organization websites, twitter feeds, and Facebook pages while the debate also continues on local news, editorial pages, political speeches, and government meetings. It is this ongoing, ever-growing and heated discourse that is the object of analysis in articles 1 and 2 and which serves as the platform for the pedagogical illustration of article 3.

### **Corpus Linguistics**

Modern corpus linguistics is often noted as beginning in the early 1960s with the development of the Brown Corpus, a 1 million word collection of 15 written genres, and its creators, Kucera & Francis, are considered foundational figures in the field (Tognini-Bonelli, 2004). These first years of corpus linguistics following the creation of the Brown Corpus were primarily concerned with developing techniques for large corpus construction, storage, and analysis as building ever-larger corpora seemed the primary mission. However, the 1980s would see two major developments in corpus linguistics: 1) increased ease of corpus building, primarily due to the scanner (Tognini-Bonelli, 2012) and 2) increased use of corpus data in pedagogy, namely COBUILD (Sinclair, 1987) and the data-driven learning method (DDL) (Johns, 1986 & 1994). With 20 to 30-million word corpora a reality and equipped with increasing quantities of data, CL witnessed a “theoretical and qualitative revolution” (Tognini-Bonelli, 2012, p. 17) marked by John Sinclair’s presentation of the Idiom Principle (1991) and the importance of

collocation. Increased sophistication in extraction techniques and corpus software tools enhanced with greater analytic functionality enabled CL to challenge entrenched theoretical positions (Tognini-Bonelli, 2012). The “second serendipity” would soon follow with the proliferation of online texts and the realization of even much larger corpora as a result of the internet and the tremendous amounts of electronic data it made accessible (Kennedy, 1998). As a result of the technological advances and the increasing ease of compilation, storage, and analysis but also growing popularity in corpus methods, the 2000s to present have witnessed both new debates and renewed discussions of corpus linguistics as theory or method, primacy of corpus-based or corpus-driven research, value of specialized and mega corpora, critiques and responses to concerns of context, the value of ready-made tools or more agile programming language, and debates concerning the best approach for integrating corpus approaches into pedagogical practices for the language learning classroom.

**Recent debates and developments in corpus linguistics.** A long-standing argument that has recently been revisited is whether corpus linguistics (CL) is a theory or methodological approach and whether CL should be pursued from a corpus-driven or corpus-based orientation. This debate has reignited with perhaps greater urgency as corpus tools, software, and extraction languages have become increasingly sophisticated in the past 10-15 years. For Hardie & McEnery (2010), CL is essentially a method, a collection of tools and procedures, for the study of language. Highlighting an important distinction between CL and other concentrations of linguistics, Hardie & McEnery note that CL is not an area of language deserving of attention and explanation in the same respect as morphology, syntax, or semantics. Instead, CL is an approach that may be applied to the various areas of linguistics as well as disciplines beyond linguistics proper (2010). However, for Teubert (2010) a self-proclaimed “old-fashioned corpus linguist”

(p. 356), CL goes well beyond “a bunch of methods for identifying, counting, measuring, and comparing real language data” (p. 356) and cautions that conceiving CL as a collection of methods ignores the many findings of CL research over the past decades. Essentially, Teubert argues that a corpus as method approach that interprets quantitative findings through a lens of a separate theoretical framework does not advance our understanding of language for it simply imposes *a priori* rules, researcher-selected annotations, and restrictive tagging systems upon the data. Similarly, Sinclair (2004) asserts that often researchers apply various annotations and mark-ups to such a level that the researcher no longer is studying the language but the preselected tags imposed on the corpus, and as a result, the potential for corpus as theory to emerge from the data is prevented. The tagging schemes, Sinclair states, cause the text to lose its “integrity” because tags are the product of a human intervention resulting in the researcher only investigating the system of tags, possibly overlooking patterns in the data not conforming to his/her selected system as he/she becomes distanced from the non-annotated text (2004, p. 191). Thus, an important distinction emerges that connects theory vs. method and corpus-based or corpus-driven debates. For Sinclair, (2004) , acceptance and alignment with the position that CL is indeed a theory initiates a corpus-driven research orientation that chooses not to use pre-tagged text but only works with the “raw text” (p. 191) to determine patterns within the “uncontaminated text” (p. 191). Sinclair states, “As long as we rely on tags we are forcing the attention on pre-corpus models of language which require only small corpora anyway” (2004, p. 192). It is the corpus-driven CL as theory tradition from which the Idiom Principle (Sinclair, 1991) and subsequent work on the “centrality of collocation” has been advanced (Hardie & McEnery, 2010).

A corpus-based CL as method stance considers CL to be a set of techniques and procedures that may be applied in various areas of language study and in support of numerous

theoretical frameworks (Hardie & McEnery, 2010). For example, corpus techniques have been increasingly applied to studies in Critical Discourse Analysis (CDA) but also within departments generally conceived of as beyond linguistics and language studies proper. Likely the strongest opponent to the corpus-driven orientation emerges from Stefan Gries' (2010) response to Teubert (2010)<sup>1</sup>. Gries comments that corpus-based linguists approach data with a theory or framework in mind and attempt to support/disprove/improve the theory under inquiry, essentially conceiving CL as "a methodological paradigm" for hypothesis testing (p. 328). Gries laments and criticizes rather strongly the corpus-driven linguists self-imposed exile from other areas of linguistics and the academy and feels CL could grow and become more visible if corpus linguists applied CL methods outside of CL proper and without such strong adherence to a corpus-driven stance. One area Gries finds appropriate for CL collaboration is within Cognitive Linguistics (CogL). He argues that CL and CogL can be quite fruitful partners and presents a detailed argument for areas where CL and CogL overlap: 1) CL frequency correlates to CogL entrenchment, 2) frequency correlates to ease of acquisition, 3) CL and CogL both agree on separation of syntax and lexis, and 4) CL patterns are similar to CogL constructions. Gries asserts that researchers beyond CL are applying CL methods and that CL must move beyond its rather descriptive approach to prediction, explanation, etc. which may only be reached through application of CL external theory. For instance, he notes CL integrated with CogL can more completely answer 'why' questions than a corpus-driven discourse-and-meaning-in-context approach. Thus, he argues that CL embed analyses in more psycholinguistically-informed usage-based linguistics. While Gries has an impressive list of publications that support his calls for a CL and CogL integration,

---

<sup>1</sup> Hardie & McEnery (2010), Gries (2010), and Teubert (2010) were all articles in a special issue of the *International Journal of Corpus Linguistics*. The authors were invited to submit articles detailing their position from an ongoing debate on the Corpora listserv.

traditional corpus linguists such as Sinclair and Teubert have been reluctant to move beyond a corpus-driven CL-as-theory orientation. However, it is undeniable that the majority of CL research is indeed corpus-based, and even Teubert (2010) has acknowledged that while “Stefan Gries’ brand of corpus linguistics may well be our brave new world” that “it is, however, not John Sinclair’s corpus linguistics” (Hardie & McEnery, 2010, p. 387). While Teubert is certainly correct that CL has moved beyond traditional corpus driven doctrine, there is no reason to consider corpus-based research as somehow deficient in comparison to “Sinclair’s corpus linguistics”(Hardie & McEnery, 2010, p. 387). But as Hardie and McEnery claim, the value of the corpus-based tradition is its “general theoretical openness” (p. 389) and ability to be adapted to various contexts and purposes. Thus, a corpus-based approach with a toolbox for all to access and apply for his/her individual context seems the way forward for CL.

**Mega or specialized and the search for context.** Implications of the CL based or driven and theory or method arguments connect to recent developments and discussions in corpus size and design, as the last 10-15 years has experienced dramatic growth in both size, quantity, and types of corpora. The well-known Brigham Young University (BYU) suite of corpora hosts 7 English corpora ranging in size from 50 million, The Strathy Corpus of Canadian English, to as large as 1.9 billion, the Global Web-Based English corpus. However, likely the most widely used is the Corpus of Contemporary American English (COCA). Declared to be the first monitor corpus of American English, the BYU team led by Mark Davies adds approximately 20 million new words to the collection each year. It is the only corpus to provide data on the past 15 years of English use (Davies, 2009) (The British National Corpus was last updated in 1993. However, Lancaster University announced plans in July, 2014 to construct the BNC2014, a spoken corpus of British English.) The COCA’s year-to-year balance in composition allows diachronic analyses

to be pursued with the confidence that each year in the corpus is essentially equivalent in make-up. These mega-corpora are important, for Sinclair (2004) writes that corpus-driven CL requires large corpora to reveal frequencies of occurrences without the use of tagging, although the BYU corpora are part-of-speech (POS) tagged. As Sinclair famously asserts, “There is no virtue in being small. Small is not beautiful; it is simply a limitation” (p. 189). Perhaps a more direct statement against small corpora could not be offered. Sinclair asserts that with large corpora the “underlying regularities have a better chance of showing through the superficial variation” (2004, p. 189). As instances are repeated, the researcher is able to notice repetition and regularities within the feature and its structure can be noted. He does state though that if through using corpus techniques on a small corpus, you can reach the results you wish, then “your methodology is above reproach” (2004, p.189). However, the critique for large mega corpora, especially as web-as-corpus potential manifests, is the perception that large corpora present output which can neither be properly nor completely interpreted for the context from where it was produced cannot be fully recreated. Thus, while the race for ever-larger corpora has continued, a growing concern for lost context has prompted an increased amount of work and attention to the value of small, specialized corpora for applications from corpus-informed genre analysis, discourse analysis, and classroom pedagogy.

In recent years, there has been a clear response and shift in the field towards the analysis of small, specialized corpora. A specialized corpus<sup>2</sup> is ““delimited by a specific register, discourse domain, or subject matter” (Beaugrande, 1996, p. 11) and has immense value for discourse analysis (Hyland, 2009) generally and the investigation of academic and professional

---

<sup>2</sup> A specialized corpus is not necessarily a small corpus. A small corpus is generally considered to be fewer than 250,000 words (Flowerdew, 2004). However, a specialized corpus can and often does exceed this threshold.

English texts and contexts specifically (Flowerdew, 2004). Acceding that small corpora are not appropriate for all tasks, Koester, 2011 asserts several notable advantages of small, specialized corpus research. First, he states with small corpus studies that all instances of a feature may be analyzed instead of just a random selection of tokens, or in worst cases, researcher selected tokens displaying desired patterns or meanings. Often, the sheer quantity of data resulting from large corpus inquiries may make data difficult to engage and interpret for both researchers and students as the volume of tokens extracted can become unwieldy (Beaugrande, 1996). While a wealth of data may well be desired in lexicography, dictionary constructing, and general language research, the overwhelming amount of data limits qualitative, context-sensitive interpretations. Large corpora enable general, quantitative analysis and generalizations; however, small corpora make possible more qualitative interpretations of data (Flowerdew, 2004). As Koester (2011) asserts, the primary advantage of small corpus studies is the close connection between context and the corpus output. While context may not be retrievable from a large corpus, with smaller specialized corpus, it is more likely the researcher has familiarity with the context from which the corpus was compiled. Thus, small corpora are better suited for producing insights into connections between contexts and linguistic patterns. A final advantage forwarded for small corpora are the insights made possible when corpus data from a large corpus are compared to data from small, specialized corpora. This type of contrastive corpus analysis helps reveal “vital differences among general and specialized meanings of important terms” (Beaugrande, 1996, p. 9), allowing better understanding of specific contexts.

Attention to context has advanced corpus approaches in discourse analysis. Hyland states, “Corpora and discourse are perfect bedfellows” (2009, p. 110) and notes that corpus analysis moves discourse analysis away from the study of “individual preferences” to a larger

focus on communities of users. Thus, corpus-based discourse analysis of specialized corpora allow for more complete understanding of the relations between authors and audiences and provides insights into linguistic choices and their context (Hyland, 2009). Flowerdew has also been quite outspoken in her assertions for the value of specialized corpora and their ability to mitigate the concerns of context noted by corpus critics such as Widdowson. Flowerdew (2004) asserts that small, specialized corpora composed of specific registers, genres, or discourses enable greater recontextualization and more grounded interpretations of the linguistic patterns. Flowerdew (2008) displays examples where the application of contextual understanding is key to interpretation of concordance output while also showing when corpus data may be used to offer insights into the context. She investigates the semantic prosody of “associated with” and epistemic stance through devices such as “it is expected that...”. In this example, she notes how an understanding of the context of the use of “associated with” may be “read off” the concordance lines, as the item occurred frequently and across the full range of texts, indicating general practices of the community. In a second example, a consultation of a “specialist informant” was required to reconstruct the context (p. 124). The insights from the informant based on his/her understanding of the context allowed for the interpretations. With smaller specialized corpora, the researcher likely has greater familiarity with the context in which the discourse was produced, and can, therefore, offer greater insight into the data. Thus, specialized corpora make possible more qualitative interpretations of data, as researchers have greater access to “co-text and extra-linguistic features” (Flowerdew, 2004, p. 16); they are better able to offer qualitative complements to quantitative analysis. Issues of context have certainly been an area which has challenged CL over the past 10-15 years, but small, specialized corpora which enable context to be connected to linguistic features have been a major development. As Biber et. al.

comment, ““a crucial part of the corpus-based approach is going beyond the quantitative patterns to propose functional interpretations explaining why the patterns exist” (1998, p. 9). Many researchers now believe these “functional interpretations” are best attained through the analysis of small, specialized corpora; however, the debate continues.

Related to issues of context are methodological advances designed to further resolve critiques of absent context. Most notable among these are integrations of CL with genre analysis but enhanced with ethnomethodological practices such as observations and interviews of insider participants (Flowerdew, 2004). An early example of this progression is Flowerdew and Wan (2006) corpus-informed genre analysis of 25 tax computation letters from a Hong Kong-based accounting firm which applies the ethnographic method to gain deeper insight and understanding of a particular professional business genre, the tax letter, and the high frequency lexicogrammatical features which occur in the letters. For the ethnographic element of the study, the researchers observed regularly scheduled meetings in which accountants discussed various cases on which they were working. After the corpus analysis of the letters was complete, the accountants were presented the information and interviewed in a semi-structured format regarding interactions in the meetings but also the corpus data. The questioning of the accountants in regards to use of templates, awareness of moves, and explanations of lexicogrammatical patterns common in the genre provided a contextual knowledge to the researchers, strengthening the merit of their qualitative interpretations. Flowerdew and Wan (2010) further exhibits the value in ethnographic methods in corpus-based discourse analysis of a professional genre. Again, an ethnographic element through observation and reflective interviews is pursued to produce a deeper explication of the genre than what is possible through what the author critiques as “narrow” (p. 79) analysis. However, as with many research projects,

the authors note the lack of longitudinal data is an important weakness of the study. This challenge is perhaps an avenue for future corpus work to pursue. Nonetheless, Flowerdew and Wan (2006, 2010) display an important methodological development within corpus-informed discourse analysis, as they clearly show how the ethnographic dimension may enrich traditional corpus and genre analysis.

**Pedagogical challenges and developments.** With the research produced by corpus-based discourse analysis of small, specialized corpora, it would be expected that corpus-informed pedagogical practices would also experience change. These influences have been realized most obviously in developments of top-down discourse-based methods in English for Specific Purposes (ESP) and English for Academic Purposes (EAP) classrooms. A turn to more applied pedagogical ends for corpus research was evolving in the mid 1990s (see Henry & Roseberry, 1996), but it was Flowerdew's (1998) challenge which sparked corpus researchers to be more explicit with the pedagogical value of corpus findings for ESP and EAP curriculum and materials design, but also to "look below the surface of collocational patterns" towards the investigation of "functional patterns" (p. 547). Asserting the need for greater attention to the pedagogical implications of findings and the need for more discourse-oriented practices, Flowerdew (1998) ushered in a new era of corpus pedagogy. The discourse-oriented approaches largely resolved critiques of the bottom-up quality of corpus instruction as more rhetorically-sensitive top-down functional approaches were developed.

Beginning the trend to top-down practices, Henry & Roseberry (2001) reported the findings of a corpus-informed genre analysis of a small, specialized corpus with particular attention to pedagogical best practices. This study determined the prominent linguistic features present in the moves of the letter of application. The efficacy of a corpus-informed genre

analysis of moves and strategies for subsequent instruction was clearly displayed. Henry (2007) also assessed the effectiveness of the corpus-informed genre analysis approach in a classroom setting and investigated whether effectively presented corpus data can lead to the successful transfer of moves and corpus-attested features by the learners to successful production of a genre, in this case the letter of application. After study of the online corpus materials, students did include and successfully realize a greater number of moves. Also, keyword analysis of the second round of letters showed considerable changes and displayed the students' usage of lexical items and discourse patterns presented through the corpus instruction.

Studies have also assessed student attitudes of corpus approaches in writing classrooms. A precursor to important work from Charles (2007, 2011), Yoon & Hirvela assessed ESL students' perceptions of corpus use in a 10-week L2 writing class through an end-of-course survey instrument and semi-structured interviews. A majority of students reported that using the corpus was beneficial for learning the usage and meaning of vocabulary, a finding common in corpus research. Overall, students indicated quite positive attitudes towards corpus use, claimed that using the corpus improved their writing ability, students attributed an increase in confidence in L2 writing to using the corpus. Yoon & Hirvela also discuss how genre-based corpus approaches allow students more focused study of features present within specific professional & academic genres of interest. The authors state that corpus study benefits learners through exposure to authentic language from targeted discourses. Again, these features and the benefits derived from the instruction evolve from prior critiques of the bottom-up nature of corpus pedagogy. Yoon & Hirvela display the value a corpus-informed genre approach can have for writers in specific contexts and with specific purposes.

Most developed of the top-down pedagogical movement is work from Charles (2007, 2011). She notes that a common critique and a likely reason corpus approaches have not been more fully integrated into curricula is that corpus consultation focuses primarily on lexicogrammatical features and has not been applied to the teaching of higher-level rhetorical functions. Her approach though uses a small corpus of master's theses to highlight rhetorical features, particularly concession strategies where writers anticipate and minimize potential criticisms, for 40 graduate student writers. This approach begins with emphasis on rhetorical/discourse functions and then follows with a closer look at the linguistic features that typify and often signal moves, i.e. instruction moves from the top down. Charles (2007) also responds in length to critics of corpus approaches and their claims that corpus study is too often lacking attention to context. In rebuttal to the charge, Charles (2007) provides several examples that show how corpus/concordance afford greater levels of contextualization than would be possible in the average textbook. A more refined top-down approach is exemplified in Charles (2011). Again, the approach begins with discourse based activities that highlight rhetorical functions before transitioning to hands-on corpus consultation by the students. While the 2007 article focused on strategies of concession and highlighted the use of 'while' in these moves, this study instructs learners on strategies for offering criticism. As with the earlier study, a small focused corpus produces a limited set of concordance lines to enable more transparent and accessible interpretation. Overall, students did indeed report that they learned discourse features through the concordance approach. Charles (2007, 2011) displays the pedagogical value of the top-down rhetorically-informed approach and show that while early corpus pedagogy may have indeed promoted more atomistic bottom-up instruction, corpus pedagogues have responded and developed more dynamic pedagogical practices.

The previous research on top-down approaches generally begins with compilation of a corpus by a course instructor. However, as corpus tools have become more common, arguments for equipping students with the skills to build and analyze their own specialized, often discipline-specific corpora have been made. Displaying the potential for student-led corpus building and analysis projects, Lee and Swales (2006) report a 13-week advanced-level corpus-based writing course for 4 NNS doctoral students. The course spent the opening weeks training students in corpus techniques and also taught students to build their own specialized discipline-specific corpora. Lee and Swales note the course achieves its objectives as it created a decentered classroom where students produced their own insights into language and did not rely on native speaker intuition for confirmations. Students reported the corpus approach to be empowering and even commented that a corpus is more useful than reference & grammar books for writing because corpora provide more discipline-specific information than is typically available in other resources. Lee and Swales comment that their corpus approach represents a “technology-enhanced rhetorical consciousness-raising” (p.72). Working with a small corpus of authentic, discipline specific examples may help students notice meanings and contexts not available through large corpus study or an encounter with a sole token in text (Beaugrande, 1996). Again, the pedagogical research review exemplifies developments in corpus pedagogy.

**Ready-made corpus tools and programming.** Gries and Newman (2013) state that likely one of the most important factors contributing to the increased popularity of CL is the methodological progress made possible through the development of corpus software tools. While initial interfaces were not user-friendly and often lacked the functionality of most current tools, recent generations of Wordsmith Tools, Antconc, ConcGram, and MonoConc are much more intuitive in their design but also offer greater analytical capability through procedures such as

keyword analysis, cluster analysis, collocational profiling, and even some simple visualizations, e.g. dispersion plots. However, these tools do not come without their critics. For instance, Gries & Newman (2013) do comment that the freely-available corpus tool Antconc is likely the best tool for its robust use of regular expressions, ability to handle multiple languages, and ability to work with a range of annotation schemes but also argue that the corpus researcher is bound by the functionality of the program and is limited in the operations and searches he/she may perform. Therefore, corpus linguists are increasingly urged to learn and use programming languages such as Perl, Python, and R for CL research. Gries has been so eager in his calls for increased use of programming languages that he organizes and leads summer workshops teaching R to corpus linguists and has published instructional texts on the topic as well.

**Corpus approaches to discourse analysis.** As referenced in the previous section, corpus linguistic methods are increasingly applied for discourse analysis. While the debate detailed previously regarding issues of context and corpus size was furthered by scholars central to the development of corpus linguistics, in recent years, corpus methods have been adopted by researchers across the academy for investigations of various linguistic features, genres, communities, issues etc. Much of this research, as referenced previously, has been of academic genres and discourses for insights into academic registers for the benefit of EAP and ESP learners. However, corpus-based discourse analysis research has proliferated and the topics investigated are exceptionally diverse. Areas that have received much attention are political discourse (e.g. Partington, 2003; Partington & Morley, 2004; Milizia, 2009, 2010, & 2012; Ferrari, 2010; Conoscenti, 2011) and social issues such as immigration (Morley & Taylor, 2012; Fitzsimmons-Doolan, 2009; Blinder & Allen, 2014; Incelli, 2013) Other work, e.g. Baker (2010) and Baker, Gabrielatos, & McEnery (2012), have reported the representation of Muslims in the

British newspapers and tabloid press, the language of extremism (Prentice, Rayson, & Taylor, 2012), and the discourse of news reporting of war (Kutter; 2012).

Business and corporate discourse has also attracted attention from discourse analysts using corpus methods. One genre receiving particular attention are corporate annual reports as well as the related, and often accompanying, letter to the shareholder. Many of these studies focus on how institutions promote successes or obscure failures while also projecting corporate ideology (Abrahamson & Park, 1994; Clatworthy & Jones, 2003; Poole, in press; Rutherford, 2003; Subramanian, 1992; Thomas, 1997). For example, Hyland (1998) investigated letters to shareholders using a corpus analysis approach to analyze the meta-discourse present in the president's letter of publicly listed Hong Kong-based companies, finding that "the CEO's letter clearly moves beyond passive disclosure to what amounts to the marketing of corporate ideology" (p. 232).

Other less-explored but nonetheless valuable and engaging studies report the discursive production of tourist locations (Jaworska, 2013), the discourse of tourism on the internet (Maci 2012a, 2012b), English-only ideologies in the United States (Subtirelu, 2013), and attitudes and representations of feminism in the UK press (Jaworska, S. & Krishnamurthy, 2012).

Forensic linguistics, e.g. Tabbert (2013) and the investigation of how German and United Kingdom press represent and construct offenders, and studies of discourses of health and hospital settings (e.g. Trigg, 2007; Maci, 2012c & 2012d, Staples & Biber, 2014) are areas attracting greater attention in recent years.

These referenced selections are not intended to be an exhaustive list of corpus-based discourse analysis studies but rather to display the volume and diversity of studies in this area. While these studies are unified by their application of corpus methods for textual analysis, they

often apply an eclectic selection of theoretical frameworks from cognitive linguistics, conceptual metaphor theory, critical discourse analysis, evaluation and appraisal, to name just a few.

Omitted in the present section are corpus studies of environmental issues such as genetically-modified foods and global warming; these will be discussed in much greater detail in the first article.

### **Ecology and Ecolinguistics**

The term “oecology” was first employed by the German zoologist Ernst Haeckel in the 1860s (Kingsland, 1991, p. 1) in an attempt to capture and explain the complex relationships Charles Darwin elaborated in his 1859 *On the Origin of Species* (from Macintosh, 1985, cited by Kingsland, 1991, p. 1). This coinage and his work helped launch ecology as a science that was made distinct through its “application of experimental and mathematical methods to the analysis of organism-environment relations, community structure and succession, and population dynamics (Kingsland, 1991, p. 2). In the decades following Darwin’s classic and Haeckel’s elaboration, naturalists, botanists, and zoologists took a more “rigorous approach to natural history” (p. 2) and at the turn of the 20<sup>th</sup> century, ecology was described as a “dynamic, experimental approach to the study of adaptation, community succession, and population interactions”( p. 2). Increasingly, an ecology-driven approach was adopted by prominent naturalists in the United States who rejected more traditional, descriptive methods for the more quantitative and theoretical principles offered by the emerging science. In this time, Hutchinson’s seminal piece “The Lake as Microcosm” was published, an article that remains a frequently cited piece in ecological work, and the conceptualization of the complex interactions present within an ecosystem was further developed. Interestingly, considering the location of the current project, in the early 1900s the Carnegie Institution funded the Desert Botanical

Laboratory in Tucson, Arizona and in this region, the first applications of ecology to language through works of Boaz, Sapir, Whorf, and the Voegelins would emerge.

While the roots of the biological applications of ecology are undoubtedly European in origin, Sapir's work in the American southwest is often regarded as foundational in the area of language ecology for its description of the interrelationships between environment and language. In Sapir's well-known essay, he cautioned against producing a view of human culture and human life in which all differentiation may be allocated to environmental influences. Indeed, he states that deterministic arguments that explain features of human culture through direct correlation to physical environment "rest on fallacy" (1912/2001, p. 13). Nonetheless, he did define language as "a complex of symbols reflecting the whole physical and social background in which a group is placed" (1912/2001, p. 14) and that language is indeed "influenced by the environmental background of its speakers" (1912/2001, p. 14). In perhaps the most profound yet critiqued statement of the essay, Sapir wrote, "The complete vocabulary of a language may indeed be looked upon as a complex inventory of all the ideas, interests, and occupations that take up the attention of the community, and were such a complete thesaurus of the language of a given tribe at our disposal, we might to a large extent infer the character of the physical environment and the characteristics of the culture of the people making use of it" (1912/2001, p. 14). While the strong form of this statement certainly attracted its critics, it seems that in recent years this view has reemerged and its essence has become much less controversial. The evidence of his influence on language ecology remains and in perhaps the most comprehensive volume of essays on ecolinguistics, *The Ecolinguistics Reader: Language, Ecology, and Environment*, his essay is the first included in the text.

The American southwest and languages of native peoples would also influence the work of C.F. and F.M Voegelin and their collaboration with Noel Shutz Jr. Their 1967 publication in Dell Hymes' "Studies in Southwestern Ethnolinguistics" is of import for its application of ecology to the description of the languages used within a geographical area from a sociocultural perspective. This work moved beyond census-like reporting of number of speakers of a given language to a much more in-depth analysis of the functional interactions between languages within an area and the complexity of interactions between the language varieties present. Thus, this often overlooked pre-cursor to Haugen has an important place in the development of language ecology.

The scholar most often cited as the founder of language ecology is Einar Haugen; he coined the term *ecolinguistics* and his 1972 text *The Ecology of Language* is often marked as the beginning of modern ecological approaches. In his text, Haugen (1972) defines language ecology as "the study of interactions between any given language and its environment" (p. 325) and critiques biological, instrumental, and structural metaphors, asserting the ecology metaphor presents a more dynamic conceptualization of language and its interrelationship with the environments in which it exists. Within an ecological approach, Haugen charges that one must explain the context of the language but also report the effect the language has on the context. Thus, in Haugen's view, language ecology is characterized by reciprocity and interaction. Haugen also critiques many studies in language description for offering brief, and as he states, "perfunctory" (1972, p. 324) information on the ecological environment in which languages are situated. He claimed the basic comments do little to actually explain the status and function of a language and argued that linguists should pay more attention to explaining much more

comprehensively the ecology of a language. Three main strands of research have emerged in the years following Haugen's seminal text.

**The strands of ecological research.** Within ecology and ecolinguistics, there have generally been several dominant conceptualizations of what may be deemed a language ecology with each conceptualization yielding different research focuses. The first of these strands presents language to exist in a *symbolic* ecology in which languages and their co-existence with other systems are investigated (Fill & Steffensen, 2013). This conceptualization forwards the dialectical and reciprocal relationship between language and context/environment/ecosystem and most reflects the arguments forwarded by Haugen (1972):

Part of ecology is therefore psychosocial: its interaction with other languages in the minds of bi- and multilinguals speakers. Another part is sociological: its interaction with the society in which it functions as a medium of communication (p, 324).

From this position, much research in language policy has been generated for, as Mülhäusler writes, "language planning needs to be ecological in the sense of being informed by interconnections" (p. 9) within communities of speakers. If language is conceived through an ecological lens, then issues such as standardization and corpus building lose prominence to concerns of speakers, variation, attitudes, and a variety of other factors (Mülhäusler, 2003). Fill detailed work emerging from the Haugen ecology metaphor and noted that concepts from biological ecology such as conservation, interaction, endangerment, & environment have been transferred to language.

An additional conceptualization of the environment of a language considers language to exist in a *natural* ecology (Fill & Steffensen, 2013). Present within this strand is work that questions whether linguistic patterns and features function within discourse to produce unsustainable relationships between humans and the environment (Steffensen & Fill, 2013). This

work originates from views that hold “there is some sort of misfit between the contours of our language and the contours of our natural environment” (Mülhäusler, 1999, p. 22). Likely the most well-known work within ecological discourse analysis is Halliday’s discussion of features embedded within English language in what he calls the “cryptogrammic fourth level” (1990/2001, p. 193). Halliday’s argument asserted that “categories and concepts are construed by grammar, and thus, language is not an arbitrary system which represents and encodes pre-existing environmental realities” (p. 180). Thus, he forwards four features which construe the environment in ways he marks as unsustainable: 1) unbounded, non-count nouns present items as limitless, 2) the grammar of ‘good’ as the grammar of ‘big’ with grammar of ‘small’ representing ‘bad’, 3) inanimate objects are prohibited agency, 4) personal pronouns permitted only for animate entities.

Similarly, Chowla (2001) argued that features of language prevent speakers of English from having holistic views of the natural environment, while Shultz (2001) stated that properties of language may be exploited to convey particular impressions and feelings about our environment. Shultz also asserted the language of western society is the “language of commercial users of the environment” (p. 109) and even criticized activists and environmentalists for using exploitive language even when the negative influence has been highlighted. In particular, Shultz (2001) noted the use of “neutral words” (p. 109) such as *development*, *progress*, and *sustainable*, euphemisms, and the labeling of neutral items negatively as three linguistic features that favor commercial uses of environment. Reflecting Chowla (2001) and Shultz (2001), Fill (2002) commented that humans use nature as a resource which is reflected by the language, and thus, language normalizes and perpetuates continued resource depletion. Additional work in this area emerges from Alexander (2002, 2009) and his

corpus-based CDA investigations of texts produced by corporate interests relevant to the ecological crisis, Gerbig (1997) and her study of agency, causality, and responsibility through variation in verb usage, namely ergative constructions, and Goatly (1996, 2002) in his work on anthropocentrism and his call for a green grammar that more accurately represents reality.

In addition, while these works aligned with ecological discourse analysis may on the surface seem less informing of classroom practice, Haig (2001) reported an application of ecolinguistics and critical discourse analysis (CDA) for the teaching of eco-literacy through the study of complex, environment-related texts where ideology is often nuanced and subtle. He displayed how a theory-informed and linguistically-aligned approach may be used to help language learners understand not only what ecological texts mean but also how they create such meaning. He stated that a simplified systemic functional linguistics (SFL) & critical discourse analysis (CDA) approach in the classroom can help students develop “a critical consciousness of key issues within their natural and social environment” (p. 209) and asserts his method to be a “useful, systematic way” (p. 209) to help students raise critical awareness of the ideology present within ecological texts.

A third conceptualization of language ecology represents language as existing in a *sociocultural* ecology. From this conceptualization, much work pertaining to language education and language acquisition has emerged. Perhaps most well-known within this thread of ecology and ecological approaches is the work of Claire Kramsch. In Kramsch’s opening chapter on ecological perspectives on acquisition and socialization, she forwards an ecology metaphor for understanding language development and language users’ interactions with their contexts. While she does not discard the LEARNER-AS-COMPUTER metaphor nor the LEARNER-AS-APPRENTICE metaphor, she does assert the need for a new overarching metaphor, the ecology

metaphor, “which captures the dynamic interaction between language users and the environment as between parts of a living organism” (2002, p. 3). She suggests the ecology metaphor allows a new means for understanding the complexity of language. Thus, this metaphor does not replace the existing ones, but instead offers a new way for viewing the relationship between “the dancer and the dance”(2002, p. 5), i.e. the language user and their environment. This new conceptualization of the interrelationship “characterized by interdependence, interconnectedness, and interaction allows the merging of socialization and acquisition through the avoidance of traditional dichotomies” (Kramsch & Steffensen, 2008). She asserts, “An ecological perspective on language development opens up possibilities of embracing the paradoxes, contradictions, and conflicts inherent in any situation involving semiotic activity”(p. 22).

This sociocultural thread of ecology further extends to the domain of SLA and language education. Leather & Van Dam (2003) detailed the dominant assumptions of SLA and the need for a more holistic, ecological rethinking and comment that too much SLA research still originates from a learner as a “closed system” orientation where variables may be controlled and manipulated (p. 1). Asserting the need for increased scrutiny towards an “acquirer’s extensive interaction with their environment--spatial, social, cultural, and so on” (p.1), they challenge long-entrenched assumptions that languages are distinct, monolingualism is the natural order, language acquisition is innate, verbal forms clearly represent semantic meaning, acquisition develops separately than socialization, as well as several other strongholds of traditional SLA theory. They counter that in a globalized, multilingual world that representations of native speaker norms marginalize bi- and multilinguals. Concurring, Kramsch and Steffensen (2008) note that in addition to global mobility, the proliferation of the internet and English around the world lead to a necessary questioning of illusory native speaker constructs. Further, Leather and

Van Dam (2003) assert language is not acquired as we activate an innate universal system and set abstract syntactic parameters but rather language acquisition occurs through complex developmental and cognitive mechanisms. Language learning becomes less about learning structures and formal rules but about learning how to exist in a multitude of contexts (Leather & Van Dam, 2003), to align to new social spaces (Kramsch, 2002), and to “negotiate paradoxes” (Kramsch, 2002). Additionally, Leather and Van Dam state the existing SLA assumptions about the static quality of verbal codes ignores contexts and environments and does not acknowledge the fluidity and context-dependency of language. The general premise is that language acquisition and language use is much more complex than can be isolated and analyzed from a single transcription. The authors assert ecological approaches deny a tendency towards normative interpretations and instead see language development in individuals mediated in complex, contextualized and situated settings, a recognition of the integration rather than a fragmentation of socialization and acquisition. Thus, ecology accepts variation and complexity and does not attempt to over-simplify the acquisition process and instead views language development as emergent in the sense that it cannot be predicted but rather arises from situatedness in specific social, cultural, and physical environment.

Evident in the discussion of the work of Kramsch, Van Lier, and Leather & Van Dam are the implications for an ecological reconceptualization of language acquisition and potential rethinking of language teaching and learning. For SLA, the dominant assumptions of language acquisition and still operational constructs of native and non-native speakers and the image of the language learner must continue to be questioned. As Kramsch & Steffensen (2008) note, considering our increasingly global, technology-enhanced, multilingual reality, the models simply no longer hold true for countless contexts and situations. Thus, we must no longer

perceive the individual as language learner but instead as “newcomer” into a context in which he/she must learn to act and perform (Kramsch & Whiteside, 2007, p. 910). This raises a series of questions for language practitioners especially considering the linear conception of curriculum, classroom objectives, and institutionalized testing regimes. Further, within an ecological approach, we must redefine what constitutes success in language learning environments (Kramsch, 2002). If we do indeed ascribe to the ecological notion that language learning is a dynamic process of socialization, how may teachers facilitate and promote the confluence of factors that may afford language development? Or in Van Lier’s terms, how may we create an environment rich in the affordances necessary for facilitating the emergent process of language development?

The conceptualization of a symbolic ecology has had clear and considerable influence of language policy and planning as well as revitalization efforts and the Hallidayan-influenced natural ecology conception has had great import on system-critical and text-critical ecological discourse analysis. Further, the sociocultural strand has raised serious challenges towards dominant assumptions of second language acquisition and has pushed research towards more holistic, context-sensitive frameworks for SLA while simultaneously empowering learners. While the influence of ecological discourse analysis in the Hallidayan-tradition is perhaps most apparent in the following research, an ecological orientation informs each of the three articles.

### **Integrating Ecology and Corpus Linguistics**

This dissertation offers a model for integrating corpus linguistic techniques with an ecology-informed theoretical framework. The ecological orientation assumed in this dissertation is one that views the ongoing environmental crises as intricately connected to language practices and discourses that are situated and embedded within a complex of cultural, social, historical,

and geographical networks. In southern Arizona and other locales discussing the proposed Rosemont Copper Mine, conceptualizations and relationships to the mountain, the mine, and the environment are reflected, reproduced, and normalized through discursive actions partly influenced by proximity to the Rosemont site. In this debate, a culturally-dominant discourse of the ever-present need for economic development and resource independence produce a linear conceptualization of progress and development, a conceptualization marked by the underlying themes of legitimacy, dominion, and inevitability. In contrast, an oppositional and mostly localized discourse reflects an intimate relationship with the landscape through the naming of places, both obscure and prominent, and through other lexicogrammatical patterns which produce a discourse grounded in sustainability, stewardship, and aesthetics. These contrasting discourses are typified by recurring discursive patterns that allow each stakeholder to elicit, manipulate, and sustain reactions from supporters. By revealing the linguistic and discursive phenomena at play, I hope that citizens may become more informed and skillful communicators and advocates for the environmental challenges faced by our communities.

Another mission of this dissertation is to map place-name mentions and semantic representations of places through the integration of methods from geographical information systems (GIS) and CL. This process of spatialization enables the mapping of discursive patterns and displays how proximity and/or distance to a referent, e.g. the Santa Rita Mountains, allows and enables but also constrains and prevents access to particular discourse actions. As SSSR is embedded within the social, historical, and cultural ecology much more intimately and completely than external stakeholders, the group has license to discuss and present the mountain and the mine in a manner distinct from activists or proponents in Phoenix, the state capital, Denver, the regional headquarters of Augusta Resources, Vancouver, the home of Augusta

Resources, the EPA regional and national offices, and the Toronto-based Hudbay. In these groups' discourses, various types of deictic relationships present in texts represent a key discursive element and highlight the importance of context in the debate. This integration of ecology with corpus linguistics and GIS presented in Chapter 3 enables insights into the interrelationship of context and language and the importance of place within environmental discourse. Further, the integration of GIS and CL contributes to the development of compelling visualizations for the dissemination of findings from corpus-based discourse analysis research. CL continues to expand and researchers from a variety of domains are increasingly applying CL methods; however, CL must develop techniques to share findings in more dynamic and engaging ways to compel "intuitive understanding of complex linguistic developments" to users (Hilpert, 2011, p. 446). From my perspective, the development of intuitive, compelling, and meaningful visualizations for the representation and dissemination of corpus data represents a new frontier in CL.

Ecological thinking also informs the pedagogical intervention of chapter 4 and its application of the principles of ecology for the teaching and learning of L2 writing for second language learners. While many will likely assume the pedagogical illustration is a form of ecoliteracy designed to heighten students' awareness of environmental issues similar to the previously discussed Haig (2001) illustration, this is not actually the mission of the unit. Certainly, raising ecological consciousness is a positive and desirable by-product, but the ecological orientation informing the intervention detailed in chapter 4 is grounded in Bawarshi's (2012) conceptualization of genres as rhetorical ecosystems and Dobrin and Weisner's (2002) work in ecomposition. The ecological approach to writing pedagogy pursued sees writers as intimately and intricately situated and embedded in spaces, whether natural or man-made, and

that understanding and appreciation of the interrelationships between genres, discourses, and language in these spaces can lead to more critical writing that is more responsive to rhetorical situations and purposes. Thus, the intervention does not model how to design and implement a course on eco-literacy, a worthwhile goal and the focus of future research, but instead attempts to integrate corpus and genre approaches through an ecological lens to help students identify and understand the interconnections between speakers, writers, and texts engaged in the discourse of the Rosemont Mine.

This work, while oriented within ecology and informed significantly by corpus linguistics, is driven by M.A.K. Halliday and his work detailing the features of English which produce an unsustainable relationship with our natural environment and heeds his call for increased attention to environmental discourse by the applied linguistics community. His groundbreaking work explicating features of the language and their effects on our how we exist in the natural world and contribute to the current ecological crisis remains a major force in ecolinguistics and a catalyst which has motivated many applied linguists to investigate environmental discourse. While Halliday's work may not be cited or discussed in each of the subsequent chapters, it is his seminal work and his call for a transdisciplinary movement to address the environmental crisis that informs and underpins much of the following research.

This introduction has provided a general overview to the Rosemont Copper Mine debate, CL, and ecology/ecolinguistics. The following articles draw from different aspects of CL and ecolinguistics; thus, each article provides a more focused review of the relevant literature to frame each study.

## **A CORPUS-BASED ECOLOGICAL DISCOURSE ANALYSIS OF THE ROSEMONT COPPER MINE DEBATE**

### **Abstract**

This article reports a corpus-based ecological discourse analysis of press releases from Rosemont Copper Company (RCC) and blog posts from the Rosemont Mine Truth (RMT) pertaining to RCC's proposed development of an open-pit copper mine in the Santa Rita Mountains of southern Arizona. The ecolinguistic analysis details linguistic patterns and their functions within the localized discourse of a particularly controversial environmental issue and how these grammatical and semantic features form rhetorical constellations, i.e. patterns of linguistic features performing a shared rhetorical purpose, within the interest groups' texts. Findings show that RCC produces a rhetoric of authority, certainty, and dominion through deployment of particular constellations of lexicogrammatical features while the linguistic elements in the RMT corpus construe uncertainty, doubt, aesthetic value, and environmental stewardship. The corpus analysis reveals that the RCC rhetoric of inevitability and certainty perpetuates and advances a discourse which places humans in a role of dominion and authority over nature and the environment while the oppositional discourse forwarded by RMT projects the aesthetic value of the land and a need for responsible environmental stewardship. The article details the ecolinguistic framework, provides a review of corpus-based research in the study of environmental discourse, explains the methodology for the corpus analysis, and discusses the findings and implications of the rhetorical constellations.

# **A CORPUS-BASED ECOLOGICAL DISCOURSE ANALYSIS OF THE ROSEMONT COPPER MINE DEBATE**

## **Introduction**

In recent years, corpus-based discourse analysis has proliferated and the topics covered have become increasingly diverse. Areas that have received much attention are political discourse (Conoscenti, 2011; Ferrari, 2010; Milizia, 2009, 2010, & 2012; Partington, 2003; Partington & Morley, 2004), social issues such as immigration (Blinder & Allen, 2014; Fitzsimmons-Doolan, 2009; Morley & Taylor, 2012), representation of Muslims in British newspapers and tabloid press (Baker, 2010; Gabrielatos, McEnery, & Baker, 2013), the language of extremism (Prentice, Rayson, & Taylor, 2012), the discourse of news reporting of war (Kutter; 2012), the emerging area of forensic linguistics (Tabbert, 2013), and the discourse of health and hospital settings (e.g. Trigg, 2007; Maci, 2012a & 2012b, Staples & Biber, 2014).

During this period, corpus-based discourse analytic approaches to the study of environmental discourse has also expanded to the field of ecolinguistics, which has been defined generally as “the study of the interrelationships between language and its environment” (Couto, 2013, p. 1) and more specifically as “the study of the impact of language on the life-sustaining relationships among humans, other organisms, and the physical environment” (Alexander & Stibbe, 2013, p. 2). Under these broad definitions of ecolinguistics, Bang and Trampe (2013) identify two general approaches: 1) explication of “theories of language inspired by a holistic paradigm of ecology” (p.1) and 2) the application of linguistic analysis for the investigation of ecological texts and discourses. This second concentration, often referred to as ecological discourse analysis (EDA), was largely prompted by M.A.K. Halliday and his seminal keynote to the AILA convention in 1991 in which he famously asserted that the environmental crisis is not a

problem for only the biologists and physicists to engage in, but that applied linguists also have a role to play (Halliday, 1990/2001). In closing, he stated, “I do not suggest that we [applied linguists] hold the key. But we ought to be able to write the instructions for its [language] use” (Halliday, 1990/2001, p. 199). This challenge to applied linguists prompted greater attention to the features of language that normalize and reproduce dominant, often unsustainable, conceptualizations of the relationships between humans and their environments. As Harré, Brockmeir, and Mühlhäusler (1999) asserted and Bang and Trampe (2013) later echoed, environmental studies took a “linguistic turn” (Harré, Brockmeir, & Mühlhäusler, 1999. pg. 3), as ecolinguistics “adopted the view that perceptions of nature are mediated through language and that in turn such perceptions and lifestyles feed back into the structure of discourse” (Mülhäusler, 2003, p. 12). Thus, in the past 25 years, ecological discourse analyses studying various features of language and their functioning in the discursive production of environment and environmental crises have proliferated.

In line with the research in ecolinguistics, this article reports a corpus-based ecological discourse analysis of texts from two interest groups involved in the now 9-year debate in southern Arizona regarding the construction of the Rosemont Copper Mine in the Santa Rita Mountains, which, if built, would become one of North America’s largest open pit mines. The proposed copper mine, though centered on privately-owned property, is situated on the eastern boundary of the Coronado National Forest and the western edge of the Las Cienegas National Conservation Area, a region well-known for its abundant birdlife, countless recreation trails, rich ranching history, deep cultural importance to native peoples, and the approximately forty local wineries which flourish on the Sonoita Plain. For these reasons, local opposition towards the mine has been strong, and in 2008, the first organized opposition to the mining plan materialized

as the Las Cienegas Conservation Group made public their objections to the proposal. Soon after, other groups formed with Save the Scenic Santa Ritas Organization (SSSR) emerging as the most active and influential; SSSR is also the primary litigant in challenges to the mine. This corpus-based ecolinguistic analysis investigates the discourse produced by the two primary interest groups in the contentious debate, SSSR and Rosemont Copper Company, and details the grammatical and semantic features that form rhetorical constellations, i.e. patterns of linguistic features performing a shared rhetorical purpose, whose confluence produces a rhetoric of authority, certainty, and dominion in RCC discourse but construe uncertainty, doubt, aesthetic value, and environmental stewardship in SSSR texts.

### **Literature Review**

A discussion of ecolinguistics and ecological discourse analysis must begin with the features cited by Halliday as producing unsustainable relationships with the environment. In *New Ways of Meaning* (2001), the publication based on the aforementioned AILA speech, Halliday cited several features in what he termed the “cryptogrammatic fourth level” (p. 193) that he argued construe reality in a way no longer sustainable for the long term survival of the planet. The first of the features discussed by Halliday is the construal of items such as *air, water, soil, coal, iron* etc. as unbounded, non-count resources, implying these non-renewable resources are limitless. Echoing Halliday’s assertion, Chowla (2001) similarly contends that the practice by English speakers of individualizing nouns in this manner creates the “tendency to perceive resources in isolation rather than holistically” (p. 115). Halliday also asserts that English organizes entities according to their ability to be agents in processes while inanimate objects are rarely conceived of as having agency. Halliday’s example of “What’s the forest doing?” (2001, p. 194) is impermissible, or at least awkward, in the English language and is not attested in the

450-million word Corpus of Contemporary American English (the highest frequency *-ing* verb collocate with *forest* is *clearing*). Thus, he offers the critique that since we are unable to grant agency to the forest, we are unable to respond, “The forest is holding water” or “It’s cleaning the air” (p. 194), representations more aligned with current scientific thinking and more sustainable conceptualizations of nature. Finally, Halliday states the English pronoun system allows pronouns *he/she* for conscious beings only, and therefore, the Earth cannot be represented in a holistic manner and conceived of as a living entity. While Halliday may have brought these features to light within the ecolinguistics program, other researchers have further investigated these features and others within ecological texts.

An additional feature receiving attention within ecological discourse analysis is the linguistic feature of nominalization. Halliday, along with Martin, was central in detailing the emergence of these features within scientific discourse, and while the explication of these phenomena from scientific discourse may seem peripheral to a discussion of ecological discourse analysis, they have been central to discussions of the fragmentation of humans from their environment. Although nominalization, Halliday and Martin (1993) argue, may indeed be the feature which grants scientific and academic writing its “distinctive quality” (p. 3), it has become pervasive across many registers of discourse, functioning to remove or obscure agents in ecological texts as readers/listeners are distanced from the reality of the event/action described. For example, Martin (1986) first highlighted how nominalizations function within environmental discourse in his study of texts from the Canadian Wildlife Federation (CWF) and an Australian environmental advocacy group. Martin details the use of nominalization within the texts and asserts that these usages are incongruent with reality as they present verbs and adjectives as

nouns, influencing the level of abstraction within the text and producing a disconnect between the reality of the event and the grammatical structure which represents it.

Beyond nominalization, transitivity has been an additional feature to receive much attention. In his work, Goatly (2002) applies frameworks from systemic functional grammar (SFG) and critical discourse analysis (CDA) to investigate representations of nature on the prestigious, well-known, and purportedly objective BBC World Service. His study investigated choices of lexicogrammar that pattern within the 2.5 million-word corpus from the early 1990s to instantiate representations of nature in the nearly 140 million listeners. Asserting a weak form of the Whorfian Hypothesis, he argues that language “predisposes us to perceive, think, and act in certain ways, and makes it more difficult to perceive and think in alternative ways” (p. 2). By extracting concordance lines of various categories of animals (e.g. insects, birds, land animals, aquatic animals) and other lexical items from nature (e.g. water, land and landscape, weather, disease) and classifying the role of each item in the concordance lines (e.g. actor in transitive clause, actor in transitive etc.), he concluded that on the BBC there are “frames of consistency” (p. 6) in how nature is represented and notes that with great frequency the BBC presents nature and animals as acted upon rather than actors. Related to transitivity, Gerbig (2001) investigated variation in the usage of ergative verb constructions as well as reporting verbs within a corpus of approximately half a million words categorized into sub-corpora based on the interest groups discussing Ozone depletion. Her work is quite important for her extensive use of corpus methods for the investigation of agency, causality, and responsibility attribution in a corpus of ecological texts.

Researchers have also investigated the use of euphemism that obscures practices harmful to the environment. Shultz (2001) argues that properties of language may be exploited to convey

particular impressions and feelings about our environment and asserts that the language of western society is the “language of commercial users of the environment” (2001, p. 109); she even criticizes activists and environmentalists for using exploitive language even when the negative influence is highlighted. In particular, Shultz notes the use of neutral words, euphemisms, and the labeling of neutral items negatively as three linguistic features that favor commercial uses of the environment (2001, p. 110). In her discussion of euphemism, she highlights the prevalent use of these features within forestry where terms such as *harvest*, *clear*, *value-adding*, *prescribed burning* and others are common. For example, she argues that *harvest* implies the forest/trees are a crop and that trees are removed as a result of “labour justly rewarded” (p. 111). Jung (2001), however, critiques ecolinguistics and what he calls their “preoccupation with euphemisms” and the claim often made that a particular word choice “glosses over or falsifies a fact of reality” (p. 276) and thus limits more ecological language use. He strongly states, “the accusation of euphemizing is therefore about all the expression of self-righteous moralizing intended to immunize against the consideration of differing views” (p. 278), and notes we should be attentive to and critique public discourse but to avoid the accusations of euphemizing.

Corpus-based ecological discourse analysis has also pursued more lexically-oriented investigations of ecological texts. Alexander’s work (2002, 2009) is notable for its integration of corpus methods and CDA to study lexical features of industry/corporate texts dealing with environmental topics. His work generally focuses on strategies of “greenwashing”, a term coined by Plec in which a corporation “conceals environmentally harmful actions with the rhetoric of environmental friendliness to entice and manipulate the consumer” (2012, p. 464). Through text-critical analysis employing concordance programs to highlight salient lexical patterns, Alexander

charges that companies such as British Petroleum (BP), whose actions few would likely consider as environmentally friendly, present themselves positively as stewards, as knowledge sources to help solve the problem, and as leaders in environment friendly actions; i.e. they greenwash their actions. The discourse of BP, and many corporations, as Alexander asserts, is framed through economic interest and displaces oppositional discourses. Corporations position themselves as part of the solution while speaking against regulation. For example, Alexander (2002) analyzes the use of the term *sustainable development* in the 1999 and 2000 reports from Shell Corporation. He focuses on the concept of *sustainable development* and how it is applied and deployed within corporate discourse as part of greenwashing strategies and asserts that corporations have manipulated the term into a state where the concept is essentially meaningless (2002). He argues that ideas such as *sustainable* and *sustainability* have drifted and become normalized into areas of social and business discourse, but he also notes that environmentalists continue to use the term, which has reduced its meaningfulness.

More recently, there has been research employing much larger corpora for identifying meaningful lexical patterns and collocational profiles for environmental terms within particular contexts and across specified periods of time (Augoustinos, Crabb, & Sheperd, 2009; Bevitori, 2011; Grundman & Krishnamurthy, 2010; Grundman & Scott, 2012, Wild et. al. 2013). For example, Augoustinos et. al. (2009) apply a CDA approach to a corpus of texts (editorials, commentaries, opinion, letters to the editor) collected from six British newspapers over a three month period in 2004 referencing the growing controversy surrounding genetically modified food in the United Kingdom (UK). They specifically targeted the constructions and representations of different stakeholders (British public, government, science, bio-tech companies) within the GM debate in UK media and employed a combined CDA and corpus

approach. The article displays how the science of GM farming became contested and how the debate became centered on public trust/mistrust of government rather than rational discussion of empirical findings. Similarly, Bevitori (2011) investigated the construction of the word *green* across various media genres and how its patterning reveals underlying messages and conflicting ideologies across the different media genres, while Grundman & Krishnamurthy (2010) report a corpus-based investigation of keywords surrounding the issue of climate change in popular media from the USA, UK, Germany, and France. Through a similar methodology and for a similar topic, Grundman & Scott (2012) provide an in-depth and comprehensive investigation of terms *global warming*, *greenhouse effect* and *climate change* and their equivalents in French and German in the media of the same four developed countries from 2000-2010.

Also focused on UK discourse through a corpus approach to CDA, Wild, et. al. (2013) analyzes a mega web-corpus compiled through the use of a web-crawler and several small specialized corpora to investigate how the natural environment is represented in public UK discourse. The researchers analyzed collocates of a list of over 100 environmental terms and through integrated manual and automatic analyses and methods from CDA and corpus linguistics (CL) produce what they argue represents a useful methodological approach for enhancing the rigor and validity of corpus-based critical discourse studies. They found evidence for multiple conceptualizations of nature: 1) nature presented as distinct and separate from humans, 2) as interconnected with humans, 3) as a commodity with value. Within their public corpus, there was more focus on problems related to nature (e.g. floods) but the government corpus represented nature as valuable and in need of development. Wild et. al. (2013) is of particular value for its comprehensive, integrated CDA-CL analysis of a massive collection of environmental discourse.

This review has reported several linguistic features which applied linguists within the ecolinguistics program have often analyzed as well as the contributions corpus approaches have made to ecological discourse analysis. Through system-critical to text-critical critiques and analyses of small, specialized corpora and large corpora, researchers have detailed linguistic patterns that often conspire to normalize and perpetuate, as Halliday (1990/2001) asserted, unsustainable relationships between humans and their environment. However, although several of these studies analyzed texts originating in particular countries, e.g. Grundman & Krishnamurthy's comparison of climate discourse in four developed nations, or from particular interest and industry groups, e.g. Alexander's analysis of texts from Shell Corporation (2002) and Gerbig's investigation of the Ozone Corpus, this article will display that greater attention to the discursive production of environment and ecological issues in localized, specific ecological spaces and in connection to specific environmental topics can produce valuable insights into environmental discourse. This article does not investigate an issue within a broad national context nor analyze a global issue such as climate change but instead focuses on a particular environmental controversy in a specific location to enable greater connection between context and language use. Thus, this article analyzes the discursive depictions of a particular localized controversy in the Santa Rita Mountains of southern Arizona, USA surrounding the potential construction of one of the world's largest copper mines to show how two interest groups produce oppositional perspectives towards the environment and the mine. Through a corpus-based analysis of the texts produced by the two groups, the article answers the following research questions.

1. How do keywords present in the corpora of the two interest groups, the Rosemont Mine Company and the Rosemont Mine Truth environmental organization, index oppositional ideological orientations towards the Rosemont Mine and the Santa Rita Mountains?
2. Which part of speech (POS) and semantic features occur at significantly different frequencies between the two interest group corpora?
3. What rhetorical and functional purposes do the constellations of lexicogrammatical patterns achieve and produce in the discourse?

## **Methodology**

### **The Corpus**

The corpus for the present study consists of two small, specialized corpora containing blog posts from Rosemont Mine Truth (RMT) and press releases from Rosemont Copper Company (RCC). The RMT website was developed by the Tucson-based environmental group Save the Scenic Santa Ritas (SSSR) to persuade Arizonans to challenge the proposed copper mine. The second interest group whose texts are included in the corpus is the company, RCC, which plans to construct the mine. Thus, the corpus is organized and investigated according to the interest groups of SSSR and its RMT blog and RCC and their press releases.

While the distinctiveness of internal linguistic features may indeed vary due to the differing genre conventions and constraints, the classification scheme and rationale applied in this study views the purpose of the RCC press releases and RMT blog posts in this context as similar. Both genres, the press release and the blog post, are the mediums of choice by each interest group for their management of public relations. Thus, these texts pursue a rather unified rhetorical purpose of public relations and perception management, as the primary purpose of both groups of texts is to influence public perception of the proposed mine.

RMT produces texts at a greater frequency and generally of greater length, and thus, the RMT corpus is larger in word count and number of texts. As noted in table 1, the corpus of RCC press releases consists of approximately 17,000 words while the RMT blog corpus has slightly over 70,000 words; the tables included in the appendix display the frequency lists of words, parts of speech (POS) items, and semantic tags of each corpus. The word lists were generated using the corpus tool Antconc (Anthony, 2014) and POS and semantic tag frequencies were created through the online corpus tool WMatrix (Rayson; 2008, 2009).

Table 1: Corpora of RCC and RMT

Interest Group	Genre	# of texts	Time Period	Tokens #	Token Types
Rosemont Copper Company	Press Release	33	06/28/ 2007* to Present	16,624	2,463
Rosemont Mine Truth	Blog Posts	130	08/30/2010** to Present	71,619	5237

\*06/2007 was approximately one month before RCC's second mining plan submission.

\*\*Rosemont Mine Truth posted its first blog release on September 30, 2010.

### Keywords and Log Likelihood Analysis

Comparing differences in raw frequencies between words, POS, and semantic tags enables only limited contrastive analysis of corpora. Instead, it can be more insightful to produce quantitative measures of significant difference between items in the corpora under investigation. One such contrastive process is a keyword analysis. A keyword analysis produces a list of words which are generated from a comparative analysis of a node text/corpus to a larger reference corpus (Tribble, 2000). This process determines which words are particularly “key” in the node text/corpus through the comparison to the reference corpus. This process was completed using the corpus tool Antconc (Anthony, 2014).

To extend beyond the analysis of keyword lists, a log likelihood (LL) measure was used to test significance between frequencies of POS and semantic tags. LL is often used within corpus linguistics to test similarities and differences between corpora because it does not assume a normal distribution of data, thereby enabling a comparison of features of differently sized corpora (Rayson and Garside, 2000). Also, while frequencies of individual lexical items may occur at lower numbers within small corpora, POS and semantic tag frequencies appear at higher frequencies and thus, the LL test is less susceptible to overestimating significance (Rayson & Garside, 2000). LL statistics have been used to determine variation in use of interactional pragmatic markers in office hours consultations (Reinhardt, 2013) and has been applied to measure overuse/underuse of particular POS categories within learner corpora as compared to native speaker corpora (Granger and Rayson, 1998). For LL analysis, scores greater than 3.84 or less than -3.84 are deemed significant; these scores produce p-value less than .05 (Reinhardt, 2013). A score greater than 3.84 indicates that the POS or semantic tag was used with a significantly higher frequency in the node corpus. For example, if a LL score of greater than 3.84 is calculated for the semantic tag “geographical terms” when the RMT corpus is analyzed, this indicates that instances of “geographical terms” occur at a significantly higher frequency in the RMT corpus than in the RCC corpus.

Following the quantitative assessments of significant difference between the interest group corpora, qualitative analysis was performed to code the items according to their rhetorical purpose. Codes were applied following analysis of each tags representative tokens and close study of concordance lines and were grouped into *constellations*, multiple lexicogrammatical and semantic features which produce a specific rhetorical purpose within a corpus of texts. For example, a variety of features occur within the RMT corpus at a significantly higher frequency

than they appear within the RCC corpus. These distinctive features were organized into *rhetorical constellations*, i.e. functional groups, which represent particular rhetorical purposes and objectives that they collectively produce. These identified constellations of features are the focus of the subsequent section.

## Results

### Rosemont Mine Truth

The keyword analysis of the RMT corpus reveals several lexical features of interest. As noted in the methods sections, the lexical items present in the RMT keyword list occur at a significantly higher rate than they occur within the RCC corpus. A first item of particular interest is the keyword *proposed* which is used within the RMT texts as a pre-modifier to the nominal *Rosemont Mine*; *proposed* is used 252 times in the interest group texts as opposed to only 10 instances in the company releases. A contrasting word, *project*, is used consistently within the RCC text as part of the noun phrase structure *the Rosemont Mine Project*. Evident in these contrasting keywords are oppositional stances towards the mine. As the official review and approval process is ongoing and no official decision has been reached, *proposed* would be a more accurate depiction than *project*, yet the company, RCC, instead uses a term that creates an image of the decision as a foregone conclusion and the mine as a project that has begun. Additionally, RMT deploys several other items that display a similar stance of incompleteness and uncertainty toward the mine. For instance, the modals *could* and *would* and the subordinators *if* and *whether* also depict the mine and the review process as incomplete and possible rather than certain and decided. Thus, there is evidence within the keyword list of an emerging rhetoric of doubt and uncertainty regarding the mine and its possible construction.

Also noticeable within the keyword list for RMT are the multiple references to geographical spaces and features as well as animals through items such as *jaguar*, *mountain*, *Santa Rita*, *forest*, *Coronado*, and *endangered*. While a reader may reason that the company would place little emphasis on discussions of jaguars and endangered species, it does seem surprising that they would not be referencing the Santa Rita Mountains, the Coronado Forest, and

other southern Arizona locations with much frequency. The keywords indicate that the environmental group extends greater effort in situating their discourse within their community and local environment while also advocating for the protection of species located in the area. While the keywords discussed are revealing, the POS and semantic tag analysis to follow can both confirm and extend initial insights provided through the keyword technique. For example, the keyword analysis compares only individual lexical items and their frequencies between the two corpora. In contrast, the POS and semantic tag analysis will display not only whether individual modals, e.g. *would*, or single lexical items, e.g. *mountains*, differ in frequency but rather if all aggregated modals or geographical terms contrast significantly.

Table 2: Keywords in the Rosemont Mine Truth Corpus

<b>Rank</b>	<b>Token</b>	<b>Freq</b>	<b>Keyness</b>	<b>Rank</b>	<b>Token</b>	<b>Freq</b>	<b>Keyness</b>
<b>1</b>	<i>mine</i>	837	99.29	<b>26</b>	<i>coronado</i>	137	21.99
<b>2</b>	<i>it</i>	439	82.71	<b>27</b>	<i>daily</i>	50	20.87
<b>3</b>	<i>admin</i>	129	53.85	<b>28</b>	<i>army</i>	140	20.53
<b>4</b>	<i>proposed</i>	252	53.66	<b>29</b>	<i>takeover</i>	49	20.46
<b>5</b>	<i>its</i>	449	53.01	<b>30</b>	<i>rk*</i>	48	20.04
<b>6</b>	<i>stated</i>	173	47.35	<b>31</b>	<i>seeking</i>	47	19.62
<b>7</b>	<i>hubbay</i>	301	47.29	<b>32</b>	<i>objections</i>	75	19.44
<b>8</b>	<i>that</i>	1044	46.50	<b>33</b>	<i>bid</i>	46	19.20
<b>9</b>	<i>jaguar</i>	97	40.49	<b>34</b>	<i>would</i>	217	18.64
<b>10</b>	<i>could</i>	129	35.23	<b>35</b>	<i>last</i>	132	18.21
<b>11</b>	<i>mountains</i>	128	34.86	<b>36</b>	<i>because</i>	60	18.18
<b>12</b>	<i>corps</i>	186	34.75	<b>37</b>	<i>regulators</i>	43	17.95
<b>13</b>	<i>santa</i>	154	33.15	<b>38</b>	<i>repeatedly</i>	42	17.53
<b>14</b>	<i>mile</i>	93	31.09	<b>39</b>	<i>until</i>	58	17.41
<b>15</b>	<i>tailings</i>	74	30.89	<b>40</b>	<i>shares</i>	97	17.28
<b>16</b>	<i>augusta</i>	1015	29.00	<b>41</b>	<i>if</i>	110	16.29
<b>17</b>	<i>whether</i>	66	27.55	<b>42</b>	<i>filings</i>	39	16.28
<b>18</b>	<i>investors</i>	64	26.72	<b>43</b>	<i>endangered</i>	53	15.51
<b>19</b>	<i>rita</i>	124	26.15	<b>44</b>	<i>feis</i>	129	15.34
<b>20</b>	<i>report</i>	91	25.35	<b>45</b>	<i>hostile</i>	36	15.03
<b>21</b>	<i>forest</i>	429	23.17	<b>46</b>	<i>shareholders</i>	81	14.91
<b>22</b>	<i>massive</i>	55	22.96	<b>47</b>	<i>issue</i>	97	14.77
<b>23</b>	<i>reported</i>	72	22.83	<b>48</b>	<i>told</i>	35	14.61
<b>24</b>	<i>cash</i>	83	22.38	<b>49</b>	<i>however</i>	50	14.37
<b>25</b>	<i>star</i>	53	22.13	<b>50</b>	<i>shareholder</i>	50	14.37

\* *rk* is an abbreviation for Red Kite, an London-based financial group.

The log likelihood analysis of the variation between semantic and grammatical tags of the RMT and RCC corpora revealed 72 out of approximately 200 possible tags occurred at a significantly higher frequency within the environmental texts than in the industry texts. Of these 72 items, closer analysis of the collocational patterns and concordance lines of these items revealed several distinct rhetorical constellations patterning through the discourse from the environmental group. The first of these constellations, and the most dominant within the data,

was labeled as uncertainty/doubt. In this constellation, 23 items present in this rhetorical constellation commonly occur to sow and produce a rhetoric of doubt and uncertainty concerning the construction of the mine, the stability of the company's finances, and the potential for harm to result from the mine's development. While on the surface some of these items may not seem inherently tied to a rhetoric of uncertainty and doubt, closer inspection of concordance lines and texts (see table 5) reveals their rhetorical function in the discourse. For example, key to the group's rhetoric is producing doubt in the stability of RCC finances. Thus, semantic features of *money and pay*, *success and failure*, *evaluation: bad*, *evaluation: false*, and *qualities: little* all collocate in patterns that question the strength of RCC's financial standing. These features and examples for each are displayed in table 3.

Table 3: Rhetorical Constellation of Uncertainty/Doubt in Rosemont Mine Truth Corpus

Tag	Tag Feature	RMT	RCC	LL	Examples
I1.1	money and pay	1533	209	60.74	<i>Augusta, also funds, invest, investors, investing, shareholder, asset, capital, taxpayer</i>
A7+	likely	784	95	43.10	<i>may, could, potentially, can, would, clearly, may, likely, confirmed, potential, certainly</i>
CS	Sub conj	371	41	25.14	<i>if, when, whether</i>
Z6	negative	424	54	20.49	<i>not, no way, no</i>
X9.2-	success and failure	97	4	20.33	<i>failed, failure, fails, unsuccessful, lost, defeats</i>
N5.1-	part	208	19	19.88	<i>shares (investment), section 404, part, run out, incomplete, devoid, segment</i>
Z7	if	112	8	14.83	<i>if, whether or not</i>
A5.2-	evaluation: false	71	3	14.66	<i>lie, false, misleading, unfounded, falsely</i>
XX	not, n't	321	42	14.39	<i>negation</i>
Z4	discourse bin	104	10	9.15	<i>rather, in other words, however, in this case, on the other hand, at best, obviously</i>
A5.1-	evaluation: bad	44	2	8.66	<i>severe, flawed, adversely, bad, severity, shortcomings, terrible, detrimental</i>
S1.2.1 -	formal/unfriendly	52	3	8.51	<i>formal review, takeover, hostile (referencing hudson takeover)</i>
X2.1	thought/belief	161	20	8.31	<i>wonders, considers, presumably, outlook, presumed</i>
S8-	hindering	105	11	7.92	<i>disruption, blow, adverse, oppose, prevent, frustrate, obstacles, setbacks</i>
N5-	quantities: little	136	17	6.91	<i>slight, deficit, inadequate, reduce, insufficient, decreased, underestimate, depletion, scant, underestimated</i>
S7.4+	allowed	690	126	6.76	<i>permit, permitting, permits, allow, approve</i>
VDD	did	38	2	6.72	<i>all 38 are 'did not' accusations against the company</i>
VDZ	does	60	5	6.56	<i>58 of 60 instances "does not" in relation to RCC acts/shortcomings</i>
A12-	difficult	65	6	6.11	<i>problems, challenges, crisis, difficult, complexity, problematic, complicate</i>
X7-	unwanted	43	3	5.84	<i>reject, unsolicited, unacceptable, ill-conceived, rejected</i>
RRQ	wh- general	76	8	5.68	<i>primarily rhetorical questions</i>

T4-	time: late	73	8	5.03	<i>delay, delayed, delays, delaying, late</i>
-----	------------	----	---	------	---

A second constellation consists of 13 items whose confluence produces a rhetoric of aesthetic value and environmental stewardship. The environmental texts, with much greater frequency than the mining company, reference living creatures and plants present in the Santa Rita Mountains while also referring to a multitude of geographical spaces such as *wetlands*, *ponds*, and *canyons* within their discourse. This density with which geographical terms and living creatures are referenced displays an intimacy with the land and the mountains and forwards a rhetoric of stewardship. The group repeatedly deploys semantic evaluations of beauty and importance while impressing the importance of stewardship of the mountains and their inhabitants.

Table 4: Rhetorical Constellation of Aesthetics/Stewardship in Rosemont Mine Truth Corpus

<b>Tag</b>	<b>Tag Feature</b>	<b>RMT</b>	<b>RCC</b>	<b>LL</b>	<b>Examples</b>
L2	living creatures	277	11	59.40	<i>wildlife, fish, endangered species, insects, birds, animals, prey, jaguar, ocelot, amphibians, invertebrates, frog, cuckoo, bald eagles, predator</i>
A1.1.2	damaging and destroying	132	3	36.58	<i>damage, destruction, destroying, obliterate, charred, violate, smash, harm, irreparable, collapse, broken, rips, devastating</i>
W3	geographical terms	543	65	30.75	<i>aquifer, mountains, creek, desert, land, stream, wetland, pond, canyon, lake, grasslands, slope, wells, valley, tributary, river, forest, mount, butte, crest</i>
O4.2+	judgment of appearance: beautiful	146	10	20.24	<i>attractive, clean, scenic, appealing, spectacular, lush</i>
A11.1+	important	334	40	18.89	<i>meaningful, key, serious, significant, well known, major, value, integral, substantive, urgency, crucial</i>
E3-	violent/angry	123	8	17.99	<i>threatens, toxic, threatened, poisonous, acrimonious, poison, harass</i>
B2-	disease	34	1	8.51	<i>contaminate, contamination, high blood pressure, Cancer, lesions, syndrome, fractures (side effects of the mine toxins)</i>
L3	plants	57	4	7.70	<i>trees, willows, cottonwoods, vegetation, mesquites, (hedge fund), grass, sycamore (smelters), saguaro</i>
APPG E	possessive pronoun	617	110	7.18	<i>its, our</i>
H4	residence	107	13	5.84	<i>residents, lived, habitat, occupying, accommodate</i>
K5.1	sports	57	5	5.79	<i>anglers, hunters, fishing (bats/animal, final/decision, exercise/decision)</i>
S6+	strong obligation/necessity	277	45	5.53	<i>have to, promises, essential, needed, should, must, need</i>
O1.2	substances: liquid	344	59	5.12	<i>water, waters, waterways, seepage,</i>

These keyword, POS, and semantic tag features and the rhetorical effects they produce can be verified by a close analysis of a RMT text in table 5. This text displays the density of geographical features also reflected in the semantic tag analysis as multiple references to places implicated in the debate are included. From naming the national forest, orienting the reader to the mine's location in relation to Tucson, mentions of the Santa Rita Mountains, as well as particular geographical features such as *canyons*, the text emphasizes the geographical location of the proposed mine in an effort to create a bond and relationship between the mine, the mountain, and the reader. In addition, the importance of protecting the habitat of the endangered male jaguar is repeatedly referenced as the animal is referred to as the *only known jaguar*, *male jaguar*, *endangered jaguar*, *jaguar recovery*, *photos of the jaguar*, and once again *only known jaguar*. The keyword and its collocational windows imply the importance of the animal and impresses our duty to protect it. Also evident is the sowing of doubt noted in previous discussion as the keyword *proposed*, modal *would*, and subordinate *if* are clearly on display in contexts implying a decision regarding the mine has not yet been reached.

Table 5: A Sample RMT Text

Rosemont Copper mine's fate may hinge on **Jaguar**  
January 24, 2013

The New York Times published a story and photograph today on the impact of the nation's only known **jaguar** and the **possible** development of the massive Rosemont open pit copper mine on the **Coronado National Forest southeast of Tucson**.

A **male jaguar** has been photographed several times in close proximity to the **proposed** copper mine that Vancouver, B.C.-based Augusta Resource Corporation is seeking state and federal permits to build through its U.S. subsidiary, Rosemont Copper Company.

The appearance of the **jaguar** near the mine site has now given the U.S. Fish & Wildlife Service a critical role in the ultimate fate of the copper project. The Service is in the process of establishing critical habitat for the **endangered jaguar**.

**If** the Fish & Wildlife Service includes the **proposed** Rosemont mine site as critical habitat, the copper project will be dealt a severe blow because the mine intends to destroy 3,000 acres of national forest with 70-story high piles of waste rocks and mine tailings.

"That area is key for **jaguar** recovery," Randy Serraglio of Center for Biological Diversity told the Times. "**If** we follow the law, as laid out under the Endangered Species Act, they will come back. They always lived there, until the U.S. government either killed them or chased them out." Mr. Serraglio told the Times that a series of photos of the **jaguar** taken this year in close proximity to the **proposed** mine site should be "the nail in the coffin of Rosemont Mine," adding, "There's no way the mine **proposal** can go forward **if** this land is designated as critical habitat." The Times reported that the public comment period for the Fish and Wildlife Service's critical habitat **proposal** published in August will reopen this spring. A final critical habitat determination is expected by August.

**If** the **northern Santa Ritas** are designated as critical habitat, it **would** put into play the legal concept of "adverse modification" — which **would** prohibit the federal government from approving a project on public lands that **would** adversely modify critical habitat. While the mile-wide, half-mile deep copper pit **would** be on private land, the mine **would** use public land as a massive dump site, **destroying canyons, streams, springs** and other **habitat** that is now being frequented by the nation's only known **jaguar** (another jaguar was euthanized after being captured in southern Arizona in 2009.)

The Rosemont project **would** be among the top five largest copper mines in the US. It **would** be located in the middle of an intersection of three major wildlife corridors in the **proposed jaguar** critical habitat with visual evidence of a **jaguar** occupying the area. These factors combine to present a significant challenge that Rosemont will find extremely difficult to overcome.

## Rosemont Copper Company

Most obvious from the RCC keyword list are the multiple business and stock market related acronyms present on the list. As their press releases are generally intended for an audience of investors whose primary interest is return on investment, it is not surprising that items such as the New York Stock Exchange (NYSE), the American Stock Exchange (AMEX) the Toronto Stock Exchange (TSX), and their own stock market abbreviation, AZC, would feature prominently within their output. However, whereas the RMT keywords list displayed a number of geographical terms anchoring their discourse to the local environment, the RCC list does not present similar geographic references; the only place name on the list is Denver, the US headquarters of RCC's parent company. Additionally, the RMT list was populated with modals and if/whether subordinators reflecting a rhetoric of doubt and certainty, but RCC instead reflects a more assertive and decisive tone indicating a sense of inevitability regarding the construction of the mine. For RCC, they seem *pleased* with the *process* and their *project is advancing forward*. And as referenced previously, in contexts where RMT used the premodifier *proposed*, RCC instead uses *project*. Interestingly, while RCC does not refer to the mountains where the mine will possibly be built, they do refer to the *community*, its *students*, and the *scholarships* they are providing. It seems that while RMT texts create connections to the mountain RCC texts instead produce connections to human groups and institutions, e.g. *schools*, *chamber of commerce*, *agencies*, *district*, and *foundation*. The RCC texts produce relationships with business and community interests rather than with the land, the mountains, and animals.

Table 6: Keywords in the Rosemont Copper Company Corpus

Rank	Freq	Keyness	Token	Rank	Freq	Keyness	Token
1	26	86.801	<i>nyse</i>	26	12	40.062	<i>mkt</i>
2	29	78.156	<i>program</i>	27	12	40.062	<i>students</i>
3	569	76.787	<i>and</i>	28	16	38.094	<i>looking</i>
4	23	76.785	<i>pleased</i>	29	32	37.903	<i>agencies</i>
5	38	76.764	<i>usfs</i>	30	55	37.798	<i>we</i>
6	33	70.283	<i>azc</i>	31	24	37.578	<i>forward</i>
7	27	67.968	<i>school</i>	32	11	36.723	<i>chamber</i>
8	32	67.433	<i>com</i>	33	11	36.723	<i>denver</i>
9	24	66.857	<i>tsx</i>	34	13	36.613	<i>foundation</i>
10	25	61.863	<i>molybdenum</i>	35	36	36.384	<i>local</i>
11	30	61.776	<i>district</i>	36	14	35.517	<i>announce</i>
12	18	60.093	<i>amex</i>	37	16	35.07	<i>team</i>
13	18	60.093	<i>augustaresource</i>	38	12	33.428	<i>robust</i>
14	26	55.617	<i>acoe</i>	39	10	33.385	<i>drill</i>
15	16	53.416	<i>symbol</i>	40	10	33.385	<i>producer</i>
16	61	52.728	<i>process</i>	41	10	33.385	<i>solid</i>
17	25	50.247	<i>www</i>	42	10	33.385	<i>tier</i>
18	20	46.813	<i>deposit</i>	43	37	32.715	<i>eis</i>
19	129	44.157	<i>project</i>	44	29	32.69	<i>mineral</i>
20	21	44.009	<i>supply</i>	45	31	32.489	<i>us</i>
21	13	43.4	<i>advancing</i>	46	16	32.451	<i>drilling</i>
22	13	43.4	<i>scholarships</i>	47	215	32.044	<i>copper</i>
23	18	40.901	<i>results</i>	48	32	30.514	<i>president</i>
24	26	40.85	<i>community</i>	49	11	30.257	<i>income</i>
25	84	40.451	<i>or</i>	50	11	30.257	<i>reserve</i>

The LL analysis of POS and semantic tags within the RCC texts revealed contrasting constellations from the RMT texts. While RMT texts were marked by frequencies of features producing a rhetoric of uncertainty and stewardship, the first constellation within RCC texts reflects certainty and authority. For example, the semantic tag at the top of the RCC list displays this certainty/authority about the copper mine. Coded as *general actions/making*, the list of example items for this feature include the previously mentioned *project* in addition to other items

projecting certainty towards the building of the mine, e.g. *production, operating, pursue, construct, done, installed, implementation, and drilled*. The examples of the second tag item also present the group's authority as they repeatedly deploy the phrases *we are, we will, we have, we continue, and we know*. Other tags from the list (see Table 7) are *toughness: strong/weak, comparing: similar, comparing: usual, expected, success*, as well as *future*. These tags and their representative tokens display RCC's authority towards the mine and what they present as its imminent construction. The future-oriented nature of their messages as well as the *science and technology* features reflect a rhetoric of authority and inevitability.

Table 7: Rhetorical Constellation of Certainty/Authority in Rosemont Copper Company

Tag Code	Tag Description	Tag Freq in RCC	Tag Freq in RMT	LL	Examples
A1.1.1	general actions/ making	464	1216	75.10	<i>project, to close, closing, production, to create, operating, made, pursue, construct, closure, committed, done, operations, installed, drill, drilled, drilling, implementation, projects, conducted</i>
PPIS2	1st person plural pronoun	53	66	40.64	<i>We are, we will, we have, we expect, we continue, we know</i>
S1.2.5 +	toughness: strong/weak	24	12	39.13	<i>robust, robustness, strength, strict</i>
A6.1+	comparing: similar	52	100	19.82	<i>agreement, consistent, agreed, compliant, in keeping with, agree, consistently, compliance, compatibility</i>
A6.2+	comparing: usual	37	64	17.35	<i>common, customary, normal, basic, generally</i>
A1.8+	inclusion	81	207	14.34	<i>include, including, includes, contains, consist, consists, comprehensive</i>
X2.6+	expected	65	157	13.85	<i>expected, expects, due, anticipates, look forward</i>
M7	places	151	470	11.05	<i>international, regional, area, local, province, site, district</i>
X7+	wanted	180	577	10.99	<i>required, strategy, policy, schedule, plan, target</i>
X9.2+	success	20	37	8.25	<i>achievement, success, on track, win, successfully</i>
N5.1+	entire, maximum	77	225	7.88	<i>maximum, all, 100%, complete, total, maximizing, thorough, every, full, as a whole, to the full</i>
VBN	been	38	102	5.62	<i>been completed, been submitted, been notified</i>
S5+	belonging to a group	138	481	4.41	<i>federal, organization, institutions, joint, public, team</i>
T2+	time: beginning	59	184	4.27	<i>starting, commencement, foundation, beginning</i>
T1.1.3	future	153	543	4.10	<i>will, going to, future</i>

A second constellation within the RCC corpus was identified as marking a discourse of dominion and economics; these features are listed in table 7. These items display results of the mine's construction and the economic value the mine will bring to the region. For example, the tag *education* displays the density of references to the positive effects the mine will bring to the area's universities, schools, and students. The construction of the mine and the environmental degradation it will undoubtedly cause is removed by the economic results the mine will engender. The next tag on the list identifies the economic value to be generated followed by a list of items that the mine will extract from the Rosemont site. The discourse from the industry group first notes the many minerals that can be removed from the earth but interestingly follows with semantic items that reflect enhancements, developments, and positives from the mine. Not surprisingly the list continues with tagged features noting the *jobs*, *careers*, *staff*, and *workforce* that will be a result of the mine and the technical expertise the company possesses for successful exploitation of the mountain. It is also valuable to note the differing nature of place references between the two groups. The place name mentions within the RCC texts are overwhelmingly financial centers of London, New York City, Frankfurt, Shanghai and Sydney, but there are far fewer mentions of local geographical places and terms similar to those in the RMT corpus. The anchoring of the RT texts to the local community is contrasted to the positioning of the RCC texts within global financial centers and the authority and power held by these locations.

Table 8: Rhetorical Constellation of Dominion/Economics in Rosemont Copper Company

Tag Code	Tag Description	Tag Freq in RCC	Tag Freq in RMT	LL	Examples
P1	education	108	89	125.73	<i>school, education, students, Community college, university, Ph.D., study, alumni, scholars, educate</i>
FO	formula	71	82	59.53	<i>US \$40, 100%-owned, 4.5%, 1.1 billion</i>
O1.1	substances and materials	236	547	56.87	<i>copper, metals, silver, gold, deposit, diamond, ore</i>
A5.1+	evaluation: good	58	89	33.28	<i>enhances, enhanced, merit, defensible, developed, positive, favorable, great, progress, world class, improve, positive, defensible</i>
I3.1	work and employment	67	158	15.35	<i>work, role, teamwork, jobs, career, roles, employ, payroll, staff, personnel, working, workforce</i>
I2.1	business	185	564	15.23	<i>corporation, company, offices, offices, stock exchange, infrastructure, Ltd, businesses</i>
Z2	geographical names	252	845	11.18	<i>Vancouver, Bermuda, New York, Shanghai, Sydney, Canada, Quebec, United States, San Francisco, Arizona, Denver, University of Arizona, Arizona State University, Pima County, Nevada, Venezuela, Ecuador, Vancouver, Toronto, Frankfurt, London, Washington DC, South America, Davidson Canyon, Ontario</i>
A1.5.1	using	34	71	10.77	<i>used, using, consuming, utilize</i>
Y1	science and technology	60	163	8.46	<i>engineers, technical, technologies, hydrology, engineering, state-of-the-art, scientific</i>

There is, however, another difference between the corpora that is worthy of discussion.

Within the RMT constellations, there are more features whose confluence produce the effects of uncertainty and stewardship than there are features which convey certainty and dominion in the

RCC corpus. One may assume that the additional features are a result of the larger size of the RMT corpus than the RCC corpus; however, another explanation is possible. It is unquestioned that RCC held the upper hand and much of the power through the years of the debate, and although SSSR has fought against the mine, the law, the review process, and much of the state and local government is in favor of and supports the mine's construction. From this position of power and authority, it was not necessary or even wise for RCC to engage in extensive public debate. Thus, the relative absence of markers of interest may reflect a shrewd public relations tactic to proceed with caution and to avoid any rhetorical risk. Thus, length of press releases, infrequency of press releases, and lack of distinct and revealing semantic fields may also index the company's powerful stance within the debate.

### **Discussion**

The keyword lists and POS and semantic analysis clearly reflect the oppositional stances of RCC and RMT to the mine, the mountain, and the environment. For RMT, their keyword list is dominated by items reflecting an intimacy and knowledge of the land as well as numerous items forwarding the uncertainty of the mine's construction. The organization repeatedly references the wildlife of the region, often mentions particular geographical features, and deploys modals such as *if* and *whether* to create doubt and elicit concerns for the well-being of the mountains and its many inhabitants. In contrast, the keyword list from the mining company indexes rather different relationships towards the mine and the mountain. In fact, few connections and references are made to the local flora and fauna or to geographical features of the region. Instead, the company speaks of various international stock exchanges and financial centers and the economic value and opportunities that will be generated through the exploitation of the mine site. Quite clearly, the keyword list from the environmental group indexes a

relationship of stewardship and knowledge of the land while the company keywords reflect a stance that views the mine and the mountain as an object for exploitation and economic gain.

The POS and semantic tag analysis, which combined multiple lexical items into their POS and semantic categories for significance testing, affirmed many of the patterns noted in the keyword lists from the two groups. Within the list of POS and semantic features that occurred at a significantly higher rate within the environmental group's texts, constellations of features producing similar rhetorical effects were evident. These high frequency items conspired to produce a rhetoric of uncertainty/doubt while also forwarding a message of stewardship/aesthetics for the land and its many inhabitants. However, while the environmental group with great frequency referred to the beauty of the land and the potential harm that could come to the environment, the industry group made few mentions to the environmental concerns as they instead asserted their expertise and authority to successfully extract copper and other minerals from the mountain. The company chose not to discuss the endangered jaguar, the rare ocelot, or to situate their message within the region, but instead they discuss the jobs that will be produced and the revenue that will be generated while locating their message within global financial and political centers and the power afforded by these locations.

As Nash and Mühlhäusler (2013) assert, the challenge is to “create functional interconnections between philosophical and empirical approaches to ecolinguistics and to apply such an integrated approach to practical problems faced by users of language” (p. 8). To apply this challenge to the issue of the Rosemont Mine debate requires not only a revelation of the discursive practices of the two groups but a statement of the implications such findings have for environmental communication. However, proclamations of how the environmental group may subvert the dominant discourse of development and dominion are offered cautiously. In many

ways, the environmental group has had great success in its campaign against the mine and are seemingly more challenged by legal and regulatory realities than any rhetorical shortcomings of their messages. They draw power and strength from their intimacy and connection to the land but could display greater assertiveness in their message. As noted, the group is keen to deploy modals that create uncertainty and doubt and forecast the potential devastation that could result from the mine's construction. The group could instead opt for more assertive modality structures that display the destruction that *will* occur rather than damage that *could* occur. The risk though is that adoption of too many of the rhetorical strategies and sources of power that are also employed and accessed by the industry group may make their messages indistinguishable from their counterparts. For example, if the environmental group were to more frequently reference international financial centers and political capitals, perhaps their message would then disillusion and alienate their audience and create a distance between their organization and the community they represent. However, by highlighting the rhetorical patterns at work in this environmental debate, the article hopes to produce insights that will benefit language users as they face their own "practical problems" (Nash & Mühlhäusler, 2013, p. 8).

**Limitations.** The primary limitation of the study is the size of the interest group corpora. Although they are composed of all press releases from RCC and all blog posts from RMT during the defined period, the size of the corpora were small, and the power of the findings would be strengthened if the corpora were larger. The combining of POS and semantic tags for LL analysis helps address this concern, but the issue is still present. Also, although the texts were the most frequently produced by the groups and were their primary means of public relations, the genres of press release and blog post are distinct and there is a possibility that high or low frequencies of particular grammatical or semantic features is a reflection of the genre affordances and

constraints rather than an underlying rhetorical agenda. The selection of items for the rhetorical constellations, however, did not include all items of significance because some items, e.g. past tense verbs in blog releases, were deemed to reflect genre variation rather than rhetorical distinctiveness. Finally, the company press release is mostly directed to an audience of investors while the blog posts are written primarily for citizens of Arizona. Thus, while the research compiled and analyzed the most appropriate and accessible textual data for the purposes of the study, these limitations were understood and efforts to mitigate their effects were taken when possible.

**Future research.** Future research that constructs larger corpora of environmental texts and performs additional statistical tests of significance, e.g. Wilcoxon's rank sum test or bootstrap tests as recommended by Lijffft et. al. (2014), for testing relationships between POS and semantic tags will be of value. In addition, while a defined third interest group who regularly produced a particular genre was not present in the current debate, an investigation of a more national environmental issue in which multiple groups produced texts would be interesting for comparison purposes. For instance, a larger corpus composed of texts from multiple interest groups could enable greater triangulation of the findings. As these patterns from the two sharply oppositional groups likely exist on the poles of the debate, further research that includes other groups, e.g. government agencies, could provide further insights.

### **Conclusion**

This corpus-based ecolinguistics discourse analysis revealed keywords and rhetorical constellations within the interest group texts that reflect stances and relationships to the mine and the environment. For RCC, the theme pervasive in their productions is one of certainty as they project confidence to their audience of investors that although the legal process for approval

required by the National Environmental Policy Act (NEPA) is yet to be resolved, approval of the mine is indeed inevitable and forthcoming. In contrast, and to perhaps balance the overt confidence of RCC, RMT produces messages that emphasize that the outcome is far from decided. Thus, this study identified the lexicogrammatical items key to the messages of these two entities and how these frequently co-occurring items index particular ideological positions and relationships to the environment. However, the study moves beyond keyword analysis to a more comprehensive and detailed investigation of the functional lexicogrammatical patterns in which these indexical items occur and the patterns in which they frequently colligate. These constellations of grammatical and semantic features construe and produce the functional effects of inevitability, authority, dominion, and certainty in the RCC corpus while an oppositional discourse of sustainability and stewardship manifests in the RMT corpus.

## **INTEGRATING GIS AND CORPUS LINGUISTICS FOR THE ANALYSIS OF ENVIRONMENTAL DISCOURSE**

### **Abstract**

This article details an integration of geographical information system (GIS) and corpus linguistics (CL) techniques with an ecolinguistics-informed interpretative framework for the analysis of a contentious, environmental debate in southern Arizona. The application of GIS and CL procedures enabled the mapping of place name mentions present within two interest group corpora as well as the frequency of particular semantic tags and semantic tag sets that co-occur with specific places prominent in the debate. This approach to the study of environmental discourse highlights the discursive representation of place and the importance that references to places serve within environmental debates. The findings and the GIS visualizations exhibit how different interest groups refer to and represent geographical places within their discourse and how these references to places index ideological positions towards the environment. The article will reports the GIS-CL approach, displays and interprets the findings and visualizations, and explains implications for further research in environmental discourse studies.

# **Integrating GIS, Ecology, and Corpus Linguistics for the Analysis of Environmental Discourse**

## **Introduction**

In recent years, corpus-based discourse analysis has proliferated with attention directed toward a variety of social issues. A few areas receiving attention are political discourse (Conoscenti, 2011; Ferrari, 2010; Milizia, 2009, 2010, 2012; Partington, 2003; Partington & Morley, 2004), social issues such as immigration (Blinder & Allen, 2014; Fitzsimmons-Doolan, 2009; Morley & Taylor, 2012), representation of Muslims in British newspapers and tabloid press (Baker, 2010; Gabrielatos, McEnery, & Baker, 2013), the language of extremism (Prentice, Rayson, & Taylor, 2012), the discourse of news reporting of war (Kutter, 2012), the emerging area of forensic linguistics (Tabbert, 2013), and the discourse of health and hospital settings (Trigg, 2007; Maci, 2012a & 2012b, Staples & Biber, 2014).

During this period, corpus-based discourse analytic approaches to the study of environmental discourse have also received attention. This work may broadly be placed within ecolinguistics which has been defined generally as “the study of the interrelationships between language and its environment” (Couto, 2013, p. 1) and more specifically as “the study of the impact of language on the life-sustaining relationships among humans, other organisms, and the physical environment” (Alexander & Stibbe, 2013, p. 2). Under these broad definitions of ecolinguistics, Bang and Trampe (2013) identify two general approaches: 1) explication of “theories of language inspired by a holistic paradigm of ecology” (p.1) and 2) the application of linguistic analysis for the investigation of ecological texts and discourses.

This second concentration, often referred to as ecological discourse analysis (EDA), was largely prompted by M.A.K. Halliday and his seminal keynote to the AILA convention in 1991

in which he famously asserted that the environmental crisis is not a problem for only the biologists and physicists to engage with but that applied linguists also have a role to play (Halliday, 2001). This challenge to applied linguists prompted greater attention to the features of language that normalize and reproduce dominant and unsustainable conceptualizations of the relationships between humans and their environments. As Harré, Brockmeir, and Mühlhäusler (1999) asserted and Bang and Trampe (2013) later echoed, environmental studies took a “linguistic turn” (Harré, Brockmeir, & Mühlhäusler, 1999. p. 3), as ecolinguistics “adopted the view that perceptions of nature are mediated through language and that in turn such perceptions and lifestyles feed back into the structure of discourse” (Mülhäusler, 2003, p. 12). Thus, in the past 25 years, ecological discourse analyses studying various features of language and their functioning in the discursive production of environment and environmental crises have proliferated.

Ecolinguistics research has investigated particular discourses from select interest groups, e.g. a corpus-based study of interest groups discussing Ozone depletion (Gerbig, 2001) and have integrated critical discourse analysis (CDA) with corpus methods (Alexander, 2002, 2009) for the study of corporate “greenwashing” in which a corporation “conceals environmentally harmful actions with the rhetoric of environmental friendliness to entice and manipulate the consumer” (Plec, 2012, p. 464). While the corpora analyzed in Gerbig (2001) and Alexander (2002, 2009) may be characterized as small and specialized, i.e. “delimited by a specific register, discourse domain, or subject matter (Beaugrande, 1996, p. 11) there has also been work employing much larger corpora for analyzing the discursive production of the environment and environmental issues. For example, Augoustinos, Crabb, and Sheperd (2009) applied a CDA approach to a corpus of articles (editorials, commentaries, opinion, letters to the editor) collected

from six British newspapers over a three month period in 2004 referencing the growing controversy surrounding genetically modified food in the United Kingdom (UK). Similarly, Bevitori (2011) investigated the construction of the word *green* across various media genres and how its patterning reveals underlying messages and conflicting ideologies across the different media genres. Also in 2010, Grundman and Krishnamurthy report a corpus-based investigation of keywords surrounding the issue of climate change in popular media from the United States, United Kingdom, Germany, and France. Through a similar methodology and for a similar topic, Grundman and Scott (2012) investigated the terms *global warming*, *greenhouse effect* and *climate change* and their equivalents in French and German in the media of four developed countries (USA, UK, France, & Germany) from 2000-2010. Also focused on UK discourse through a corpus approach but unique for its compilation of mega web-corpus, Wild, Church, McCarthy, and Burgess (2013) analyzed a corpus compiled through the use of a web-crawler as well as several small specialized corpora to investigate how the natural environment is represented in public UK discourse.

However, while corpus approaches to the study of environmental issues have often been situated within particular national contexts (e.g. UK discourse in Wild et. al. (2013), corpus-based research into the discursive production of place, i.e. how places are discussed, referenced, and represented, has received less attention. The following section reports several studies in which the focus of inquiry is the discursive production of an actual physical location, albeit without applying corpus methods.

### **Studies of Place**

Investigations into the discursive production of place have been an area of interest by previous researchers. Perhaps most informing to the present work though are Haig (2001),

Carbaugh (2001) and Carbaugh and Rudnick (2006) which can all be situated within the ecolinguistics paradigm. While not strictly concerned with place, Haig (2001) noted the frequent use of “location circumstantials” (p. 254) within his study of texts from a Japanese Whaling Research Institute and Greenpeace and asserted the institute and environmental group’s texts were especially concerned with “pinning down their accounts” of the confrontation between the whaling ship and the environmental group within a specific “time and space” (p. 254) While Haig (2001) applied close textual analysis informed by systemic functional grammar to the two opposing texts, Carbaugh (2001) and Carbaugh and Rudnick (2006) compiled their data through ethnographic methods of observations and interviews. In the former study, Carbaugh (2001) discussed the “dueling depictions” (p. 125) of nature present in the debate surrounding a project in eastern Massachusetts that aimed to build ski facilities and condominiums near Mt. Greylock, the tallest peak in Massachusetts. The variation in geographical referencing produced various motives for acting in relation to the environment and revealed several practices which, Carbaugh claimed, indexed differing stances towards the proposal. He asserted, “[w]hen one makes a reference to this land, in some communicative context, a selection is made from a set of terms” (p. 139) with selections such as “the mountain” attaching an ecological aesthetic while “the project” highlights meanings of economic value and interest. For Carbaugh, “[c]ommunicating about natural space is a way of anchoring messages in physical space, but in so doing, more than a physical space is involved” (p. 139) as “the communicative process of depicting nature carries with it a potent complex of socio-cultural messages” (p.139).

In the second investigation, Carbaugh and Rudnick (2006) compared and analyzed the tour presentations of park guides and native American Blackfoot guides in Glacier National Park. The analysts noted the place names used in the talks and how a reference to either “the park” or

“the reservation” indexed views of the storyteller and situated the audience to the area. The native Blackfoot guides provided an elaborate history of ancestors, contact with settlers, and the many difficulties faced as they forwarded a situated history that references historical, spiritual, and sacred events which transpired at specific places. In contrast, the non-native guides presented a discourse of discovery and development as features are referenced from a settler point of view and with little mention of the region’s history before the park’s establishment. Powerfully, the authors assert that “to name a place, or to refer to a place, is to make a move in a cultural political game” (p. 182) and that “how one refers to a place produces a particular view of the place” (p. 183).

These place-based studies and their discussion of the importance of place in environmental discourse are insightful for their detailing of how referring to or naming particular places invokes and indexes particular relationships and modes of action towards the environment. In the next section, I will highlight how this area of research may be advanced through an integration of GIS and corpus linguistics procedures.

### **Integrating CL and GIS for the Study of Place**

Geographical Information Systems (GIS), a process of assigning latitude and longitude coordinates (also referred to as *georeferencing*) to data points for subsequent projection onto a map, allows users to “discover relationships that make a complex world more immediately understandable by visually detecting spatial patterns that remain hidden in texts and tables” (Bodenheimer, Corrigan, and Harris, 2010, vii). Through a process of spatialization in which non-geographic data is converted to spatial forms for visual analysis, texts can be transformed to “tell stories” (Yuan, 2010, p. 111). Through mapping, data can be visualized and a variety of features can be analyzed to reveal “potential geographic influences” (p. 113) on a text. GIS

enables different and additional analyses “to discover the meaning of text and reveal implications and significance of places to determine proper geographic mapping of these places” (p.116). Recently, GIS has received increased attention from the humanities, for although humanities scholars have often applied metaphors of space, they have been less keen to engage with quantitative data (Bodenheimer et. al, 2010). However, in this era of big data, humanities research is more often employing quantitative and computational methods to complement its traditionally hermeneutic and interpretive approaches.

Corpus linguistics (CL) and its affinity for data is a well-suited partner for the type of GIS-informed textual analysis discussed by Yuan. Traditionally though, CL has been concerned with ‘what’ is in the text/corpus (Gregory & Hardie, 2011) as researchers analyze corpora through a variety of semantic and syntactic analyses to interpret meaning and function of language. However, CL has recently begun to explore GIS techniques for the study of discourse (e.g. Gregory & Hardie, 2011; Grover, et. al. 2010). Gregory, noted for his work in spatial humanities, and Hardie, well-known within corpus linguistics, contend that CL has ignored spatial elements within texts while GIS has not attended to the qualitative textual data, which is the most common object of inquiry within the humanities (Gregory & Hardie, 2011). In their work, they develop and display a method for mapping place name mentions extracted from a corpus of historical texts but also how the semantic representations of these places may also be spatialized for visualization.

Informed by the methodology of Gregory and Hardie (2011), this study produces a corpus-based ecological discourse analysis integrated with and enhanced by GIS to enable intuitive, context-sensitive, spatial representations of environmental discourse. This study views space not simply as “a physical backdrop, container, or stage to human life”; rather, space is

conceived “as a complex social formation, part of a dynamic process” (Soja, 2001, as cited in Ayers, 2010, p.1). Drawing from previously discussed place-based work yet extending through the integration of CL and GIS, this corpus-based study investigates what places are referred to within the contentious debate but also how these places are discussed and referenced. Through the production of a complex of visualizations, this study displays the relevance of place name mentions within environmental discourse but also how the semantic representations of places index relationships to the environment. The article will show the potential for integrating GIS and CL techniques for studies of place in environmental discourse while also highlighting potential for further GIS-CL applications is corpus design and place-based pedagogy.

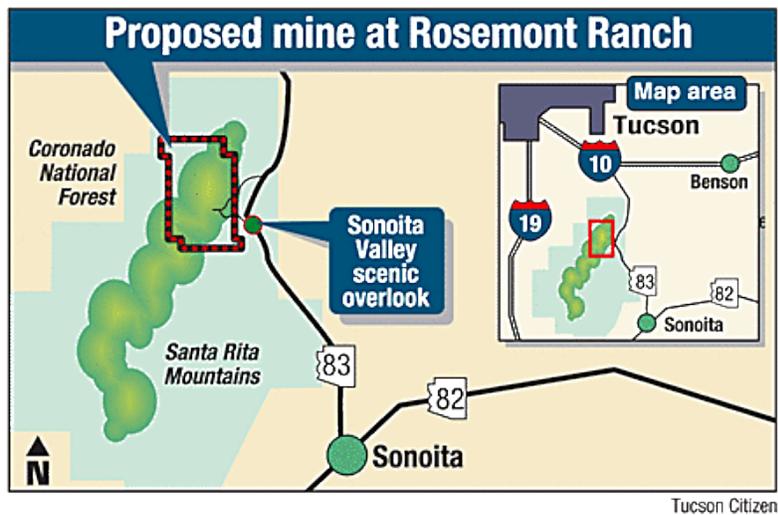
## **Methods**

### **The issue**

The environmental debate to be analyzed began in 2005 when Augusta Resources, a Canadian-based mining corporation, bought the rights to the Rosemont Ranch area in the Santa Rita Mountains approximately 20 miles southeast of Tucson, Arizona and soon after submitted an official mining plan as part of the National Environmental Protection Act (NEPA) project protocol (see image 1 & 2). The proposed copper mine site, though centered on privately-owned property, is situated on the eastern boundary of the Coronado National Forest and the western edge of the Las Cienegas National Conservation Area, a region well-known for its abundant birdlife, countless recreation trails, rich ranching history, deep cultural importance to native peoples, and the approximately forty local wineries which flourish on the Sonoita Plain. For these reasons, local opposition towards the mine has generally been strong, and in 2008, the first organized opposition to the mining plan materialized as the Las Cienegas Conservation Group made public their objections to the proposal. Soon after, other groups formed with Save the

Scenic Santa Ritas Organization (SSSR) emerging as the most active and influential; SSSR is also the primary litigant in challenges to the mine.

Figure 1: The Proposed Mine Site



Tucson Citizen (2008). Retrieved 03/15 from archives. No longer in publication.

Figure 2: The Rosemont Mine Site



This is a photo of the proposed mine site on the eastern side of the Santa Rita Mountains.

It seemed for several years the primary strategy of SSSR was to delay the construction of the mine through legal maneuvers and repeated calls for additional environmental impact studies in hopes that the delays would produce a financial burden too significant for Augusta Resources and their subsidiary Rosemont Copper Company (RCC) to sustain. Indeed, in the final months of 2013, this strategy seemed to be succeeding as RCC and Augusta's stock prices lost approximately 60% in value as stocks decreased from \$1.69 per share to \$.63 in less than one hour resulting in the rapid loss of nearly \$57 million in company value following the SSSR's posting of a confidential letter written by the EPA suggesting the permit be denied (Parry, 2013). In the spring of 2014, Augusta and RCC again faced major obstacles in their efforts to secure Environmental Protection Agency (EPA) and Army Corps of Engineers (ACOE) approval as new images surfaced of a rare and endangered jaguar and ocelot near the proposed mine site; additional delays were likely. Many activists were exceedingly optimistic that they had finally won in their battle to protect the Santa Rita Mountains. However, in the summer of 2014, RCC and Augusta Resources were acquired by the larger and more financially secure Hudbay Minerals Inc. and the debate began anew.

### **The corpora**

The corpora for the present study are two small (less than 100K words), specialized corpora containing blog posts from Rosemont Mine Truth (RMT) website and press releases from Rosemont Copper Company (RCC). The RMT website was developed by the Tucson-based environmental group Save the Scenic Santa Ritas (SSSR) to persuade Arizonans to challenge the proposed copper mine, and the RMT sub-corpus contains all blog posts from its inception in 2010 to August of 2014 (the RMT website remains an active site with regularly posted updates as of February, 2015). The second interest group whose texts are included in the analysis is the

company, RCC, which plans to construct the open-pit copper mine. The RCC corpus contains all press releases from RCC prior to its takeover by Hudson Bay Minerals in 2014; RCC ceased releasing press releases at this time. Thus, the corpora are organized and investigated according to the interest groups of SSSR and its RMT blog and RCC and their press releases.

The classification scheme and construction rationale applied in this study views the purpose of the RCC press releases and RMT blog posts in this context as similar. Both genres, the press release and the blog post, are the mediums of choice by each interest group for their management of public relations. These texts pursue a rather unified rhetorical purpose of public relations and perception management, as the primary purpose of both groups of texts is to influence public perception of the proposed mine. Similar to Gerbig (2001), texts were organized into separate corpora and analyzed to reveal group-specific linguistic practices reflective of their opposing positions on the topic. The corpora were specialized and exhaustive, i.e. all possible texts were included, and designed to reveal differing ideological stances towards the construction of the mine.

RMT produces texts at a greater frequency and generally of greater length, and thus, the RMT corpus is larger in word count and number of texts. As noted in table 1, the corpus of RCC press releases consists of approximately 17,000 words while the RMT blog corpus has approximately 70,000 words. The corpora were tagged for parts of speech (POS) using the CLAWS tagger and were ascribed semantic tags using the USAS semantic analysis framework through the W-Matrix online platform (Rayson; 2008, 2009). The USAS tagger applies tags to all lexical items through the application of a framework of twenty-one discourse fields, e.g. the field *I* refers to money, industry and commerce. Each of the twenty one major discourse fields

can be sub-divided into more specific categories, e.g. *I.1.1+* equals *money: affluence* and *I.1.1-* equals *money: debt*.

Table 9: Corpora of RCC and RMT

Interest Group	Genre	# of texts	Time Period	Tokens #	Token Types
Rosemont Copper Company	Press Release	33	06/28/ 2007* to August 2014	16,624	2,463
Rosemont Mine Truth	Blog Posts	130	08/30/2010** to August 2014	71,619	5237

\*06/2007 was approximately one month before RCC's second mining plan submission.

\*\*Rosemont Mine Truth posted its first blog release on September 30, 2010.

The log likelihood (LL) statistic was used to test significance between frequencies of semantic tags between the two corpora as a means to focus on a limited set of data. LL is often used within corpus linguistics to test similarities and differences between corpora because it does not assume a normal distribution of data, thereby enabling a comparison of features of differently sized corpora (Rayson and Garside, 2000). Also, while frequencies of individual lexical items may occur at low numbers within small corpora, part of speech (POS) and semantic tag frequencies appear at higher frequencies, and thus, the LL test is less susceptible to overestimating significance (Rayson & Garside, 2000). LL statistics have been used to determine variation in use of interactional pragmatic markers in office hours consultations (Reinhardt, 2013) and has been applied to measure overuse/underuse of particular POS categories within learner corpora as compared to native speaker corpora (Granger and Rayson, 1998). For LL analysis, scores greater than 3.84 or less than -3.84 are deemed significant; these scores produce p-values less than .05 (Reinhardt, 2013). A score greater than 3.84 indicates that the POS or semantic tag was used with a significantly higher frequency in the node corpus. For example, if a

LL score of greater than 3.84 is calculated for the semantic tag *geographical terms* when the RMT corpus is analyzed, this indicates that instances of “geographical terms” occur at a significantly higher frequency in the RMT corpus than in the RCC corpus.

### **Place names and GIS**

All place name mentions were extracted from the corpora through the use of Wmatrix as each place name receives a tag as a proper noun. However, this step requires manual intervention as not all extracted items from these operations are points which can be ascribed a latitude and longitude position. For example, a proper noun extraction will extract all place names present in the corpus but will also include names of people. Following the manual removal of false items, each place name mention was geo-referenced, i.e. assigned a latitude and longitude in the database, through a relational joining of a gazetteer and the list of place name mentions. A gazetteer is a list of place names, GPS data, and additional information such as postal code, legislative district, school district, etc. Thus, geo-data from the gazetteer was aligned with the respective place names. Only place name and GPS data from the gazetteer were retained for the present study. The database was then uploaded to the online mapping platform CartoDB which allows a variety of mapping and visualization options for GIS data. The aforementioned operations allowed the first series of maps which display all place name mentions in the interest group corpora.

The second series of maps was produced to move beyond “what” places were mentioned in the interest group corpora but “how” these places were reference and discussed. This operation first required the use of the Wmatrix USAS semantic tagger to assign a semantic tag to each item in the corpora. Then, all proper nouns were extracted and the data manually cleaned to remove non-place names, e.g. person’s names. A database was then produced which consisted of

place name, GPS coordinates, and cumulative frequencies for each tag within a 5L to 5R collocational window. Table 1 displays how the data appeared in the database. The table shows a total of 813 references to “Rosemont” from RMT (the row of data continues; this displays only the first four tags due to space limitations). The tag frequencies indicate how many times within the 5L-5R collocation window a tag of *evaluation: bad*, *evaluation: false*, etc. appeared with the geographical location. This data could then be mapped on the CartoDB platform to show semantic meanings ascribed to the different locations referenced in the debate.

Table 10: Sample Data

Place	Freq	Author	Latitude	Longitude	Evaluation: bad	Evaluation: false	Likely	Money and Pay
Rosemont	813	RMT	31.851	-110.757	29	44	111	131

## Results and Discussion

The log likelihood analysis of the variation between semantic tag frequencies of the RMT and RCC corpora revealed that approximately 50 semantic tags occurred at a significantly higher frequency within the environmental texts than in the industry texts. Table 1 displays the semantic tags of ecological and geographic relevance that appeared at a significantly higher frequency within the RMT corpus than in the industry RCC corpus; the full list of semantic tags that occurred at a significant frequency within RMT texts is provided in appendix A.

Table 11: Rosemont Mine Truth High frequency/Significant Semantic Tags

Tag	Tag Feature	RMT	RCC	LL*	Examples
L2	living creatures	277	11	59.40	<i>wildlife, fish, endangered species, insects, birds, animals, prey, jaguar, ocelot, amphibians, invertebrates, frog, cuckoo, bald eagles, predator</i>
A1.1.2	damaging and destroying	132	3	36.58	<i>damage, destruction, destroying, obliterate, charred, violate, smash, harm, irreparable, collapse, broken, rips, devastating</i>
W3	geographical terms	543	65	30.75	<i>aquifer, mountains, creek, desert, land, stream, wetland, pond, canyon, lake, grasslands, slope, wells, valley, tributary, river, forest, mount, butte, crest</i>
O4.2+	judgment of appearance: beautiful	146	10	20.24	<i>attractive, clean, scenic, appealing, spectacular, lush</i>
E3-	violent/angry	123	8	17.99	<i>threatens, toxic, threatened, poisonous, acrimonious, poison, harass</i>
L3	plants	57	4	7.70	<i>trees, willows, cottonwoods, vegetation, mesquites, (hedge fund), grass, sycamore (smelters), saguaro</i>

\*LL accounts for variation in corpus size; thus, norming for per million is not required.

The high frequency semantic tags present in Table 1 display a confluence of items which connect to the local landscape and environment and produce a situated, locally-informed discourse. The high frequency tags of *living creatures (L2)*, *geographical terms (W3)*, and *plants (L3)* exhibits an intimacy with the land and its living inhabitants as the group repeatedly discusses the endangered jaguar of the Santa Ritas, names the cottonwoods and willows of the Coronado National Forest, discusses canyons and watersheds that support Las Cienegas National Conservation Area, and speak of grasslands, valleys, washes and aquifers. These items index a value system of aesthetics and stewardship as each reference displays a connection to the mountains and a knowledge of the environment only possible through the every day lived

experiences afforded by proximity. These repeated references are not arbitrary but are selections which index and reflect an ecological aesthetic in much the same manner as discussed by Carbaugh (2001). As *living creatures* and the *geographical places* in which they reside are frequently mentioned, so too are the destructive consequences of the proposed mine through evaluations of *damaging and destruction* and references to *violence and anger*. Thus, the discourse which emerges is one which is highly situated in the landscape of southern Arizona and the Santa Rita Mountains that makes repeated efforts to list the many living creatures present and the beauty of the area while claiming the deleterious and harmful effects of the proposed mine. The importance of these references to places is highlighted through their absence in the contrasting industry corpus. Thus, while we see an anchoring of the environmental groups message within the local ecology, the company instead references a much different set of locations.

The company's discourse is unsurprisingly distinct from the environmental group but does display similar attempts at anchoring a message within a geographical space. As noted previously, the environmental group displays an intimacy and proximity with the land that is reflected in their mentioning of local tree species, wildlife, and geographical features such as canyons, washes, and watersheds. In contrast, RCC uses proper *geographical names*, i.e. places with specific GPS coordinates (see tag Z2 in table 2), at a higher frequency. However, these places are generally beyond the region of southern Arizona as RCC more often references financial centers such as New York, Shanghai, Frankfurt, and metropolitan and government centers in the United States such as San Francisco and Washington D.C. The company is also much more likely to reference the University of Arizona (located 26 miles from the proposed mine) and Arizona State University (approximately 200 miles from the mine) and also high

schools and local school districts. Closer investigation of the references to local educational institutions revealed an initiative on the part of the company to offer financial support to local districts.

The frequency of the tag *places* in the RCC texts (see tag M7 in table 2) is also relevant for its anchoring of its message. However, *international*, *regional*, *local*, and *province* are deployed in contexts of governments and institutions and the authority and power they provide. These mentions reflect the dominance of the industry's discourse as it is aligned with groups of authority in centers of influence rather than the ecological aesthetic reflected in the *geographical terms* commonly deployed in RMT texts. Finally, rather than evaluations of beauty or references to environmental features, the company discusses the substances and materials which may be extracted from the mountain and how the company's operations will advance education (tag P1) and increase local employment. Thus, evident is a discourse from the company that conceives of the mountain as a resource to be exploited for human benefit with little consideration offered to the environmental ramifications for the local ecology. (See Appendix B for a full list of significant semantic tags within the RCC corpus).

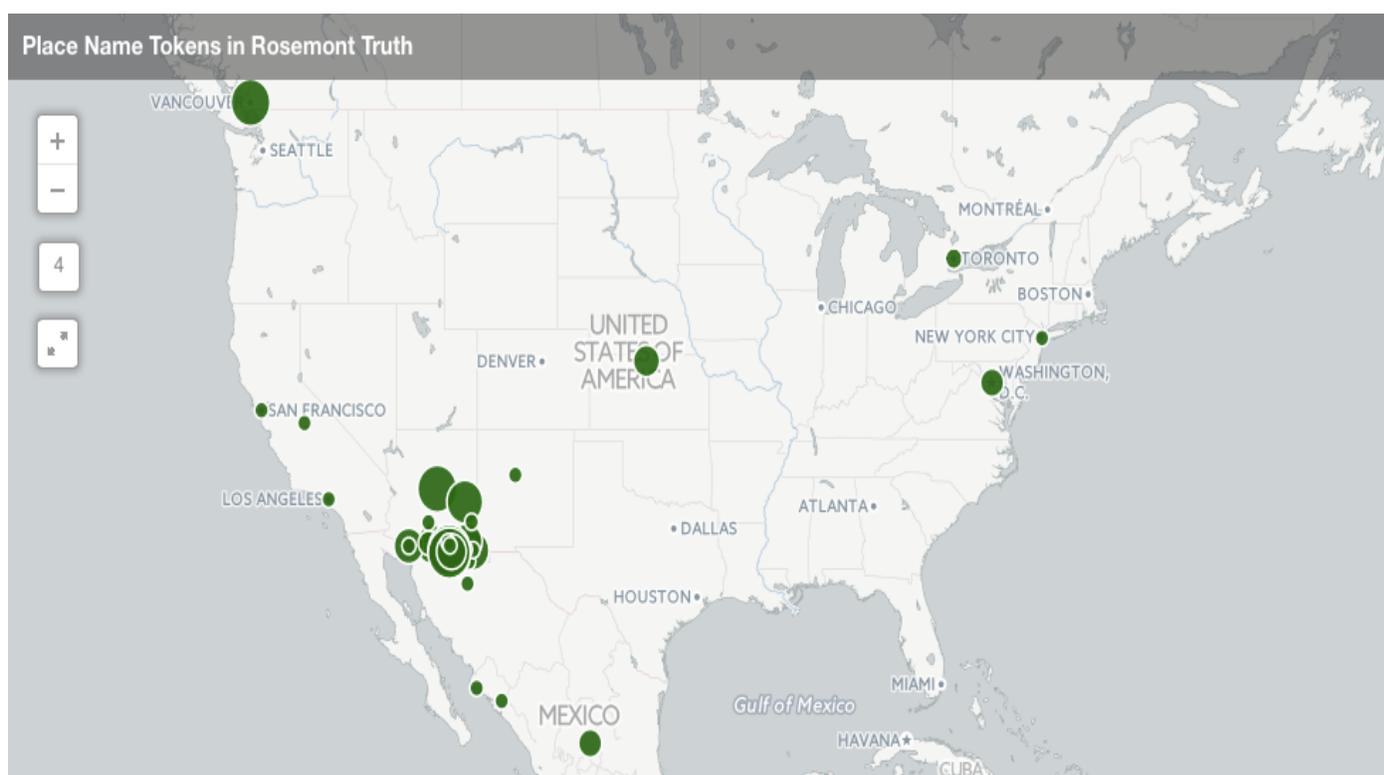
Table 12: Rosemont Copper Company High frequency/Significant Semantic Tags

Tag	Tag Feature	RMT	RCC	LL	Examples
M7	places	151	470	11.05	<i>international, regional, area, local, province, site, district</i>
P1	education	108	89	125.73	<i>school, education, students, Community college, university, Ph.D., study, alumni, scholars, educate</i>
O1.1	substances and materials	236	547	56.87	<i>copper, metals, silver, gold, deposit, diamond, ore</i>
A5.1+	evaluation: good	58	89	33.28	<i>enhances, enhanced, merit, defensible, developed, positive, favorable, great, progress, world class, improve, positive, defensible</i>
Z2	geographical names	252	845	11.18	<i>Vancouver, Bermuda, New York, Shanghai, Sydney, Canada, Quebec, United States, San Francisco, Arizona, Denver, University of Arizona, Arizona State University, Pima County, Nevada, Venezuela, Ecuador, Vancouver, Toronto, Frankfurt, London, Washington DC, South America, Ontario</i>
Z2	geographical names	252	845	11.18	<i>Vancouver, Bermuda, New York, Shanghai, Sydney, Canada, United States, San Francisco, Arizona, Denver, University of Arizona, Arizona State University, Pima County, Venezuela, Ecuador, Vancouver, Toronto, Frankfurt, London, Washington DC, South America, Ontario</i>

Moving beyond discussion of the variation in tag frequencies, the current study also explores the potential insights that could be gained from applying GIS techniques to the study of place in the corpora. The situated quality of the environmental group's discourse is also reflected in a mapping of place name mentions present in the group's writing. As displayed in figure 3, the highest concentration of place name mentions are within close proximity to the mine. RMT's

writing identifies locations such as Davidson Canyon, Empire Gulch, Las Cienegas Creek and many other locations in close proximity to the proposed mine site. The frequency of these mentions further displays the rhetoric of stewardship and ecological value. The naming of the locations grants them relevance and importance for naming a place is an important discursive event which instantiates a location and provides a deictic grounding of the discourse while indexing a relationship to the land that can only be experienced through close proximity to the location.

Figure 3: Geographical place mentions in the RMT Corpus (Raw frequency)

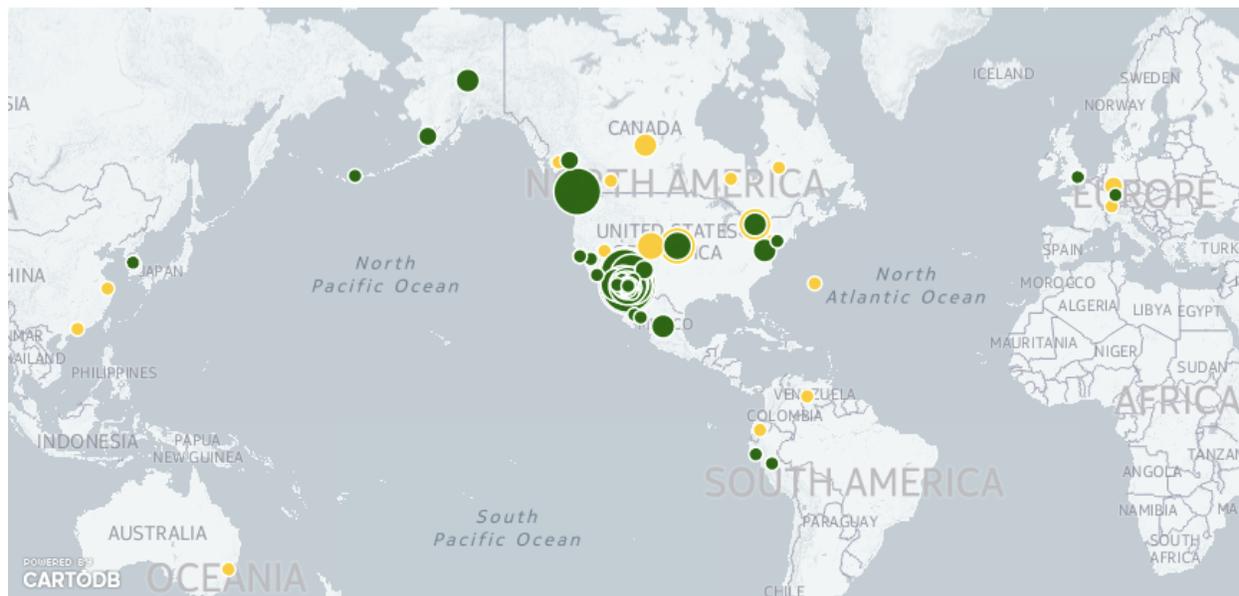


**This map was produced using the online mapping software CartoDB. On the site, a viewer can zoom in/out and click on each bubble to receive information about the marker such as place, author, GPS, frequency.**

The difference in place naming practices between the two groups is also evident in figure 4 which provides an overlay of place name mentions by both the company and the environmental group. As the two interest group corpora were of different size, the place name mentions were normed for occurrences per million to allow for comparison between the two groups. This figure

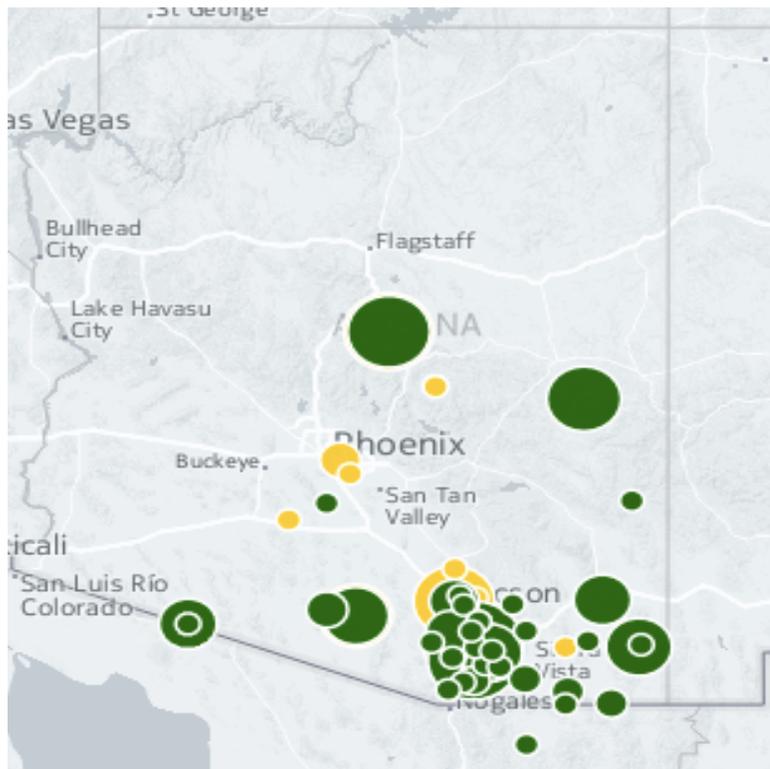
displays the density of references within southern Arizona within the RMT corpus. While the environmental group does mention places beyond the Arizona borders, figure 5 clearly displays that the greatest density of mentions fall within close proximity to the proposed mine site. It is important to note that each bubble represents places which may be georeferenced with specific GPS coordinates. However, as the previous discussion noted, many references to the local landscape are not able to be georeferenced for they are discussing non-specific canyons, watersheds, etc.

Figure 4: Place Name Mentions in RMT and RCC



Legend: Green=Rosemont Mine Truth; Gold=Rosemont Copper Company

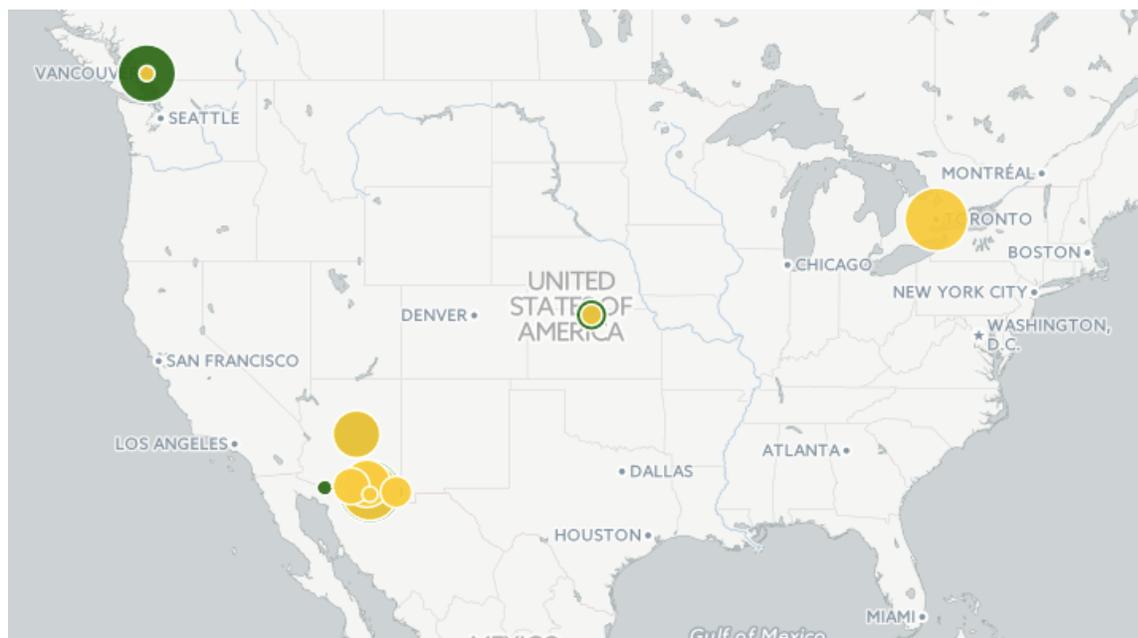
Figure 5: Place name mentions in the state of Arizona



Legend: Green=Rosemont Mine Truth; Gold=Rosemont Copper Company

The LL analysis and mapping of place name mentions was further supplemented by mapping of semantic tags which occurred within a 5L-5R collocational window of each place name. Thus, while the previous maps display what place names were mentioned, the subsequent maps exhibit how each place name was discussed. In other words, the previous maps display how many times the Santa Rita Mountains, Pima County, Tucson, and other locations were mentioned in the corpora, but the following maps display, for instance, how many times the Santa Rita Mountains were discussed within a semantic window of, for example, *money/pay* (tag I1.1), *evaluations of beauty* (tag O4.2+) etc. This procedure was completed to attempt to understand and then visualize the different conceptualizations of the mountains, the mine, and, most importantly, the environment. The map in figure 6 displays the locations of particular importance (Rosemont, Santa Rita Mountains, Pima County, Arizona, Toronto, Vancouver, United States) which appear at high frequency within the corpora with the tags money and pay, business, employment. Thus, the markers on the map do not represent the frequency of place name mentions but rather the frequency of mentions of money, business, and employment within the 5L-5R collocational window.

Figure 6: References to Money, Pay, and Business



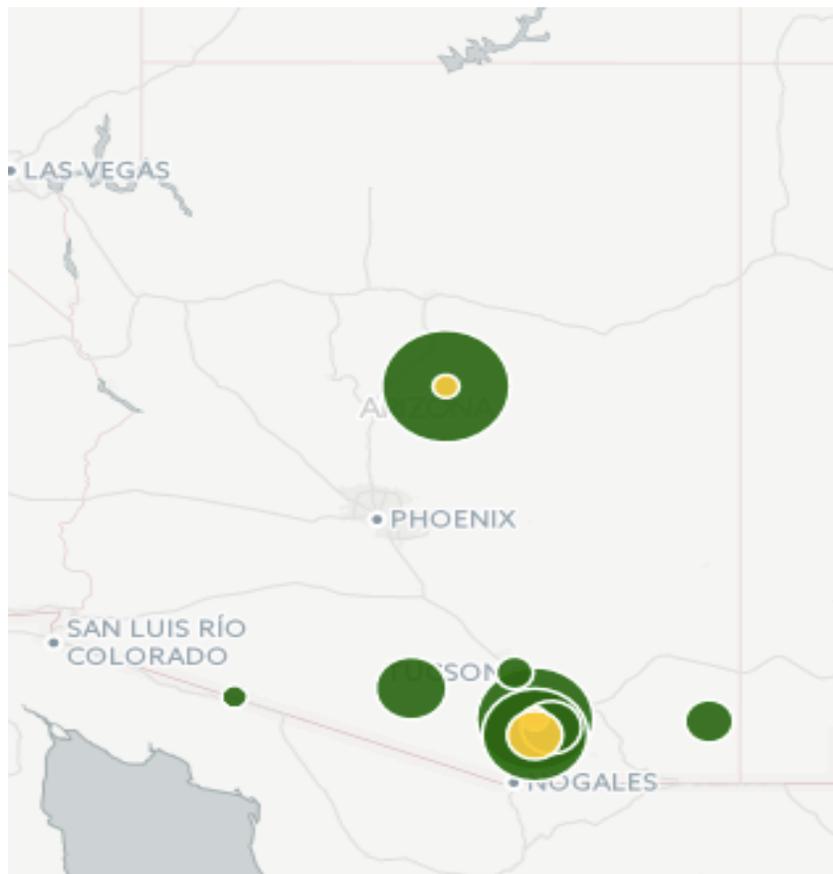
**Green=Rosemont Mine Truth; Gold=Rosemont Copper Company**

While on first study the image may not seem insightful, it does reveal an interesting trend in the discourse. As the bubble visualizations show, the copper company speaks of Pima County, Arizona, and the Santa Ritas from a resource orientation. Vancouver, the home of RCC's parent company is the only location in which mentions of economic relevance are more dense for the environmental group. While the environmental group references Vancouver in a context of economics, the company presents the mountains and the mine in context of resources, financial gain, and economic opportunity. The density of economic mentions within Arizona from the company far exceed similar mentions from the environmental group and display the opposing representations of place between the interest groups. This pattern within the Rosemont Copper Company data reflects a resource-driven, economic-focused discourse.

Previously displayed in table 11 were several semantic features (e.g. *evaluations of beauty, plants, living creatures*) that occurred significantly more frequently within the RMT corpus. These tagged items displayed the situated quality of the RMT discourse and indexed

values of stewardship and ecological awareness. These cumulative frequencies of the tagged items were calculated and normed for per million for both corpora (the tag *geographical names* was removed). These references are RMT and RCC and are displayed in figure 6.

Figure 7: Ecology-informed References



**Green=Rosemont Mine Truth; Yellow=Rosemont Copper Company**

While the visualization of figure 5 represents the mapping of per million frequencies, it is worthy to note that within the complete RCC corpus, the company only twice referred to the Santa Rita Mountains in a context of beauty and only twice mentions living creatures inhabiting the mountains. Further, the company only mentions living creatures one time in the context of Arizona. In contrast, the visualization shows that the environmental group speaks of local places not in the context of economic value but rather for the beauty of the land and the many plants and animals present. Thus, this second map when contrasted with the previous map of semantic

tags of economic relevance further displays the divergence in how place names important to the issue are represented.

The visualizations reveal the relevance of place within the discourse of each of the two groups and also indicate the sources from which each group draws legitimacy. Within the RMT texts, legitimacy is established through their displays of intimate knowledge of the landscape evident in the density of references to streams, watersheds, washes, canyons, aquifers, etc. This knowledge results from the group's proximity to the Rosemont Mine site and is a knowledge gained from years of walking the land, surveying its contours, and traversing its many ridges. The RMT discourse names the many living creatures which live in the Santa Rita Mountains at a significantly higher frequency than their industry counterpart but also, and with similar significance, names sycamores, cottonwoods, mesquites, and saguaros. From this knowledge, RMT exhibits an ecological awareness, displays an intimacy with the landscape, and creates a connection with the land. In contrast, the company does not name these local places nor list the many flora and fauna whose habitat would be affected by the mine's construction. Their discourse does not display such intimacy for the landscape because the company does not have the knowledge and connection with the land that their counterparts at RMT possess. Instead, RCC accesses and produces an economics-focused, resource-driven discourse much more likely to speak of financial centers such as Frankfurt and Hong Kong and government capitals like Washington D.C. while speaking of the mountain in terms of the minerals which may be extracted.

### **Conclusion**

These maps display the potential for integrating GIS and corpus linguistics, although this current iteration has one primary limitation: the size of the corpora. The corpora composed for

the study are exhaustive, i.e. they contain as all the blog posts and press releases from the two groups that were published, but remain rather small in order to fully demonstrate the potential of the approach. However, many debates, environmental or otherwise, may not be represented through copious amounts of textual data. A limitation on text is a reality but future research could pursue a greater scale of investigation through linking multiple environmental debates into a larger study and corpus. Through uniting data from debates such as the Pebble Beach Mine of Bristol Bay, Alaska, fracking across the Marcellus Shale field, oil drilling in the Gulf of Mexico following the Deepwater Horizon spill and other debates, the potential of GIS and corpus linguistics for the study of environmental discourse may be more fully realized.

Although environmental studies may have taken a linguistic turn in the mid-1990s following Halliday's seminal talk (1991), the absence of an open-access environmental corpus for more comprehensive studies of environmental discourse must be addressed. An open access corpus that combines multiple genres, texts from numerous interest groups, but most importantly, consolidates textual data from multiple environmental debates could offer contributions to our understanding of the discursive production of environment and environmental issues but could also contribute much to areas beyond environmental studies. As evident in the Rosemont Mine debate analyzed in this study, the local environmental group and their message is overwhelmed by a dominant economic discourse that seemingly overpowers local groups and their voices. However, only cursory review of the Pebble Beach Mine controversy and other contentious debates reveals a repetition of discursive frames used to forward these ultimately destructive projects. While the CartoDB spatializations are projections of only one debate, a larger corpus representing multiple debates coupled with functionalities of GIS could provide an engaging and productive space for the analysis of environmental issues.

An environmental corpus could develop and integrate novel ecolinguistic-informed annotation schema to enable different types of ecolinguistics analyses. For example, deixis markers of space and time could be tagged within an environmental corpus to reflect relationships with and stances to the environment. These deixis features could be tagged, extracted, analyzed, and mapped through named entity recognition, georeferencing procedures, and the use of GIS software to reflect the text's position/stance to referential objects such as mining projects, dam construction, and other environmental debates. Further research into deictic features such as cardinal points, directional indicators, geographical terms, and place names could expand our understanding of how environmental spaces are discursively realized in discourse.

An environmental corpus could also inform pedagogy in a number of educational contexts. For example, an environmental corpus could be exploited by students in environmental communication, environmental sciences, and environmental policy. It is possible that through a better understanding of the discursive processes surrounding environmental issues that writers with environmental interests could more effectively present issues of ecological relevance to the general public. Finally, an environmental corpus integrating GIS visualizations could also enable critical pedagogies of place for place-based curriculum. Through integrating corpus data, GIS techniques, and game design technologies such as Aris, a robust, place-based learning platform could be produced. Through integrating CL and GIS and utilizing place-based design systems, there exists a true and attainable possibility for an online place-based curricular platform.

This interdisciplinary study and its integration of corpus linguistics and GIS represents an advancement in empirical methods for ecolinguistics, a field often criticized for the sometimes philosophical nature of its work. Through the spatialization of corpus data, intuitive

interpretations of data are made possible and a story perhaps occluded within the rows and columns of a database is revealed. This process of mapping both place names as well as their semantic representations offers exciting potential for investigations into place-based environmental discourse, highlights the need for a larger, more inclusive corpus, and displays opportunities for further developments in place-based critical pedagogy. Future research will create a much larger corpus of environmental texts, develop ecolinguistic annotation schema and taxonomies for spatializing other textual features present within texts and corpora, and produce platforms for place-based pedagogy

**A CORPUS-AIDED APPROACH FOR THE TEACHING AND LEARNING OF  
RHETORIC IN AN UNDERGRADUATE COMPOSITION COURSE FOR L2 WRITERS**

**Abstract**

In this study, twenty-one international students in the second course of an undergraduate writing program sequence at an U.S. university studied an ongoing local debate regarding a proposed copper mine as a means to increase their ability to identify, understand, and analyze rhetorical strategies and their patterning within texts and discourse. The participants analyzed texts from two primary interest groups, a local, environmental group and an international mining company, and engaged in a series of corpus-aided activities using corpus data derived from collections of texts from the opposing groups. The contrastive analyses made possible through the study of texts and corpus data from the two sharply distinct groups enabled students to notice, analyze, and discuss the meaningful and purposeful variation in word choice and rhetorical strategies present in the texts, the corpus data, and the debate. This implementation of local, specialized corpora comprised of texts with immediate relevance to the students' campus and community provides a means to incorporate corpus study in the writing classroom by offering multiple opportunities for the analysis and discussion of meaningful, purpose-driven language choices. The approach to be presented provides an outline for how corpus data can be integrated into writing classrooms for advancing students' abilities to analyze language and increase their rhetorical awareness. This article details the rationale and the principles guiding the design of the approach, explains the corpus-aided activities, reports students' attitudes to the use of corpus data in their academic writing classroom, and offers suggestions for implementing similar activities in L2 writing classrooms.

## **A CORPUS-AIDED APPROACH FOR THE TEACHING AND LEARNING OF RHETORIC IN AN UNDERGRADUATE COMPOSITION COURSE FOR L2 WRITERS**

### **Introduction**

Data-Driven Learning (DDL) (Johns 1986, 1991) for direct in-class corpus study, the corpus-based CoBuild dictionaries and materials (Sinclair, 1987), and the concordancing method (Tribble & Jones, 1990) are generally considered the first applications of corpus linguistics for language learning classrooms and materials design. In the years following these early developments in corpus-aided pedagogy, there has been an “explosion of studies devoted to various aspects of the use of corpora in language learning” (Chambers, 2005, p. 111) with the teaching and learning of vocabulary through corpus consultation most often the focus (e.g., Bernadini, 2002; Chambers, 2005 & 2007; Cobb, 1997 & 1999; Horst, Cobb, & Nicolae, 2005; Johns, 1997; Tribble and Jones, 1990; Varley 2009).

While these studies have indeed been insightful and the benefits of corpus study and consultation have generally been confirmed, much of this work also reflects the characteristics of corpus pedagogy that have been criticized: bottom-up practices with little attention to context and too much attention to linguistic features. However, Flowerdew (1998) first noted an emerging shift in corpus pedagogy as researchers increasingly completed genre and discourse analyses, moving beyond “the surface of the collocational patterns” to investigations of “functional patterns” for more applied pedagogical ends (p. 547). Several years later, Flowerdew (2005) further refuted the critiques against corpus-based methodologies and cited the increased use of top-down strategies influenced by ESP genre theory as well as the use of small, specialized corpora where context is more easily accessed and authenticated by language learners. Thus, in recent years, while valuable grammatical and lexical based study continues,

research in corpus analysis and pedagogy has increasingly shifted to more discourse and genre-specific approaches, resulting in deeper insights into particular genres of EAP and ESP interest and more comprehensive explanations of the pedagogical implications (e.g. Charles, 2007; Hyland, 2009; Henry and Roseberry, 1996, 2001; Flowerdew, L. 2003; Flowerdew, J. & Wan, 2006, 2010; Flowerdew, J. & Forest, 2009; Upton & Connor, 2001).

Despite the continued advancement of corpus approaches for language teaching and learning, Römer (2010) comments that it would be false to claim that corpora and corpus tools have changed the “pedagogical landscape” (p. 18) and Boulton (2010) laments that corpus research “seems largely invisible downstream to teachers and learners” (p. 129). Further, while corpus approaches have been assessed for a variety of purposes and contexts, investigations of corpus approaches for writing instruction have been few (Ädel, 2010). Ädel comments that although corpus consultation for developing writing skills seems clearly beneficial, corpus study in L2 writing contexts “seems not to be widely practiced”, especially beyond the teaching of vocabulary and collocation (p.40). Reflecting Ädel’s comment, Boulton’s (2011) meta-analysis of 27 studies assessing learning outcomes from corpus study includes only two studies with writing as a focus.

This slow uptake of corpus pedagogy in the writing classroom has prompted many to rethink corpus approaches, and in recent years, researchers have positioned corpus pedagogy as a supplement to existing methods and curriculum rather than a stand-alone pedagogical approach. Conrad asserts that if corpora and/or corpus data are to be utilized in the classroom, corpus approaches should be integrated within existing pedagogy rather than presented in isolation (2000), serving to “complement traditional language learning resources” (Chambers, 2005, p. 111). Reinhardt suggests, “(corpus) approaches are commensurable with, and can be scaffolded

into, more familiar approaches that focus on learning through meaningful language use and the development of critical thinking and autonomous learning skills” (2010, p. 247). While Reinhardt was writing specifically with SL/FL contexts and classrooms in mind, this statement alludes to the potential for the integration of corpus approaches in undergraduate L2 writing courses.

Informed by corpus findings in EAP/ESP pedagogical contexts and guided by the suggestions of Conrad (2000), Chambers (2005), and Reinhardt (2010), this study develops and details the use of corpus-aided activities for an undergraduate L2 writing class engaged in the study of rhetoric, a common curricular topic in first year composition which requires students to analyze the rhetorical strategies present in texts. While some language teachers and writing instructors have knowledge of corpus pedagogy, it seems few instructors possess the strategies for integrating meaningful corpus study into their courses. The approach presented in this study provides a model for incorporating corpus study in the L2 writing classroom as it illustrates how corpus-aided pedagogy can be effectively introduced into a writing classroom by contextualizing the instruction in a local issue of immediate relevance. As will be discussed, this approach’s use of corpora of texts from opposing groups in a contentious debate facilitated the noticing, analysis, and discussion of meaningful and purposeful language use by highlighting the sharp variations in language and rhetorical practices from the two groups. The contrastive corpus approach modeled in this study makes accessible a variety of insights into language and rhetorical practices for L2 writers in an EAP setting.

### **Corpus Approaches in ESP & EAP Writing**

Corpus linguistics has increasingly contributed to English for Academic Purposes (EAP) and English for Specific Purposes (ESP) research in recent years (Biber, Reppen, & Friginal,

2010) and corpus pedagogy has been adopted and implemented more frequently in EAP and ESP classrooms. As previously noted, a turn to more applied pedagogical ends for corpus research was evolving in the mid 1990s (see Henry & Roseberry, 1996), but it was Flowerdew's (1998) challenge which sparked corpus researchers to be more explicit with the pedagogical value of corpus findings for ESP and EAP curriculum and materials design. Asserting the need for greater attention to the pedagogical implications of findings and for more genre-informed practices, Flowerdew (1998) helped usher in a new era of corpus pedagogy. The discourse-sensitive approach helped resolve critiques of the bottom-up quality of corpus instruction as more rhetorically-sensitive top-down functional approaches were developed, especially in L2 writing contexts.

Early representatives of the shift to more functional, discourse-oriented approaches, Henry & Roseberry (2001) and Henry (2007) assessed a corpus-informed approach for the teaching and learning of a genre of particular ESP interest, the letter of application. In the first analysis, Henry & Roseberry (2001) reported the findings of a corpus-informed genre analysis of a small, specialized corpus of applications letters with particular attention to reporting pedagogical best practices. Similarly, Henry (2007) assessed the effectiveness of the corpus-informed genre analysis approach in a classroom setting and investigated whether corpus consultation can lead to the successful transfer of rhetorical moves and corpus-attested features by the learners to successful production of a letter. After study of the online corpus materials, the 13 Bruneian undergraduates did include and successfully realize a greater number of moves, and a keyword analysis of second drafts displayed increases in the students' usage of lexical items and discourse patterns presented through the corpus instruction.

Within EAP corpus pedagogy, Charles (2007, 2011) continued the pedagogical movement toward more top-down genre-informed approaches. Echoing earlier critiques of corpus approaches, she noted that corpus consultation has not been more fully integrated into teaching practices due to its focus on lexicogrammatical features with less attention to higher-level rhetorical functions. As is common in many corpus-informed writing interventions, the researcher/instructor compiles a small, specialized corpus of a genre of interest to the students in the course which is followed by either direct or indirect corpus work in the classroom. For Charles (2007), the genre of interest was the master's thesis with a corpus compiled to highlight and instruct its rhetorical features, particularly concession strategies where writers anticipate and minimize potential criticisms, for a class of 40 L2 graduate students. Her approach began with emphasis on rhetorical/discourse functions and then follows with a closer look at the linguistic features that typify and often signal moves, i.e. instruction moves from the top down. Similarly, her subsequent study (2011) also foregrounded discourse-based activities that highlight rhetorical functions before transitioning to hands-on corpus consultation by students. As with the earlier study, a small corpus produces a limited set of concordance lines to enable more transparent and accessible interpretation. Overall, students reported that they learned discourse features through the concordance approach. These Charles (2007, 2011) studies display the pedagogical value of a rhetorically-informed approach and show that while early corpus pedagogy may have promoted more bottom-up instruction, corpus pedagogues have responded and developed more dynamic pedagogical practices.

Within writing research, it can be difficult to measure and assess the usefulness of a pedagogical approach for facilitating transfer of skills to writing tasks. However, while corpus research has been critiqued for lacking quantitative assessments of learning gains, there has been

work to assess whether corpus consultation does indeed produce transfer of knowledge and skills to other writing tasks. For example, Kaur and Hegelmeir (2005) found that students successfully transferred academic vocabulary to authentic writing tasks following the use of an online concordance, and Friginal (2013) reported that professional forestry students more frequently used linking adverbials and reporting verbs while also using passive/active structures in a manner more aligned with their target discourse after a series of lessons using corpus data.

Many of the studies previously discussed also assessed learner attitudes towards corpus instruction. For example, Yoon and Hirvela (2004) produced an in-depth assessment of ESL students' perceptions of corpus use in a 10-week L2 writing class through an end-of-course survey instrument and semi-structured interviews. A majority of students reported that using the corpus was beneficial for learning the usage and meaning of vocabulary, a finding common in corpus research. Overall, students indicated quite positive attitudes towards corpus use, claimed that using the corpus improved their writing ability, and attributed an increase in confidence in L2 writing to using the corpus. Similarly, Sun (2007) investigated learner attitudes of students in an academic writing course towards the use of a concordance tool. The 20 graduate-level Taiwanese writers first identified moves present in the writing of their academic fields and then created a sample-sentence corpus of the previously identified moves. The participants reported through both a questionnaire and a reflection interview that the concordance tool was indeed helpful and facilitated their writing development. And finally, the student writers in Friginal (2013) indicated that the corpus techniques were helpful as they wrote and edited their research papers.

While the aforementioned studies display a movement towards an integration of genre and corpus pedagogy, the studies generally focus on a particular EAP/ESP genre of interest (e.g.

introductions in research articles, letters of application) for their student population (typically advanced L2 graduate students) with instruction focused on linguistic features present in the target genres (e.g. linking adverbials in forestry writing; Friginal, 2013). However, the tasks and essays present in most first-year undergraduate writing courses are more varied and the objectives likely much different from an ESP courses in business English or an advanced writing class developed for L2 graduate students seeking publication. Undergraduate composition courses though are a crucial step in undergraduate international students' academic writing development and are often these students' first encounter with university-level academic writing. Although Ädel (2010) indicated the potential for corpus methods in L2 writing classrooms, pedagogical applications attempting to use corpus data in undergraduate composition beyond vocabulary and error correction (e.g. Liu & Jiang, 2009) are largely absent from the literature. Further, while there are studies assessing the usefulness of corpus study for heightening rhetorical awareness, the research has generally focused on advanced writers learning to write specialized genres within their disciplines (e.g. Lee & Swales, 2006). Studies of the effectiveness of corpus-informed genre approaches within EAP and ESP settings have been confined to the study of particular, specialized genres such as letters of application (e.g. Henry & Roseberry, 2001) where students perform genre analysis much in the manner outlined by Swales (1990) and Bhatia (1993) before producing their own text.

The approach described in this article foregrounds the development of rhetorical awareness and the understanding of the interrelationship between language, rhetorical purpose, and context through the implementation of several key design principles. In the course, students were prompted to think carefully and critically about how arguments and rhetorical strategies are structured and patterned within a debate through the close study of texts from opposing interest

groups supplemented with teacher-prepared corpus materials produced from corpora of the opposing groups. Through a practice of “pedagogic mediation” (Flowerdew, L. 2009, p. 404) where the instructor helps contextualize corpus findings, the approach introduced corpus data derived from a corpus of texts pertaining to a controversial local issue to facilitate understanding of the principles of rhetoric through the study of meaningful language from the students’ campus community which participants could contextualize and authenticate. Second, the text selection and corpus compilation was informed by the contrastive and continuum principles set forth by McCarthy and Carter (1994). The contrastive principle states “awareness of the operation of language in all texts is usually best stimulated when texts are compared and contrasted” (p.166). The interest groups in the present study display diametrically opposed viewpoints and interests, and thus, their texts differ sharply. As McCarthy and Carter assert, “Contrasting treatments of the same or related content enable a focus on language difference and can do much to promote language awareness” (p. 166). The oppositional stances of the groups are reflected through linguistic choices in the text and rather clear differences in rhetoric. Finally, multiple genres were studied in the unit in efforts to fulfill the continuum principle for teaching texts (McCarthy & Carter, 1994), according to which, “students’ language development is best supported when students are exposed to both literary and non-literary texts” (p. 167). The variety of texts provides multiple examples of meaningful language use but all in relation to the same topic (McCarthy & Carter, 1994). Thus, this approach illustrates how the use of locally-sourced texts regarding an issue of immediate relevance to the students’ campus and community can be compiled into small, specialized corpora and exploited for a variety of corpus activities which operationalize the contrastive and continuum principles for the study of rhetoric and the raising of rhetorical awareness.

The following details the course and the approach, explains the corpus activities, reports student attitudes, and offers implications and suggestions for implementing corpus activities in L2 writing classrooms.

### **A Corpus-Aided Approach to Undergraduate L2 Writing**

#### **The Participants**

The participants for the study were international students enrolled in the second of a two-course undergraduate writing sequence at a large public state university in the southwestern United States. To enroll in the university, students must display English proficiency by receiving a minimum score of 70 on the TOEFL or a 6.0 on the IELTS. Also, students in the course must receive a satisfactory grade within their introductory writing course before matriculating to this second course. The author was the instructor for the group of 21 students; eighteen students were L1 Chinese speakers, two were L1 Arabic speakers, and one was an L1 Korean speaker.

#### **The Course**

The 16-week course, the second and final course of the first year writing sequence that all undergraduate students at the university must complete, met twice per week for 75 minutes per session; the first unit and focus of this article was six weeks in duration. The first year writing program's standardized syllabus emphasizes "the skills of rhetorical analysis, research, persuasion, reflection, and revision", and the course's curricular focus on rhetoric, rhetorical analysis, and research is similar to composition courses taught in many U.S. university writing programs. The theme of the course was environmental rhetoric with the first unit focused on the proposal to build a massive open-pit copper mine in the Santa Rita Mountains approximately 20 miles from our campus. The contentious debate clearly displayed many of the rhetorical concepts discussed in the course and the proximity of the proposed mine to our campus and its potential

effects on our community granted relevance, authenticity, and context to readings and classroom discussions.

### **The Texts and the Corpus**

As previously noted, the selection of texts was informed by the contrastive and continuum principles set forth by McCarthy and Carter (1994). Students first read and discussed a company brochure, interest group blog posts, and an opposition group's video to introduce the students to the issue and begin discussion of rhetoric. Following the study of these introductory texts, the students began reading a collection of texts responding to an important event in the ongoing debate. In early November of 2014, RosemontMineTruth.org released a letter written by the Environmental Protection Agency in confidence to the private company seeking permission to build the mine. The "leaked" letter was an integral event in the ongoing debate, and many speculated that it was deliberately released by the opposition group to cause financial troubles for the copper mining company. Indeed, following the release of the letter, stock prices for the company fell so dramatically that the company lost approximately \$60 million in value over a 24-hour period. The texts (listed in Table 13) that received scrutiny from the students were those responding to the "leaked" letter.

Table 13: The Texts

<b>Title</b>	<b>Date of Publication</b>	<b>Genre</b>	<b>Author</b>
<b>(Catalyst text)</b> Analysis of Updated Draft	11/07/2013	formal business letter (the “leaked” letter)	Environmental Protection Agency
EPA Recommends Against Federal Water Permit for Rosemont Mine	11/20/2013	press release	Save the Scenic Santa Ritas (SSSR)
Augusta Comments on SSSR’s Inaccurate Press Release	11/20/2013	press release	Augusta Resources
Augusta Resources Shares Collapse After Group Leaks Report	11/20/2013	financial news report	Stockhouse
Rosemont Shouldn’t Get Federal Permit to Build Mine	11/23/2013	Opinion-Editorial	Arizona Daily Sun
Save the Scenic Santa Ritas Distorts EPA Report on Rosemont Mine	11/21/2013	Opinion-Editorial	Arizona Independent

The close study of the texts listed in table 14 was supplemented by two small, specialized corpora; the first included all blog posts from The Rosemont Mine Truth (RMT) website and the second consisted of all press releases from Rosemont Copper Company (RCC). The corpora reflect the contrastive principle (McCarthy & Carter, 1994) as the sharp contrast between the two collections allowed a clearer and more student-accessible view of the rhetorical strategies and language choices of the two groups. While the two groups employ different genres, the general purposes each group and their text pursue are quite similar. Both text groups engage in public relations management as each attempt to persuade their respective audience of the veracity of their claims and the legitimacy of their positions.

The teacher prepared the corpus data used in the instructional modules; no student was asked to use a corpus program, perform a corpus query, or generate a single concordance line. While direct corpus consultation can be effective, for students new to corpus analysis, the corpus-informed, teacher-prepared materials avoided noted obstacles limiting corpus pedagogy asserted by Ädel (2010): “the corpus as maze” (p. 45) where students have difficulty identifying items of interests while “drowning in the data” (p.46). The data used in the class consisted of concordance lines of salient lexical items in the productions of each group and keyword lists generated from keyword analysis procedures using the free corpus tool Antconc (Anthony, 2014) (see table 2 for sample concordance lines). A keyword analysis determines which words are particularly “key” in the node text/corpus through the comparison to a reference corpus (Tribble, 2000). As Tribble states, the keywords are the “words which matter” (2000, p. 79). For example, from analysis of the corpora, I discovered through Antconc’s keyword function that the word *project* occurs at a much higher frequency in the texts from the company while *proposal* appears more frequently in the environmental group’s releases. Thus, these words are key because they appear at a high frequency in the respective interest group corpus but not in their counterpart’s texts. The corpus data supplemented the study of the texts as the keywords pattern broadly through the wider discourse of the primary interest groups, indexing rhetorical strategies of the organizations.

### **Instructional Activities**

The first few class meetings introduced students to the rhetorical strategies of *ethos*, *logos*, and *pathos*, the three elements comprising the rhetorical situations, the Rosemont Mine controversy, the major interest groups involved, and the primary arguments forwarded by each group. During these class sessions, videos concerning the Rosemont Mine were viewed and

several newspaper articles on the topic read and discussed. Students gained an understanding of the central interest groups involved in the debate and foundational knowledge of the key arguments forwarded by the different groups. Following these introductory sessions, the class transitioned to closer study of the rhetorical strategies present in the texts listed in Table 1 and to the corpus activities.

**Activity #1: Rebuilding context using concordance lines of keywords.** The first corpus activity was intended to help students discern the connections between each group's word choices and the effects these choices have on their messages. Further, the concordance study enabled students to view examples of lexical items and the “collocational, colligational, semantic, pragmatic, or textual features” with which they often co-occur (Gavioli, 2001) while also helping students to “consider the interrelations between the data and the texts they are derived from” (Gavioli, 1997, p. 88). The items indexed rhetorical strategies pursued by the interest groups, and the use of a limited number of concordance lines allowed students to focus their noticing towards the rhetorical aspects of the data. This focused analysis also yielded what Leech (1997) refers to as convergent corpus-aided learning where all students, if successful, will produce similar answers; the accessible lines and convergent quality of the activity enhanced and facilitated discussion.

For this corpus activity, I identified and extracted items that were particularly revealing of these two rhetorical patterns within the two groups' texts and leveraged these findings into the creation of the materials. From study of the corpus data and interest group texts, I noticed a number of lexical items indicating a future-oriented message that presented the mine as inevitable and its construction a certainty. In contrast, it was evident within the environmental group's texts that a rather different rhetorical undercurrent was present as the group attempted to

portray the approval process as ongoing and a decision regarding the mine as far from decided. For the company, the students were presented concordance lines for *project* and *will* for the industry group and *proposed* and *could* for the environmental group.

For the concordance activity, students were placed into groups of 3-4 and asked to read the concordance lines from the two groups and answer the questions repeated at the bottom of the handout (see table 3 and 4 for an abridged version of the handout and appendix A for the full version). As the students began discussing the questions, I moved from group to group to help them think through the questions. I prompted them to answer questions such as, “Why would this group use *project* again and again when discussing the mine?” and “What is the difference between using *will* and *could*?” I challenged them to think carefully about the words and to consider how particular word choices might serve the author’s overall purpose and how the words may function to create particular effects on the audience. This activity continued for approximately 25-30 minutes with the last 10 minutes being a full class discussion of these words. One student commented that as a non-native speaker it was not easy to determine the difference between the items but that the activity had helped him think about word choice in a manner he had not previously. The student’s comment reflects the critical principle (McCarthy & Carter, 1994) which asserts, “learners are likely to gain more interest and to be more empowered as educated citizens if they also develop a critical capacity to see through language to the ideologies and values which particular stylistic choices encode” (p. 168). Indeed, the principle motivated the design of this activity since the keywords in context were presented and discussed with regard to how they reflect the interests and values of the environmental organization on the one hand and business on the other and to how they provide a window into the rhetorical patterns and strategies of each group of each group. The activity was designed to facilitate a discussion of

the importance of word choice and how such seemingly simple choices may have a subtle yet significant rhetorical effect when used repeatedly throughout a debate.

Table 14: Student Handout #1 (Abridged)

**Stakeholder A**

Read the sentences below carefully. Think about the keyword that is written in bold in each line.

**Project**

1. a final EIS for the Rosemont Copper **project** along with the Record of Decision.
2. Our leadership has always understood that our **project** would receive the highest scrutiny from
3. answered fully and successfully and that the **project** meets all standards addressed by the
4. and robust economics of the Rosemont **project**, propels Augusta to becoming a solid
5. The Rosemont Copper **project** is located in Pima County

*(10 more concordance lines with **project** were provided to students)*

**Will**

1. Rosemont Copper **will** set a high standard for sustainable mining
2. Phase 2 of the program **will** now focus on infill drilling
3. robust economics of this project, **will** propel Augusta to become a solid
4. the mine **will** produce over 400 direct jobs and about 1,600

*(2 more sentence with will were provided to students)*

**Discussion**

1. Who do you think is the author of these sentences?
2. Who do you think the audience is for this writing?
3. What do you think is the purpose of this writing?
4. What rhetorical effect do you think the author wants to achieve through the use of these words?
5. Why would he/she choose these words and not other similar words?
6. What do you think the author wants the audience to feel or thing?
7. Are there other ways the author could reach the same effect?

Table 15: Student Handout #2 (Abridged)

**Stakeholder B**

Read the sentences below carefully. Think about the keyword that is written in bold in each line.

**Proposed**

1. the **proposed** Rosemont mine would destroy more than 3,0
2. permit the company needs to construct the **proposed** Rosemont open-pit copper mine.
3. the Cienega Creek watershed caused by the **proposed** Rosemont open-pit copper mine that will
4. conservation area is directly east of the **proposed** mine site.
5. and construction timeline for Augusta's **proposed** Rosemont copper project.

**Would**

1. That copper **would** presumably be exported to Korea.
2. the mining operations **would** be visible from Green Valley
3. that the mine **would** inflict on Southern Arizona
4. any legal challenge **would** likely take years to resolve.
5. The mine **would** also have ripple effects on water supplies...

*(The same discussion questions from Student Handout #1 were provided)*

Following the concordance activity, several groups were asked to make a list of words they felt would appear frequently in the writings of the environmental group while the others were asked to do the same for the mining company to encourage students to make connections between word choices and rhetorical strategies. The company groups listed *money, jobs, economics, protect, project, financial, investors, safe, secure, economy, and increase* while the environmental groups listed *water, animals, jaguar, environment, toxic, waste, ecosystem, save, Canadian, permit, and proposed*. With each item listed on the board, students were asked to explain why the word would appear frequently in the group's writing, how the word/s reflected a rhetorical strategy, and how the item/s would affect an audience of local citizens. Their word lists and explanations displayed their increasing understanding of the issue and the interest groups involved as well as their understanding of rhetoric and its patterning in the debate.

Further, the student-generated lists provided additional opportunities for comparison and contrast as they were soon able to compare their own keyword lists with the corpus-attested keyword lists provided in the following activity. Again, each activity was devised and implemented to foster reflection on language and rhetorical choices through comparison and contrasts of multiple features.

**Activity #2: keywords.** In the next class meeting, students were presented with texts from the opposing groups and the keywords lists generated from the corpus analysis. Again, the design imperative enabled multiple comparisons as authors, texts, and corpus data were in clear opposition. As stated, this quality of the design allowed students to notice and discuss variation in linguistic and rhetorical strategies from the two opposing groups.

In the activity, students were divided into two large groups. Group A was asked to read an editorial written to a local newspaper from a member of the Save the Scenic Santa Ritas (SSSR) organization while Group B read an editorial in a local newspaper from a mine proponent. The groups were instructed to first read their assigned text and identify the keywords/phrases that were important to the text and that could be used as evidence of a rhetorical strategy. Group A identified the keywords *scientific*, *technical*, *Forest Service*, *distorted*, *inappropriate*, *inaccurate*, *false*, *flawed*, amongst others. The group was asked to explain why these words were chosen as rhetorically salient for the texts. For example, the students explained that *Forest Service* was key because it was repeated several times in the text and could be connected to the strategy of *ethos*; they commented that referring to credible organizations could help writers establish their *ethos*. This discussion of SSSR keywords was followed by a similar discussion of keywords in the pro-mine text.

Following this initial task, the two groups were asked to create a list of keywords that would likely be used repeatedly by their side not just in the one text previously discussed but also throughout the ongoing debate. Group B listed items such as *money, jobs, economics, protect, project, financial, investors, safe, secure, economy, increase* accompanied with explanation and discussion of why these keywords would be rhetorically important and how these keywords reflect particular rhetorical strategies. Additionally, Group A listed *water, animals, jaguar, environment, toxic, waste, ecosystem, save, Canadian, permit, proposed*. Again, the group explained why they believed these keywords would regularly pattern throughout the texts from the environmental group.

Following the creation of their own keyword lists, the groups were given a handout that had corpus-generated keyword lists of the 10 most key items from each group. I explained how the list was created and what the headings on the table meant. While a keyness statistic is generally used to report keyness in the literature, I modified this term to *Rating* to make it more comprehensible and accessible to students. I explained that the rating of 51 for the first item on the environmental group list means that the advocacy texts were 51 times more likely to use the word *proposed* in their writing than the mining company. In contrast, the mining company was 103 times more likely to use *forward* in their press releases. The students were asked to discuss in groups why these words would be present on the list and why these would continue to occur in each interest group's writings. We discussed the meanings of these words together, and the students were asked to make connections between strategies and word usage and to think carefully about how these words could produce rhetorical effects to aid each side in achieving their purpose. Once again reflecting the contrastive principle, this activity was greatly facilitated by the contrasting made possible through the use of texts and corpus data from the two interest

groups. Without this ability to contrast the groups, their texts, and their data, students would have had difficulty identifying the rhetorical strategies at play within the discourse. The keywords activity handout is included in Table 16.

Table 16: The Keywords Activity Handout

<b>KEYWORDS</b>			
<b>Table 1: Environmental Group Keywords</b>			
	Keyword	# of Times	Rating
1	proposed	163	51
2	tailings	60	36
3	whether	55	33
4	objections	74	31
5	forest	296	29
6	would	170	26
7	jaguar	42	25
8	mountains	69	25
9	coronado	99	24
10	could	85	20
<b>Table 2: RCC Keywords</b>			
	Keyword	# of Times	Rating
1	forward	54	103
2	nyse	25	68
3	school	27	65
4	copper	222	57
5	economic	25	50
6	amex	18	49
7	molybdenum	25	47
8	process	61	45
9	community	28	42
10	jobs	20	42

**Activity #3: keywords and rhetoric.** Following the in-class discussion of keywords lists included in table 4, the students were given an essay prompt to complete before our next class session. The assignment (see Table 17) asked students to write a 1-2 page short essay discussing the keywords, why the items would be rhetorically important for each group, and how the keywords could be used as evidence for various rhetorical strategies. The prompt for the assignment is included in Table 18.

Table 17: Short Essay Assignment

In texts and debates, it is common to find patterns of ideas and strategies. These patterns of ideas are often represented by keywords. For your homework, analyze the keywords from the two lists and explain why these keywords are rhetorically important for each group's messages. What ideas and strategies are likely being used in the texts from the two groups? What do the keywords show us about the strategies the company and the environmental organization are using? In your writing, you should consider the rhetorical situation and use the words ethos, logos, and pathos (These are our keywords, and you should be using them.)

Students performed well on the activity and were generally successful connecting keywords from the lists to rhetorical strategies and the rhetorical situation of the authors. For example, the excerpt in Table 18 displays a growing awareness of word choice and its importance in shaping rhetorically effective messages. The student in the first excerpt notes that the keyword *proposed* is only “an idea or a suggestion for the future” and, therefore, is used repeatedly to motivate followers to continue the fight against the building of the mine. Excerpt #2 in Table 7 also includes a discussion of the keyword *proposed* but offers a slightly different discussion of the item's importance. These discussions display the students' growing rhetorical awareness and recognition of the importance of word choice in shaping effective messages. While these two students thoughtfully explain selected keywords and detail their rhetorical purpose, they were not alone. Many of the students successfully and thoughtfully connected

keywords, rhetorical strategies, and elements of the rhetorical situation in a manner similar to those included in table 18 and 19.

Table 18: Short essay excerpt #1

*Based on the data that was presented to us in class, the SSSR (Save the Scenic Santa Rita) organization used some special keywords that are completely different than their opponent's keywords. For example, the SSSR organization used the word proposed 163 times, and the word jaguar 42 times. Proposed was used 163 times because of its rhetorical importance in their writings. In my opinion, the author adopted this specific word multiple and multiple of times because by using it, the SSSR keeps their supporters in a fighting mode, and to keep them standing up against their competitors. The reason behind this enormous effect comes from the definition of the word itself, which is defined as an idea or a suggestion for the future. Since the word causes this effect on the audience, proposed is then considered to be ethos. Furthermore, we also can see that the word jaguar was used 42 times in their writings. This word is used by the SSSR organization because it causes pathos in the audience. Pathos is created because the word jaguar resembles nature and wildlife in the mountains, and with the proposed mine, wildlife will be damaged, which will eventually cause compassion in their audience.*

Table 19: Short essay excerpt #2

*Proposed is written more than fifty times than the RCC. Proposed means something is planning and not decided yet. Proposed in these letters combines two situations logos and ethos. By the logos aspect, the copper mine is exactly planning and not decided yet. Using the proposed word suggests there still have hope to against the company. To have hope extremely inspires other opponents and audience. By the ethos aspect, proposed is more correct word than project. Therefore, Proposed is more credibility.*

*The three words—forward, economic and job—are more frequent appearing in RCC's letters than the SSSR's letters. Forward is a very appropriate words in company's letter. The word is an ethos that increases the company's credibility. Forward means the future, or hope. The company using forward is leading the audience to consider about the future. The company tries to convince the audience that their copper mine can bring a good future. In some ways, the company leads audiences to believe what the cooper mine bring. They do not want their audience to consider the asked questions by the SSSR.*

### **Student Attitudes**

One week following the final corpus activity and short essay assignment, the students were asked to complete a brief 10-item survey to assess their attitudes to the use of corpus data; the ten questions with responses are listed in table 9. In the survey, approximately 81% of respondents indicated that the keyword lists helped them understand rhetorical strategies. Also, a high percentage of respondents indicated the usefulness of keyword lists and noted the concordance activity was beneficial and served as a helpful supplement to the text and other classroom materials. Students also noted the corpus activities were interesting, the activities would help them identify rhetorical strategies in texts, and that the activities could even inform their own writing practices. Although a statement of potential transfer can not be offered, the generally positive attitudes of the participants to the corpus activities for learning rhetoric is encouraging. And finally, considering the persistent concern that corpus study is often too difficult for many language learners, the survey responses were divided on whether corpus study was difficult, although a high percentage were in favor of using keywords lists in their writing class.

Table 20: Survey Questions

	<b>Strongly disagree</b>	<b>Disagree</b>	<b>Uncertain</b>	<b>Agree</b>	<b>Strongly agree</b>
1. In general, working with the keyword lists helped me learn about rhetorical strategies.	2 (12.5%)	0	1 (6.25%)	3 (18.75%)	10 (62.5%)
2. In general, working the keyword lists helped me understand patterns of rhetorical strategies in texts	1 (6.25%)	1 (6.25%)	2 (12.5%)	4 (25%)	8 (50%)
3. In general, working with the keyword lists and example sentences did NOT help me understand rhetorical strategies.	5 (31.25%)	4 (25%)	1 (6.25%)	1 (6.25%)	5 (31.25%)
4. In general, working with the keyword lists and example sentences was a useful activity that added to the textbook and other materials.	1 (6.25%)	1 (6.25%)	1 (6.25%)	6 (37.50%)	7 (43.75%)
5. In general, working with the keyword lists and example sentences was interesting.	1 (6.25%)	1 (6.25%)	2 (12.5%)	6 (37.50%)	7 (43.75%)
6. The keyword list activity helped me to understand rhetoric.	1 (6.25%)	1 (6.25%)	1 (6.25%)	3 (18.75%)	10 (62.5%)
7. Working with keyword lists helps me recognize patterns of rhetorical strategies when I am reading.	1 (6.25%)	1 (6.25%)	1 (6.25%)	4 (25%)	9(56.25%)
8. Working with the keyword lists was difficult.	2 (12.5%)	4 (25%)	3 (18.75%)	2 (12.5%)	5 (31.25%)
9. I believe studying keyword lists can help me be a better writer on other academic assignments.	1 (6.25%)	0	4 (25%)	5 (31.25%)	6 (37.50%)
10. I would like to use keyword lists in my writing class.	1 (6.25%)	0	3 (18.75%)	5 (31.25%)	7 (43.75%)

## Discussion

The success of the corpus-informed activities described here was furthered by several important elements that should be considered by any instructor attempting a similar pedagogical approach. First, the corpora were comprised of texts of a local, controversial issue of which I was quite knowledgeable but that also had particular relevance to the students' campus community. This familiarity with the issue enabled the type of "pedagogic mediation" for recontextualizing the data discussed by Flowerdew (2009, p. 404) while also engaging students in an issue of immediate local importance. Additionally, my knowledge of the issue and the debate allowed the corpus activities to be targeted and focused on specific items of salience for discussions of the rhetorical strategies present in the debate. For corpus consultation of the type outlined to be successful and effective, it is important that the instructor know the issue well, so that he/she may develop effective classroom activities that focus student attention on features of particular interest.

Secondly, the corpora compiled for such a pedagogical application need not be especially large nor the collection and analysis time-consuming. The small corpora produced a limited amount of concordance lines, enabling more transparent and accessible interpretation while avoiding the "lost in the maze" concerns of Ädel (2010, p. 45). Collecting only a small number of texts, e.g. opinion-editorials on an issue of importance to a campus community, can reveal items of interest to be studied by language learners in a variety of contexts. Thus, concerns that these activities would be especially demanding to develop are largely unfounded. Many instructors in undergraduate writing courses already collect authentic texts for students to analyze, and the activities detailed would only require a small number of additional texts and a

free corpus tool such as Antconc (Anthony, 2014). Certainly, some topics may be more appealing and engaging to undergraduate students.

Finally, and perhaps most crucial to the success of the approach, the construction of two small corpora representing two groups and their language usage allowed for contrastive analysis across multiple levels. The noticeable variation in language usage evident in the corpus data and texts facilitated engaging discussion of the rhetorical motivations for particular language choices and made obvious and accessible the concepts of rhetoric emphasized in the course. With only one corpus, the data would not have facilitated the discussion and enabled such accessible interpretations of rhetoric. Thus, to highlight rhetorical strategies at work within a debate, I suggest compilation and analysis of oppositional corpora to allow students to notice and discuss the language and rhetorical variation between two groups.

### **Conclusion**

As previously discussed, corpus approaches to language pedagogy have received increased attention from researchers and have seen marked advancement in their attention to top-down genre-sensitive practices and description of pedagogical value. However, these approaches have yet to make significant inroads into the writing classroom despite the generally favorable findings of the learning gains produced from corpus consultation and the positive attitudes from learners. The present study has detailed an innovative contrastive corpus analysis method for integrating corpus-informed activities that supplement existing materials and facilitate the achievement of course objectives. The corpus activities produced meaningful classroom discussion of word choice, rhetorical situation, and rhetorical strategies and involved students in activities that required critical thinking and engaged them in meaningful language tasks. Importantly, as the surveys indicated, students responded rather positively to the study of corpus

data and commented that they felt corpus study helped them understand the concepts of rhetoric that are the focus of the course curriculum. Although the corpus activities were likely novel to the participants, the students enjoyed the tasks and responded that the activities were beneficial to their learning. The contrastive analysis approach put forth displays the benefits of using a local, specialized corpus to support a variety of corpus-based activities in an L2 undergraduate writing classroom.

## CONCLUSION

This closing section will first detail the potential for future research in corpus-based ecological discourse analysis and corpus design, GIS and corpus linguistics, and corpus-aided place-based pedagogy before ending with some concluding remarks.

**An environmental text corpus and ecological corpus design.** Although corpus-based discourse analysis of environmental communication has more frequently been pursued in recent years, there is no large, publically-available corpus of environmental texts with the majority of studies conducted on small, specialized corpora compiled by researchers to answer specific research questions. This absence seems surprising considering that in the past fifteen years the collection, organization, and analysis of corpora has been greatly advanced by the development of numerous corpus resources. Public corpora exist for many discourse domains for numerous research purposes, yet no large comprehensive environmental text corpus is available for public use. Thus, there is a need for the construction of a corpus of environmental texts which implements a similar design, organization, and annotation schemes as popular corpora such as the COCA, Corpus of Historical American English, Michigan Corpus of Upper Level Student Papers, and others. For example, an environmental corpus could include texts from environmental advocacy groups, government agencies with environment-related missions, popular media, documentaries, international environmental agencies, environmental sciences and other environment-related research disciplines, as well as industry groups such as British Petroleum, Exxon Mobil, and Hudson Bay Mining. The inclusion of texts from multiple agencies is an important goal for the proposed corpus. The interest group texts included in the current study could be conceived as representing opposite poles on a continuum where a market-driven business discourse resides on one end and strong environmental discourse occupies the other. It

is important that a newly developed environmental corpus be constructed in a manner which includes multiple groups within this discourse continuum. It would be revealing to place other interest groups and their texts on this spectrum between the sharply oppositional groups such as those detailed in the current study to answer questions such as whether media and government are more likely to adopt either a business or environment-aligned message. This environmental corpus platform could be exploited by researchers and students from numerous academic disciplines and could inform pedagogy in environmental policy, environmental communication, etc.

While potential for a traditionally designed corpus of environmental communication exists, this conclusion would like to present the possibilities for the development of an ecology-informed corpus design and annotation schema, henceforth referred to as the Eco-Corpus. The Eco-Corpus could integrate geographically-informed annotation strategies and GIS techniques to further reveal how linguistic features pattern within place-based environmental discourse and how variation in lexicogrammatical frequencies are influenced by the dialectical ecological relationship between context and language. While perhaps the most fundamental considerations in corpus design are size and representativeness for producing generalizable conclusions about language in general, an Eco-Corpus would offer greater attention to ecological features and contextual factors to more fully investigate the relationship between context and language and how language mediates society's conceptualization of and relationships with nature. To enable context-sensitive ecology-informed analyses, the Eco-Corpus would be annotated using text-external deictic criteria to organize texts both for place and time of production as well as speaker and referential object of focus. Further, internal features indexing ecological and spatial relationships such as ecological deixis markers (e.g. cardinal points and place-name references)

and other features relevant to an ecolinguistics research agenda (e.g. tagging animals mentioned in the corpus) will be tagged.

An Eco-Corpus could also provide a platform to innovate and incorporate data visualizations to enable intuitive interpretations of corpus data, providing an engaging space for researchers, students, and the public to engage with environmental language. Amazingly, in my opinion at least, the premier journal in corpus linguistics, *The International Journal of Corpus Linguistics*, includes only one article in the past five years that explicitly focuses on the development of visualizations appropriate for displaying corpus data. The article (Hilpert, 2011) models the use of dynamic motion charts for producing visualizations of linguistic features from diachronic corpus data. The motion charts, the author asserts, facilitate “intuitive understanding of complex linguistic developments” to users and enable the analysis and interpretation of corpus data. The value of visualizations is their ability to “allow the viewer to take in and process a large amount of information in an effective and natural way” (p. 457) and may be applied for various inquiries of linguistic variation.

Further, an eco-corpus interface could integrate GIS techniques for the spatialization and visualization of features similar to those produced in the maps of article #2 in this dissertation. Hardie and Gregory (2011) discuss diexis, the linguistic forms which create distance, proximity, and spatial relations between a speaker, hearer, or object, and note the potential of these features to be reflected in GIS maps and Haig (2001) notes a high frequency of “location circumstantials” (p. 209), i.e. diexis features, as environmental texts seem concerned with situating events firmly within a specific time and place. Present within texts are directional indicators, geographical terms, and place names that could inform our understanding of how environmental spaces are discursively produced in environmental communication. The density

and nature of deictic markings reflect both spatial and ideological distance between interest groups and the referential object, i.e. the place at the center of a debate. These deixis features could be tagged, analyzed, and mapped through named entity recognition, geo-coding procedures, and the use of GIS to reflect the text's position/stance to external referential objects such as mines, dams, and pipelines.

To illustrate, using the previously discussed mine debate as a model, an Eco-Corpus could mark the proposed location of the Rosemont Mine as the referential geographical location to which all texts are connected. The first layer of the corpus would organize texts from various groups within southern Arizona. As these groups have greater geographical proximity to the Rosemont Mine site, the ecological deictic density will likely be both quantitatively greater, i.e. greater frequency of ecological markers, and qualitatively distinct, i.e. geographical references will be functionally distinct from the Mine discourse from other regions; the analysis of articles #1 and #2 display that SSSR texts have a greater density of place name references as the locally-produced texts are situated more firmly within the topography of southern Arizona. A second layer would contain texts from state-level groups but outside of southern Arizona. This layer would contain texts from state agencies, state government officials, and Phoenix and Flagstaff area media. A third text layer would be comprised of national-level texts such as those from the Environmental Protection Agency and national media. And finally, a fourth level would contain texts from international sources, primarily those originating from Vancouver, Canada, the headquarters of Augusta Resources, and Toronto, Canada, the media center of Canada. The corpus architecture would enable a macro-organization of all texts present within the geographical region but also enable micro-analyses of interest group and genre type. In other words, geographically-specific, interest group specific, genre specific searches would be made

possible in addition to a tagged schema incorporated the ecological deixis markers previously discussed. Further, additional texts representing other environmental controversies could eventually be added to the corpus. Table 21 details the geography-informed features and the tagging schema of the Eco-Corpus.

Table 21: Potential Design Features of the Eco-Corpus

External Selection	<ul style="list-style-type: none"> <li>• Texts will be selected and organized according to geographical location, temporal production, &amp; interest group.</li> </ul>
Representatives	<ul style="list-style-type: none"> <li>• Texts will be sourced and organized to allow for linguistic statements about language in particular geographical locations and time periods from particular interest groups.</li> </ul>
Balance	<ul style="list-style-type: none"> <li>• Texts will be balanced according to geographical sphere of production when possible, but it is expected that particular geographical locations will have greater textual output.</li> </ul>
Mark-up	<ul style="list-style-type: none"> <li>• Texts will include a mark-up scheme that codes time, place, interest group, &amp; genre.</li> <li>• Texts will be geo-spheres of production and tagged with GPS location.</li> </ul>
Annotations	<ul style="list-style-type: none"> <li>• POS tagging</li> <li>• Semantic tagging</li> </ul> <p>Ecological Deixis Marker</p> <ul style="list-style-type: none"> <li>• Place names</li> <li>• Geological features (e.g. river, mountain, wash)</li> <li>• Animal references</li> <li>• Cardinal points</li> </ul>
Reference	<ul style="list-style-type: none"> <li>• Reference corpora for contrastive analysis, e.g. keyword analysis</li> </ul>

This type of ecology-informed corpus platform could further corpus linguistics' search for context, an area of criticism of corpus approaches. While corpus linguistics has increased its attention to and appreciation of context, ecological approaches are defined by the value offered to context and the interrelationship between language and environment. For Kramsch (2002), an ecological approach and the attention to language and context it produces allow a new perspective for understanding the complexity of language and its reciprocal relationship with environment. Thus, as ecology analyzes relationships between humans and environment, it is natural that ecological research values context and situatedness. Here arises the opportunity for

CL and ecological approaches to connect. Thus, the Eco-Corpus and its integration of CL, GIS and ecological principles could further research into the reciprocal, dynamic relationship between language and context. A trans-disciplinary approach bridging ecological approaches with CL, in my opinion, represents a breakthrough in ecolinguistics while also making contributions to CL and GIS and could yield important insights into environmental discourse.

Another potential extension to the approach could be through the application of a critical discourse analytic (CDA) approach to the current ecological debate and ecological discussions in general. Present within the Rosemont Copper Mine debate are issues of power through the production and normalization of discourses which legitimize actions which cause irrevocable ecological damage. While this dissertation was guided by an ecolinguistic framework, employing a CDA approach could also yield important insights into environmental discourse.

It is important to note, however, that the approach to discourse analysis modeled in this dissertation may be applied to discourses beyond matters of the environment. Other discourses that are embedded and situated within particular geographical contexts could also be explored through the approach modeled in previous chapters. For example, a study of immigration discourses from the U.S.-Mexico borderlands to state and national capitals and from a variety of interest groups could be pursued through a similar corpus-based discourse analytic approach to display variation in immigration-related texts from different locations and groups. Applying the approach modeled in this dissertation could yield insights into the discursive construction of immigration and immigrants and how proximity to the U.S.-Mexico border influences how people write and speak about the issue.

**Place-based pedagogy and eco-literacy.** The corpus-aided pedagogical approach modeled in chapter 4 could be extended through the development of an interactive online

platform that integrates CL and GIS data into a learning platform. Place-based games in language education have received attention recently with games created using augmented reality platforms such as ARIS Games ([arisgames.org](http://arisgames.org)). Through the use of the ARIS online platform, games can be created, shared, and played that integrate maps, GPS codes, images, videos, narratives, etc. for the creation of interactive game spaces for a variety of learning purposes. For example, Mentira, developed at the University of New Mexico for Spanish language courses, uses the narrative of a murder mystery situated within a neighborhood near the university's campus to produce "the context and social conditions for meaningful interaction in Spanish with simulated characters, other players, and local citizens" ([mentira.org](http://mentira.org)). Another augmented reality game was developed by researchers at the University of Arizona's Center for Educational Resources in Culture, Language, and Literacy (CERCLL), American Indian Language Development Institute (AILDI), and several indigenous communities in the southwest that integrated language and cultural knowledge of indigenous communities into a mobile place-based game for language revitalization efforts (Reinhardt, et. al. 2015).

There are parallels between the game interface of these Aris games and the interface of CartoDB, the online mapping platform used in chapter 3. For example, in CartoDB, a map author can add annotations such as images, hyperlinks, video, author information, GPS location, concordance lines of keywords, etc. to the points marked on the map. Users may then click on various points, zoom in and out of locations, find information about various places, or move from various points key to the issue to learn more about the topic and the places. Similarly, the Aris game platform allows players to move between different points within the game space, find clues and information to advance the game, watch videos of members of the local community, etc. While chapter 4 evolved to become more focused on discussions of language variation in the

two interest group corpora, a core motivation and underlying impetus for the project began with a desire to engage the class of L2 writers, students not from nor perhaps especially familiar with Tucson and southern Arizona, with an issue of local importance to help them understand, engage, and enter their new campus community. The Santa Ritas looming presence over our campus and the relevance of the issue to our city and community prompted the selection of the mine for the study, and although the initial vision of a dynamic corpus-informed platform was not fully realized, further development of an augmented reality game platform enhanced with maps, images, videos, corpus data, etc. for an L2 writing context remains a research goal. For example, imagine the same L2 participants logging into a game space at the beginning of a class meeting or outside of the classroom. Students could zoom to the proposed site of the mine, take a virtual tour of the location, view images of flora and fauna, and receive links to texts pertaining to the debate and the site. Glosses present on the map could also contain corpus-attested keywords, information about rhetorical strategies, examples of rhetorical strategies from the debate, and additional corpus information derived from corpora of relevant interest groups. Further development of placed-based pedagogy seems quite obtainable.

This type of corpus-informed place-based learning platform could also be implemented to develop ecological awareness, i.e. eco-literacy, for users. While heightening eco-literacy was not an explicit goal of the study reported in chapter 4, it seems likely that a by-product of the intervention was increased ecological awareness on the part of participants. These students engaged in the study of an important local ecological issue and participated in numerous classroom discussions about the environmental impact the mine would have on the local ecosystem. Other assignments could be developed to follow the initial unit on environmental rhetoric to further student engagement with environmental issues. Thus, an integrated place-

based ecological curriculum could be designed to advance writing ability while simultaneously engaging with ecological issues and heightening ecological awareness.

Beyond the development of Aris and CartoDB informed learning platform, there are additional opportunities for development of a corpus of academic texts for use within the classroom. A newly-developed academic corpus could include texts from across disciplines and from freshman to graduate writers. The corpus could reflect an interdisciplinary structure produced in the Michigan Corpus of Upper Level Student Papers (MICUSP), a corpus no longer maintained and updated with additional texts. While MICSUP included only successful upper level writing from upper level undergraduates and graduate students, a new academic corpus could include texts from freshman on through professors from multiple departments and various text types, e.g. proposal, reports, etc. Such a corpus would facilitate research into the writing practices pursued within a university ecology while also serving as a tool for student writers.

**Final comments.** In the Rosemont Copper Mine debate, the company's texts reflect a culturally-dominant discourse of the ever-present need for economic growth that produces a linear conceptualization of progress and development, a conceptualization marked by the underlying rhetorical themes of legitimacy, dominion, and inevitability. These rhetorical themes and their linguistic manifestations go largely unnoticed and unquestioned, and evidence of their production and perpetuation can be regularly witnessed throughout environmental discourse in general and this debate in particular. For example, the Wikipedia entry for Rosemont Copper in its second sentence states, "The project is located in Pima County, Arizona in an *undeveloped* area approximately 30 miles southeast of Tucson" and the Environmental Protection Agency in a recent release stated that Rosemont Copper is permitted to "take" an ocelot. In the first sample, the eastern slopes of the Santa Rita Mountains is marked as "undeveloped", prompting a

subconscious acceptance of the need for development that shapes our perception of the mine and the mountain. In fact, the most common collocate of *undeveloped* in the Corpus of Contemporary American English (COCA) is *land*, displaying the common cultural linguistic presentation of *land* and a perceived need of its *development* for economic purposes. In the second sample, our disconnect and the incongruence of language use with reality is evident in the generally benign verb of *take*, a legal term meaning an actor is allowed to pursue actions that will result in the harassment, displacement, and potential harm of an endangered species. Interestingly, within the COCA, the most common collocate of *take* is *care*. Thus, a lexical item and its common collocate which is regularly used as a warm farewell is instead deployed in a meaning that legitimizes harassment, displacement, and possible death. These two examples further display the overt and covert discursive productions that shape our relationships and perceptions of the environment and environmental issues and are additional evidence to the rhetorical constellations presented in chapters 2 and 3 of the cultural linguistic practices that continue to produce a disconnect between humans and the environment and that conspire to limit our collective ability to think and act towards the environment in more sustainable ways.

Although this dissertation explored only one environmental issue and its associated discourse, these linguistic practices are not confined to southern Arizona and the Rosemont Mine or Rosemont Copper Company and the Save the Scenic Santa Ritas organization. Only a cursory review of texts discussing the Keystone XL pipeline in the American Midwest, hydraulic fracturing across the Marcellus Shale region, oil exploration along the Gulf of Mexico coast, or mining in Bristol Bay, Alaska will reveal rather similar discursive practices that present the environment in a manner that normalizes exploitation and dominion to serve economic interests. Within these controversies, the rhetoric of purported “development” continues essentially

unabated despite ample evidence of the damage these actions have on the environment. This dissertation was an attempt to study one environmental controversy in hopes of developing a methodology and framework that may help to advance the linguistic turn in environmental studies called for by Mülhäusler and Harre (1991, p.1) while meeting the challenge from M.A.K. Halliday (1990/2001) for increased trans-disciplinary attention to matters of the environment and increased scrutiny to language of the environment from applied linguists. This application of corpus linguistics, ecolinguistics/ecology, and GIS is an attempt to reach the trans-disciplinary perspective necessary to reach and realize “new forms of activity” (p. 176) that may help address the ecological crisis our world faces.

## APPENDIX

Appendix A: Top 70 Words in Rosemont Mine Truth Corpus

Rank	Frequency	Token	Rank	Frequency	Token
1	5279	the	34	262	project
2	2161	to	35	252	proposed
3	2097	of	36	230	this
4	1571	and	37	223	environmental
5	1384	in	38	219	at
6	1285	a	39	218	are
7	1240	s	40	217	would
8	1044	that	41	203	arizona
9	1015	augusta	42	197	u
10	956	rosemont	43	192	mining
11	837	mine	44	190	national
12	813	is	45	186	corps
13	774	on	46	180	plan
14	715	for	47	179	was
15	579	copper	48	177	impact
16	537	by	49	173	stated
17	449	its	50	171	pit
18	439	it	51	171	states
19	438	will	52	166	final
20	435	has	53	159	resource
21	429	forest	54	156	epa
22	415	be	55	155	have
23	377	company	56	154	santa
24	358	as	57	152	decision
25	348	million	58	149	also
26	331	with	59	146	based
27	312	service	60	144	or
28	311	from	61	140	army
29	301	hudson	62	138	open
30	297	an	63	138	public
31	297	water	64	137	coronado
32	294	permit	65	135	tucson
33	275	not	66	133	said
34	262	project	67	132	construction
35	252	proposed	68	132	last
36	230	this	69	129	admin
37	223	environmental	70	129	could

## Appendix B: Top 70 Words in Rosemont Copper Corpus

<b>Rank</b>	<b>Frequency</b>	<b>Token</b>	<b>Rank</b>	<b>Frequency</b>	<b>Token</b>
1	1251	the	36	52	forest
2	582	of	37	50	service
3	569	and	38	50	state
4	468	to	39	48	at
5	277	rosemont	40	48	million
6	250	in	41	46	about
7	244	a	42	46	permit
8	215	copper	43	45	county
9	212	for	44	43	plan
10	183	s	45	42	not
11	181	is	46	42	pima
12	151	augusta	47	41	draft
13	145	will	48	41	year
14	141	on	49	40	mining
15	136	that	50	39	have
16	129	project	51	39	more
17	126	with	52	39	our
18	109	by	53	38	been
19	99	as	54	38	usfs
20	95	company	55	37	decision
21	86	be	56	37	eis
22	84	or	57	37	information
23	80	arizona	58	36	local
24	78	has	59	36	than
25	70	an	60	35	corporation
26	69	this	61	35	its
27	68	are	62	35	review
28	66	from	63	35	which
29	65	mine	64	33	azc
30	64	tucson	65	33	under
31	61	process	66	32	agencies
32	58	water	67	32	com
33	57	environmental	68	32	final
34	55	resource	69	32	president
35	55	we	70	32	Public

## Appendix C: Top 50 Semantic Tags in Rosemont Mine Truth Corpus

<b>Rank</b>	<b>Tag</b>	<b>Freq</b>	<b>Rank</b>	<b>Tag</b>	<b>Freq</b>
1	Z5	20,402	26	T1.1.3	543
2	Z99	3,444	27	W3	543
3	Z8	2,146	28	Z3	484
4	I1.1	1,533	29	S5+	481
5	Q2.2	1,240	30	N4	476
6	A1.1.1	1,216	31	M7	470
7	N1	1,121	32	W5	452
8	I4	1,101	33	Z6	424
9	T1.3	1,020	34	N5++	421
10	Q2.1	1,017	35	I2.2	373
11	Z1	965	36	O2	368
12	A9+	917	37	O1.2	344
13	M6	858	38	M2	342
14	A3+	854	39	A11.1+	334
15	Z2	845	40	X2.4	331
16	Q1.2	799	41	G1.1	304
17	A2.2	789	42	I1	304
18	A7+	784	43	M1	302
19	S7.1+	697	44	N5+	296
20	S7.4+	690	45	G2.1	295
21	A9-	640	46	N5	283
22	S8+	618	47	L2	277
23	X7+	577	48	S6+	277
24	I2.1	564	49	I1.2	264
25	O1.1	547	50	A2.1+	260

## Appendix D: Top 50 Semantic Tags in Rosemont Copper Corpus

<b>Rank</b>	<b>Tag</b>	<b>Freq</b>	<b>Rank</b>	<b>Tag</b>	<b>Freq</b>
1	Z5	4,903	26	S7.1+	134
2	Z99	831	27	S7.4+	126
3	A1.1.1	464	28	N5	111
4	Z8	353	29	P1	108
5	N1	350	30	W5	98
6	Z2	252	31	N5++	96
7	O1.1	236	32	A7+	95
8	A3+	210	33	I2.2	95
9	I1.1	209	34	G1.1	90
10	T1.3	206	35	N4	87
11	A9+	197	36	Z3	85
12	A2.2	190	37	I1	84
13	Q1.2	189	38	A1.8+	81
14	Z1	188	39	O2	81
15	I2.1	185	40	G2.1	80
16	M6	183	41	A2.1+	78
17	X7+	180	42	N5.1+	77
18	Q2.1	166	43	M1	76
19	S8+	161	44	X2.4	76
20	T1.1.3	153	45	A4.1	75
21	Q2.2	152	46	I3.1	67
22	M7	151	47	M2	67
23	A9-	150	48	A6.1-	65
24	S5+	138	49	N5+	65
25	I4	135	50	W3	65

## Appendix E: Top 50 Parts of Speech Tags in Rosemont Mine Truth Corpus

<b>Rank</b>	<b>Tag</b>	<b>Freq</b>	<b>Rank</b>	<b>Tag</b>	<b>Freq</b>
1	NN1	13,577	26	VHZ	13,577
2	JJ	5,953	27	NNU	5,953
3	AT	5,329	28	PPH1	5,329
4	II	4,658	29	VBI	4,658
5	NN2	4,110	30	DD1	4,110
6	NP1	3,827	31	CS	3,827
7	VVI	2,023	32	XX	2,023
8	IO	1,884	33	IW	1,884
9	CC	1,731	34	MD	1,731
10	VVN	1,681	35	II21	1,681
11	AT1	1,541	36	VBR	1,541
12	TO	1,260	37	VBDZ	1,260
13	RR	1,193	38	NN	1,193
14	VVG	1,146	39	CSA	1,146
15	VVD	1,097	40	DDQ	1,097
16	VVZ	1,065	41	NNT2	1,065
17	MC	1,061	42	ND1	1,061
18	VM	1,039	43	NNT1	1,039
19	VV0	964	44	RG	964
20	CST	936	45	CSN	936
21	VBZ	812	46	RMT	812
22	IF	701	47	DAR	701
23	NPM1	629	48	VBN	629
24	APPGE	617	49	DD	617
25	NP2	477	50	MC1	477

## Appendix F: Top 50 Parts of Speech Tags for Rosemont Copper Company

<b>Rank</b>	<b>Tag</b>	<b>Freq</b>	<b>Rank</b>	<b>Tag</b>	<b>Freq</b>
1	NN1	3,496	26	DD1	92
2	JJ	1,285	27	VBI	86
3	AT	1,236	28	NPM1	80
4	II	943	29	VHZ	75
5	NN2	938	30	FO	71
6	NP1	912	31	VBR	68
7	CC	649	32	NP2	63
8	IO	547	33	GE	60
9	VVN	460	34	PPIS2	53
10	VVI	404	35	XX	42
11	MC	318	36	CS	41
12	AT1	302	37	CSA	41
13	VVG	272	38	NNT2	40
14	RR	271	39	DAR	39
15	TO	255	40	DDQ	39
16	VV0	252	41	VBN	38
17	VM	208	42	MD	37
18	VVZ	195	43	II31	35
19	IF	193	44	NNT1	35
20	VBZ	184	45	CSN	34
21	VVD	136	46	DB	32
22	NNU	130	47	DD2	32
23	CST	119	48	II21	32
24	IW	115	49	VH0	31
25	APPGE	110	50	MC1	29

## Appendix G: Rosemont Mine Truth Data

Tag Code	Tag Description	Freq in RMT	Freq In RCC	LogL Stat	Examples
PPH1	3rd person singular personal pronoun (it)	417	20	79.05	<i>e.g. it is likely, it is inadequate, it would be...</i>
Q2.2	speech acts	1240	152	66.23	<i>proposed, proposing, recommends, announce, predict, report, reports, assert, claim, denial, demand</i>
I1.1	money and pay	1533	209	60.74	<i>Augusta, funds, invest, investors, investing, shareholder, asset, capital, taxpayer</i>
L2	living creatures	277	11	59.40	<i>wildlife, fish, endangered species, insects, birds, animals, prey, jaguar, ocelot, amphibians, invertebrates, frog, cuckoo, bald eagles, predator, cat</i>
I4	industry	1101	135	58.77	<i>mine, copper mine, mineral, and names of other mines</i>
VVD	past tense lexical	1097	136	56.93	<i>reporting past events in the debate</i>
CST	that as conj	936	119	45.45	
A7+	likely	784	95	43.10	<i>may, could, potentially, can, would, clearly, may, likely, confirmed, potential, certainly</i>
Z8	pronouns	2146	353	40.21	
A1.1.2	damaging and destroying	132	3	36.58	<i>damage, destruction, destroying, obliterate, charred, violate, smash, harm, irreparable, collapse, broken, rips, devastating</i>
G3	warfare	236	17	30.95	<i>Army Corps of Engineers</i>
W3	geographical terms	543	65	30.75	<i>aquifer, mountains, creek, desert, land, stream, wetland, pond, canyon, lake, grasslands, slope, wells, valley, tributary, river, forest, mount, butte, crest (not included:washes)</i>
NPM1	singular month noun	629	80	30.51	<i>June, July, etc.</i>
CS	subordinating conj	371	41	25.14	<i>if, when, whether</i>
NP2	plural proper noun	477	63	20.80	<i>false; these are misidentified genitives</i>

Z6	negative	424	54	20.49	<i>not, no way, no</i>
X9.2-	success and failure	97	4	20.33	<i>failed, failure, fails, unsuccessful, lost, defeats</i>
O4.2+	judgment of appearance: beautiful	146	10	20.24	<i>attractive, clean, scenic, appealing, spectacular, lush</i>
N5.1-	part	208	19	19.88	<i>shares (investment), section 404, part, run out, incomplete, devoid, segment</i>
Q2.1	speech: communicative	1017	166	19.82	<i>noted, stated, said, notes, comment, request, discuss, says, consult, appeal, told</i>
A11.1+	important	334	40	18.89	<i>meaningful, key, serious, significant, well known, major, value, integral, substantive, urgency, crucial</i>
A1.7-	no constraint	247	26	18.46	<i>discharge, release, releasing, releases, released, regulatory</i>
E3-	violent/angry	123	8	17.99	<i>threatens, toxic, threatened, poisonous, acrimonious, poison, harass</i>
II	general preposition	4658	943	16.28	
Z7	if	112	8	14.83	<i>if, whether or not</i>
N6+	frequent	133	11	14.69	<i>annually, repeatedly, annual, again, often</i>
A5.2-	evaluation: false	71	3	14.66	<i>lie, false, misleading, unfounded, falsely</i>
XX	not, n't	321	42	14.39	
E2-	dislike	86	5	13.98	<i>objection, objections</i>
MD	ordinal number	289	37	13.77	<i>last year, last month, last time, fourth quarter</i>
II21	general preposition	246	32	11.21	<i>according to, out of, along with, due to</i>
VVZ	s form of lexical verb	1065	195	10.23	<i>many reporting verbs, e.g. exaggerates, reports, suggests</i>
Z4	discourse bin	104	10	9.15	<i>rather, in other words, however, in this case, on the other hand, at best, obviously</i>
VHD	had (past)	54	3	9.13	<i>past perfect structures</i>
A5.1-	evaluation: bad	44	2	8.66	<i>severe, flawed, adversely, bad, severity, shortcomings, terrible, detrimental</i>
S1.2.1-	formal/unfriendly	52	3	8.51	<i>formal review, takeover, hostile (referencing hudson takeover)</i>

B2-	disease	34	1	8.51	<i>contaminate, contamination, high blood pressure, Cancer, lesions, syndrome, fractures (side effects of the mine toxins)</i>
VVI	infinitive	2023	404	8.43	<i>to approve, to get, to protect</i>
X2.1	thought/belief	161	20	8.31	<i>wonders, considers, presumably, outlook, presumed</i>
VBD Z	was	189	25	8.21	<i>reporting on past events, genre sensitive</i>
AT1	singular article	1541	302	7.98	<i>a, an</i>
S8-	hindering	105	11	7.92	<i>disruption, blow, adverse, oppose, prevent, frustrate, obstacles, setbacks</i>
L3	plants	57	4	7.70	<i>trees, willows, cottonwoods, vegetation, mesquites, (hedge fund), grass, sycamore (smelters), saguaro</i>
APP GE	poss. pronoun	617	110	7.18	<i>its, our</i>
N5-	quantities: little	136	17	6.91	<i>slight, defecit, inadequate, reduce, insufficient, decreased, underestimate, depletion, scant, underestimated</i>
S7.4+	allowed	690	126	6.76	<i>permit, permitting, permits, allow, approve</i>
VDD	did	38	2	6.72	<i>all 38 are 'did not' accusation against the company</i>
VDZ	does	60	5	6.56	<i>58 of 60 instances "does not" in relation to RCC acts</i>
JJ	general adj	5953	1285	6.49	<i>Accounted for in sem tags</i>
VHZ	has	437	75	6.48	<i>present perfect structures</i>
G1.2	politics	85	9	6.28	<i>lobbying, elected, activist, democratic, policymakers, vote, voting, congressmen, senators, poll, polled, pollster</i>
Z3	other proper names	484	85	6.23	<i>references to Forest Service, Daily Star, New York Times, District, National Forest, Arizona water, Huachuca, University of Arizona</i>
A12-	difficult	65	6	6.11	<i>problems, challenges, crisis, difficult, complexity, problematic, complicate</i>
NNL 1	singular locative noun	27	1	6.03	<i>primarily "Mount Polley Mine" references</i>
X7-	unwanted	43	3	5.84	<i>reject, unsolicited, unacceptable, ill-conceived, rejected</i>

H4	residence	107	13	5.84	<i>residents, lived, habitat, occupying, accommodate</i>
K5.1	sports	57	5	5.79	<i>anglers, hunters, fishing (bats/animal, final/decision, exercise/decision)</i>
RRQ	wh- general	76	8	5.68	<i>primarily rhetorical questions</i>
M4	sailing, swimming, water	56	5	5.54	<i>port, flows, aquatic (launch)</i>
S6+	strong obligation/necessity	277	45	5.53	<i>have to, promises, essential, needed, should, must, need</i>
S9	religion & supernatural	34	2	5.47	
Z1	personal names	965	188	5.32	<i>national forest, Las cienegas, Scenic Santa Ritas, New York City, and many names</i>
VHI	have, infinitive	74	8	5.25	
H1	architecture, house and buildings	254	41	5.23	<i>construction, facilities, build, building, shed, dam, facility</i>
O1.2	substances: liquid	344	59	5.12	<i>water, waters, waterways, seepage,</i>
NN	common noun, neutral for number	169	25	5.05	<i>most are "Corps" of Army Corps of Engineering</i>
T4-	time: late	73	8	5.03	<i>delay, delayed, delays, delaying, late</i>

## Appendix H: Rosemont Copper Company Data

<b>Tag Code</b>	<b>Tag Description</b>	<b>Tag Freq in RCC</b>	<b>Tag Freq in RMT</b>	<b>LogL Stat</b>	<b>Examples</b>
P1	education	108	89	125.73	<i>school, education, students, Community college, university, Ph.D., study, alumni, scholars, educate</i>
CC	coordinating conjunction	649	1731	98.33	<i>and, or</i>
A1.1.1	general actions/making	464	1216	75.10	<i>project, to close, closing, production, to create, operating, made, pursue, construct, closure, committed, done, operations, installed, drill, drilled, drilling, implementation, projects, conducted</i>
FO	formula	71	82	59.53	<i>US \$40, 100%-owned, 4.5%, 1.1 billion</i>
O1.1	substances and materials: liquid	236	547	56.87	<i>copper, metals, silver, gold, deposit, diamond, ore</i>
Y2	information technology and computing	48	49	45.94	<i>email, program, website, online, websites, computers</i>
PPIS2	first person plural personal pronoun	53	66	40.64	<i>We are, we will, we have, we expect, we continue, we know</i>
S1.2.5+	toughness: strong/weak	24	12	39.13	<i>robust, robustness, strength, strict</i>
A5.1+	evaluation: good	58	89	33.28	<i>enhances, enhanced, merit, defendable, developed, positive, favorable, great, progress, world class, improve, positive, defensible</i>
A4.2+	particular/general; detail	43	61	27.74	

NN1	singular common noun	3496	1357 7	26.54	<i>debt, copper, fund, credit, term, sheet, loan, agreement, balance sheet, closing date, project, option, maturity, facility, market, period, pricing</i>
N3.5	measurement: weight	26	26	25.40	<i>tons, ounces, lbs.</i>
MCMC	hyphenated number	14	7	22.83	<i>44-101, 2012-2013</i>
N1	numbers	350	1121	21.48	<i>million, two, seven, one, billion, quarter, 2010</i>
N5	quantities	111	283	19.82	<i>any, 2%, portion, sufficient, half, 65%, both, net, 1/3</i>
A6.1+	comparing: similar	52	100	19.82	<i>agreement, consistent, agreed, compliant, in keeping with, agree, consistently, compliance, compatibility</i>
IO	of (as preposition_	547	1884	19.47	<i>mostly used in post-modification of NP heads, e.g. a fee of, common shares of..</i>
II31	general preposition	35	55	19.34	<i>in connection with, in accordance with, as, in addition to, in keeping with</i>
A4.1	generally kinds, groups, examples	75	170	19.24	<i>NEPA, aspect, Tucson</i>
S1.1.1	social actions, states, and processes	32	49	18.42	<i>initiation, traditional, treat (sulfide)</i>
A6.2+	comparing: usual	37	64	17.35	<i>common, customary, normal, basic, generally</i>
Q1.3	telecommunic ations	13	9	17.30	<i>telephone, email, operator, fax, inbox</i>
IW		115	316	15.48	<i>with, without</i>
I3.1	work and employment	67	158	15.35	<i>work, role, teamwork, jobs, career, roles, employ, payroll, staff, personnel, working, workforce</i>
I2.1	business	185	564	15.23	<i>corporation, company, offices, offices, stock exchange, infrastructure, Ltd,</i>

					<i>businesses</i>
NNO		12	9	15.06	<i>million, billion, thousand</i>
MC	numbers	318	1061	14.66	<i>2010, 18, 55000, seven</i>
A1.8+	inclusion	81	207	14.34	<i>include, including, includes, contains, consist, consists, comprehensive</i>
O4.4	shape	33	60	14.09	<i>power line, transmission line, angle, footprint, distorted</i>
X6	deciding	17	20	13.95	<i>estimated, estimates, estimation (referring to money to be generated)</i>
X2.6+	expected	65	157	13.85	<i>expected, expects, due, anticipates, look forward</i>
E2++	Like	9	5	13.82	<i>preferred</i>
S5-	not part of a group	13	12	13.71	<i>personal, independently, neutral, alone</i>
II33		12	11	12.73	<i>as well as</i>
X2.3+	learning	7	3	12.35	<i>learn, learned</i>
E2+	Like	13	14	11.77	<i>precious, like, enjoyment, applauding</i>
O3	electricity and electrical equipment	8	5	11.42	<i>power plant, electric, recharged, electronic</i>
Z2	geographical names	252	845	11.18	<i>Vancouver, Bermuda, New York, Shanghai, Sydney, Canada, Quebec, United States, San Francisco, Arizona, Denver, University of Arizona, Arizona State University, Pima County, Nevada, Venezuela, Ecuador, Vancouver, Toronto, Frankfurt, London, Washington DC, South America, Davidson Canyon, Ontario</i>
X5.1+	attentive	27	50	11.12	<i>concentrate, focused, focus, concentrates</i>
M7	places	151	470	11.05	<i>international, regional, area, local, province, site, district</i>

X7+	wanted	180	577	10.99	<i>required, strategy, policy, schedule, plan, target</i>
A1.5.1	using	34	71	10.77	<i>used, using, consuming, utilize</i>
K4	drama, the theatre	16	23	10.13	<i>producer, performance</i>
VVN		460	1681	8.91	<i>past participle of lexical verb</i>
Y1	science and technology	60	163	8.46	<i>engineers, technical, technologies, hydrology, engineering, state-of-the-art, scientific</i>
X9.2+	success	20	37	8.25	<i>achievement, success, on track, win, successfully</i>
X3.4	sensory: sight	15	24	8.05	<i>see, visual, looked at, sees</i>
N5.1+	entire, maximum	77	225	7.88	<i>maximum, all, 100%, complete, total, maximizing, thorough, every, full, as a whole, to the full</i>
A13.4	degree: approximators	43	109	7.83	<i>up to, approximately, about, in a way, nearly, almost, closely</i>
N5--	quantities: little	12	17	7.76	<i>less than, fewer (discussing water resources to be used)</i>
X2.5+	understanding	14	23	7.19	<i>understood, inferred, realize, understand</i>
A1.2+	suitable	12	18	7.13	<i>qualified, accordingly, relevant, appropriate</i>
NNU	unit of measure for number	130	424	7.07	<i>2%, , 6000, 100%</i>
DD2	plural determiner: these, those	32	77	6.90	<i>these, those</i>
N6-	infrequent	11	16	6.84	<i>once, rare, lightly</i>
H3	areas around houses	15	27	6.53	<i>local</i>

DAR	comparative after determiner: more, less, fewer	39	105	5.70	<i>more, less, fewer</i>
NNB	preceding noun of a title	15	29	5.65	<i>Dr., Mr.</i>
VBN	been	38	102	5.62	<i>been completed, been submitted, been notified</i>
T3-	time: new and young	42	122	4.42	<i>new, recent, innovative, modern</i>
S5+	belonging to a group	138	481	4.41	<i>federal, organization, institutions, joint, public, team</i>
T2+	time: beginning	59	184	4.27	<i>starting, commencement, foundation, beginning</i>
VBR	are	68	218	4.15	
T1.1.3	future	153	543	4.10	<i>will, going to, future</i>
IF	for (as preposition)	193	701	4.02	

## REFERENCES

- Abrahamson, E. & Park, C. (1994). Concealment of negative organization outcomes: An agency theory perspective. *The academy of management journal*, 37 (5) p. 1303-1334.
- Ädel, A. 2010. Using corpora to teach academic writing: Challenges for the direct approach. In M. Campoy-Cubillo, B. Bellés-Fortuño, and M. Gea-Valor (Eds.) *Corpus-based approaches to English language teaching*. London: Continuum.
- Anthony, L. 2014. AntConc 3.4.3 [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net/>
- Aston, G. (1997). Enriching the learning environment: Corpora in ELT. In A. Wichmann, S. Fligelstone, T. McEnery, and G. Knowles (Eds.), *Teaching and language corpora* (51-65). London: Longman.
- Aston, G. (2001). *Learning with corpora*. Bologna: CLUEB.
- Alexander, R. (2002). Everyone is talking about ‘Sustainable Development’. Can they all mean the same thing? Computer discourse analysis of ecological texts. In Fill, A. Penz, H. & Trampe, W. , *Colourful green ideas: Papers from the conference 30 years of language and ecology*. Bern: Peter Lang.
- Alexander, R. J. (2009). *Framing discourse on the environment a critical discourse approach*. Routledge, New York and London.
- Alexander, R. & Stibbe, A. (2013). From the analysis of ecological discourse to the ecological analysis of discourse. *Language Sciences*.
- Anthony, L. (2014). AntConc (Version 3.4.3m) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net/>
- Augoustinos, M., Crabb, S. & Sheperd, R. (2010). Genetically modified food in the news: media representations of the GM debate in the UK. *Public understanding of science*.
- Ayers, D. (2010). Turning toward place, space, and time. In D. Bodenheimer, J. Corrigan, & T. Harris (eds.) *The spatial humanities: GIS and the future of humanities scholarship*. Indiana University Press: Bloomington.
- Baker, P. (2010). Representations of Islam in British broadsheet and tabloid newspapers 1999-2005. *Language and Politics*, 9 (2), 310-338.

- Baker, P., Gabrielatos, C. & McEnery, T. (2013). Sketching Muslims: A corpus-driven analysis of representation around the word “Muslim” in the British press, 1998-2009. *Applied Linguistics* 34(3), 255–278.
- Bang, J. C., & Trampe, W. Aspects of an ecological theory of language. *Language Sciences*, (0) doi:<http://dx.doi.org.ezproxy2.library.arizona.edu/10.1016/j.langsci.2013.08.009>
- Bernadini (2002). Exploring new directions for discovery learning. In B. Kettemann and G. Marko (Eds.), *Teaching and learning by doing corpus analysis: Proceedings of the fourth international conference on teaching and language Corpora* (165-183). Rodopi: Amsterdam.
- Bevitori, C. (2011) ‘Jumping on the green bandwagon’: The discursive construction of GREEN across ‘old’ and ‘new’ media genres at the intersection between corpora and discourse. In *Proceedings of the Corpus Linguistics Conference*, Birmingham: University of Birmingham.
- Bhatia, V. K. (1993). *Analysing genre: Language use in professional settings*. New York: Longman.
- Biber, D., Reppen, R., & Friginal, E. (2010). Research in corpus linguistics. In R. Kaplan (Ed.), *The oxford handbook of applied linguistics* (2<sup>nd</sup> ed.). Oxford University Press.
- Blinder, S. & Allen, W. (2014). Constructing immigrants: Portrayals of migrant groups in British newspapers, 2010-2012. Centre on Migration, Policy, and Society: University of Oxford.
- Boulton, A. (2010). Learning outcomes from corpus consultation. In M. Jaén, F. Valverde, and M. Pérez (Eds.), *Exploring new paths in language pedagogy* (129-144). London: Equinox.
- Bodenheimer, D., Corrigan, J., & Harris, T. (2010). *The spatial humanities: GIS and the future of humanities scholarship*. Indiana University Press: Bloomington.
- Carbaugh, D. ‘The mountain’ and ‘the project’: Dueling depictions of a natural environment. In A. Fill & P. Mühlhäusler (Eds.), *The ecolinguistics reader*, (124-142). London: Continuum.
- Couto, H. H. Ecological approaches in linguistics: A historical overview. *Language Sciences*, (0) doi:<http://dx.doi.org.ezproxy2.library.arizona.edu/10.1016/j.langsci.2013.08.001>
- Chambers, A. (2005). Integrating corpus consultation in language studies. *Language Learning and Technology*, 9 (2), 111-125.

- Chambers, A. (2007). Popularising corpus consultation by language teachers and learners. In Hidalgo, E., Quereda, L., and Santana, J. (Eds.), *Corpora in the foreign language classroom: Selected papers from the sixth international conference on teaching and language corpora* (3-17). Amsterdam: Rodopi.
- Charles, M. (2007). Reconciling top-down and bottom-up approaches to graduate writing: Using a corpus to teach rhetorical functions. *Journal of English for Academic Purposes*, 6(4), 289–302.
- Charles, M. (2011). Using hands-on concordancing to teach rhetorical functions: evaluation and implications for EAP writing. In A. Frankenberg-Garcia, L. Flowerdew, and G. Aston (Eds.) *New trends in corpora and language learning* (26-43). Continuum: London.
- Charles, M. (2012). Proper vocabulary and juicy collocations: EAP students evaluate do-it-yourself corpus-building. *English for Specific Purposes*, 31, 93–102.
- Charles, M. (2013). Getting the corpus habit: EAP students long term use of personal corpora. *English for Specific Purposes*, 35, 30-40.
- Clatworthy, M. & Jones, M. (2003). Financial reporting of good news and bad news: evidence from accounting narratives. *Accounting and business research*, 33(3), p. 171-185.
- Cobb, T. (1997). Is there any measurable learning from hands-on concordancing?. *System*, 25(3), 301-315.
- Cobb, T. (1999). Breadth and depth of lexical acquisition with hands-on concordancing. *Computer Assisted Language Learning*, 12(4), 345-360.
- Conrad, S. (2000). Will corpus revolutionize grammar teaching in the 21<sup>st</sup> century?. *TESOL Quarterly* 34(3), 548-560.
- Carbaugh, D. ‘The mountain’ and ‘the project’: Dueling depictions of a natural environment. In A. Fill & P. Mühlhäusler (Eds.), *The ecolinguistics reader*, (124-142). London: Continuum.
- Chowla, S. (2001). Linguistic and philosophical roots of our environmental crisis. . In Fill, A. & Mühlhäusler, P. (Eds.), *The ecolinguistics reader*, (115-124). London: Continuum.
- Conoscenti, M. (2011). *The Reframer: An Analysis of Barack Obama’s Political Discourse (2004-2010)*, Roma: Bulzoni.
- Couto, H. H. Ecological approaches in linguistics: A historical overview. *Language Sciences*, (0) doi:<http://dx.doi.org.ezproxy2.library.arizona.edu/10.1016/j.langsci.2013.08.001>

- Ferrari, F. (2010). Gender and identification in political discourse: the (American) woman: as and beyond a category for analysis. *Political Linguistics* 1(3), 39-54.
- Fill, A. & Mülhäusler, P. (2001). Introduction. *The ecolinguistics reader: language, ecology, and environment*. London: Continuum.
- Fitzsimmons-Doolan, S., (2009). Is public discourse about language policy really public discourse about immigration? A corpus-based study. *Language Policy*, 8(4), 377-402.
- Flowerdew, J. & Forest, R. (2009). Schematic structure and lexico-grammatical realization in corpus-based genre analysis: the case of research in the PhD literature review. In M. Charles, D. Pecorari, and S. Hunston (Eds.) *Academic writing: At the interface of corpus and discourse* (110-129). London: Continuum.
- Flowerdew, J., & Wan, A. (2006). Genre analysis of tax computation letters: How and why tax accountants write the way they do. *English for Specific Purposes*, 25(2), 133–153.
- Flowerdew, J., & Wan, A. (2010). The linguistic and the contextual in applied genre analysis: The case of the company audit report. *English for Specific Purposes* 29 (2), 78-93.
- Flowerdew, L. (1998). Corpus linguistic techniques applied to textlinguistics. *System*, 26, 541-542.
- Flowerdew, L. (2003). A combined corpus and systemic-functional analysis of the problem-solution pattern in a student and professional corpus of technical writing. *TESOL Quarterly* 37 (3), 489-511.
- Flowerdew, L. (2004). An argument for using specialized corpora to understand academic and professional English. In U. Connor & T. Upton (Eds.), *Discourse in the professions: Perspectives from corpus linguistics*, (17-39). John Benjamins Publishing.
- Flowerdew, L. (2005). An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: Countering criticisms against corpus-based methodologies. *English for Specific Purposes*, 24 (3), 321-332.
- Flowerdew, L. (2009). Applying corpus linguistics to pedagogy: A critical evaluation. *International Journal of Corpus Linguistics*, 14(3), 393-417.
- Friginal, E. (2013). Developing research report skills using corpora. *English for Specific Purposes*, 32, 208-220.

- Gavioli, L. (1997). Exploring texts through the concordancer: Guiding the learner. In A. Wichmann, S. Fligelstone, T. McEnery, and G. Knowles (Eds.) *Teaching and language corpora*. (83-100). London: Longman.
- Gavioli, L. (2001). The learner as researcher: Introducing corpus concordancing in the classroom. In G. Aston (Ed.), *Learning with corpora* (pp. 108-137). Houston TX: Athelstan.
- Gerbig, A. (1997). *Lexical and grammatical variation in a corpus: A computer-assisted study of discourse on the environment*. New York: P. Lang.
- Gilquin, G. & Jacobs, G (2006). Elephants who marry mice are very unusual: The use of the relative pronoun who with nonhuman animals. *Society and Animals*, 14(1), 79-105.
- Goatly, A. (2002). The representation of nature on the BBC world service. *Text*, 22(1), 1-27 .
- Gregory, I. & Hardie, A. (2011). Visual GISTing: bringing together corpus linguistics and geographical information systems. *Literary and linguistic computing*, 26(3): 297-314.
- Grundman, R. & Krishnamurthy, R. (2010). The discourse of climate change: A corpus-based approach. *Critical approaches to discourse analysis across disciplines*, 4(2), 125-146.
- Grundman, R. & Scott, M. (2012). Disputed climate science in the media. Do countries matter? *Public understanding of science*.
- Haig, E. (2001). A study of the application of critical discourse analysis to ecolinguistics and the teaching of eco-theory. *Studies in language and culture (Nagoya University)* 22:2:205-226.
- Halliday, M.A.K. (1990/2001). New ways of meaning: the challenge to applied linguistics. . In Fill, A. & Mühlhäusler, P. (Eds.), *The ecolinguistics reader*, (p. 175-202). London: Continuum.
- Harré, R., Brockmeir, M. & Mühlhäusler, P. (2003). *Language of environment, environment of language: a course in ecolinguistics*. London: Battlebridge.
- Henry, A. (2007). Evaluating language learners' response to web-based, data-driven, genre teaching materials. *English for Specific Purposes*, 26, 462-484.
- Henry, A., & Roseberry, R. L. (1996). A corpus-based investigation of the language and linguistic patterns of one genre and the implications for language teaching. *Research in the Teaching of English*, 30(4), 472-489.

- Henry, A., & Roseberry, R. L. (2001). A narrow-angled corpus analysis of moves and strategies of the genre: 'Letter of application'. *English for Specific Purposes*, 20(2), 153-167.
- Hyland, K. (2009). Corpus-informed discourse analysis: The case of academic engagement. In M. Charles, D. Pecorari, and S. Hunston (Eds.) *Academic writing: At the interface of corpus and discourse*. (110-129). London: Continuum.
- Horst, M., Cobb, T., & Nicolae, I. (2005). Expanding academic vocabulary with an interactive online database. *Language Learning and Technology*, 9 (2), 90-110.
- Halliday, M. A. K. & Martin, J. R. (1993). *Writing science: Literacy and discursive power* University of Pittsburgh Press.
- Harré, R., Brockmeir, M. & Mülhäusler, P. (2003). *Language of environment, environment of language: a course in ecolinguistics*. London: Battlebridge.
- Hyland, K. (1998). Exploring corporate rhetoric: Metadiscourse in the CEO's letter. *Journal of business communication*, 35(2) p. 224-245.
- Jaworska, S. (2013) The quest for the 'local' and 'authentic': Corpus-based explorations into the discursive constructions of tourist destinations in British and German commercial travel advertising. In Höhmann, D. & Spissu, M.V. (eds.) *Tourismuskommunikation. Im Spannungsfeld von Sprach- und Kulturkontakt*. Frankfurt: Peter Lang. 75-100.
- Jaworska, S. & Krishnamurthy, R. (2012) On the F-word: A corpus-based analysis of the media representation of feminism in British and German press discourse, 1990–2009. *Discourse & Society*, 23(4), 401-431.
- Johns, T. (1986). Micro-concord: a language learner's research tool. *System*, 14(2), 151-162.
- Johns, T. (1991). Should You Be Persuaded – Two Samples of Data-Driven Learning Materials. *ELR Journal*, 4 (1), 1-16.
- Jung, M. (2001). Ecological criticism of language. In Fill, A. & Mülhäusler, P. (Eds.), *The ecolinguistics reader*, (p. 270-286). London: Continuum.
- Kaur, J. & Heglemeir, K. (2005). ESL students' use of concordance in the transfer of academic word knowledge: an exploratory study. *Computer Assisted Language Learning*, 18(4), 287-310.
- Kutter, A. (2012). *Corpus-based content analysis: A method for investigating news coverage on war and intervention*. University of Stuttgart, Dept of International Relations & European Integration International Relations Online Working Paper Series.

- Lee, Y. W. D., & Swales, J. M. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes*, 25(1), 56-75.
- Leech, G. (1997). Teaching and language corpora: a convergence. In A. Wichmann, S. Fligelstone, T. McEnery, and G. Knowles (Eds.) *Teaching and language corpora* (1-25). London: Longman.
- Lijffit, J., Nevalainen & Saily, T, T., Papapetrou, P., Puolamaki, K. & Mannila, J. (2014). Significance testing of word frequencies in corpora. *Digital Scholarship in the Humanities*.
- Liu, D. & Jiang, P. (2009). Using a corpus-based lexicogrammatical approach to grammar instruction in EFL and ESL contexts. *The Modern Language Journal*, 93(1), 61-78.
- Maci, S.M. (2012a). Fast-Track Publications: The Genre of Medical Research letters. In M. Gotti (Ed.), *Academic Identity Traits* (243-262). Bern, Peter Lang.
- Maci, S.M. (2012b). The genre of medical conference posters. In M. Gotti (Ed.) *Academic Identity Traits* (283-301). Bern: Peter Lang.
- Martin, J. (1986). Grammaticalizing ecology. The politics of baby seals and kangaroos. In T. Threadgold, E. Grosz, G. Kress and M. Halliday (Eds.) *Semiotics, ideology, language*. Sydney: Sydney Association for Studies in Society and Culture (235-267).
- McCarthy, M. & Carter, R. 1994. *Language as discourse*. London: Longman.
- Milizia, D. (2009). Migration of n-grams and congrams in political speeches. In *Forms of Migrations – Migration of Forms: Atti del XXIII Convegno Nazionale AIA*. Bari: Progedit (496-514).
- Milizia, D. (2010). Keywords and phrases in political speeches. In M. Bondi and M. Scott (Eds.), *Keyness in Text* (127-145). Amsterdam: John Benjamins.
- Milizia, D. 2012. *Phraseology in Political Discourse: A corpus linguistics approach in the classroom*. Milan: LED.
- Morley, J. & Taylor, C. (2012). Us and them: How immigrants are constructed in British and Italian newspapers. In P. Bayley & G. Williams (Eds.) *European Identity: What the media say*. Oxford: Oxford University Press.
- Mülhäusler, P. (1999). *Greenspeak: A study of environmental discourse*. London: Sage Publications.

- Nash, J., & Mühlhäusler, P. Linking language and the environment: The case of Norfolk and Norfolk island. *Language Sciences*, (0)  
doi:<http://dx.doi.org.ezproxy2.library.arizona.edu/10.1016/j.langsci.2013.08.004>
- Parry, C. (2013). Augusta resources shares collapse after group leaks report suggesting EPA refusal. Retrieved from Stockhouse.com.
- Partington, A. (2003). Corpora and discourse strategies in action: From footing to fooling. In B. Lewandowska-Tomaszczyk (Ed.) *PALC 2001: Practical Applications in Language Corpora* (179-192), Frankfurt/M: Peter Lang.
- Partington, A. & Morley, J. (2004). At the heart of ideology: Word and cluster/bundle frequency in political debate. In B. Lewandowska-Tomaszczyk (Ed.) *PALC 2003: Practical Applications in Language Corpora* (179-192), Frankfurt/M: Peter Lang.
- Poole, R. (in press). Good times, bad times: A keyword analysis of the letters to the shareholders of two fortune 500 banks. *Journal of business communication*.
- Prentice, S., Rayson, P., & Taylor, P. (2012). The language of Islamic extremism: towards an automated identification of beliefs, motivations and justifications. *International Journal of Corpus Linguistics*, 17(2), 259-286.
- Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*. 13 (4), 519-549.
- Rayson, P. (2009) Wmatrix: a web-based corpus processing environment, Computing Department, Lancaster University. <http://ucrel.lancs.ac.uk/wmatrix/>
- Reinhardt, J. (2013). An applied genre analysis of office hour consultations. *International Journal of Corpus Linguistics*. 18 (3), 301-326.
- Reinhardt, J. (2010). The potential of corpus-informed L2 pedagogy. *Studies in Hispanic and Lusophone linguistics*, 3 (1), 239-251.
- Reinhardt, J., Penfield, S., Diaz, N. Holden, C. (2015). Mobile games for heritage and foreign language learning and revitalization. Retrieved from [https://www.academia.edu/11591682/Mobile\\_Games\\_for\\_Heritage\\_and\\_Foreign\\_Language\\_Learning\\_and\\_Revitalization](https://www.academia.edu/11591682/Mobile_Games_for_Heritage_and_Foreign_Language_Learning_and_Revitalization). March 28 2015.
- Römer, U. 2010. "Using general and specialized corpora in English language teaching: Past, present, and future". In M. Campoy-Cubillo, B. Bellés-Fortuño, and M. Gea-Valor

- (Eds.) *Corpus-based approaches to English language teaching* (18-38). London: Continuum.
- Rutherford, B. (2003). Obfuscation, textual complexity and the role of regulated narrative accounting in corporate governance. *Journal of management and governance*, 7, p. 187-210.
- Scott, M. & Tribble, C. (2006). *Textual patterns: keywords and corpus analysis in language education*. Amsterdam: John Benjamins.
- Sealey, A. & Oakley, L. (2013). Anthropomorphic grammar? some linguistic patterns in the wildlife documentary series life. *Text and Talk*, 33(3), 399-420.
- Shultz, B. (2001). Language and the natural environment. In Fill, A. & Mühlhäusler, P. (Eds.), *The ecolinguistics reader* (109-115). London: Continuum.
- Sinclair, J. (1987). Collins COBUILD: Collins Birmingham University international language database: English language dictionary. London: Collins.
- Soja, E. (2001). In different spaces: Interpreting the spatial organization of societies. *Proceedings from the 3<sup>rd</sup> International Space Syntax Symposium*.
- Staples, S. & Biber, D. (2014). The expression of stance in nurse-patient interactions: An ESP perspective. In M. Gotti & D.S. Giannoni (Eds.) *Corpus Analysis for Descriptive and Pedagogical Purposes: ESP perspectives* (123-142). Bern: Peter Lang.
- Subramanian, R. (1993). Performance and readability: A comparison of annual reports of profitable and unprofitable corporations. *The journal of business communication* 30 (1), p. 49-61.
- Subtirelu, N.C. (2013) 'English... it's part of our blood': Ideologies of language and nation in United States Congressional discourse. *Journal of Sociolinguistics*, 17(1), 37-65.
- Sun, Y. (2007). Learner perceptions of a concordancing tool for academic writing. *Computer Assisted Language Learning*, 20 (4), p. 323-343.
- Swales, J. M. (1991). *Genre analysis: English in academic and research settings*. Cambridge, UK: Cambridge University Press.
- Tabbert, U. (2013). *Crime through a Corpus: The linguistic construction of offenders, victims and crimes in the German and UK press*. Doctoral Thesis, University of Huddersfield.
- Tribble, C. (1997). Improvising corpora for ELT: Quick and dirty ways of developing corpora for language teaching. In B. Lewandowska-Tomaszczyk & J. Melia (Eds.), *Proceedings*

- of the first international conference on practical applications in language corpora* (303-318). Łódź, Poland: Łódź University Press.
- Tribble, C. (2000). Genres, keywords, teaching: Towards a pedagogic account of the language of project proposals. In L. Burnhard and T. McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective: papers from the third international conference on teaching and language corpora*. Frankfurt: Peter Lang.
- Tribble, C. (2001). Small corpora and teaching writing. In M. Ghadessy & A. Henry (Eds.) *Small corpus studies and ELT: Theory and practice*. (381-408). Amsterdam: John Benjamins.
- Tribble, C. & Jones, G. (1990). *Concordances in the Classroom*. London: Longman.
- Trigg, L.J. 2007. Social construction of the patient through problems of safety, uninsurance, and unequal treatment: A discourse analysis of Institute of Medicine executive summaries. MedInfo 2007, 12th International Congress Medical (Health) Informatics.
- Upton, T. & Connor, U. (2001). Using computerized corpus analysis to investigate the textlinguistic discourse moves of a genre. *English for Specific Purposes*, 20, 313-329.
- Varley, S. (2009). I'll just look that up in the concordance: integrating corpus consultation into the language learning environment. *Computer Assisted Language Learning*, 22(2), 133-152.
- Wild, K., Church, A., McCarthy, D., & Burgess, J. (2013). Quantifying lexical usage: vocabulary pertaining to ecosystems and the environment. *Corpora*, 8(1), 53-79.
- Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing*, 13, 257-283.
- Yuan, M. (2010). Mapping text. In D. Bodenheimer, J. Corrigan, & T. Harris (eds.) *The spatial humanities: GIS and the future of humanities scholarship*. Indiana University Press: Bloomington.