34

Analyses of Second Language Text and What Can Be Learned From Them

Eli Hinkel
Seattle University

INTRODUCTION

Analyses of second language (L2) text largely examine L2 writing, while L2 spoken production is usually investigated in the field of conversational analysis (see Markee, chap. 20, this volume). Since the emergence of applied linguistics as a discipline in the 1950s and 1960s, three large domains of research have focused on various properties of L2 written text: structuring of the information flow in discourse, syntactic, lexical, and rhetorical features employed in L2 text, and to a smaller extent, L2 grammar, and lexical errors.

In general terms, the analysis of grammatical and lexical errors in L2 written text is derived from the contrastive (error) analysis that predominated in L2 learning research between the 1950s and 1970s. Error analysis was based on an assumption that many (if not most) L2 errors are an outcome of L1 to L2 transfer of syntactic and lexical regularities and language properties.

Discourse analysis accounts for global features of text and the organization of ideas in writing. Contrastive rhetoric, as a subdomain of applied linguistics (Kaplan, 1966), gave rise to and continues to promote an examination of discourse features in the L2 writing of non-native speakers of English (NNSs) (see also chap. 33). It is important to keep in mind, however, that relatively little research on L2 writing had been carried out prior to the 1980s, when the numbers of NNSs in U.S. colleges and universities began to climb dramatically.

Historically, important advancements in contrastive rhetoric were reflected in a growing body of knowledge about the order and ideational structure of discourse and discourse moves, also called Discourse Blocs (Kaplan, 1983), in L2 writing. The innovations in contrastive rhetoric studies coincided with rapid advances in text linguistics. Speaking broadly, text analysis had the goal of identifying global discourse features that can be marked by means of syntactic and lexical elements, such as verb tenses or sentence transitions (Coulthard, 1985; de Beaugrande & Dressler, 1972; Halliday & Hasan, 1976, 1989; van Dijk, 1985).

In a distinct domain of applied linguistics, investigations in contrastive rhetoric and discourse structuring in various rhetorical traditions and across different cultures have also been extended to studies of comparative uses of textual features in L2 written prose. Analyzing written discourse paradigms and text attributes became the objective of many studies that worked with L1 writing of native speakers of English (NSs) in, for example, Australia, Canada, the United States, United Kingdom, and New Zealand, and those in the English L2 writing of speakers of many other languages. A vast body of research has thus far compared discourse and textual features employed in L2 writing of speakers of such languages as (in alphabetical order) Arabic, Chinese, Czech, Dutch, Farsi, Finnish, French, Hebrew, Hindi, German, Indonesian, Japanese, Korean, Malay, Russian, Spanish, Swedish, Thai, and Vietnamese, as well as several varieties of English.

Most studies of various features of L2 discourse and text have been motivated by immediate and long-term research and curriculum development goals, as well as pedagogically driven needs of particular groups of L2 students and learners in various locations, specific interests of individual researchers, available sources of text data, and/or attempts to apply the findings of predominantly English language-based text linguistics to L2 text (e.g., Al-Khatib, 2001; Carlson, 1988; Clyne, 1987; Grabe, 1987; Hinkel, 1994; Johnson, 1992; Taylor & Chen, 1991). To date, a coherent picture of syntactic, lexical, rhetorical, or discoursal features of L2 text has yet to emerge. However, in sum total, much has been learned about features of text produced by L2 writers in different contexts and for divergent academic, social, and communication purposes.

This chapter provides an overview of the methods widely used for analyzing L2 discourse and text, and the findings of research on macro and micro features of text. Additionally, computerized databases of L2 writing and the constraints that seem to confound analyses of L2 written corpora are briefly discussed.

QUALITATIVE AND QUANTITATIVE METHODS OF L2 TEXT ANALYSIS

Analyzing L2 writing is a time-consuming and laborious process. Investigations of L2 text have varied dramatically in the size of their data samples and types of writing. Most analyses have consisted of small-scale studies that utilized written texts produced by one to a half dozen writers (e.g., Arndt, 1993; Choi, 1988; Mauranen, 1996), and only a few had access to (or resources to work on) larger numbers of texts written by dozens or hundreds of L2 writers. On the other hand, studies that focused on specific excerpts of discourse or narrowly defined textual features, such as introductions to research papers or hedges and modal verbs, include larger samples written by 10 to 25 NNSs (e.g., Basham & Kwachka, 1991; Swales, 1990).

To a great extent, the theoretical frameworks and research methodologies for analyses of L2 discourse and text rely on those developed and formulated in such domains of applied linguistics as text linguistics, discourse analysis, ethnography, and cognitive psychology. Although much research on L1 English-language writing has been carried out in such disciplines as rhetoric and composition, on the whole, the study of rhetoric has had a minimal impact on analyses of L2 text. The philosophical underpinnings of the Anglo-American and other western rhetorical traditions draw on the classical Aristotelian and Greco-Roman foundational premises (e.g., invention, organization, and language and stylistics) not easily applicable to L2 discourse and text. On the other hand, quantitative and qualitative research methods designed for applied linguistics analyses of texts, written primarily in English, have been a great deal more useful in studies of L2 discourse moves and markers, text construction, and lexicogrammatical features (Halliday, 1994; Hoey, 1991; Grabe & Kaplan, 1996; Kaplan, 1988; Nattinger & DeCarrico, 1992).

To date, the majority of investigations into L2 writing have focused on the organizational and ideational structure of L2 discourse and the features of L2 text. Comparative studies have sought to account for differences and similarities between the properties of L2 discourse and text and those identified in the L1 writing of native English speakers who can be, for example, university students or authors of published research articles. In such examinations, comparisons can be made in regard to L1 and L2 global (macro) discourse construction, arrangements of ideas, cohesion, and coherence. Additionally, researchers can scrutinize textual (micro) features that have the function of marking discourse organization and aiding in the development of cohesive and coherent prose. As with the research on L2 discourse, the primary objectives of practically all L2 text analyses and comparative studies have stemmed from the pedagogical needs of L2 writing instruction for university students and academically bound language learners and professionals.

L2 DISCOURSE AND MACRO FEATURES OF TEXT IN BROAD STROKES

The contrastive rhetoric hypothesis (Kaplan, 1966) is largely concerned with discourse structuring and logical organization of information in various rhetorical traditions. The crosscultural analysis of discourse postulates that systematic differences exist in how topic continuity is established and coherence is developed in rhetorical traditions and writing in different languages and cultures. In the case of literate and educated speakers of various languages, L1-specific ways of organizing discourse and information may be transferred to the discourse organization in L2 writing. Thus, the overarching goal of the contrastive rhetoric hypothesis was to assist teachers and academically bound ESL students who needed to learn to write in L2 by identifying the differences among patterns in written discourse.

Global features of L2 written discourse, such as discourse moves, organization, structuring, as well as attendant issues of clarity, explicitness, fluidity, and contents of writing, represent broader and more abstract constructs than those commonly examined in analyses of text (Grabe & Kaplan, 1996; Hinkel, 1997, 1999; Indrasuta, 1988; Johns, 1997; Kaplan, 2000). For example, a number of studies has been devoted to stylistic properties of various types of L2 writing, such as textual indirectness in academic essays or narrative personalization.

In the 1980s and 1990s, research on the L2 writing of university students in a number of English-speaking countries established that discourse construction and rhetorical paradigms differ in consistent and important ways in the L1 writing of NSs and L2 written prose. Additionally, investigations devoted to discourse construction across languages and cultures have been able to determine that, for instance, marked similarities exist in the rhetorical development of text written in some European languages, such as Czech and German, or Asian languages, such as Chinese, Japanese, and Korean. For example, in German and Czech academic writing, as well as in the L2 prose of German learners of English and scholars, digressions and deviations from the main topic are considered to be acceptable, as are repetitions, recapitulations, and restatements, abstract argumentation, and broad generalizations (e.g., Clyne, 1987; Cmejrkova, 1996).

On the other hand, classical rhetoric and discourse construction in Chinese, Japanese, and Korean writing, as well as in the L2 prose of speakers of these languages, have been shown to include predictable organizational paradigms with the main idea at the end, indirect argumentation, allusions, and references to history and authority as evidence (Hinds, 1983, 1987, 1990; Matalene, 1985; Park, 1988; Scollon, 1991). Cai (1999, p. 294) points out that classical Chinese rhetoric and style has an indelible effect on the academic writing of L2 students in U.S. colleges and universities. According to the author, in L2 writing instruction, Anglo-American discourse strategies, topic development, rhetorical and linguistic norms, as well as the "sociocultural contexts in which these norms are embedded" should be explicitly taught because these "are essential in English academic writing."

The division of discourse organization paradigms into what has become known as "reader-responsible" or "writer-responsible" text was originally proposed by Hinds (1987, 1990), based on his research on Japanese, Korean, and Chinese writing. Hinds noticed that in the written discourse in these languages, the main point or the thesis is not necessarily presented to the reader at the beginning or, alternatively, can remain implicit throughout the text. In such texts, the responsibility for determining the writer's main idea is left to the reader, who then needs to deduce the writer's position and the central argument from context. On the other hand, in writer-responsible prose, explicitness, clarity, and lexical precision are considered to be requisite. Hence, it is the writer's job to construct discourse and text in which the purpose of writing is directly stated at the outset, and the writer's discourse and argument is expected to be clear and easy to follow. At present, in L2 writing instruction, the concepts of reader- or writer-responsible texts have become commonplace. Today, it would be difficult to find an ESL writing textbook that does not mention that in Anglo-American writing, it is the writer who is responsible for making the text transparent and explicit to the reader.

Other studies examined discourse construction and rhetorical organization in such diverse languages as Arabic and Spanish. Specifically, Ostler (1987) found that the L2 writing of Arabic speakers included a significantly higher rates of parallel and coordinate constructions, as well as greater numbers of discourse moves and rhetorical support elements than were found in the writing of NSs. Sa'adeddin (1989) further explains that in argumentation and rhetorical persuasion, colloquial Arabic discourse relies on parallelism, repetition, and broad generalizations, as well as ornate and elaborate vocabulary. Thus, according to Sa'adeddin, when writing in English, Arabic speakers may simply transfer from L1 the usage of coordinate and parallel constructions prevalent in interactive rhetorical style and persuasion. In his study of modern Arabic writing, Hatim (1991) similarly found that considerations of audience and interaction with audience play an important role in how Arabic written discourse, persuasion, and rhetorical moves are constructed.

In regard to the L2 writing of Spanish speakers, research has consistently demonstrated that they write longer essays and longer, more complex, and elaborated sentences than NSs do (e.g., Carlson, 1988; Montano-Harmon, 1991). In addition, Spanish speakers use significantly higher rates of coordinate clauses and phrases, long abstract words, and broad generalizations when compared to those in the writing of NSs of similar age and educational levels. In fact, Reid's (1992) study of writing of English, Spanish, and Arabic speakers demonstrated that the prose of Spanish L1 writers exhibits coordination patterns similar to those that Ostler (1987) identified in the L2 writing of Arabic-speaking university students.

According to some researchers, however, the divergences between L1 and L2 discourse structuring can also be attributed to L2 writers' developmental constraints and inexperience rather than the transfer of L1 rhetorical paradigms (Mohan & Lo, 1985). Additionally, the studies of published articles written by Chinese and English speakers (Taylor & Chen, 1991), and essays written by Korean students in a U.S. university (Choi, 1988) demonstrated that discourse structuring in L1 and L2 writing can show both differences and similarities. In both cases, the authors note that due to the internalization of scientific discourse and the effects of English writing instruction in many countries, the structural divergences between the Anglo-American discourse organization patterns and those in other rhetorical traditions have been continuously diminished over time and are likely to become even less pronounced in the future.

On the whole, however, the influence of L1 discourse and rhetorical paradigms in organizing information represents an established venue in numerous investigations on writing and text across cultures and languages. To a great extent, a large body of research on discourse construction patterns in writing in various rhetorical traditions has led to a greater understanding of many issues that confound ESL writing and its teaching and learning.

In particular, Silva (1993) highlights the most pronounced differences between practically all facets of writing in L1 and L2. In his synthesis of 72 published research reports and empirical studies, Silva points out that L2 writing is crucially distinct from L1 writing in regard to the writing process, such as composing and revision, and macro features of discourse organization. Based on the findings of dozens of studies carried out prior to 1993, Silva emphasizes that the processes of writing in L2 are fundamentally different from those entailed in writing in L1, L2 writers engage in less discourse and text planning, reviewing, and revising than L1 basic writers, while producing L2 text is far more work- and time-consuming, and revision is demonstrably more difficult.

To summarize Silva's conclusions, compared to the discourse structuring and development in L1 basic and student writing, L2 writers:

- Organize and structure discourse moves differently.
- Take a logically and conceptually different approach to rhetorical argumentation, persuasion, and exposition/narration.
- Over- or underestimate the amount of readers' background knowledge and the need for textual clarity, explicitness, and specificity.
- Differently orient the reader, and introduce and develop topics.
- Employ different strategies for extracting/citing information from sources, as well as paraphrasing, quoting, and including source material in their writing.
- Develop text cohesion differently, with weak lexical/semantic ties and theme connections, and a preponderance of overt conjunctive markers.

Silva concludes his overview of research by saying that "L2 writing is strategically, rhetorically, and linguistically different in important ways from L1 writing" (p. 669). In light of these fundamental differences, Silva points out that the learning needs of L2 writers are distinct from those of L1 writers, whether basic or skilled, and that teachers who work with L2 writers require special and focused training to deal with cultural, rhetorical, and linguistic differences of their students.

To this end, research into how L2 discourse and text are constructed, as well as contrastive analyses of discourse, have proven to be very useful in the teaching of L2 writing and creating more appropriate curricula (e.g., Leki, 1992; Reid, 1993). In particular, an important outcome of research into L2 written discourse is the increased knowledge about discourse and text in writing traditions other than Anglo-American, including such written genres as news reports, academic publications, student writing, e-mail messages, and business correspondence.

ANALYSES OF MICRO FEATURES OF L2 TEXT

In addition to the analysis of discourse in L2 writing, a large number of studies have been devoted to the comparative analyses of lexical, syntactic, and rhetorical features of L1 and L2 texts, usually produced in the contexts of their academic endeavors. Much research, for example, investigated the uses of discourse markers, cohesion and coherence devices, modal verbs, hedges, and modifiers in L1 and L2 prose (Connor & Johns, 1990; Field & Oi, 1992; Flowerdew, 2000; Hinkel, 1995, 2001a; Johns, 1984, 1990; Johnson, 1992; Khalil, 1989; Mauranen, 1996; Reid, 1992; Swales, 1990).

A typical study of L2 text features may undertake to determine, for example, how explicit cohesive devices are used in L1 and L2 academic essays. For this purpose, researchers may compare the frequencies and contexts of sentence conjunctions (e.g., *furthermore, however,* and *thus*), coordinating conjunctions (e.g., *and, but, yet,* and *so*), and/or summary markers (e.g., *in short* and *in sum*) (e.g., Field & Oi, 1992; Hinkel, 2001b; Johns, 1984; Khalil, 1989; Schleppegrell, 1996). Similarly, to analyze the uses of modal verbs, usage measurements can be computed separately or together for possibility and ability modals (e.g., *can, may, might, could*) or obligation and necessity modals (e.g., *must, should, need*).

A study methodology can entail counting the number of conjunctions or modal verbs by type in each essay, followed by obtaining a mean (or median) value of occurrences of a particular conjunction in all essays in a sample of texts or in a single essay. In quantitative analyses, for instance, descriptive values for uses of similar features can be obtained for NS and NNS texts. Then these are usually compared and analyzed statistically to determine whether they are used similarly or significantly differently in the two samples.

For example, a short excerpt from an L2 text on the topic of international sports events and competitions is presented here. In this 94-word passage, the writer employs the ability modal *can* six times and the obligation modal *should* twice, and it is easy to notice that the text probably relies on the usage of *can* to excess.

> We *can* see that sports make people show their love of their own country. You *can* hear special team songs in every game. This is not just about individuals, but nations. We *can* see passion, happy or sad around audiences and players because sports bring us energy and living. In the sports world, everyone is equal, every player *can* have the same condition to do what he or she *should* do the best. Also, we *can* understand every country is equal. So, we *should* be proud of our country, and we *can* enjoy the success together.

In this case, a computation of the frequency rate of *can* is approximately 6.4% (6/94) and *should* is 2.1% (2/94). Of course, no generalizations about the writer's uses of these two modals or their frequency rates can be made based on such a short excerpt. However, in larger and representative samples of L1 and L2 writing, comparisons of descriptive measurement values by means of appropriate statistical tests can allow researchers to gauge whether in an L2 writing sample of, say, 15,000–20,000 words, the overall usage of specific syntactic and lexical features approximates that in a parallel L1 writing sample. In L2 writing instruction, such comparisons can (and often do) lead to fine-tuning course curricula, added attention to specific areas of teaching, or individualized assistance for L2 learners.

In analyses of L2 text, the degrees of fluency, for example, can be assessed by means of measuring relative text lengths, as well as lengths of sentences, clauses, or words, in combinations with supplemental measures of accuracy, and lexical and syntactic complexity (e.g., Carlson, 1988; Hinkel, 2003a; Park, 1988; Schleppegrell, 2002). In large-scale assessments that involve hundreds or thousands of writers, developmental indexes have also been driven by research into specific attributes of L2 writing and text (e.g., Basham & Kwachka, 1991, Hamp-Lyons, 1991a; Hamp-Lyons & Kroll, 1996; Reid, 1993; Weigle, 2002).

Analyses of L2 text have further delved into various textual genres commonly associated with various L2 writing tasks. These include formal essays, university term and diploma projects, business letters, recommendations letters, e-mail messages, and journals produced by L2 writers (Al-Khatib, 2001; Bouton, 1995; Choi, 1988; Hinkel, 2001b; Jenkins & Hinds, 1987). Furthermore, research has examined the properties of L2 text produced by adult L2 writers in colleges and universities with an English medium of instruction (e.g., in Hong Kong, India, or Singapore), as well as the writing of young language learners in the course of their schooling.

Among the investigations of ESL children's writing, Edelsky (1986) and Hudelson (1988) found that texts produced by young ESL writers are very similar to those of young native speakers, although L1 culture has a definitive influence on L2 children's view of writing and its functions and purposes. On the other hand, Maguire and Graves (2001, p. 588) reported that school-age L2 learners "use different genres, rhetorical styles, modalities, and semantic and syntactic structures" and that their writing should be viewed as a means of facilitating the development of children's language proficiency. Other researchers who carried out empirical investigations of writing development indicate that it takes years of persistent, knowledgeable, and attentive teaching for L2 school-age learners to attain sufficient language control to produce academic writing effectively similar to that of L1 children of matching ages (Hakuta, Butler, &Witt, 2000; Scarcella & Chunok, 1989; Valdes & Sanders, 1999).

It is important to note, however, that relatively little research has been specifically devoted to the proficiency, development, and text features in the L2 writing of school-age children. A great majority of published reports deal with the discourse and linguistic features of text produced by adults in academic, professional, language learning, or literacy contexts.

FINDINGS OF L2 TEXT ANALYSES

As was mentioned earlier, Silva's (1993) synthesis of research on L2 writing processes and discourse also addresses the research findings that deal with morphosyntactic and lexical features of L2 text. Specifically, according to Silva's summary, L2 writers employ simpler sentences with more authoritative warnings, admonitions, personal references and narratives, and repetitions of ideas and vocabulary. In regard to its linguistic features, NNS prose contains fewer syntactically complex constructions, such as subordinate clauses, descriptive adjective phrases, hedges, modifiers of most types, compound noun phrases, and possessives, but more coordinators, sentence transitions, and pronouns[1]. To summarize, compared to L1 writers, L2 writers have a restricted syntactic and stylistic repertoire, as well as severely limited range of accessible lexis that can be used in writing. In all, Silva points out that "L2 writers' texts were less fluent . . . , less accurate . . . , and less effective" (p. 668) than those of NSs, and "in terms of lower level linguistic concerns, L2 writers' texts were stylistically distinct and simpler in structure."

Based on earlier research, Silva (1993, p. 669) calls for a reconceptualization of how L2 writing and text production are taught in U.S. colleges and universities: "The prevalent assumption that L1 and L2 writing are, for all intents and purposes, the same" is unexamined and has remained largely unvalidated, and the many findings of research "make this assumption untenable." Silva emphasizes that L1 "composition theories, which are, incidentally, largely monolingual, monocultural, ethnocentric, and fixated on the writing of NES [native English speaker] undergraduates in North American colleges and universities" are inapplicable to teaching L2 writing, do L2 writers a great disservice, and are, quite possibly, counterproductive.

Almost a decade later, a large scale empirical analysis of 68 lexical, syntactic, and rhetorical features of L2 text was carried out by Hinkel (2002), who examined 1,457 (around 435,000 words) NS and NNS placement essays written in several universities across the United States. The L2 text corpus included texts written by speakers of six languages: Arabic, Chinese, Indonesian, Japanese, Korean, and Vietnamese. All NNS students (1,215) were advanced and trained L2 writers, a large majority of whom were holders of U.S. academic degrees. Hinkel reports that, even after years of ESL and composition training, L2 writers' text continues to differ significantly from that of novice (first-year) NS students in regard to most features examined in her study. The results of her analysis indicate that even advanced and trained L2 writers have a severely limited lexical and syntactic repertoire that enables them to produce simple texts restricted to the most common language features that occur predominantly in conversational discourse. Hinkel concluded that L2 texts do not approximate those in L1 basic academic writing because NSs of English already have a highly developed (native) language proficiency that a majority of NNSs require years to develop, in most cases as adults. Like Silva (1993), Hinkel calls for changes in the methodologies for teaching L2 writing that are based on the pedagogy intended for teaching composition to NSs.

COMPUTER TEXT BASES AND ANALYSES OF L2 WRITTEN CORPORA

Beginning in the late 1980s, due to the advancements in computer technology that permitted analyses of large amounts of printed or typed text, corpus-based studies dramatically altered the methodologies for analyzing syntactic, lexical, and collocational features employed in various written genre in English (e.g., Biber, 1988; Leech, Rayson, & Wilson, 2001; Renouf & Sinclair, 1991; Sinclair, 1991; Stubbs, 1996). The rapid changes in text linguistics and the new knowledge obtained from the analyses of large text corpora have also influenced the methods and scope of research in L2 writing.

As an outcome, the studies of the syntactic, lexical, and other micro features of L2 text have similarly changed their methodological approaches, as well as the sizes and types of L2 written corpora. To illustrate, the new genres of L2 text that lend themselves to computerized investigations, NS and NNS e-mail messages can be examined in regard to their grammar structures, syntactic accuracy, lexical ranges and complexity, and common politeness formulae (Chang & Hsu, 1998; Gonzalez-Bueno & Perez, 2000; Li, 2000).

Other research projects investigate L2 academic writing in corpora that range from 50,000 to 500,000 words and include texts produced by tens or hundreds of students (Flowerdew, 2001; Granger, 1998; Hyland & Milton, 1997). In L2 writing assessment, the size and scope of analyses evolved from exclusive reliance on impressionistic ratings of essays by specially trained readers to computations of frequency and percentage rates with which particular syntactic, lexical, and discourse features are employed in learner prose (Frase et al., 1999; Hamp-Lyons, 1991b; Reid, 1993; Weigle, 2002).

It is important to note, however, that despite the increase in the size of L2 written corpora, the investigations of discoursal, syntactic, and lexical features largely remained focused on the attributes of text similar to those examined in the numerous studies carried out prior to the technological and methodological innovations in text linguistics (e.g., Choi, 1988; Field & Oi, 1992; Hinkel, 2001a, 2002, 2003a; Johns, 1984; Ostler, 1987). In particular, language errors, coherence and cohesion markers (such as coordinating conjunctions and demonstrative pronouns), clause subordinators, modal verbs, personal pronouns, prepositions, adjectives, hedges, and intensifiers remained among the mainstays of research on computerized textual databases (Flowerdew, 1998, 2000; Granger, 1998, 2002; Granger & Tribble, 1998; Green, Christopher & Lam, 2000; Hyland & Milton, 1997; Lorenz, 1998; Reid, 1992; Ringbom, 1998; Tribble, 2001).

To some extent, the reason that computer analyses of written learner corpora have continued to focus on syntactic, lexical, and rhetorical features very similar to those extensively examined in a large number of earlier studies is that, compared to published English-language texts, most L2 writing includes a limited range of linguistic constructions (Granger, 2002; Hinkel, 2001b, 2002, 2003b). Another issue with the current L2 learner corpora is that such databases as International Corpus of Learner English consist exclusively of L2 texts written by speakers of European languages, such as Dutch[2], Finnish, French, German, Spanish, and Swedish.

On the other hand, the innovations brought about by the advances in computerized L2 text analyses permit insights into the characteristics of L2 text that cannot be attained by means of manual studies (Granger, 2002; Granger & Rayson, 1998), for example:

- Frequent co-occurrences of words in different genres of L2 text.
- Overuse or underuse of particular lexis or grammar constructions, for example, *find*, *want*, and *know* are by far the most common verbs in the L2 writing of Swedish speakers (Ringbom, 1998).
- L1 to L2 transfer of specific lexical and syntactic patterns, for example, distinctive phrases and patterns can be found, for example, in the writing of the speakers of French but not Dutch and German (Granger, 2002).
- Classification of errors by types (e.g., grammatical, collocational, or stylistic) and frequencies of their occurrence.

Meunier (1998) points out, however, that the typical computations of type/token ratio of words in a text used to measure the amount of lexical variation do not seem to reflect a relative quality of L2 texts because lexically varied prose is not necessarily of good quality. For example, if some NNS writers have a good vocabulary range but poor discourse organization or grammar skills, their text would not be of high quality, even when it is lexically rich. To be specific, Lorenz (1998) found that in the case of advanced German speakers with substantial and developed L2 English vocabulary repertoires, it was not so much a lack of accessible lexis that made their L2 writing appear non-native, but the ways in which particular lexical items were used (see also Lewis, 1993, 1997, and Nattinger & DeCarrico, 1992, for additional discussion on idiomaticity, collocations, and formulaic lexical phrases in written text).

Similar to the findings of studies carried out earlier based on smaller samples of various genres in L2 writing, the results of the computerized analyses of L2 writing have also clearly demonstrated that L2 written discourse and text are crucially and significantly distinct from those produced by L1 writers in English.

ISSUES AND COMPLEXITIES IN L2 TEXT RESEARCH

Technological advances in computerized text analyses of English-language corpora have allowed researchers to shed light on the real-life uses of syntactic, lexical, and collocational features in native speaker language production. In the past several decades, some of the large native English corpora have worked with spoken and written text bases as large as 30 million words (Biber et al., 1999; Leech, Rayson, & Wilson, 2001; Stubbs, 2001) (see also Collins COBUILD corpus-based learner dictionaries developed by researchers at the University of Birmingham).

One of the key differences between analyses of corpora of L1 English-language corpora and L2 text analyses lies in the fact that, for instance, the corpora of written English include large amounts of published (i.e., polished and edited) texts. On the other hand, L2 texts are either handwritten or typed by NNSs, whose language usage (e.g., spelling, word formation, or phrasing) is distinct from that in practically any type of native speaker prose.

One of the stumbling blocks in investigations of L2 handwritten texts is that current software for converting handwriting into typed text is highly unreliable, and to overcome software problems, L2 texts are keyed after they are written (e.g., Meunier, 1998; Shaw & Liu, 1998). However, to allow computer programs to count a number of occurrences of, for example, particular words or constructions, decisions need to be made whether to correct spelling, grammar, or lexis to make them uniform or leave the original text intact.

Although studies of L2 text have mainly been driven by the growing body of findings and publications in computerized analyses of English language corpora, to date a consistent methodology for L2 corpus research has not yet emerged. Some researchers, such as Ferris (1993), claimed that computer analyses of L2 text may not be possible in the near future. Computerized analyses of typed L2 essays obtained in large scale assessments (e.g., the Test of Written English) reported significant proportions (up to 21%) of misidentified L2 textual features (Frase et al., 1999).

Developers of L2 written corpora also reported additional confounding complexities associated with building and analyzing text bases (Granger, 2002; Meunier, 1998):

- When L2 text is keyed or scanned, a number of errors are routinely introduced (e.g., omissions, additions, and misrepresentations), and proofreading typed L2 texts against the originals is an extremely laborious, time-consuming, and iterative process.
- Computerized analyses of L2 corpora do not permit analyses of global and abstract discourse features (e.g., sufficient/explicit rhetorical support), many grammar features (e.g., referential/nonreferential pronouns, as in *they tell me* . . .), sentence or T-unit counts, or measurements of lexical density; these have to be carried out manually.
- In analyses of L2 errors, error tagging is a manual and time-consuming process, and the classification of errors is often subjective.
- Word counts of the same texts can diverge significantly (up to 10%), when different types of software are used (e.g., hyphenated words or contractions can be counted as either one or two words). In such cases, statistical analyses, fundamentally dependent on baseline word counts, can be invalidated.

The following example is excerpted from a L2 university essay on taking risks in order to succeed to illustrate the difficulties of computerized parsing and tagging of L2 text.

> I think that the adecatte combination between this aditudes in life is the key for open the succed's door. In one way, take a risk can be important to rich some relevant chance, for example, "shakira", who is a very important pop singer, take a risks when she and her family recordered her first Long play. They spend all the money that they had in the bank and now they are rich and she is a important artist. but in and other way, people have lost all them fortune for a bat business so many people preffer work hard and work for sure. In conclusion. In all the success cases we will discoves that the combination between take risk and take care is the key.

In many cases, even if the writer's spelling is corrected, computer tagging of syntactic and lexical features, as in this example, would be very difficult. In fact, this particular text may need to be sounded out at least to understand what it says.

In light of the fact that reliable automatic analyses of L2 text still lag behind those of published or transcribed L1 English language corpora, most L2 text studies published to date have relied on manual analyses of discourse, syntactic, lexicalized, and rhetorical features. Tagging and hand-counting features are extremely work- and time-consuming processes that impose limitations on the amounts of text that can be analyzed by a single researcher or even a group of researchers.

SUMMARY AND CONCLUSION

In the past half century, numerous studies have examined discoursal, lexical, syntactic, and rhetorical properties of L2 writing. Practically all research on L2 written prose has focused on the features of discourse and text identified in research in text linguistics that accounts for how lexical, syntactic, stylistic, and rhetorical elements combine to create particular types of prose and genres. Investigations into L2 text have identified the important and significant differences that exist between L1 and L2 writing. In part, these differences stem from divergent written discourse paradigms valued in various rhetorical traditions and often transferred from L1 to L2. Such global features of L2 discourse as organization and information structuring, topic development and continuity, as well as text cohesion, coherence, and clarity appear to be greatly influenced by the rhetorical and text construction norms that differ widely across languages and cultures.

Research has also demonstrated that other crucial factors that confound L2 writing and text have to do with shortfalls of writers' language proficiencies and restricted linguistic repertoire that significantly undermine L2 writers' ability to produce high quality texts. Based on the results of their studies, many researchers of L2 learning and development have emphasized that even school-age children or highly educated adult L2 learners require years of language training to attain the levels of proficiency necessary to create effective written prose.

ACKNOWLEDGMENTS

My sincere thanks to Robert B. Kaplan and Rodney Hill for their insightful comments and suggestions on an earlier draft of this chapter. Through the years, their unfailing help and support have become indispensable.

NOTES

1. For similar findings of large-scale empirical investigations, see also, Carlson (1988), Flowerdew (2000, 2001), Hinkel (2001a, 2001b, 2002, 2003a, 2003b), Johns (1997), Reid (1992), Ringbom (1998), Schleppegrell (1996, 2002), and Shaw and Liu (1998), to mention just a few.
2. Other L2 corpora, e.g., Longman's Learner Corpus and Hong Kong University of Science and Technology Learner Corpus, are proprietary and accessible only to the researchers affiliated with these institutions.

REFERENCES

Al-Khatib, M. (2001). The pragmatics of letter-writing. *World Englishes, 20*(2), 170–200.

Arndt, V. (1993). Response to writing: Using feedback to inform the writing process. In M. Brock & L. Walters (Eds.), *Teaching composition around the Pacific Rim: Politics and pedagogy* (pp. 90–116). Clevedon, UK: Multilingual Matters.

Basham, C., & Kwachka, P. (1991). Reading the world differently: A crosscultural approach to writing assessment. In L. Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts* (pp. 37–49). Norwood, NJ: Ablex.

Biber, D. (1988). *Variation across speech and writing*. Cambridge, UK: Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow, Essex: Pearson.

Bouton, L. (1995). A crosscultural analysis of the structure and content of letters of reference. *Studies in Second Language Acquisition, 17*(2), 211–244.
Cai, G. (1999). Texts in contexts: Understanding Chinese students' English compositions. In C. Cooper & L. Odell (Eds.), *Evaluating writing: The role of teachers' knowledge about text, learning, and culture* (pp. 279–299). Urbana, IL: NCTE.
Carlson, S. (1988). Cultural differences in writing and reasoning skills. In A. Purves (Ed.), *Writing across languages and cultures: Issues in contrastive rhetoric* (pp. 109–137). Newbury Park, CA: Sage.
Chang, Y.-Y. & Hsu, Y. P. (1998). Requests on e-mail: A crosscultural comparison. *RELC Journal, 29*(2), 121–151.
Choi, Y. (1988). Text structure of Korean speakers' argumentative essays in English. *World Englishes, 7*(2), 129–142.
Clyne, M. (1987). Cultural differences in the organization of academic texts. *Journal of Pragmatics, 11*(1), 211–247.
Cmejrkova, S. (1996). Academic writing in Czech and English. In E. Ventola & A. Mauranen (Eds.), *Academic writing: Intercultural and textual issues* (pp. 137–152). Amsterdam: John Benjamins.
Connor, U., & Johns, A.M. (Eds.). (1990). *Coherence in writing: Research and pedagogical perspectives.* Alexandria, VA: TESOL.
Coulthard, M. (1985). *An introduction to discourse analysis* (2nd ed.). London: Longman.
de Beaugrande, R. & Dressler, W. (1972). *Introduction to text linguistics.* London: Longman.
Edelsky, C. (1986). *Writing in a bilingual program: Habia una vez.* Norwood, NJ: Ablex.
Ferris, D. (1993). The design of an automatic analysis program for L2 text research: Necessity and feasibility. *Journal of Second Language Writing, 2*(2), 119–129.
Field, Y., & Oi, Y. L. M. (1992). A comparison of internal conjunctive cohesion in the English essay writing of Cantonese speakers and native speakers. *RELC Journal, 23*(1), 15–28.
Flowerdew, L. (1998). Integrating expert and interlanguage computer corpora findings on causality: Discoveries for teachers and students. *English for Specific Purposes, 17*(4), 329–345.
Flowerdew, L. (2000). Investigating errors in a learner corpus. In L. Burnard & T. McEnery (Eds.), *Proceedings in the teaching and language corpora conference* (pp. 145–154). Frankfurt: Peter Lang.
Flowerdew, L. (2001). The exploitation of small learner corpora in EAP material design. In M. Ghadessy, A. Henry, & R. Roseberry (Eds.), *Small corpus studies and ELT* (pp. 363–380). Amsterdam: John Benjamins.
Frase, L., Faletti, J., Ginther, A., & Grant, L. (1999). *Computer analysis of the TOEFL Test of Written English* (Research report 64). Princeton, NJ: Educational Testing Service.
Gonzalez-Bueno, M., & Perez, L. C. (2000). Electronic mail in foreign language writing: A study of grammatical and lexical accuracy, and quantity of language. *Foreign Language Annals, 33*(2), 189–198.
Grabe, W. & Kaplan, R. B. (1996). *Theory and practice of writing.* London: Longman.
Grabe, W. (1987). Contrastive rhetoric and text-type research. In U. M. Connor & R. B. Kaplan (Eds.), *Writing across languages: Analysis of L2 text* (pp. 115–137). Reading, MA: Addison-Wesley.
Granger, S. (1998). *Learner English on computer.* London: Longman.
Granger, S. (2002). A bird's-eye view of computer learner corpus research. In S. Granger, J. Hung, & S. Petch-Tyson, (Eds.), *Computer learner corpora, second language acquisition, and foreign language teaching.* (pp. 3–36). Amsterdam: John Benjamins.
Granger, S., & Rayson, P. (1998). Automatic lexical profiling of learner texts. In Granger, S. (Ed.), *Learner English on computer* (pp. 119–131). London: Longman.
Granger, S., & Tribble, C. (1998). Learner corpus data in the foreign language classroom: Form-focused instruction and data-driven learning. In Granger, S. (Ed.), *Learner English on computer* (pp. 199–209). London: Longman.
Green, C., Christopher, E., & Lam, J. (2000). The incidence and defects on coherence of marked themes in interlanguage texts: A corpus-based enquiry. *English for Specific Purposes, 19*(2), 99–113.
Hakuta, K., Butler, Y., & Witt, D. (2000). *How long does it take English learners to attain proficiency?* Retrieved July 3, 2003 from http://www.stanford.edu/~hakuta/Docs/HowLong.pdf
Halliday, M. A. K. (1994). The construction of knowledge and value in the grammar of scientific discourse, with reference to Charles Darwin's *The Origin of Species.* In M. Coulthard (Ed.), *Advances in written text analysis,* (pp. 136–156). New York: Routledge & Kegan Paul.
Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English.* London: Longman.
Halliday, M. A. K., & Hasan, R. (1989). *Spoken and written language.* Oxford, UK: Oxford University Press.
Hamp-Lyons, L. (1991a). Issues and directions in assessing second language writing in academic contexts. In L. Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts* (pp. 323–329). Norwood, NJ: Ablex.
Hamp-Lyons, L. (1991b). Reconstructing "academic writing proficiency." In L. Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts* (pp. 127–153). Norwood, NJ: Ablex.
Hamp-Lyons, L., & Kroll, B. (1996). Issues in ESL writing assessment: An overview. *College ESL,* 52–72.
Hatim, B. (1991). The pragmatics of argumentation in Arabic: The rise and fall of text type. *Text, 11*(1), 189–199.
Hinds, J. (1983). Contrastive rhetoric. *Text 3*(2), 183–195.
Hinds, J. (1987). Reader versus writer responsibility: A new typology. In U. Connor & R. B. Kaplan (Eds.), *Writing across languages: Analysis of L2 text.* (pp. 141–152). Reading, MA: Addison-Wesley.

Hinds, J. (1990). Inductive, deductive, quasi-inductive: Expository writing in Japanese, Korean, Chinese, and Thai. In U. Connor & A. Johns (Eds.), *Coherence in writing* (pp. 87–110). Alexandria, VA: TESOL.

Hinkel, E. (1994). Native and non-native speakers' pragmatic interpretation of English text. *TESOL Quarterly, 28*(2), 353–376.

Hinkel, E. (1995). The use of modal verbs as a reflection of cultural values. *TESOL Quarterly, 29*, 325–343.

Hinkel, E. (1997). Indirectness in L1 and L2 academic writing. *Journal of Pragmatics, 27*(3), 360–386.

Hinkel, E. (1999). Objectivity and credibility in L1 and L2 academic writing. In E. Hinkel (Ed.), *Culture in second language teaching and learning*. (pp. 90–108). Cambridge, UK: Cambridge University Press.

Hinkel, E. (2001a). Matters of cohesion in L1 and L2 academic texts. *Applied Language Learning*, 12(2), 111–132.

Hinkel, E. (2001b). Giving examples and telling stories in academic essays. *Issues in Applied Linguistics, 12/2*, 149–170.

Hinkel, E. (2002). *Second language writers' text*. Mahwah, NJ: Lawrence Erlbaum Associates.

Hinkel, E. (2003a). Simplicity without elegance: Features of sentences in L2 and L1 academic texts. *TESOL Quarterly, 37*(2), 275–301.

Hinkel, E. (2003b). Adverbial markers and tone in L1 and L2 students' writing. *Journal of Pragmatics, 35*(7), 1049–1068.

Hoey, M. (1991). *Patterns of lexis in text*. Oxford, UK: Oxford University Press.

Hudelson, S. (1988). *Write on: Children's writing in ESL*. Englewood Cliffs, NJ: Prentice-Hall.

Hyland, K., & Milton, J. (1997). Qualification and certainty in L1 and L2 students' writing. *Journal of Second Language Writing* 6(2), 183–205.

Indrasuta, C. (1988). Narrative styles in the writing of Thai and American students. In A. Purves (Ed.), *Writing across languages and cultures: Issues in contrastive rhetoric* (pp. 206–227). Newbury Park, CA: Sage.

Jenkins, S., & Hinds, J. (1987). Business letter writing: English, French, and Japanese. *TESOL Quarterly, 21*(2), 327–349.

Johns, A. (1984). Textual cohesion and the Chinese speaker of English. *Language Learning and Communication, 3*, 69–74.

Johns, A. (1990). Coherence as a cultural phenomenon: Employing ethnographic principles in the academic milieu. In U. Connor & A. Johns (Eds.), *Coherence in writing* (pp. 211–225). Alexandria, VA: TESOL.

Johns, A. (1997). *Text, role, and context: Developing academic literacies*. Cambridge, UK: Cambridge University Press.

Johnson, P. (1992). Cohesion and coherence in compositions in Malay and English. *RELC Journal, 23*(2), 1–17.

Kaplan, R. B. (1966). Cultural thought patterns in intercultural education. *Language Learning, 16*(1), 1–20.

Kaplan, R. B. (1983). Contrastive rhetorics: Some implications for the writing process. In A. Freedman, I. Pringle, & J. Yalden (Eds.), *Learning to write: First language/second language* (pp. 139–161). London: Longman.

Kaplan, R. B. (1988). Contrastive rhetoric and second language learning: Notes towards a theory of contrastive rhetoric. In A. Purves (Ed.), *Writing across languages and cultures: Issues in contrastive thetoric* (pp. 275–304). Newbury Park, CA: Sage.

Kaplan, R. B. (2000). Contrastive rhetoric and discourse analysis: Who writes what to whom? When? In what circumstances? In S. Sarangi & M. Coulthard (Eds.), *Discourse and social life* (pp. 82–102). Harlow, UK: Longman.

Khalil, A. (1989). A study of cohesion and coherence in Arab EFL college students' writing. *System 17*(3), 359–371.

Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English*. London: Longman.

Leki, I. (1992). *Understanding ESL writers*. Portsmouth, NH: Boyton/Cook.

Lewis, M. (1993). *The lexical approach*. Hove, UK: LTP.

Lewis, M. (1997). Pedagogical implications of the lexical approach. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 255–270). Cambridge: Cambridge University Press.

Li, Y. (2000). Linguistic characteristics of ESL writing in task-based e-mail activities. *System, 28*, 229–245.

Lorenz, G. (1998). Overstatement in advanced learners' writing: Stylistic aspects of adjective intensification. In S. Granger (Ed.), *Learner English on computer* (pp. 53–66). London: Longman.

Maguire, M., & Graves, B. (2001). Speaking personalities in primary school children's writing. *TESOL Quarterly, 35*(4), 561–593.

Matalene, C. (1985). Contrastive rhetoric: An American writing teacher in China. *College English, 47*, 789–807.

Mauranen, A. (1996). Discourse competence: Evidence from thematic development in native and non-native texts. In E. Ventola & A. Mauranen (Eds.), *Academic writing: Intercultural and textual issues* (pp. 195–230). Amsterdam: John Benjamins.

Meunier, F. (1998). Computer tools for the analysis of learner corpora. In S. Granger (Ed.), *Learner English on computer* (pp. 19–37). London: Longman.

Mohan, B., & Lo, W. A. (1985). Academic writing and Chinese students: Transfer and developmental factors. *TESOL Quarterly, 19*(3), 515–534.

Montano-Harmon, M. (1991). Discourse features of written Mexican Spanish: Current research in contrastive rhetoric and its implications. *Hispania, 74*(3), 417–425.

Nattinger, J. & DeCarrico, J. (1992). *Lexical phrases and language teaching*. Oxford, UK: Oxford University Press.
Ostler, S. (1987). English in parallels: A comparison of English and Arabic prose. In U. Connor & R. Kaplan (Eds.), *Writing across languages: Analysis of L2 text* (pp. 169–185). Reading, MA: Addison-Wesley.
Park, Y. M. (1988). Academic and ethnic background as factors affecting writing performance. In A. Purves (Ed.), *Writing across language and cultures* (pp. 261–273). Newbury Park, CA: Sage.
Reid, J. (1992). A computer text analysis of four cohesion devices in English discourse by native and non-native writers. *Journal of Second Language Writing, 1*(2), 79–107.
Reid, J. (1993). *Teaching ESL writing*. Englewood Cliffs, NJ: Prentice-Hall.
Renouf, A., & Sinclair, J. (1991). Collocational frameworks in English. In K. Aijmer & B. Altenberg (Eds.), *English corpus linguistics* (pp. 128–143). New York: Longman.
Ringbom, H. (1998). Vocabulary frequencies in advanced learner English. In S. Granger (Ed.), *Learner English on computer* (pp. 41–52). London: Longman.
Sa'adeddin, M. A. (1989). Text development and Arabic-English negative interference. *Applied Linguistics, 10*(1), 36–51.
Scarcella, R., & Chunok, L. (1989). Different paths to writing proficiency in a second language? A preliminary investigation of ESL writers of short-term and long-term residence in the United States. In M. Eisenstein (Ed.), *The dynamic interlanguage: Empirical studies in second language variation* (pp. 137–154). New York: Plenum.
Schleppegrell, M. (1996). Conjunction in spoken English and ESL writing. *Applied Linguistics, 17*(3), 271–285.
Schleppegrell, M. (2002). Challenges of the science register for ESL students: Errors and meaning-making. In M. Schleppegrell & M. Colombi (Eds.), *Developing advanced literacy in first and second languages* (pp. 119–142). Mahwah, NJ: Lawrence Erlbaum Associates.
Scollon, R. (1991). Eight legs and one elbow: Stance and structure in Chinese English compositions. In (Ed.), *Proceedings of the 2nd North American Conference on Adult and Adolescent Literacy* (pp. 26–41). Ottawa: International Reading Association.
Shaw, P. & Liu, E. T. K. (1998). What develops in the development of second language writing. *Applied Linguistics, 19*(2), 225–254.
Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL Quarterly, 27*(4), 657–677.
Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford, UK: Oxford University Press.
Stubbs, M. (1996). *Text and corpus analysis*. Oxford, UK: Blackwell.
Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Oxford, UK: Blackwell.
Swales, J. (1990). Non-native speaker graduate engineering students and their introductions: Global coherence and local management. In U. Connor & A. Johns (Eds.), *Coherence in writing* (pp. 189–207). Alexandria, VA: TESOL.
Taylor, G., & Chen, T. (1991). Linguistic, cultural, and subcultural issues in contrastive discourse analysis: Anglo-American and Chinese scientific texts. *Applied Linguistics, 12*, 319–336.
Tribble, C. (2001). Small corpora and teaching writing. In M. Ghadessy, A. Henry, & R. Roseberry (Eds.), *Small corpus studies and ELT* (pp. 381–408). Amsterdam: John Benjamins.
Valdes, G., & Sanders, P. (1999). Latino ESL students and the development of writing abilities. In C. Cooper & L. Odell (Eds.), *Evaluating writing: The role of teachers' knowledge about text, learning, and culture* (pp. 249–278). Urbana, IL: National Council for Teachers of English.
van Dijk, T. (Ed.). (1985). *Handbook of discourse analysis*. (4 vols) London: Academic Press.
Weigle, S. (2002). *Assessing writing*. Cambridge, UK: Cambridge University Press.