

DOCUMENT RESUME

ED 452 734

FL 026 707

AUTHOR Jabbhour, Georgette
TITLE Corpus Based Instruction.
PUB DATE 2001-00-00
NOTE 23p.
PUB TYPE Opinion Papers (120)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Computational Linguistics; Computer Uses in Education;
*Databases; Discourse Analysis; Grammar; Language Patterns;
Psycholinguistics; Second Language Instruction; Second
Language Learning; Text Structure

ABSTRACT

Corpus-based instruction, also referred to as corpus-based linguistics, is essentially the study of genre texts for the production of materials that fit a specific group of second language learners. For example, if a group of students is learning the language in order to read medical articles, a specific teaching program can be set up based on the text analysis of the genre and corpus linguistic tools. The strength of this procedure is that word associations represent meanings, or functions. Because language teaching is essentially relating language forms to language functions, corpus-based instruction provides a rich learning environment. The opportunities provided by improved technologies make it possible for a few large collections of texts, or corpus texts, to be computer processed. Computer tools for text analysis reveal significant patterns that can be used instantly by learners. Such significant patterns are generally hidden when the texts are analyzed manually. Furthermore, because the texts forming the corpus represent a specific genre of writing, language forms and language functions relate grammar to discourse. This additionally draws attention to the presence of specific grammar for specific types of discourse. (Contains 20 references.) (Author/KFT)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 452 734

CORPUS BASED INSTRUCTION

Paper prepared
by
Dr. Georgette Jabbour

Submitted to ERIC
04/18/2001

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL HAS
BEEN GRANTED BY

Georgette Jabbour

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

- This document has been reproduced as received from the person or organization originating it.
- Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

New York Institute of Technology
ESL Program - Department of English
P.O.Box 8000
Old Westbury, NY 11568-8000

Phone: (516) 686-7713/7557
Fax: (516) 686-7760
Email: gjabbour@nyit.edu

BEST COPY AVAILABLE

26 707



ABSTRACT

Corpus-based Instruction

Corpus-based instruction also referred to as corpus linguistics is essentially the study of genre texts for the production of materials that fit a specific group of second language learners. For example, if a group of students is learning the language in order to read medical articles, a specific teaching program can be set up based on text analysis of the genre and corpus linguistic tools. The strength of this procedure is that word associations represent meanings, or functions. Since language teaching is essentially relating language forms to language functions, a corpus-based instruction provides a rich learning environment.

The opportunities provided by improved technologies make it possible for a large collection of texts, or a corpus of texts, to be computer processed. Computing tools for text analysis reveal significant patterns that can be used instantly by learners. Such significant patterns are generally hidden when the texts are analyzed manually. Furthermore, because the texts forming the corpus represent a specific genre of writing, language forms and language functions relate grammar to discourse. This additionally draws the attention to the presence of a specific grammar for specific types of discourse.

Contents:

TITLE: CORPUS-BASED INSTRUCTION

1. Introduction: Technology and the Written Word.....	4
2. Reading Discipline Specific Texts	5
2.1. Overview of The Medical Research Article	5
2.2. Reading Processes and Text Types	7
3. Corpus linguistics and Teaching	10
3.1. The Theory and its Methodology	10
3.2. Learners as Researchers: the Use of Technology	13
4. Criteria to Develop a Language Program	15
6. Focus on Discourse and on Language.....	16
7. Conclusion.....	17

Works Cited

Corpus-based Instruction

1. Introduction: Technology and the Written Word

Personal computers have started to be popular in the seventies. Since then, investigations concerning the mental processing of language (Rumelhart 1980) have blossomed. As a result, word-processing, laser printing and desktop publishing have all affected the development and dissemination of the written word.

Information and computer technology are no doubt an extension to the advent of printing in the 15th Century which had a tremendous impact on the dissemination of the written word by providing printed copies of the same text, instead of relying on hand-written copies (Baron 1989). Similarly, the computer revolution provides writers with a handy tool for their writing and for the transfer of knowledge in forms accessible for retrieval by other members of the discourse community. For example, the use of in-house printing, transfer of information on floppy disks, and certainly the International Computer Network (Internet) and the World Wide Web, have an effect on our communication trends. With the turn of this new century, the word 'writer' may well be taken to mean 'user of computer for text editing'. Information and communication technology are terms that refer to the whole system of improved devices that allow for the transfer of information and knowledge in a lapse of time remarkably short in comparison to that required by other means of communication. The basis of this computer-led information technology is by no means the written word or text.

Corpus linguistics falls into this category of improved means for the investigation of the written word. This branch of linguistics consists of the storage of texts belonging to text genre and type on the computer hard drive. The database thus formed is investigated using specific text analysis software. The unit of investigation is the word, as a single element, in combination with what precedes it and what follows it, otherwise its collocates. Corpus linguistics is ultimately the study of the discourse of the texts forming the database, but starting with the word and moving upward to the larger discourse level. The argument that is advanced in this paper is the possibility of relying on corpus linguistics to design second language teaching materials.

The paper first reviews the concept of reading discipline specific texts, and seeks to apply reading theories to the medical research article (MRA). It also provides background information about the medical research article, and then argues that teaching materials for reading purposes are best designed if corpus linguistics research is performed.

2. Reading Discipline Specific Texts

2.1. Overview of The Medical Research Article

The form of a MRA is specific, as well as its language. This makes the language easy to analyze while the content matter is no doubt complicated, allowing for an English language instructor to control the process of learning in a content-based area, such as the writing of medical reports and articles for publication.

To start with a surface description, a MRA consists of a Title, an Abstract and four sections: Introduction, Methods, Results and Discussion. These sections are followed by a Reference list and when applicable by Appendices (Polgar and Thomas 1991). An awareness of the form facilitates the identification of text functions.

Article sections are most importantly identified in terms of language use. The language used in an introduction is specific and differs tremendously from the language used in a method, a result, and undoubtedly a discussion section. The function of the introduction is to create a "research space" for the writer, or "niche" (Swales 1990), in order to ground the argument. To ensure a research space and niche for themselves, writers use the language resources available, among which, most importantly, are superscripts, tenses, and collocations. Superscripts refer the reader to the reference list of the article, while tenses connect the elements of time and action. Collocations refer to rhetorical phrases, or patterns, common in all texts and more specifically in genre texts. The most frequent tense is that which connects past action to a present time. Tenses play a role in bringing together the research argument, through specific verbs like "report", "show" and "require", and structures like "it is clear that". Collocations play a role in the creation of specific meanings.

In fact, much has been stated about the variety of tenses used in medical research article sections and their implications. For example, Salager-Meyer (1992 and 1994) talks about the importance of citations in introduction and discussion sections of the MRA in terms of hedging. Swales, (1990), also shows that methods and results sections are written differently from introduction and discussion sections that are more argumentative. Method and result sections represent a series of narrative statements that are less "explicitly cohesive" than introduction and discussion sections.

Typical collocations, or patterns, in medical research are notable with the use of citations, and with the use of prepositions. When the writer makes reference to research results performed by other researchers, specific language patterns are used, such as "*it has been reported that*", "*the study shows that*", and other typical constructions. Similarly, in

methods and results sections, specific patterns are used such as an intensive use of prepositions, exemplified by a raw frequency of the preposition "*of*". These patterns become regular and somewhat routine, and appear in all texts that report medical research. This regularity in the language is a clue that such patterns are useful to learners.

In addition to citation patterns and prepositional patterns, corpus linguistic tools provide word frequencies of texts. For example a list of frequent words in a corpus of medical research articles shows that in medical texts there are grammar words, research words, and medical terminology (Jabbour 1998). Research words associate with specific grammar words to form the interactive elements that provide context to the medical terminology. The high frequency of certain prepositions such as "*of*", "*in*" and "*to*" and other grammar words is necessary for the processing of both research activities and medical terminology. Word combinations, or word associations, gathered from a corpus of medical research articles can therefore be classified for their role in organizing text, or for expanding phrases, and limiting meanings. Most representative of such expansions are the phrases with the preposition "*of*".

2.2. Reading Processes and Text Types

After gathering patterns valuable to learners, it is necessary to understand what reading strategies improve learners' understanding and retention of information. According to Carrell et al (1988), reading involves interaction between top-down and bottom-up views of the text. In other words, reading involves both knowledge of text discourse and knowledge of linguistic elements. Text discourse and linguistic elements are interrelated. Discourse, or

top-down reading, controls linguistic elements and linguistic elements, or bottom-up reading, define the discourse.

According to Carrell et al (1988), interaction between the top and the bottom views builds the coherent knowledge that instruction seeks to instill as learners develop their language proficiency, relating words to meanings, or forms to functions. When processing a text, linguistic elements and developing subject knowledge interact. The output of this interaction is text understanding, reading comprehension and ultimately progress in language acquisition.

When on task, learners interpret linguistic elements, such as words, phrases or patterns within the context of their occurrence. For example, the phrase '*is required*', in the introduction section of an MRA is likely to be interpreted as a question that needs to be examined in a research context. By contrast, the same phrase in a discussion section implies direction for future research. An introduction section provides context for current research, a discussion section provides context for further research. Learners must be given a clear idea of how a target discourse is likely to develop, and the language associated with it. As one question is answered, or one stage of the discourse is established, learners should be able to predict what is likely or bound to follow (Willis 1990).

On the other hand, the focus on form in ESL is building momentum. This has given importance to the notion of fixed phrases. As stated earlier, the correspondence between article sections and word recurrence in specific combinations, or collocations, is enlightening. Computerized article sections offer recurrent words and combinations of representative words. Consequently, the availability of discourse knowledge, and of linguistic patterns and phrases specific to one type of discourse will make it convenient for learners to build their

discourse and linguistic knowledge interactively. This type of instruction does not build on decontextualized memorization of patterns as in the structural era, but is a conscious progression from the use of patterns to understanding and internalizing discourse and its associated expressions.

Text types condition the reading process. For example, introduction and discussion sections of a MRA imply communication in the form of a dialogue between the writer, the reader and the research community concerning the exposition of causal relations the research writer advances. Methods and results sections are narrative genres, built on events well grounded in space and time, providing the physical context of the experiment. Thus, the writing of a MRA embodies causality, and evidence, against a background of interaction involving the writer, the reader and the community. But to be accepted by the community, research writers must assert a link between their contribution and previous research in the field, thus the importance of citations in medical research. The research article reader must recognize how the link to other articles and researchers is expressed linguistically.

Expository and narrative writings are distinct. Grabe (1996) recognizes the importance of distinguishing between text-types in reading. The distinction between expository and narrative writing depends on language features. Grabe uses the two terms 'information' and 'events' to refer to expository and narrative texts and says that in each type there are language features, which show the reader how the text is structured.

Major genre distinctions such as narrative and expository represent extremely important and useful modes for organizing information and language use in discourse. It is obvious that there are major divisions between types of texts that primarily convey new information and texts which relate events and tell stories. Certainly there seems to be an aesthetic mode which incorporates unusual combinations of formal language features for conscious effects upon the reader. (Grabe 1996)

This suggests that there is a link between the purpose of a text and the kind of language used to realize that text. It also suggests that meaning is a function not only of the form of words, but also of the textual environment in which those words occur. The relationship between the structure of an MRA and the language associated with elements of that structure is essential. Nattinger and DeCarrico (1992) refer to certain language items, lexical phrases, as 'form-function composites'. These phrases are largely or partly fixed and have a predictable occurrence in discourse structure.

Furthermore, there are grammatical patterns, phrases, or collocations that occur in a regular fashion to perform specific functions in the discourse. Phraseology is a term associated with corpus linguistics denoting frequent word associations and word combinations, essentially based on phrases. The term has been increasingly used with a stronger implication that may conflict with the concept of syntax.

Reading a MRA, and practically covering any text that pertains to the content of a scientific subject should therefore take more account of word frequencies and collocation. Certain vital words create patterns representative of the discourse. A research corpus can be investigated regularly for the presentation of materials that fit the needs of the learners. Teaching materials can thus be built with reference to the collocations which are found to be representative of the discourse. At a preliminary stage of the instruction, the patterns can be chosen from the most central. The less frequent and more varied patterns may be considered at a later, more advanced stage.

3. Corpus linguistics and Teaching

3.1. The Theory and its Methodology

The initial procedure of corpus linguistics is the compilation of representative texts of the discipline the English teacher is involved in, such as medicine. Because the unit of study is the word and its combinations, the larger the size of the corpus, the larger the number of occurrences of research words, and the more evidence there is to support conclusions. Recommendations for a corpus size in specific areas, like medicine, are however scarce in the literature.

The frequency of a word on a wordlist is indicative of the texts collected in the corpus (Renouf 1992). For example, a corpus of medical research texts shows that the occurrence of the preposition "of" and of the form "have/has" is larger than in a corpus of general texts. By contrast, a corpus of second language learners' writings shows a significantly higher occurrence of the pronoun "I" and of the conjunction "and" (Jabbour 1999¹). A further look at the frequency list of a corpus of medical texts provides the following information. There are grammar, research, and medical words. Grammar words are the most frequent, followed by research words. Grammar and research words provide the context for the medical terminology. The preposition "of" provides a frame for semantic networks, such as *the presence of, the absence of, the prevalence of, the probability of*, etc. The forms "have/has" associate with the forms "V+ ed" and "been" to relate a past action to a present time, as in *recent studies have demonstrated that, it has been proposed that*, etc. Therefore, starting with the highly frequent words, it is possible to identify the common frameworks that govern the use of relatively less frequent words. This in turn leads to specify the occurrence of medical words denoting particular semantic categories such as "symptoms", "diseases", or "treatment".

¹ Paper entitled "Visibility and Distance in ESL Writing", presented at NYSTESOL Annual Conference in Melville, New York

The design of teaching materials for medical students, therefore, begins with the identification of the most frequent words in a corpus, in other words the grammar words, which leads into the use of research words and the terminology that pertains to the discipline. The next stage is to look at the patterns these words create. In some cases it is possible to identify a 'frame' by characterizing the items associated with a particular word or phrase, such as "it + is + adjective + that", as in "*It is noteworthy that*", and "in + noun + with", as in "*in patients with*".

A more expanded study of "*in patients with*", for example, shows that the words which follow this phrase refer to disease in most cases, such as "*in patients with diabetes mellitus*"; "*in patients with sickle cell disease*"; or "*in patients with peripheral vascular disease*". They may also be more specifically oriented towards the patients' history as in the methods section, such as "*in patients with known allergy*"; or "*in patients with solid tumors*". In results sections, there may even be a tendency for the disease to be more detailed as in "*in patients with inactive SLE*".

The new option opened up by computers is to evaluate actual instances and select the most typical. A complete set of instances of this kind should exemplify the dominant structural patterns of the language without recourse to abstraction, or indeed to generalization. (Sinclair 1986)

In corpus linguistics, two paths are available for the text analyst. One involves looking at the word and its combinations and then moving up to the sentence level where meaning is constructed and reflected at the level of the discourse of the section. The second is to look at the text starting with the highest level. This involves the recognition of the functions of the sections and the recognition of the functions within the sections. In this manner, the analyst is identifying interactive elements that connect functions and text at every level of the text construction. The text in this sense means the consideration of the individual

lexical items in the direct context of their environment and in the larger context of the section in which they occur. In a pure lexical approach, the reader would be involved with the level of word. In a discourse or functional approach, the reader would start with the highest level.

To summarize, the preparation of teaching materials generated by corpus linguistics tools requires that the teacher first perform an analysis of the discourse of the texts forming the corpus in order to find out their typicality, and then carry out a search of the most frequent words. This search yields significant collocations that can be exploited in teaching. The behavior of words and their combinations is enlightening, and obviously, the larger the size of the corpus the larger the number of occurrences of specific words there are to choose from.

3.2. Learners as Researchers: the Use of Technology

The above example with "*in patients with*" makes it clear that teaching material based on corpus research brings to the teaching experience a richer view of how language is used in specialized contexts. In the past, new audio-visual facilities have been made available to the teacher. The language laboratory, for example, provided valuable resource for the teaching of the sixties and the seventies. At the same time audio and videocassette players became widely available in the classroom. This had a profound effect on teaching, for example by making it possible to look much more closely at the spoken form of the language.

The impact of corpus linguistics on language description and language teaching is, however, potentially far greater than the technologies of the sixties and seventies. In the early eighties, Johns (1981), in the context of teaching ESP, argued that the computer provided the ESP teacher with the means to produce more flexible ESP teaching materials. Now, with the advent of the 21st Century, there are even greater possibilities for reliance on the computer

for language teaching and practice. Central to these possibilities is the notion of the learner as researcher (Johns and King 1991). This notion envisages the possibility of learners being presented with concordanced citations of the target language and working from these citations to create their own picture of the language.

The new paradigm² that considers the learner a researcher, is, however, evolving mainly as a result of corpus linguistics and the advent of corpora and corpus studies. This paradigm is making its shift towards specifying linguistic items in texts, thus emphasizing the forms used in language teaching.

Because a corpus-based instruction is based on language description, it goes beyond the confines of the traditional structural syllabus in that it provides the 'natural environment' of language use. Willis (1990) states that the lexical syllabus, an example of corpus-based teaching, involves both the structural and the communicative syllabuses.

The lexical syllabus does not identify simply the commonest words of the language. Inevitably it focuses on the commonest patterns too. Most important of all it focuses on these patterns in their most natural environment. Because of this, the lexical syllabus not only *subsumes* a structural syllabus, it also indicates how the 'structures' which make up that syllabus should be exemplified. It does this by emphasising the importance of natural language. (Willis 1990) (Emphasis in original)

The importance of Willis' quotation is that the "commonest words" mean the most frequent words sorted out through computerized text processing, and that the "commonest patterns" refers to the patterns in which the commonest words occur sorted out by concordances and collocations.

In a lexical approach, the texts are the means to an end, which is the acquisition of the language through genre texts. A corpus approach provides more insight about the language

² Title of a presentation by Tim Johns at the 3rd North American Corpus Linguistics Symposium, Boston, 23-25 March, 2001

because the most typical "lexical phrases" (Nattinger and DeCarrico 1992) and patterns of a specific text-genre will come more readily to the fore, in their original context. The strength of a corpus approach in genre specific texts is that the texts are transformed into huge data samples, each category of which can be specified at two levels, the discourse level and the linguistic level.

4. Criteria to Develop a Language Program

The identification of the linguistic items used in patterns and phrases can most usefully be highlighted to learners. The methodological procedure to do so would be by providing a series of tasks that will involve comprehension of target texts. These texts can then be analyzed by learners in such a way as to focus on the linguistic items.

What corpus linguistic tools assist in is collecting such recurrent patterns through word and word combination searches, which otherwise would have been impossible to identify. With corpus linguistic tools, the patterns not only become clearer, but larger instances can be gathered. These patterns can additionally be customized depending on the target level of learners. In general, learners benefit tremendously from exposure to examples of typical text patterns, mainly if these are structured the way a corpus linguistics tool generates them, as we shall see later in the section on design of teaching materials.

Four stages are involved in the design of corpus-based materials. The first stage involves analyzing the corpus, a sub-corpus or a representative text, to determine the structure of the discourse and the communicative acts involved. The second stage involves mapping functions to linguistic exponents in the text, or performing an analysis of the genre. The third stage is the provision of a list of the most frequent words in the corpus. The fourth stage involves relating frequent patterns with their function in discourse and their occurrence in the discourse framework.

To develop a language teaching program, consideration must be given to external and internal elements of the process. Developments in theories pertaining to language, psychology, pedagogy and outcomes of their applications are external factors that influence the creation of a teaching program. Approach and methodology form the intrinsic part of the process and of its day to day implementation. Approach is more theory based while methodology is seen as the practical aspect of the teaching. A quality program therefore must connect field knowledge to teaching theories, and their associated technologies.

Knowledge of the field relies mainly on text investigation. Teaching theories are generally outcomes of studies performed on language description and language acquisition. Defining the language of science has, since the early 60's, included the compilation of lists of the lexical items, and core terms, in a manner similar to West's *General Service List of English Words* (West 1953). At present, a more current inclination is to view language as formed of chunks, rather than single words, and to view acquisition through a cycle that starts with memorization, followed by acquisition of grammar, and a reinforcement stage of lexis and grammar. Studies performed on field-specific texts have shown that scientific language indeed relies on word frequencies and phraseology. If text analysis is performed using corpus linguistic tools, and if representative phrases are gathered in a structured manner, material designers would undoubtedly offer students, in both content-based and ESP courses, the best a language course can offer.

6. Focus on Discourse and on Language

Discourse analysis has influenced ESP course design because there is a tacit assumption that text discourse, or rhetorical patterns and text organization differ from one specialist area to the other. The basic notion in the theory of discourse and genre analysis is

that any text is represented at a higher level of organization than just the sentences forming that text. This has led to a conflict between a focus on discourse and a focus on forms in classroom instruction, leading, at best, to a strict separation of grammar instruction from the study skill component, if not to a complete cancellation of grammar classes. As is well known, the cancellation of grammar classes has led to students weak ability in reading and writing.

Unless corpus linguistics is used, once the focus of attention is on a study skill or text discourse, favoring a top-down view of text, it becomes difficult to restore a balance that offers a realistic focus on language code. On the other hand, when focus on code is the selected approach, grammar and syntax become the central focus. Corpus linguistics offers the possibility of presenting the grammar of the language without losing touch with the discourse of the selected texts. All this implies a two-level structure for a unit of interaction, the discourse level and the grammar level. Although each of these two levels has a divergent unit of analysis, corpus linguistics allows the simultaneous contextualization of both.

7. Conclusion

The role of the computer in English studies originated in lexical studies, in the early sixties, and corpus linguistics is one of the outcomes of such studies. The Brown corpus, of Brown University, with its one million words is the first collection of texts intended for large-scale text analysis purposes. With the COBUILD³ project in the eighties, and the BNC⁴, prospects of using the computer for larger scale text processing is coming to fruition. Linguists working with corpora aim at describing the language with an emphasis on empirical evidence and naturalness (Sinclair 1991).

³ Acronym for the Collins-Birmingham University Lexical Database, also referred to as the Bank of English

The benefit of such computerized studies to language teaching is threefold. First, a large number of texts can be investigated as one unit under the same conditions, thereby providing possibilities for constant generalizations to be made. The second important point is that a corpus study enables teachers to look at language in a range of contexts, starting with words and phrases, and moving upward to larger units such as chapters and sometimes whole books (Phillips 1989). The third point is that enough information is available to determine not only the immediate environment of a word, but also its developing function in text. The practicality of a research corpus is that teachers can take the corpus into the classroom and call up relevant material for student' analysis.

A corpus-driven design of a teaching program that is associated with discourse analysis caters for both features of meaning and features of structure. It reconciles language function and language form, something the structural and the general notional/communicative approaches have failed to do. A corpus-based program represents a higher surrender value course.

Additionally, context is the key element that provides strength to a corpus-driven program. This means that the language item is not to be equated with a function, but forms a contribution to that function. In the same manner, the language function itself is not an abstract notion, and does not exist separately from the words that form it. In a hierarchical multi-layered structure, the item does not play its role individually. At whatever level, higher or lower, the language item can only offer incomplete meaning until all levels of its occurrence are involved. Text linguistics offers the possibility of defining the levels of analyzed texts. The look at language in this manner coincides with top-down and bottom-up

⁴ Acronym for British National Corpus

reading strategies. A bottom-up reading strategy focuses on the language used in text, and seeks to show how language creates and determines the message at the top.

WORKS CITED

- Baron, N. S. (1989) "From Print Shop to Desk Top: Evolution of the Written Word", GURT '88 Synchronic and Diachronic Approaches to Linguistic Variation and Change, T. J. Walsh (ed.), Washington DC: Georgetown University Press, 8-21
- Carrel, P. et al (eds) (1988)Interactive Approaches to Second Language Reading, Cambridge University Press
- DeCarrico, J. and Nattinger, J.R. (1988) "Lexical Phrases for the Comprehension of Academic Lectures" in English for Specific Purposes, Vol.7, pp 91-102
- Grabe, W. (1996) "Textual Coherence and Discourse Analysis", paper presented at the AILA '96 Symposium, Finland, August 4-6, 1996, mimeo
- Jabbour, G. (1998) "Corpus Linguistics, Contextual Collocation and ESP Syllabus Creation: A Text Analysis Approach to the Medical Research Article", PhD Thesis, University of Birmingham
- Johns, T. and P. King (eds.) (1991) Classroom Concordancing, Centre for English Language Studies, ELR Journal, Vol. 4, The University of Birmingham
- Johns, T. (1981) "The Uses of an Analytic Generator: The Computer as Teacher of English for Specific Purposes", in The ESP Teacher: Role, Development and Prospects, ELT Documents: 112, The British Council
- Nattinger, J.R. and DeCarrico, J.S. (1992) Lexical Phrases and Language Teaching, Oxford University Press
- Philips, M. (1989) Lexical Structure of Text, Discourse Analysis Monograph No. 12, English Language Research, University of Birmingham
- Polgar, S. and S.A. Thomas (1991) Introduction to Research in the Health Sciences, Churchill Livingstone
- Renouf, A. (1992) "What Do You Think of That: A Pilot Study of the Phraseology of the Core Words of English" in G. Leitner (ed), New Direction in English Language Corpora, Mouton de Gruyter
- Rumelhart, D. E. (1980) "Schemata: the Building Blocks of Cognition", in R.J. Spiro et al (eds), 33-58
- Salager-Meyer, F. (1992) "A Text-Type and Move Analysis Study of Verb Tense and Modality Distribution in Medical English Abstracts", English for Specific Purposes, Vol.11, No.2, 93-113
- Salager-Meyer, F. (1994) "Hedges and Textual Communicative Function in Medical English Written Discourse", English for Specific Purposes, Vol.13, No.2, 149-170
- Sinclair, J. McH. (1986) "First Throw Away Your Evidence", in G. Leitner (ed), The English Reference Grammar, Tübingen: Max Niemeyer Verlag
- Sinclair, J. McH. (1991a) Corpus, Concordance, Collocation, OUP
- Sinclair, JM and Renouf, A.(1988) "A Lexical Syllabus for Language Learning" in Vocabulary and Language Teaching, Carter, R. and McCarthy, M. (eds), Longman
- Swales, J.M. (1990) Genre Analysis: English in Academic and Research Settings, Cambridge University Press
- West, M. (1953) A General Service List of English Words, Longman, Green and Company
- Willis, D. (1990) The Lexical Syllabus, London: Collins

THE ROSE

1

There are those to whom place is unimportant,
But this place, where sea and fresh water meet,
Is important—
Where the hawks sway out into the wind,
Without a single wingbeat,
And the eagles sail low over the fir trees,
And the gulls cry against the crows
In the curved harbors,
And the tide rises up against the grass
Nibbled by sheep and rabbits.

A time for watching the tide,
For the heron's hieratic fishing,
For the sleepy cries of the towhee,
The morning birds gone, the twittering finches,
But still the flash of the kingfisher, the wingbeat of the scoter,
The sun a ball of fire coming down over the water,
The last geese crossing against the reflected afterlight,
The moon retreating into a vague cloud-shape
To the cries of the owl, the eerie whooper.
The old log subsides with the lessening waves,
And there is silence.

I sway outside myself
Into the darkening currents,
Into the small spillage of driftwood,
The waters swirling past the tiny headlands.
Was it here I wore a crown of birds for a moment
While on a far point of the rocks
The light heightened,
And below, in a mist out of nowhere,
The first rain gathered?

22

196

2

As when a ship sails with a light wind—
The waves less than the ripples made by rising fish,
The lacelike wrinkles of the wake widening, thinning out,
Sliding away from the traveler's eye,
The prow pitching easily up and down,
The whole ship rolling slightly sideways,
The stern high, dipping like a child's boat in a pond—
Our motion continues.

But this rose, this rose in the sea-wind,
Stays,
Stays in its true place,
Flowering out of the dark,
Widening at high noon, face upward,
A single wild rose, struggling out of the white embrace of the morning-
glory,
Out of the briary hedge, the tangle of matted underbrush,
Beyond the clover, the ragged hay,
Beyond the sea pine, the oak, the wind-tipped madrona,
Moving with the waves, the undulating driftwood,
Where the slow creek winds down to the black sand of the shore
With its thick grassy scum and crabs scuttling back into their
glistening craters.

And I think of roses, roses,
White and red, in the wide six-hundred-foot greenhouses,
And my father standing astride the cement benches,
Lifting me high over the four-foot stems, the Mrs. Russells, and his
own elaborate hybrids,
And how those flowerheads seemed to flow toward me, to beckon me,
only a child, out of myself.

What need for heaven, then,
With that man, and those roses?

197

23



*U.S. Department of Education
Office of Educational Research and Improvement
(OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)*



Reproduction Release
(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: <i>Corpus-based Instruction</i>	
Author(s): <i>Georgette Tabbour</i>	
Corporate Source:	Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, Resources in Education (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign in the indicated space following.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY <hr/> TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY <hr/> TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	<input checked="" type="checkbox"/>
Level 1	Level 2A	Level 2B
↑ <input checked="" type="checkbox"/>	↑ <input type="checkbox"/>	↑ <input type="checkbox"/>
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g. electronic) and paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only
Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.		

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche, or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Signature: <i>G. Jabbour</i>	Printed Name/Position/Title: <i>Dr. Georgette Jabbour</i>	
Organization/Address: <i>NY Institute of Tech. P.O.B. 8000 Old Westbury, NY 11576-2165</i>	Telephone: <i>(516) 686-7713</i>	Fax: <i>(516) 687760</i>
	E-mail Address: <i>gjabbour@nyit.edu</i>	Date: <i>4/18/01</i>

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:

Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706
Telephone: 301-552-4200
Toll Free: 800-799-3742
e-mail: ericfac@inet.ed.gov
WWW: <http://ericfac.piccard.csc.com>

EFF-088 (Rev. 9/97)