Charles C. Fries, Linguistics and Corpus Linguistics

Peter H. Fries, Department of English Language and Literature,
Central Michigan University

Halliday has long claimed that information concerning the relative frequencies of the various options within a system should be considered part of the system itself. Such a position entails that linguists have some basis for describing these frequencies. Hence SFL has made considerable use of corpora and of corpus linguistics. Of course, the field of corpus linguistics is commonly regarded as a brand new approach to linguistics which has developed and become popular over the past forty years—since the development of computers. Like all new fields, however, its roots lie in earlier forms of the discipline. This paper addresses one of the forebears of this field, Charles C. Fries. He thought of himself simply as a linguist (not a corpus linguist), yet his theory and practice have much in common with current versions of corpus linguistics and SFL

His approach grew out of his background in the history of English and his work as editor of the *Early Modern English Dictionary*. As a part of his very conscious effort to make linguistics a science, he gradually developed a theory of grammar which emphasized the signals in the language code which led listeners to interpret the language in the way they did. His theory placed primary emphasis on contrast — the paradigmatic relations among the entities described.

I will address 5 trends in his work which relate to current corpus work:

All conclusions should be based on the analysis of a clear body of data gathered to represent the language of the community he intended to analyze.

The analysis must be systematic and exhaustive.

The analyses should address relative frequency counts of patterns in contrast.

The basis for comparison of forms was the meanings they expressed.

The meanings of words are regularly distinguished by looking at their 'lexical sets' (his term for what we now call collocations).

Finally, his work will be compared with selected aspects of Systemic Functional Grammar.

KEYWORDS: Charles C. Fries, corpus linguistics, history of linguistics, goals of linguistics

1   INTRODUCTION

From the first stirrings of the set of concepts that were to develop into Systemic Functional linguistics it has been considered important to gather examples of the language really used by people as they interact. The title of a book published by James Benson and William Greaves in 1973 illustrates our attitude at the time. It was called *The language people really use*. Sometimes this interest has been informal, and satisfied simply by listening to the speech around us and noting various examples of interest. At other times the interest is satisfied only by carefully gathering a sample of the language to use as a source of examples—a corpus—and working systematically with that corpus. These days, with the advent of computers, the more careful approach to gathering examples has been expanded to gathering large corpora of millions of words, and the discipline which has developed as we process these extremely large corpora has been termed "corpus linguistics".

But in fact, the size of the corpus and the use of a particular tool, the computer, should not define a discipline.[1] Rather, it seems to me that what defines the discipline of corpus linguistics should be the assumptions concerning the nature of language, what one considers to constitute evidence concerning the nature of the language being described, and the principles which underlie the gathering and use of a corpus in linguistic analysis. Of course the size of the corpus and the tools used DO affect the sorts of results one may obtain. (In this case, large differences in size make a qualitative difference in the sorts of results that can be obtained.) But the fundamental approach, the fundamental assumptions about science and about the nature of language, the issues that are considered interesting, and the methodology used to explore those issues should remain roughly constant regardless of the corpus size or the tools used. If you grant me this, at least with a 'willing suspension of disbelief', you will agree that corpus linguistics has roots which extend at least back to the beginnings of modern linguistics. Certainly by the time I was being initiated into linguistics in the late 1950's, the gathering of representative samples of language and the development of techniques to efficiently recover relevant information from the data (now called 'data retrieval') was an important part of our education. At that time we were not learning corpus linguistics, we were merely learning how to be linguists.

Because of the importance of corpus linguistics to the Systemic Functional endeavor, and because the organizers of this conference felt it useful to include corpus linguistics as one of the underlying issues for discussion at this meeting, I thought it might be useful to discuss how corpora were used in one tradition of early work using corpora[2], and to note some of the issues that were encountered. Specifically, I want to

---

[1] This view contrasts sharply with Simpson and Swales (2001: 1) statement: "Corpus linguistics is essentially a technology…."

[2] I do **not** claim that his work was typical of the time. Indeed, his work differed significantly from that of other linguists of his generation, particularly those in the US. (As partial support for my position, note the exclusion of any significant discussion of Fries's work from the otherwise extensive history of American structuralism in Hymes and Fought (1981), in spite of the fact that they considered him an important figure in the linguistics of that time (personal communication from Fought).)

examine the work of one figure from the first half of the 20'th century, Charles C. Fries, who consistently worked with corpora. Before I discuss his approach to corpora I want to say a bit about his personal development and his development as a linguist, for his experiences greatly affected what he considered important in linguistics, what goals he chose as a linguist, and the methodolgies he used to achieve these goals. Since his approach to corpora depended greatly on the goals he chose, in this presentation I will organize the paper around his fundamental assumptions about language and his goals, and then discuss how these principles and goals affected his approaches to specific corpora.

Section 1.1 provides a bit of his personal history. 1.1.1 shows his relation in age to other linguists of his generation. 1.1.2 lists some selected events, interests and publications in his life.

1.1   HISTORY:

1.1.1   Charles Fries and selected other linguists of his generation:

Charles C. Fries   1887 - 1967

Leonard Bloomfield	1887 - 1949	Roman Jakobson	1896 - 1980
J. R. Firth	1890 - 1960	Daniel Jones	1881 - 1967
Louis Hjelmslev	1899 - 1965	Edward Sapir	1884 - 1939

1.1.2   Selected events and publications in Fries's career.[3] (Dates associated with specific titles indicate dates of publication of those works. Titles with asterisks involved the analysis of some specific corpus of data and were begun several years before publication.)

~1911   Teaching classical Greek (5 years). (This experience stimulated the initial development of his 'Oral Approach' for teaching foreign languages.)

~1916   Moved to teaching English literature and composition. Became interested in problems of teaching Literature, and in teaching English in schools (English to native speakers of English).
Became interested in the history of English and its relevance for the language features his students were writing.

~1918   Went to the University of Michigan to study rhetoric with F. N. Scott. Later he moved into historical linguistics.

1922   Received Ph. D. in English at the University of Michigan. Dissertation: The periphrastic future with *shall* and *will* in Modern English

1926   *The Teaching of Literature* (with Hanford and Steves.)

1925   *The periphrastic future with *shall* and *will* in Modern English.

1920's & 1930's   His duties at the University of Michigan included teaching English composition and literature at the University High School associated with the University of Michigan school of Education.

---

[3] See R. W. Bailey 1985b for a more extensive discussion of CCF's life.

1927	the expression of the Future
	The teaching of the English Language
1928	Became editor of the *Early Modern English Dictionary*
1940	Became Director of English Language Institute at the University of Michigan
1940	*On the development of the structural use of word order.
	English Word Lists (with Eileen Traver)
	**American English Grammar*
1945	*Teaching and Learning English as a Foreign Language*
1952	**The Structure of English*
1954	Meaning and linguistic analysis
1961	*Foundations for English Teaching* (with Agnes C. Fries)
1963	*Intonation of yes-no questions.
	Linguistics and Reading
1970	*The time "depth" of coexisting conflicting grammatical signals in English.

2 ASSUMPTIONS CONCERNING LINGUISTICS AS A SCIENCE, THE NATURE OF LANGUAGE, AND THE USE OF A CORPUS

2.1 THERE IS A CLOSE RELATION BETWEEN THEORETICAL AND APPLIED LINGUISTICS:

Like most linguists of his generation, Fries began his professional life teaching language —in his case, teaching Classical Greek. As a result of his struggles and research to improve his teaching he found the works of Otto Jespersen and Henry Sweet and they attracted him into a more careful study of language. When Greek was no longer to be a required course in highschool, he moved into teaching English composition and literature. His interest in teaching composition attracted him to the University of Michigan where he worked for a time with F. N. Scott, a professor of Rhetoric. This sequence of events was typical of his approach throughout his life. He encountered practical problems and then carefully and systematically brought to bear all the theoretical knowedge he could find to address the problem. Indeed he saw a close relation between theoretical and applied linguistics. In a letter to Albert Marckwardt (Fries 1944) in which he discussed the relation of theoretical linguistics and teaching English as a foreign language he said "… linguistic theory must be tested by practical applications and practical teaching will help to develop that theory, …"

In fact, most of Fries's theoretical projects arose out of issues that he encountered first in some aspect of his language teaching.

2.2 LANGUAGE MUST BE APPROACHED SCIENTIFICALLY: THE ESSENCE OF SCIENCE IS PREDICTION OF DISPARATE PHENOMENA.

He wanted to make linguistics a science. He believed that the basis of science was prediction of disparate phenomena—for example, one uses the law of gravity to predict

how objects fall. In language, he wanted to describe those aspects of the forms of language which best predicted the responses (particularly the recognition responses) of the listeners. In other words, he focussed on the signals in the language which led people to interpret the language the way they did. As he said (1967: 668),

> [structural] grammar aims not at definitions and classifications but at such a description of the formally marked structural units as will make possible a valid prediction of the regular recognition responses that the patterns will elicit in the linguistic community.

Given the importance of 'listener response' to his whole enterprise, we should see that, for him, a corpus was not merely a set of forms which had been uttered, but it also involved the reactions engendered by those forms in the interaction. This aspect is explicitly mentioned in his methodological discussion in the *Structure of English*, where one of the steps of analysis he describes is to group the single free utterances "… in accord with the responses that followed them. All the evidence that appeared in our records *concerning the nature of the response* was used for this purpose." (1952: 41)[4].

Given his goals for linguistics, it is no wonder that one of his major criticisms of traditional grammar was that it aimed primarily at providing a taxonomy of the language (e.g. classifying sentences into questions, statements and commands, etc. and words into the eight parts of speech —nouns verbs, adjectives, etc.). As he often said, traditional grammar does not address the question of how listeners KNOW that a given sentence is a statement, command or question.

He even considered the goal of transformational grammar described in Chomsky (1957: 13) to be a taxonomic goal.

> The fundamental aim in the linguistic analysis of a language L is to separate the *grammatical*[5] sequences which are the sentences of L from the *ungrammatical* sequences which are not sentences of L and to study the structure of the grammatical sequences.

Clearly, advocates of the formalist approach, even in its later more sophisticated forms, never really addressed his issue of describing the signals in the language that cued listeners to interpret the language the way they did. From his point of view, formalist descriptions of languages which approached the task by first equating a language with a set of sentences, and then tryied to describe that set by pointing out parallels (even very abstract and sophisticated parallels) among the various members of that set were simply being taxonomic.

---

[4]   In this as in all following quotations, emphases through italics or through underlining were in the original. If I [PHF] wish to emphasize some portion of a quotation, I will use bold face.

[5] All italicized words and phrases in the quotations in this paper were italicized in the originals.

2.3   The paradigmatic aspect of language is essential.

He consistently argued against treating language as a set of disconnected items. Rather, in his view it was the RELATIONS among these items that was important. One can see implications of this view in the fact that he distrusted phonetic similarity as anything more than a useful field technique for making a phonemic analysis. What was critical was the role of the sounds in the system as a whole. (See Peter Fries 1982 for a more careful discussion of this point.) In his book *Linguistics and Reading* Fries (1963: 62), we find Fries speaking more generally about the importance of contrast.

> There is power or force in the structural system itself. The habits that constitute the control of one's native language are not habits concerning items as items, but habits concerning contrastive items as functioning units of an *ordered system of structural patterns.*

And a few pages later (1963: 64):

> From our structural point of view, items such as these ["items of sounds that must be pronounced, the individual words that must be identified with the meanings, the parts of sentences that must be classified"] have no linguistic significance by themselves. Only as such items contrast with other items in the patterns of an arbitrary system do they have linguistic significance. In other words, all the significant matters of language are linguistic features in contrast.

His emphasis on contrast and the underlying paradigmatic relations is evident in his treatment, in *Structure of English* (Fries 1952: 79)*,* of the words that belonged to the major classes.

> It is not enough for our purposes to say that a Class I word [~ noun] is any word that can fill certain positions in the structure of our sentence, even if we enumerate all these positions. We want to know what the special characteristics of these words are that make them recognizably different from the words used in other positions. To discover these characteristics we need to explore these other positions and form comparable lists of words that can fill these positions. Significant formal characteristics of each class will appear then in the contrasts of one class with another.

As a means of achieving this goal, his chapter 7 of that book (*the Structure of English*) addresses the formal characteristics of parts of speech. While the chapter begins by addressing an aspect of the morphological make-up of the words belonging to the major classes, he does not simply provide a morphological analysis. Rather, he assumes that morphological analysis and he provides lists of word classes in contrast. These lists demonstrate differences in the internal structures of the major word classes that lead to the recognition that the word belongs to one or another major class. Thus a portion of

a table headed "class 1 contrasting with class 2" lists examples such as those in Table 1 (from Fries 1952: 113)[6].

Table 1: Sample of Fries's lists showing contrasts in the forms of words which belong to the major classes

	Class 1	Class 2		Class 1	Class 2
1.	arrival	arrive	3.	delivery	deliver
	refusal	refuse		discovery	discover
	denial	deny		recovery	recover
	acquittal	acquit	4.	acceptance	accept
2.	departure	depart		acquaintance	acquaint
	failure	fail		admittance	admit
	erasure	erase		annoyance	annoy

Further implications of his emphasis on contrast will be encountered in the discussion in section 2.7.

2.4   THE SPOKEN LANGUAGE IS PRIMARY.

Like many linguists of his time, Fries felt that the 'real' language was the spoken language of the people.[7] As a result, in all of his work he made every attempt to discover what that spoken language was. His discussion (Fries 1927: 137) of the teaching of English to native speakers implied a typical goal.

> … the schools seem to be committed to the program of equipping the pupils with the language habits of those we have called the socially acceptable group. …

but he notes a few pages later (1927: 140)

> There has never been an adequate scientific survey of the spoken language in English…

---

[6] It should be noted that he "assumes that the morphemes have been identified" (1952: x), thus he is not merely looking at similarities in sound.

[7] Taking the spoken language as primary should not be equated with regarding the **sounds** of speech as primary. Rather CCF's interest consistently lay in the grammatical patterns used. He knew that the words and grammar of the spoken language differed greatly from the language that was written.

Indeed, his *American English Grammar* (1940: 27) was intended to provide just such a survey. He clearly wished to describe the spoken language.

> The ideal material of course for any survey of the inflections and syntax of present-day American English would be mechanical records of the spontaneous, unstudied speech of a large number of carefully chosen subjects.

However, before the mid 1940's it was very difficult to obtain a record of the spoken language, so he had to be satisfied with APPROXIMATIONS of spoken language in the written records. Thus his discussion (1940: 27) continues,

> The practical difficulties in the way of securing a sufficient number of records of this kind from each of a large number of subjects, sufficient to make possible the kind of study necessary in charting the field, seem to make it prohibitive as a preliminary measure.
>
> The use of any kind of *written* material for the purpose of investigating the living language is always a compromise, but at present an unavoidable one and the problem becomes one of finding the best type of written specimines for the purpose in hand.

Similarly, his 1922 study of the development of *shall* and *will* used dramas as the evidence because

> The language of drama is probably nearer to actual usage than that of other types of literature since the drama carries its effects through the speaking of actor to actual hearers. At the least, the language of the drama is perhaps the best compromise between the living spoken English and the written English of literature. (Fries 1925: 987)

His interest in obtaining approximations of the spoken language is also responsible for the fact that the data he used for the history of the structural use of word order contained only examples taken from prose. (He assumed that the language used in poetry was likely to deviate from the spoken language more than did the language of prose.)

2.5   LINGUISTIC ANALYSES SHOULD BE RELIABLE AND REPLICABLE.

In each of his projects, he wanted his descriptions to be based on a body of evidence which should, in principle, be available to other investigators for their inspection. Thus his descriptions are based on examinations of explicit corpora.[8] Table 2 lists the major projects he engaged in and the data used in each.

---

[8]   While his analyses were based on explicit, well defined corpora, as far as I know he never made the corpora available to other linguists by, say, placing copies in a public collection.

Table 2: Fries's projects which directly involved the gathering and use of a specific corpus of data.

	Project	Data source/size	Purpose
A.	**Corpus as an informal source of data**		
1.	*Early Modern English Dictionary*: a	~2,000,000 slips gathered for OED (1928) plus ~300,000 slips gathered at U of M. Readers asked to find special or new or otherwise noteworthy uses of words.	Discover changing patterns in the Early Modern English vocabulary.
B.	**Systematic analysis**		
1.	**Gathering data systematically, and exhaustive analysis**		
a.	*Early Modern English Dictionary*: b	~700,000 slips gathered from an "intensive" reading of 69 dated texts. Gathered essentially every instance of all major category words in these texts.	Ensure that all uses of each word are accounted for. Develop some evidence for describing a word use as usual or unusual.
b.	Structure of English	~50 hours (= >250,000 running words) of recorded phone conversation involving ~300 speakers.	Discover language features of spoken English.

2.	Gathering data systematically, exhaustive analysis, *and* counting coexisting conflicting signals of meaning		
a.	Periphrastic use of *shall* and *will*	English drama: 50 plays from every decade of British Literature, 1560 - 1915. 18 plays each from American and British Literature from 1902-1918. ~ 20,000 instances of *shall* and *will*.	Examine the actual uses of *shall and will* in English during the early modern English period with a view to evaluating the accuracy of the rules found in school grammars and their use in teaching English.
b.	History of English	English from 10'th century to mid 20'th century. 20,000 words from each time period, Samples taken roughly each 50 years. 400,000 words total.	Discover the stages by which a grammatical system characterized by the use of inflections changed into a system where the order of elements was a major signal of grammatical function.
c.	American English Grammar	2,000 complete letters, plus excerpts from about 1,000 additional letters, all written to the war department during WW I.	Compare the language habits of uneducated with those of educated. (First contrastive analysis)
d.	Intonation of yes-no questions	39 episodes of a TVshow ('What's my line?'). 2,561 yes/no questions	Check accuracy of previous descriptions of intonation of yes/no questions

In addition to the replicability issue, he also felt that corpora needed to be reliable samples of the normal language used as people interacted with one another. He was deeply suspicious of the conscious judgements of speakers when they were FOCUSSING on their language rather than USING their language for some immediate social purpose. (Note the adjectives in the phrase *the spontaneous, unstudied speech* in his description of the ideal data to use for his survey of American English grammar above.) He wanted samples of the language as actually used by speakers when they were occupied with the immediate demands of communicating in a situation. In his view, speakers do not really know what they actually say, and often provide inaccurate information when they are thinking consciously about the language they use. Fries (1964: 245) expressed his attitude toward mechanical recordings made in artificial laboratory situations.

> There must be mechanical records of a substantial body of materials which can provide any number of exact repetitions for analytical study. But we no longer believe that we can accept as satisfactory evidence the recordings, made in a laboratory, of specially constructed conversations read or recited by those who are aware that their language is being recorded. Such conscious recordings inevitably show many important differences from those live conversations, made when the participants do not at all suspect that recordings are in progress.

The need for systematically examining some corpus of examples runs through his work either explicitly or implicitly (1925, 1940a, 1940b, 1952, 1964), but he put the issue most clearly in a letter he wrote to me in 1959.

> Introspection, I believe is useful only as a source of suggestions or hunches that must be verified by an "objective" examination of a systematically collected body of evidence. Evidence, to be completely satisfactory, should be in such a form that it can be checked and re-examined by other workers. … In my own experience, I have found that I've been wrong so often in conclusions (especially concerning frequency) based upon impressionistic and casual observation, that now I'm never satisfied until I've been able to <u>record</u> systematically some definite body of evidence and list and <u>count</u> the occurrences comparatively. My conclusions may still be wrong, but at least they are good for the body of material examined and can be supplemented and corrected by others. [Underlining in the original]

Because of his attitude toward data and toward introspection, it is no surprise to find that late in his life, he was very critical of some of the assumptions and attitudes that formalists displayed toward the notion of grammaticality and toward gathering data. He felt that their approach was significantly weakened by their assumption that membership in the set of grammatical sentences of the language (the notion of grammaticality) was a non-controversial notion that needed no discussion. In conversation, for example, when he reacted to the notion of grammaticality used in formalist grammars, he often said, "You can SAY anything you want. The question that interests me is how will what you say be UNDERSTOOD." Given his attitude and interest, it is no wonder that he greatly mistrusted the goals, assumptions and results of

the transformational generative grammars of the time. One can glimpse this mistrust when in Fries (1963: 91) he wrote:

> In the discussions of those who have tried to understand these new approaches a number of fundamental questions have been raised for which adequate answers do not seem to be available in the published materials. Valid criteria for the judgments of "grammaticality" as applied to sentences are essential for a "generative" grammar. The theoretical and practical principles upon which the criteria now used depend seem hard to find.

2.6 THE DATA USED FOR ANALYSIS SHOULD BE A REPRESENTATIVE SAMPLE OF THE LANGUAGE OF SOME COMMUNITY.

Of course the notion of 'representative sample' is a vexed concept and what is considered to be representative depends greatly on what one is trying to represent. It seems to me that two means of sampling language have typically been used by linguists. One is to take a casual approach. Here we simply collect all examples that happen to catch our notice into some location.[9] These examples may be bits of conversation that we have heard, whole advertisements or letters or other texts that have caught our attention, or perhaps it is some text that we intend to use as an example for teaching some topic. Then, when we come to look for data of some type for our analysis, we search our store of examples and work with them. I compare this approach to the approach of a person who likes to sew, and collects all sorts of pieces of cloth just in case one might come in handy. Such an approach to data gathering is something all linguists do, and is quite valuable. The examples collected are all valid examples of language in use. They may not represent the full range of phenomena available, however, and they are very likely to misrepresent the frequencies with which certain language features are encountered.

An alternative approach to the gathering of data is to systematically gather a representative sample of some facet of the language, paying careful attention to what the sample is intended to represent and also to the techniques used to gather the sample. Two examples of careful samples are the old Brown and the LOB corpora which claim to be statified random samples of written British and American English.

Let me follow Matthiessen (2006:107) in using the term *archive* to represent the informal collection of data which happpen to be convenient, and *corpus* to refer to a more careful sampling of language. With this distinction in mind, Fries tried to gather corpora. While this was his goal, I suspect that most of the corpora that he used were compromises.[10] I have already mentioned that he wanted to examine the spoken

---

[9] All linguists have experienced, at one time or another, the joy of encountering in written or spoken language an example of some predicted but as yet unencountered structure (say a single verb form that uses all the secondary tenses possible). We then make a special effort to record or remember what was said and its context—sometimes to the exclusion of paying attention to what we should be concentrating on at the time.

[10] Perhaps all corpora are compromises between the demands of representativeness, the demands of the sampling procedures, and the practical demands of obtaining usable data

language, and yet he had no means to do so regularly until the last years of his life. Thus, his 1925 study of *shall* and *will* was based on drama, which he considered an approximation of the spoken language. His study of the growth of the structural use of word order excluded poetry in the belief that, of written language, poetry deviated most from the spoken language. His study comparing Vulgar English with educated English used letters written to the War Department during World War I because that was a convenient source of unedited data produced with limited objectives by educated and uneducated writers with known backgrounds. Only when we get to his study that was published as *the Structure of English* do we find him using recorded material. But even that was a compromise in that because of practical limitations he had to limit himself to recording conversations on a single phone line. The result was that a few individuals were over represented in the data. While I do not believe that fact affected his results, it does mean that these data are not a true random sample of Spoken American English, nor even of the English spoken in Ann Arbor at that time.

2.7 THE METHODOLOGY USED AS ONE ANALYZES A CORPUS SHOULD BE EXHAUSTIVE AND SYSTEMATIC.

Once one has gathered a corpus, one may take one of several approaches to analyzing it. These range from (a) an 'informal' use in which one searches the corpus for examples that suit ones purposes and then reports those examples; (b) a systematic, exhaustive analysis of all the relevant examples in the data; and (c) a systematic and exhaustive counting of conflicting contrastive features of some aspect of the language. Fries used all three of these approaches, though he clearly preferred the third. The three sections of Table 2 classify Fries's projects according to the way he used his corpora.

Fries took the first, informal and non-exhaustive approach to using corpora as part of his work on the *Early Modern English Dictionary*.[11] This dictionary was to be one of several period dictionaries which would supplement the *Oxford English Dictionary* (the *OED*). To begin with Fries, obtained the slips for Early Modern English from the Oxford University Press. In addition, he supplemented these slips with a reading program quite similar to the one used at the Oxford University Press for the *OED*. The reading program for the *OED* did not attempt to be exhaustive. That is, readers were not asked to find EVERY example of the words they were searching for. Rather they were asked to focus on uses of the words which were likely to be new uses, or late survivals, or of particular interest in some other way.

(both the practical demands of simply gathering the data, as well as the legal issues of being able to use the data once gathered). Thus the choice made by the editors of the Brown corpus to use only 2,000 word samples of each text included in that corpus increased the comparability of the text samples in word count, but at the same time decreased the comparability in that no controls were imposed on which portions of the text structures of each text were chosen. We know that the language of introductions differs from the language of conclusions. Given the choices made by the editors, the Brown corpus provides us with a poor tool to examine how the language of these two text portions differ.

[11] A more careful and extensive account of Fries's work on the EMED can be found in Bailey 1985a.

Fries took the second, more systematic, but non-quantitative approach to corpora in a second aspect of his reading program for the *Early Modern English Dictionary* and also in his work with the *Structure of English*. He realized that, given the type of reading program he inherited from the *OED*, he could not possibly tell the difference between what was a normal use of a word and what was unusual. Indeed the OED slips would have represented primarily unusual—remarkable—uses of the words. Further, the *OED* slips would have typically illustrated uses that were unambiguous. (I think of this as the 'good example' phenomenon. If you are looking for a word used to express a particular meaning, you look for a clear example that illustrates that usage with minimal ambiguity or room for dispute.) It was Fries's contention that much of the development of the vocabulary moved from one meaning to another through instances that were ambiguous. As a result of the *OED* reading program, these ambiguous usages would be systematically underrepresented. Therefore he instituted a program of what he called "intensive reading". He selected 69 texts of representative dates within the Early Modern English period, made multiple photocopies of the texts (to eliminate scribal error in copying them) and then made slips for essentially every instance of the major vocabulary items in those texts. These slips provide information on the distinctive environments that indicated that one or another of the meanings of the target word were being used in that instance. Fries called these distinctive environments the 'lexical sets' for the various meanings. With its focus on other words in the environment, the term *lexical set* clearly indicates that he was interested in and used collocational information about the individual words and their various meanings. However, I also believe (without much evidence so far) that he was also interested in the colligational information as well.

Fries took the third, quantitative approach to data in most of his larger projects. This third approach to using corpora involves counting systematically and exhaustively the conflicting contrastive features of some aspect of the language found in the corpus. In other words it emphasizes paradigmatic relations in the analysis of the data. Counting contrastive features allows one to identify patterns in the use or development of the language. He expressed his reasoning in a lengthy passage in *American English Grammar* (1940: 34) where he describes how he intends to anayze the letters that constitute his data.

> In the attempt to gather, analyze, and record the significant facts from any such mass of material as the specimines here examined, one cannot depend upon some general impressions and note only the special forms that attract attention. If he does, the unusual forms and constructions or those that differ from his own practice will inevitably impress him as bulking much larger in the total than they really are. Those forms that are in harmony with the great mass of English usage will escape his notice.

And after discussing distortions of the representation of Vulgar English by writers such as H. L. Menken and comic writers he says (1940: 35-36):

> In order to avoid errors of this kind we have in the study of this material tried first to record *all* the facts in each category examined. For example every preterit and past participle form was copied on a separate slip of paper in order that we

might determine not only the kind of variety that existed in actual usage but also something of the relative amounts of that variation. … We do not assume that the absolute frequency of occurrence of particular forms in the limited material here examined is in itself significant; we have simply tried to make sure of the *relative* frequency of the language usages appearing here in order to give proportion to our picture of actual practice and to prevent a false emphasis upon unusual or picturesquely interesting items.

A simple example of this approach comes from Fries's study of the intonation of yes-no questions. He was bothered by descriptions that said that yes-no questions normally used rising intonation, but that falling intonation was used in special circumstances. He knew that his data for the *Structure of English* contained many yes-no questions with falling intonation. He therefore decided to explore this issue by recording 39 programs of *What's My Line*, a TV program in which a panel of four judges used yes-no questions to determine the profession of a contestant. Recording this program provided him with a high concentration of yes-no questions which he could then analyze. Table 3 summarizes the most important of his results as described in Fries 1964: 248-249.

Table 3: Distribution of rising and falling intonation on yes-no questions in 39 programs of *What's my Line?*.

General results		
rising	falling	**total**
981 (38.3 %)	1580 (61.7 %)	**2561**

Range of usage of falling intonation			
Ranges by program		Ranges of individuals	
Minimum	Maximum	Minimum	Maximum
31.6 %	77.5%	57.2 %	72.7 %

Columns 1 and 2 in the top portion of table 3 show that over 60% of the 2,561 yes-no questions in his data have falling intonation. Counting only rising intonations or only falling intonations would not have produced information which could be used to interpret the results. Such partial information would only repeat what was already known—that yes-no questions sometimes were expressed with rising intonations and sometimes with falling intonations. Only by counting the incidence of both intonations on these questions could he establish the patterns of choices.

Similar instances of counting contrastive conflicting features come from his historical work (the first place he used this approach to the analysis of data.) Aside from his dissertation, i.e. his study of *shall* and *will*, and the work on the *Early Modern English Dictionary*, his historical work focussed on the changing means of signalling grammatical functions in English. In Old English, the major signal of grammatical functions was the

inflectional form of words, while in Modern English the major signal is the physical order of elements in the sentence. Fries was interested in how this change came about. Thus, in Fries 1940: 206 he presents the results of a study of the changing patterns of expression of a head-modifier relation in which the modifier is what he calls a "genitive". In order to explore this change he had to locate the various options which were used at one or another time to express this relationship. He found three: (a) the inflected genitive placed before the noun it modifies (the pre-positive genitive as in *the boy's hat*, *the table's leg*); (b) the inflected genitive placed after the noun it modifies (the post-positive genitive as in OE *xghwylc ymbsittendra*[12]), and (c) the periphrastic genitive (the 'of' construction as in *the mother of the children*, *the leg of the table*). Table 4 (from Fries 1940: 206) presents the shifting relative frequencies of these three constructions from 900 to 1300.

Table 4: Frequency of three placements of the English genitive construction through 400 years

	Post-positive genitive	**'Periphrastic' genitive**	**Pre-positive genitive**
c. 900	47.5%	0.5%	52.0%
c. 1000	30.5%	1.0 %	68.5%
c. 1100	22.2 %	1.2 %	76.6 %
c. 1200	11.8 %	6.3 %	81.9 %
c. 1250	0.6 %	31.4 %	68.9 %
c. 1300	0.0 %	84.5 %	15.6 %

Again, raw numbers are unimportant other than to ensure that sufficient numbers of instances exist in each cell to be reliable. What is relevant is the changing patterns of relative frequencies of the various alternatives of expression used. In the earliest data we see that the periphrastic genitive is almost non-existant and the pre-positive and post-positive genitives are used roughly equally. By 1300 we see that the post-positive genitive has disappeared from use and the most frequent means used to express the genitive relation is the periphrastic genitive.

3   EVALUATION AND IMPLICATIONS

Let me now turn to evaluate Fries's work and also point out some implications I see of his work for Systemic linguistics.

---

[12] *xghwylc* prn. indef. 'each', *ymbsittendra* gen. pl. 'neighbouring people'. The construction means 'each of the neighbours'. Of course the post-positive genitive construction cannot be illustrated using Modern English since it has gone out of use. Michael Cummings provided this example from OE.

3.1 CONTEXT OF USE AND REGISTER:

Fries had no formal notion of register. Of course he knew that we change our language as we engage in different activities. And indeed, when he compared Standard English with Vulgar English in his *American English Grammar* it was important to him, as he chose his data, that the writers of the letters be engaging in the same sorts of activities. Fries (1925: 987-988) states this idea explicitly in his study of the usage of *shall* and *will*.

> But one type of literature is here used to permit the maximum use of comparisons both of statistics and of instances. Because of the fact that the numerical distributions of the uses of the various grammatical persons differs in the several types of literature, statistics to be comparable must be from the same type.

In his description of the data used in his *American English Grammar*, Fries (1940: 28) assumes the restrictions inherent in letters written to the war department concerning social services, and focusses on the variety inherent in the letters.

> The correspondence must cover a wide range of topics. The material here used was largely made up of intimate descriptions of home conditions (family activities, family needs, domestic troubles, financial difficulties sickness, ambitions, accidents) all offered as reasons for appeals of one kind or another. This material was limited, however, by the fact that all the letters were very serious in tone. Nowhere was there anything of a light or humorous feeling.

Despite a strong practical sense of language variation associated with the purposes it is being used to achieve, Fries's intuitive knowledge of register variation never was expressed (so far as I have noticed) in his theoretical statements, and indeed one can criticise him for choosing data sources that don't represent the range of language available. For example, his data for the study of yes-no questions were taken from episodes of *What's my Line*, a TV show which was organized in such a way that panelists asked yes-no questions of a contestant. Each panelist was able to keep asking questions provided they received a 'yes' answer, but the minute a 'no' answer was received a different panelist would begin to ask questions. In retrospect, this fact could have skewed the data in that it is quite possible that questioners who had a theory of the contestant's profession and were following a line of logic would tend to use falling intonation, while if they were mystified and had no real idea as to the profession they might use rising intonation.

Related to the fact that Fries had no theory of register is his constant focus on obtaining samples of spoken language (or at least as close to spoken language as he could manage at the time). The quotations under assumption 4 in part II adequately document his attitude that the spoken language is primary and that written language is a reflection of the spoken language[13]. These days we have a more balanced approach

---

[13]   Of course in taking his position he was reacting to the discussions current during his life, in which the written language, particularly literary language, was considered the 'real' (or the best) language.

granting independent but related status to all registers of language regardless of whether they are written or spoken.

3.2 The importance of paradigmatic oppositions:

Fries's emphasis on investigating paradigmantic oppositions is in my view well taken. I take his point that when we count the instances of only a portion of the potential we reduce the value of our counts. I think, in particular, of the various counts used to investigate the notion of Theme and how many of them examine only a portion of the data. We have many instances of counts of the lexico-grammatical or semantic characteristics of the clause Themes of a body of texts, without comparable information for what is occurring in the Rhemes of the clauses in the same texts. As a result, we have no information as to whether the language of the Themes is or is not distinctive.

In a similar vein, I recently read a paper that explored the use of Theme predications to introduce new sections within a text. The paper made a number of interesting points, but I was left wondering if the results demonstrated special features of just those Theme predications which began new sections or was this typical of all Theme predications and the author merely examined those that began text segments. While we may be primarily interested in only one of the terms in a system, we need to examine all the terms within that system in order to gain a full interpretation of the results.

3.3 The need for a combination of quantitative analysis and qualitative analysis:

The places where Fries's counting techniques led him into trouble also provide lessons for us. I have mentioned at several points his study of the intonation of yes-no questions, most recently criticising it for perhaps not locating a representative sample. This defect might have been combatted had he been more persistent in his counting techniques. I am suggesting, that is, that in this case he did not follow his principles far enough. Specifically he could have examined more carefully the contexts in which falling and rising intonations were used. Now, he did explore these contexts to some extent when he explored alternations of intonation pattern used on repetitions of the same question (1964: 249). The relevant pasaage is provided below.

> There were some occasions when a question was repeated because it was not heard clearly or understood. On these occasions it was *the same question asked twice by the same panelist,* and *directed to the same person,* but almost immediately after the first had been put. The point of special interest here is the fact that in many of these instances *the intonation pattern of the question as uttered the second time was the opposite of that used the first time.* If this change had all been in one direction — if, for example, a question with rising intonation had always received a falling intonation in repetition, one would suspect that the repetition itself constituted an instance of the 'special circumstances' that are said to attach to the falling intonation used with yes-no questions. But this was not the case. Questions with falling intonations were repeated with rising intonations, questions with rising

intonations were repeated with falling intonations, and some questions were repeated with the same intonations.

However, a more detailed analysis of the context of use (even a simple noting of which intonation type was more frequent on the first question asked of a contestant), and a full count of all the instances of repeated questions, and their contexts might have helped decipher the difference in interpretation of rising and falling intonations on yes-no questions.

3.4   THE NOTION OF SYSTEM:

While Fries emphasized the paradigmatic aspects of language, he had no notion of system as used within systemic linguistics. Rather his notion of 'contrasting conflicting signals' focussed on examining the various structural resources used to express similar grammatical relations. However, having said that, it is useful to note that because he emphasized grammatical resources that expressed similar grammatical relations, much of his work is reinterpretable within a systemic framework. In many cases, the meanings that are expressed, such as the expression of the future with *shall* and *will* are interpretable within systemic terms as deriving from a single system within the lexico-grammar. In other cases, the relation is not quite so easy. For example, the various realizations of the genitive relation studied by Fries do not derive from a single point in the lexico-grammar. On the one hand the relation between the pre-positive and post-positive genitive is one of position and thus is likely to be best addressed in our terms as an aspect of the textual metafunction. On the other hand, the use of an 'of' construction instead of an inflected genitive, while also involving a textual component, is likely to involve more than that.

But notice that this sort of visualizing features from disparate portions of the lexico-grammar within a single larger system is necessary for other features of language which we are already addressing. I am thinking here of the work of Peter White on engagement where he elegantly demonstrates that modal verbs and adverbs, quotative forms, expressions of point of view and other radically different looking constructions can be seen to express different options within a single semantic system. We systemicists have theoretical resources such as our stratal approach to language which we can and do use to account for issues such as this.

3.5   FRIES'S HISTORICAL RESULTS AND PROBABILITIES WITHIN SYSTEMS:

Given Fries's results for the history of the language, in which he noted gradual progressions from expressing a given structural relation by using one form to using a different form (for example, from the use of inflections to signal elements of structure in the clause to the use of word order to signal the same elements of structure) we should expect a range of probabilities with these systems. In other words, we should expect some departures from Halliday's (2005: 80 and 96) general hypothesis that systems tend to have either equiprobable options or options that are related as 9:1. Specifically, systems which are undergoing periods of major change should be expected to have intermediate frequencies.

It is interesting to examine Table 4 in light of Halliday's hypothesis. The initial and final states in Table 4 are not too far from the predictions made by Halliday. In 900 we have essentially a two term system with the two options being roughly equiprobable. In 1300 we again have essentially a two term system, but now the two options approach Halliday's prediction for a skew system. In between those two dates, however, we still have essentially two term systems, but the frequencies range from 30%-68%, 22%-76%, 12% - 82%, and 31%-69%.[14] One suspects that if we had figures separated by 25 years instead of 50 or 100 years we might find frequencies intermediate between those given in the table. For those intermediate time periods Halliday's predictions do not work very well.

3.6   CORPUS AS INFORMATION ON HOW LANGUAGE IS UNDERSTOOD:

Finally, let me mention one of Fries's attitudes toward corpora that I believe is quite important to us now. Specifically, he used his corpus of conversational data not merely as evidence for the language that was produced. He also used it as evidence as to how what was said was understood.

Corpora, particularly spoken corpora, contain examples where the speakers make mistakes (e.g. spoonerisms), correct themselves, or pause and change in midstream the structure being produced, etc. The linguist has no need to include these special cases on an equal footing with other data that are more representative of the intents of the speakers. Examining the contexts in which the language is produced and particularly the behavior of the speakers and the listeners (e.g. noting which utterances are associated with various types of repair behavior) will provide significant aid in judging which portions of the corpus are more important to address first.

Further, noting responses such as providing information, complying with requests, responding to information given, etc. gives linguists information about the ways the language produced was understood in context.

In other words, addressing how the language is understood is fully as important to corpus linguistics as looking at what is produced. Admittedly, in the case of monologic texts it is difficult to obtain participant reactions. However, even in the case of written data, at least the linguist himself/herself is available to provide interpretations. These interpretations are critical to the use of the corpus.

Let me end by saying that, although Fries worked with corpora by hand, beginning about 90 years ago and ending about 40 years ago, much of what he did is still relevant to the present. We use many of his techniques of analysis. We encounter many of the same problems that he did. I hope you believe with me that it is useful to examine what he did, and the ways in which he was or was not successful.

---

[14]   Notice that my discussion here largely ignores the difference in form involved in the evolving system. The issue is that there is, in some stratum, a system which has two terms. The realizations of the terms may differ as time passes but the system is relatively steady.

REFERENCES

Bailey, Richard W. 1985a. Charles C. Fries and the *Early Modern English Dictionary*. In Peter H. Fries and Nancy M. Fries (eds.), 1985: 171-204.

Bailey, Richard W. 1985b. Charles C. Fries: The life of a linguist. In In Peter H. Fries and Nancy M. Fries (eds.), 1985: 1 - 15.

Benson, James D., and William S. Greaves. *The language people reallyuse*. Agincourt Canada: The Book Society of Canada, Ltd.

Fries, Charles C. 1925. The periphrastic future with *shall* and *will* in Modern English. PMLA 40 (4): 963 - 1024.

Fries, Charles C., James Holly Hanford, and Harrison Ross Steeves. 1926. *The teaching of Literature*. New York: Silver Burdett and Co.

Fries, Charles C. 1928. *The teaching of the English language*. New York: Thomas Nelson & Sons.

Fries, Charles C. 1940a. *American English grammar*. New York: Appleton Century Crofts.

Fries, Charles C. 1940b. *English word lists: A study of their adaptability for instruction*. Washington, DC: American Council on Education.

Fries, Charles C. 1940c. On the development of the structural use of word-order in Modern English. *Language* 16 (3): 199 - 208.

Fries, Charles C. 1944. Letter to Albert Marckwardt. April 26, 1944.

Fries, Charles C. 1945. *Teaching and learning English as a foreign language*. Ann Arbor, Michigan: University of Michigan Press.

Fries, Charles C. 1952. *The structure of English*. New York: Harcourt Brace.

Fries, Charles C. 1954. Meaning and linguistic analysis. *Language* 30: 57-68.

Fries, Charles C. 1959. Letter to Peter H. Fries, November 15, 1959.

Fries, Charles C., and Agnes C. Fries. 1961. *Foundations for English Teaching*. Tokyo, Japan: Kenkyusha Ltd.

Fries, Charles C. 1963. *Linguistics and Reading*. New York: Holt, Rinehart and Winston.

Fries, Charles C. 1964. On the intonation of yes-no questions in English. In David Abercrombie, D. B. Fry, P. A. D. McCarthy, N. C. Scott and J. L. M. Trim (eds.), *In honour of Daniel Jones: Papers contributed on the occasion of his eightieth birthday*, 12 September 1961, 242-254. London: Longmans.

Fries, Charles C. 1967. "Structural Linguistics." *Encyclopaedia Britannica*. Chicago: Encyclopaedia Britannica, Inc.

Fries, Charles C. 1970. The time "depth" of coexisting conflicting grammatical signals in English. In A. Graur (ed.), Actes du Xe Congrès International des Linguistes. Bucarest, 28 August - 2 September, 1967. Bucarest: Éditions de l'Academie de la République Socialiste de Roumanie.

Fries, Peter H. 1983. C. C. Fries, signals grammar and the goals of linguistics. In John Morreall, (Ed.), *The Ninth LACUS Forum 1982*, 146 - 158. Columbia, SC: Hornbeam.

Fries, Peter H. and Nancy M. Fries (eds.). 1985. *Toward an understanding of language: Charles Carpenter Fries in perspective*. Amsterdam: John Benjamins.

Halliday, M. A. K. 2005. *Collected works of M. A. K. Halliday, Volume 6: Computational and Quantitative Studies*, London: Continuum.

Hymes, Dell, and John Fought. 1981. *American structuralism*. The Hague: Mouton.

Matthiessen, Christian M. I. M. 2006. Frequency profiles of some bassic grammatical systems: an interim report. In Geoff Thompson and Susan Hunston (eds.), *System and Corpus: Exploring Connections*, 103 - 142. London: Equinox.

Nesbitt Christopher and Gunther Plum. 1988. Probabilities in systemic grammar: The clause complex in English. In Robin Fawcett and David Young (eds.), *New developmentws in systeemic linguistics*, *2: Theory and application,* 6-38. London: Pinter.

Peter H. Fries
Professor Emeritus,
Department of English Language and Literature
Central Michigan University
Mount Pleasant MI 48859
USA

peterhfries@cs.com