Charles C. Fries, linguistics and corpus linguistics*

Peter H. Fries, Central Michigan University

Abstract

The field of corpus linguistics is commonly regarded as a 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.

Fries's approach to linguistics and to the use of corpora in linguistics grew out of his background, as well as the goals and functions of theory that he adopted for his linguistics. This background and these goals led him to adopt a number of themes that permeated his work throughout his career:

1. Because his interests in language and linguistics grew out of his early work teaching classical Greek and then English composition, his theory had a very practical basis in that it grew out of the problems encountered in pursuing these (and similar) practical tasks.
2. He valued the scientific goals of making predictions of disparate phenomena. In linguistics, he took as his goal the description of language in such a way that he could identify the signals in the language that would lead listeners to interpret the language the way that they do. That is, he tried identify the signals that would allow him to predict listener responses.
3. He emphasized items in contrast – the paradigmatic relations among linguistic elements.
4. He considered the spoken language to be primary.
5. In science, all analyses should be reliable and replicable. If linguistics was to be scientific, this required that all descriptions be based on some explicit body of evidence – a corpus, or body of language being used by people for real purposes.
6. The data gathered and analyzed should represent the language of some community.

7. The analysis must be exhaustive and systematic. Preferably these analyses should include relative frequency counts of patterns in contrast, where the comparison of the patterns is based on similarities of meaning.

   The paper ends with an evaluation of several aspects of Fries's work.

1 Introduction

The field of corpus linguistics is now generally considered to be a subfield of linguistics that has developed largely since the early 1960's when Henry Kučera and W. Nelson Francis created the now famous Brown corpus of one million words. Further, it is now regularly presented as developing in opposition to the then current dominant formalist version of linguistics, which distinguished an idealized 'competence' from actual 'performance', and which consistently emphasized the defective nature of performance while taking the study of competence as its goal. One consequence of the formalists' focus on competence has been the regrettable (in the view of corpus linguists) neglect and devaluation of actual language use (and samples of actual language use) as they created their descriptions and theoretical positions.

In what follows, I do not want to contradict the general features of the above sketch of the genesis of corpus linguistics as a discipline, for I recall a number of conversations with Kučera and Francis in which they criticized the absence of 'real language' from formalist accounts, and also recounted the overt opposition of the then-dominant advocates of transformational grammar to their project of gathering the Brown corpus. However, I would like to suggest an additional interpretation of the rise of corpus linguistics. In this interpretation, corpus linguistics is a reassertion of older traditions in the study of language that were current before the rise of formalist approaches. A number of these older traditions required the gathering and study of some coherent body of data – a corpus. They included at least philology, dialect geography, and anthropological linguistics. (Another discipline that required the use of carefully gathered representative samples of language, socio-linguistics as conceived by William Labov and his colleagues, developed about the same time as the initial stages of corpus linguistics.) Indeed, a significant part of my training at the University of Michigan and with Kenneth Pike during the late 1950's consisted of techniques to gather and record information in such a way that I would be able to retrieve relevant portions of the information with relatively little effort. In modern parlance we were being taught a kind of data gathering and data retrieval as part of our training as ordinary linguists. A further bit of evidence that corpus linguistics constitutes a continuation of a tradition is the age of the initial proponents. Kučera and Fran-

cis were well established in their careers when they began work on the Brown corpus. Randolph Quirk began his work on the survey of English usage (a corpus project, though not one that used computers at the beginning) well on into his career. In other words the early corpus linguists were linguists who had been trained in the older approaches to linguistics and then had the professional status and the independence and vision to maintain their beliefs and fundamental ways of working in the face of considerable opposition.

These days, with the advent of larger computers and their accessibility, what in those old days was often merely a casual gathering of examples from various convenient sources has transformed itself to a much more careful approach to large corpora of millions of words, and the discipline which has developed as we process these extremely large corpora has been termed 'corpus linguistics'. However, 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.

Given the current importance of corpus linguistics within linguistics generally, 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 examine the work of one figure from the first half of the twentieth 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 methodologies he used to achieve these goals. Because of this, 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 using corpora in linguistics.

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 C. Fries and selected other linguists of his generation:

Charles C. Fries	**1887–1967**
Leonard Bloomfield	1887–1949
J. R. Firth	1890–1960
Louis Hjelmslev	1899–1965
Roman Jakobson	1896–1980
Daniel Jones	1881–1967
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 the history of English with Professor Samuel Moore.

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.

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 Fries's 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.[4] When Greek was no longer to be a required course in high school, 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.[5] 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 in which he discussed the relation of theoretical linguistics and teaching English as a foreign language (Fries 1944a: 1–2) he said "… linguistic theory must be tested by practical applications and practical teaching will help to develop that theory, …"

We can also see the type of intimate connection that he made between theory and practice in the following excerpt from a letter he wrote to his dean (Fries 1944b) reporting on the activities of the English Language Institute (the ELI) over the previous few years. In it he described the functions of the ELI in the following terms:

(2) As I see it, the present function of the English Language Institute is primarily investigation and research leading to the development of materials and techniques for the teaching of the English language.

(A) The research and investigation include:

(a) Original research in the English language. Samples of such research, which we have already accomplished, are the work done on the intonation of American English, the work on English sentence patterns, and the work that I am doing on formal clues for syntactic relations in English.

(b) Original research in those aspects of other languages for whose speakers materials in English are being prepared – aspects which are necessary to sound work and for which the necessary research has not yet been done. An example of such research is the work done on the segmental phonemes of Brazilian Portuguese.

(c) The compilation and analysis of pertinent research in English and in the languages of those for which English materials are being prepared.

(d) Compilation and analysis of research in matters of general linguistic import that bear upon the practical problems of teaching English.

> (B) The development of materials and techniques, including those to be used in the teaching of the English language at various levels and for students of various linguistic backgrounds.
>
> [[This section of the letter continues with a description of the materials already produced and in progress at the ELI.]]

This quotation is indicative in a number of ways. First it is significant of his attitude that the research functions of the ELI are described first. However, a close examination of the exact research projects described shows that they can easily be seen to be grounded in and derived from the practical problems at hand of determining information that will be used in developing teaching materials, and techniques of teaching as well as, ultimately, in evaluating student performance.

It seems to me that this passage encapsulates his way of working. He first encountered some practical problem that he wanted to address (e.g. language teaching either to native English speakers or to speakers of other languages, or teaching reading to children, etc.). The nature of the problem framed the particular questions that had to be addressed whether through descriptions or through development of the linguistic theory. He then searched the literature for anything relevant to his problem as he conceived it, and if other work was non-existent or inadequate, he developed a description that satisfied him. In other words, most of Fries's theoretical and descriptive projects arose out of issues that he encountered first in some aspect of his teaching.

2.2 Language must be approached scientifically: The essence of science is prediction of disparate phenomena

Fries 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 his model, the term *recognition response* is closely associated with meaning. For example, in *Linguistics and reading* (Fries 1963: 73) he refers to "the basic contrastive arrangements of the patterns of grammatical structure that regularly elicit **recognition responses of grammatical or structural meanings.**" In other words, Fries focused 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.

Because of the importance of listener response to his whole enterprise, he regarded a corpus not merely as a set of forms which had been uttered, but also as a source of information concerning the participant 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)[6].

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* 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 trying 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. Certainly the goal of classifying sequences into 'sentence' vs. 'not sentence' is at heart taxonomic. But further, simply saying that a certain surface structure (and wording) was related to a particular deep structure (and interpretation) did not address, in Fries's eyes, the core issue of describing the aspects of the wording that cued listeners, in the stream of speech, that that particular wording was to be interpreted in a particular way.

2.3 The paradigmatic aspect of language is essential

Fries 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 P. Fries 1983 for a more careful discussion of this point.) In his book *Linguistics and reading* (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 *The structure of English* (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 the morphemes have been identified" (1952: x) and he provides lists of word classes that contrast in their morphemic make up. These lists demonstrate differences in the internal structures of the major word classes that lead to the recognition that one word belongs to one major class while a second word with a contrasting structure belongs to different 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).

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 under point 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] In part, his interest in the spoken language arose in reaction to teachers and grammarians of the previous generation who generally treated written English, and most particularly literary language, as the 'real' language. For example, the great grammarians of the early twentieth century such as Poutsma and Jespersen devoted the bulk of their efforts to describing literary language, ignoring for the most part not only the spoken language, but also non-literary genres such as scientific or technological language. Even F. N. Scott (1926: 25), the teacher who attracted Fries to Michigan, compared the language that children brought to school to "the language of the animals from which they are descended" and described it as consisting in large part of "modulations of simple primitive sounds that probably go back to the infancy of the race."

By contrast, Fries, in all of his work, made every attempt to discover the characteristics of the language spoken in the communities while the speakers were focused on what they were communicating not how they were communicating it. In other words, he was looking for the vocabulary choices and grammatical patterns used in unedited spoken language. His discussion (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…

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

> 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 continues (1940a: 27):

> 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 Ph. D. study of the development of *shall* and *will* used dramas as a source of data 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. (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, Fries 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 (see Section 3.7 below).

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 record systematically some definite body of evidence and list and count 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.

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. I think, for example of Otto Jespersen, who gathered thousands of examples from an unspecified body of English literature.[10] We cannot replicate or test his work, because we do not know the exact sources he used, and he certainly did not sample these sources exhaustively. However, the examples he collected are all valid examples of language in use and he was able to use them effectively as a basis for his landmark descriptions of English grammar. However examples gathered informally, as he did, may not represent the full range of phenomena available, 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 stratified 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 happen 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.[11] I have already mentioned that he wanted to examine spoken language, and yet he had no means to do so regularly until late in 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 minimally edited 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 the data represented the language practices of only a few people. While I do not believe that fact seriously compromised 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 one's 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 (see Section 3.7 below).

Fries took the first, informal and non-exhaustive approach to using corpora as part of his work on the *Early Modern English dictionary.*[12] 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.

Fries took the second, more systematic, 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. Although the term *lexical* in *lexical set* seems to imply that he was particularly interested in the other words that appeared in the context (i.e. what we would now call 'collocations'), in fact the discussions in Fries (1963: 104–105)[13] demonstrate that he also included phenomena which we would describe as colligations. In other words he did not distinguish carefully between our two concepts.

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* (Fries 1940a: 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 specimens 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 (1940a: 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* s*tructure 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		Range of usage of falling intonation			
			Ranges by program		Ranges of individuals	
Total	Rising	Falling	Minimum	Maximum	Minimum	Maximum
2561	981 (38.3 %)	1580 (61.7 %)	31.6 %	77.5%	57.2 %	72.7 %

He found 2,561 examples of yes-no questions. Columns 2 and 3 show that over 60 percent 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 (1940c: 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*[14]); and (c) the periphrastic genitive (the 'of' construction as in *the mother of the children*, *the leg of the table*). Table 4 (from Fries 1940c: 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 nonexistant 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 comment on a few aspects of Fries's work.

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* (1940: 28) Fries 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 do not 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 profes-

sion 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 in Section 2.4 adequately document his attitude that the spoken language is primary and that written language is a reflection of the spoken language. These days we have a more complicated approach to language variation. With the development of the theoretical concepts of register (e.g. Gregory 1967; Halliday 1985, 2007) and the quantitative analyses of people such as Douglas Biber (1988) we have statistical evidence that demonstrates that a variety of factors affect the nature of the language used in a particular social interaction. The difference between spoken and written language is but one of these factors.

3.2 The importance of paradigmatic oppositions

Fries's emphasis on investigating paradigmatic oppositions is in my view well taken. When we count the instances of only a portion of the potential we reduce the value of our counts. If, for example, a theory divides clauses into two grammatical functions: Theme and Rheme, it is not enough merely to study the lexis, or structures or meanings that are found within one of those functions but not the other. If we discover that certain features, say, characterize the Themes, we still cannot say if those features are distinctive within the texts considered. That is, we do not know if we have found those characteristics in the Themes because the **entire text** contains those characteristics in great abundance, or whether the Themes contain a greater than normal concentration of those characteristics. In Fries's view, even though 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 ground a full interpretation of the results.

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

Much of Fries's work was done at a time when most linguists – particularly in the U.S. – saw no relevance of quantitative analyses for linguistic description. Chomsky (1962: 128, n9) continued this earlier tradition in part when in the process of discussing the concept of grammaticality he said:

> It seems clear that probabilistic considerations have nothing to do with grammar, e.g., surely it is not a matter of concern for the grammar of

> English that "New York" is more probable than "Nevada" in the context "I come from –."

Certainly few synchronic descriptions of languages published in the journal *Language* during those years used quantitative analyses in serious ways.

By contrast, most corpus linguists now accept the need for quantitative and qualitative analyses. Indeed, most current introductory textbooks on corpus linguistics (e.g. Biber, Conrad and Reppen, 1998; Gries, 2009; and Hunston, 2002) contain prominent discussions of quantitative methods and their appropriate uses for analyzing corpora. Further, computer concordancing programs often come integrated with sophisticated tools for statistical analysis to apply to the results of the various searches performed.

Of course merely counting phenomena does not automatically make a study good, and it may be instructive to examine several instances where Fries's analyses could have been improved. 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. Actually, 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 passage 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 formalized notion of system (such as is used within Systemic Functional linguistics and Stratificational grammar). Rather his notion of 'contrasting conflicting signals' focused 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 meanings, much of his work is reinterpretable within linguistic theories that do formalize descriptions of the choices available to speakers.

3.5 Fries's historical results and the notion of a single unified system for the language

In the 1940's and 50's, when Fries was doing some of his most important work, the dominant view within structuralism was to assume the existence of a single coherent system for each language. Admittedly, the linguists of the time were very much aware that dialects existed, and, further, that no two individuals used exactly the same language patterns. For the most part these facts were ignored in theoretical discussions; however in a few cases some linguists attempted to deal with them by developing the notion of idiolect (the dialect used by an individual).

Because of his historical corpus work, and particularly because he counted the relative frequencies of conflicting contrasting items, Fries took a very different view. One can see why from the pattern presented in Table 4. Every time Fries looked at the texts in his historical data he discovered instances of old systems that were being displaced, as well as instances of new systems that were in development and about to take over. As he said (1970:924):

> …there is no time at which one can make a cross section of the grammatical forms actually in use, in an adequate sample, without finding the evidence for an exceedingly complicated 'time depth' of approximately 100 years.[15]

The conclusion that he drew from these facts was that there was no time in the history of written English when there was a single coherent grammatical system.[16] Later Fries and Pike (1949:29) expressed a similar view concerning the phonology of a language in the following terms:

> The speech of monolingual native speakers of some languages is comprised of more than one phonemic system; the simultaneously existing systems operate partly in harmony and partly in conflict. No rigidly descriptive statement of the facts about such a language accounts for all the pertinent data without leading to apparent contradictions.

3.6 Fries's historical results and probabilities within systems

As described above, Fries was regularly interested in counting the relative frequencies with which the 'conflicting contrastive items' occurred. These counts gave him a feeling as to which of the contrasting elements were, so to speak, normal, and which were unusual. It is interesting that, so far as I can tell, he never took the next step to say that when he found a great difference in the relative frequencies of two options, that very difference implied that the two options expressed different meanings. That is, he never claimed that the relative frequency with which an option was used was directly part of its meaning. Clearly more modern approaches (for example Systemic Functional grammar) take that step, and further, they indeed attempt to integrate frequencies into their descriptions of grammar. (See the comments on frequency in Halliday 2005.)

In addition to Halliday's belief that relative frequencies constitute an important part of the description of any system, he has also suggested (2005: 80 and 96) that the relative probabilities of the options of any two-term system within language are likely to center either on options that are roughly equi-probable, or options that occur in roughly a 9:1 ratio. Fries's results for the history of English provide both some support as well as some challenges for this suggestion. As illustrated in Table 4, Fries found 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). The support for Halliday's suggestion comes from the fact that the initial stage in his data indicates a two-term system in which the options are roughly equi-probable. Similarly the final stage in his data illustrates a system in which one option occurs roughly nine times more frequently than the other term. Both of these are probabilities that Halliday predicts. However, the stages that intervene between the initial and final stages illustrate intermediate probabilities that fall outside of the ranges

predicted by Halliday.[17] In most cases we still have essentially two term systems, but the frequencies range from 30–68 percent, 22–76 percent, 12–82 percent, and 31–69 percent.[18] 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. One is tempted to posit two sorts of systems: those that are relatively stable (such as the initial and final stages of Table 4), as opposed to systems which are undergoing relatively rapid change. Halliday's suggestion may work for the relatively stable systems but not for the ones undergoing rapid change. In any case, Fries's results make clear that Halliday's suggestion concerning frequencies needs to be examined carefully in light of historical change.

3.7 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.[19] 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 listeners (or readers) react to and understand the language 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, (and it is always possible to gather groups of readers and ask questions concerning their interpretations of the text.) 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 and many of his assumptions are still being used and are relevant to the present. We encounter many of the same problems that he did. We use many of his techniques of analysis. Certainly modern corpus linguistics has developed significantly since his time – not merely in the technological tools that have been developed, but also in reconceptualizing basic issues such as exactly what constitutes a corpus and how corpora can be used with insight.

It seems to me that one way that we have of discovering where we are as a discipline is to look back at our early stages to see what it was like at that time. In this light, I hope you believe with me that it is useful to examine what C. C. Fries attempted, and the ways in which he was or was not successful.

Table 2: Fries's projects which directly involved the gathering and use of a specific corpus of data.[20]

	Project	**Data source/size**	**Purpose**
A.	**Corpus as 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. Also, the shifting frequencies of word uses might provide some evidence of change in progress.
b.	The 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.
3.	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)
4.	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

Notes

* An earlier version of this paper was presented at the Nineteenth Euro-International Systemic Functional Linguistics Conference and Workshop, July 23–25, 2007. Saarbrücken, Germany. I would like in addition to acknowledge the comments of Elizabeth Berriman and Richard Forest who commented on an earlier version of the paper.

1. This view contrasts sharply with Simpson and Swales's (2001: 1) statement: "Corpus linguistics is essentially a technology…."
2. I do **not** want to claim that Fries's work was typical of the time. Indeed, I would claim that 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).

3. See R. W. Bailey (1985b) for a more extensive discussion of Fries's life.
4. In conversations Fries often referred to being converted to linguistics. This is the conversion he meant.
5. At some point Fries transferred to the English Department and did his Ph.D. under Prof. Samuel Moore, who was interested in the history of English. Scott remained an influential figure in Fries's thinking, however.
6. 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.
7. Taking the spoken language as primary should not be equated with regarding the **sounds** of speech as primary. Rather Fries's interest consistently lay in the grammatical patterns used. He was very much aware that the words and grammar of spoken language differed greatly from the language that was written.
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.
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. While Jespersen certainly had a carefully planned program for reading his corpus of literary works, his definition of that corpus was, in my view, casual. He did not describe the principles for choosing those works, nor so far as I know did he list the total corpus from which he extracted his examples. (Of course he did provide references for the provenience of each of his examples.)
11. Perhaps all corpora are compromises between the demands of representativeness, the demands of the sampling procedures, and the practical demands of obtaining usable data (both the practical demands of simply gathering the data, as well as the legal issues of being able to use the data once gathered). For example, 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.

12. A more careful and extensive account of Fries's work on the EMED can be found in Bailey (1985a).
13. Fries (1945: 55–56) also addresses the issue of lexical sets although without using that term.
14. *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.
15. The particular data summarized in Table 4 do not support the 100 year timeline mentioned in this quotation; however most of the other changes that he studied were accomplished in the roughly 100 year period mentioned here. In addition to the 100 year timeline for individual changes, he also noted that many of the individual changes seemed to achieve the same sort of end (e.g. the reduction of the inflectional endings) in different parts of the language. He used the term 'slope' to describe these sets of related changes.
16. It is useful to notice that he was not simply addressing dialect differences here, nor 'dialect mixture'.
17. An interesting feature of this example is that, technically speaking, only the last data sample, from 1300, actually illustrates a true two-term system. All the others contain evidence of a third option. But in four of the five sample periods the third option (whichever it is) accounts at most for 1.2 percent of the data, and so it is reasonable to treat these systems as two term systems. The only exception is the data set from 1200, where the two options used least frequently are 11.8 percent and 6.3 percent compared to the dominant option that is used 81.9 percent.
18. 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 itself is relatively steady.
19. See Sinclair and Mauranen (2006) for a recent account of a principled way to address some of the dysfluencies mentioned above.
20. One corpus gathered by Fries has been omitted from this table. At about the time when he was completing his *American English grammar*, Fries became interested in exploring variation in grammar that correlated with geographical dialects. He purchased a number of local newspapers from around the United States published in a particular week. Unfortunately he found that the stories printed in those papers contained few regionalisms. As a result he abandoned that project.

References

Bailey, Richard W. 1985a. Charles C. Fries and the *Early Modern English dictionary.* In P. H. Fries and N. M. Fries (eds.), 171–204.

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

Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, Douglas, Susan Conrad and Randi Reppen. 1998. *Corpus linguistics: Investigating language structure and use.* Cambridge: Cambridge University Press.

Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton.

Chomsky, Noam. 1962. A transformational approach to syntax. In A. A. Hill (ed.). *Third Texas conference on problems of linguistic analysis in English*, 124–169. Austin, TX: The University of Texas.

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. 1927. *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. 1944a. Letter to Albert Marckwardt. April 26, 1944.

Fries, Charles C. 1944b. Letter to Dean Clarence S. Yoakum. April 20, 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 D. 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*, 923–926. Bucarest: Éditions de l'Academie de la République Socialiste de Roumanie.

Fries, Charles C. and Kenneth L. Pike. 1949. Coexistent phonemic systems. *Language* 25: 29–50.

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.

Gregory, Michael J. 1967. Aspects of varieties differentiation. *Journal of Linguistics* 3 (2): 177–198.

Gries, Stefan Thomas. 2009. *Quantitative corpus linguistics with R: A practical introduction.* Abington: Routledge.

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

Halliday, M. A. K. 2007. *Collected works of M. A. K. Halliday. Volume 10: Language and society.* London: Continuum.

Halliday, M. A. K. and Ruqaiya Hasan. 1985. *Language, context and text: Aspects of language in a social-semiotic perspective*. Victoria, Australia: Deakin University Press.

Hunston, Susan. 2002. *Corpora in applied linguistics.* Cambridge: Cambridge University Press.

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

Matthiessen, Christian M. I. M. 2006. Frequency profiles of some basic grammatical systems: An interim report. In G. Thompson and S. 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 R. Fawcett and D. Young (eds.). *New developments in systemic linguistics*, *2: Theory and application,* 6–38. London: Pinter.

Sinclair, John McH. and Anna Mauranen. 2006. *Linear unit grammar: Integrating speech and writing*. Amsterdam: John Benjamins.

White, Peter. 2003. Beyond modality and hedging: A dialogic view of the language of intersubjectivity. *Text* 23: 259–284.