ANOTHER LOOK AT GENRE:
CORPUS LINGUISTICS VS. GENRE ANALYSIS

ANNA MAURANEN

University of Joensuu

In current talk of text typology, the term "genre "has superseded most other formerly popular terms, such as "text type" and "register". We speak of genres not only in genre theory, but also in the rapidly expanding domain of corpus studies. However, as often is the case, terms which are widely used become polysemous and vague in meaning. Using the same term helps conceal the fact that the sociocognitive approach which characterises much current genre theory (e.g., Swales 1990; Bhatia 1993; Kamberelis 1995; Berkenkotter – Huckin 1995) is widely different if not downright incompatible with the notions of genre entertained among those linguists who amass large corpora for empirical research. The former approach sees genres as social, dynamic, interactive processes, which get realised in verbal interaction, while the latter treats genre as a label for many, often vaguely defined, kinds of variation in discourse types, the main use of which seems to be to ensure wide coverage for corpora.

To illustrate the difference in the approaches, I shall show an influential formulation from both sides; one by Berkenkotter and Huckin (1995), and one by Biber (1988). After that, I shall take up some issues which I find problematic in one or both of these approaches. Most of them centre around the issue of external vs. internal criteria in defining genres. After discussing the issues briefly, I shall then consider the possibilities of a convergent future.

1. Two conceptions of genre

Recent genre theory generally favours conceptions of language which emphasise its nature as social action. The following characterisation of genre is presented by Berkenkotter and Huckin (1995), which sums up the sociocognitive view

in its current understanding, and has been widely adopted by many scholars working in the framework of genre analysis:

1. *Dynamism*. Genres are dynamic rhetorical forms that are developed from actors' responses to recurrent situations and that serve to stabilize experience and give it coherence and meaning. Genres change over time in response to their users' sociocognitive needs.

2. *Situatedness*. Our knowledge of genres is derived from and embedded in our participation in the communicative activities of daily and professional life. As such, genre knowledge is a form of "situated cognition" that continues to develop as we participate in the activities of the ambient culture.

3. *Duality of Structure*. As we draw on genre rules to engage in professional activities, we constitute social structures (in professional, institutional, and organizational contexts) and simultaneously reproduce these structures.

4. *Community Ownership*. Genre conventions signal a discourse community's norms, epistemology, ideology, and social ontology.

5. *Form and Content*. Genre knowledge embraces both form and content, including a sense of what content is appropriate to a particular purpose in a particular situation at a particular point in time.

By looking at this list, could we identify a genre? If we already have a particular genre in mind, like a scientific article or a dinner conversation, the list might assist us in seeing it as genre, and assessing its "genericity". But without such a starting-point, there is little to guide us in construing an entity that in real life might pass as a genre, or orienting us to the size or level of unit to look for. For instance would research reports and review articles be variants of the same genre, or two genres, and are scientific journals one genre and popular science journals another, and how do popular science journals relate generically to popular science books? Or how stable must a conversation type be to merit the status of genre? And is all language use covered by genres?

Secondly, as an example from the corpus camp, let us look at what Douglas Biber (1988: 67) took as his point of departure in his famous study of variation in speech and writing. He based his work on the standard corpora available in the English language at the time. New and much larger corpora have since been compiled, but not on essentially different principles.

GENRE	NUMBER OF TEXTS
Written – genres 1-15 from the LOB corpus	
Press reportage	44
Editorials	27
Press reviews	17
Religion	17
Skills and hobbies	14
Popular lore	14
Biographies	14
Official documents	14
Academic prose	80
General fiction	29
Mystery fiction	13
Science fiction	6
Adventure fiction	13
Romantic fiction	13
Humor	9
Personal letters	6
Professional letters	10
Spoken – from the London-Lund Corpus	
Face-to-face conversation	44
Telephone conversation	27
Public conversations, debates, and interviews	22
Broadcast	18
Spontaneous speeches	16
Planned speeches	14
Total	481
Approximate number of words	960,000

This is a motley list of kinds of texts. Some of the "genres" are quite incommensurate: "religion" (4) is one genre, as is "academic prose" (9), while there are five different genres of fiction, none of them poetry or drama. And as a genre of its own, we again have "humour" (15). Personal and professional letters are wisely distinguished on the written side (this was Biber's addition to the LOB list), but on the spoken side we have all "telephone conversations" lumped into one. It is obvious that this list can hardly be based on a good

theory of genre, if on a theory at all. Other corpus linguists treat genre in similar but more offhand ways, as for example McEnery and Wilson, who talk about "... genres such as newspaper reporting, romantic fiction, legal statutes, scientific writing, and so on." (1996: 65).

Matti Rissanen (1992: 193) expressed his concern about the problem, pointing out that text type categorisation is frustrating because there is no theoretically satisfactory classification available yet. This may still be true, even given that Rissanen did not seem to be much acquainted with the theoretical work done in genre analysis. However, he is clearly right in asserting that we cannot expect to have a perfect theory before we can get down to empirical work.

It is clear from the examples above that the approaches are very far from each other, and in view of current trends the two fields seem to be diverging even further. Genre theory appears to be developing increasingly towards qualitative research and "thick descriptions" of individual genres (see, e.g., Swales 1996; Huckin 1997; Mauranen forthcoming), drawing from social theory, ethnography and anthropology, without attempting to provide comprehensive maps of genres, or indeed to validate empirically the genres they postulate in linguistic terms.

Corpus studies, on the other hand, seem to be developing towards larger and larger corpora, running now into hundreds of millions of words. The largest electronic corpora are what are known as "general" corpora, i.e. aimed at representing particular languages as a whole. A good deal of work goes into developing standards for their encoding and annotation. Smaller corpora (a few million words or less) are also being compiled, usually with more restricted scope and purpose, and often with special tagging. These can be restricted as to mode, domain or user groups, or they can be bi- or multilingual.

Despite their differences, these two approaches nevertheless subscribe to some similar goals and ideals in language study, notably the empiricist objective of describing language as it can be observed in use, as opposed to the rationalist methodology based on introspection. The sociocognitive view is wholly committed to a conception of language as social interaction. A similar need to study language in its social context is often emphasised among corpus linguists, especially those in the British tradition. For example, in the editorial of the first issue of the journal Corpus Linguistics, Wolfgang Teubert writes that corpus linguistics deals with language as a social phenomenon (1996: vi). By this he means that it studies concrete acts of communication as opposed to ideal language systems. The empiricist view is, then, seen as inextricably linked with language in the social world. Such thinking can easily be traced back to J.R. Firth (e.g., 1951, 1957 [1968]), whose work is behind both the socio-functionally oriented branches of contemporary British linguistics, and also corpus studies. For example his notion of collocation, or the "company words keep", has been

developed into one of the key concepts of present-day corpus linguistics, starting from Sinclair's work in the 1960's (Sinclair 1966 [1996]) when computers were not quite up to the challenge yet.

But is there really anything else the two lines of research involving genre have in common? Is there any way in which they might profit from each other's work?

In principle, we might hope to find ourselves in a situation where genre theory provides us with a principled way of sampling genres, as well as with good hypotheses about the linguistic properties of genres, which could be tested with empirical study of large bodies of linguistic data. And we would certainly wish to see large corpora compiled with the help of a sound genre theory, firmly anchored in the social reality of language use, so as to be able to account for both variation and commonalities in the behaviour of items in a reliable and meaningful way. But in practice there are major obstacles, perhaps insurmountable, on our way to this happy goal.

2. Internal criteria?

Should we rely on text-external or text-internal criteria in defining genres? This question presents problems to both corpus linguists and genre analysts in determining genres, or typologising texts. Corpus linguistics has traditionally based its text selections on what have been called "external" criteria, that is, criteria which in one way or another relate to what Firth calls the "context of situation" – criteria such as topic, medium, authorship (e.g., gender, nationality, age of author or speaker), and what is vaguely called "genre". The parameters of describing the context of situation have been most systematically developed by Halliday (e.g., 1978) and his followers. In this they would seem to be close to genre theorists, who firmly tie their notions of genre to situational and social context parameters such as purpose, function and discourse community.

Recently, however, there have been moves towards "internal" description of genres, by relying on inductive methods (e.g., Biber 1988; Sinclair 1992; Nakamura – Sinclair 1995) and the automatic analysis of very large corpora. The idea is to apply various highly automatic analytical and statistical procedures to very large and representative corpora, until patterning emerges from the material that can be related to the origins of the text material. The patterning seems to take the shape of feature clustering rather than single out individual features which could be taken to indicate genres or text types directly. One of the goals of such procedures is automatic typologising, which would mean that any new text material fed into a typologiser would be assigned to a genre. Biber used a factor analysis in his study (1988) to extract six dimensions on which the genres varied: "Informational vs. Involved Production"; "Narrative vs. Non-Narrative Concerns"; "Explicit vs. Situation-Dependent Reference" etc.

Nakamura and Sinclair (1995) explored the possibilities of using collocational patterns to fix genre variation statistically. Their study detected variation in the collocational patterns of one lexical item in different subcorpora of the Bank of English, but because of its narrow scope was more an exercise in methodology than a discovery of important genre patterns.

McEnery and Wilson (1996) welcome the taking of internal criteria on board on account of two advantages they take it to confer:

> First, it provides an incentive to stylistic analysis which is not only empirical and quantitative but which also takes greater account *of the general stylistic similarities and differences of genres* and channels *rather than how they differ on individual features*. Second, it should be influential in developing more representative corpora. *Culturally-motivated, and hence possibly artificial, notions of how language is divided into genres can be replaced or supplemented by more objective language-internal perspectives* on how and where linguistic variation occurs. (McEnery – Wilson 1996: 103, italics mine)

The first point I entirely agree with; it seems indeed that if we can achieve multifactorial or multi-perspective descriptions of language variation (be they types, styles, genres or registers), this is bound to be more informative than distributions of individual features taken one at a time. And the fruitfulness of this line of study has already been shown by people like Biber and his associates, and for instance Kaj Wikberg (1992).

The second point I find problematic. To begin with, because it appears that if we start looking for language-internal parameters for genre differentiation, we seem to be adopting a view of language which is excessively systemic – seeing language as a collection of systems, rules, structures and features with a logic of their own, severed from its users. On this point I think some corpus linguists are really taking the wrong path: moving further away from an understanding of genre as discourse in social contexts.

In addition, insofar as we manage to extract results from text-internal analyses, I do not see them as representing genres as I understand it, but some other kind of text typology. If we agree as a starting point that language is a form of social behaviour and that both genre analysts and corpus linguists take this as a common foundation, then obviously it makes sense to take a social view of genre, perhaps not unlike the one sketched by Berkenkotter and Huckin (1995), or earlier Swales (1990) or Miller (1984). Some recent developments have added another dimension to this kind of view: genre as a frame or script. As Kamberelis puts it: "genres are historically constituted open-ended sets of discourse expectations or discursive etiquettes" (Kamberelis 1995: 120). If genres are construed as schemata or frames, this implies understanding them as social facts, as shared expectations of the initiated in some discourse community.

Can we expect language-internal analyses to produce anything like genres thus defined?

If we attempt this, what we are actually doing is testing the functional hypothesis of language: if language adapts to its uses, purposes and functions, then by looking into the forms we ought to be able to infer the uses, and by looking into the uses/functions, we ought to be able to get to the corresponding variation in language. This is by and large the Hallidayan view (e.g., Halliday 1978).

But there are complications. For one thing, if there is no theory-based hypothesis about functionally differentiated genres, we cannot put the functional hypothesis to a real test, but can maintain it whatever the results. And the inductive kinds of procedures so far do not appear to yield results which would match a theoretically justifiable set of genres. If we feed in texts to a corpus, however large, on an atheoretical basis, we are in great danger of getting out what was put in, or not quite knowing what it is that comes out.

This mismatch might be overcome by wholly new ideas of what we might want to be looking for in language as genre differentiators – and corpus studies do show some promise in this area. For instance, the study of co-selectional patterns in texts questions established notions of units of meaning (Sinclair 1996), and by discovering new functional units, redefines borderlines between syntax and semantics, and between semantics and pragmatics (see also Tognini-Bonelli 1996). Such emerging functional units possess potential for coming to grips with socially significant variation in language perhaps better than the more traditionally investigated tenses, passives, nominalisations, connectors, etc. But whether such features co-vary with genres is yet to be established, and their potential in actually distinguishing between genres is an entirely open question. Thus the role of text-internal features is more appropriately to provide confirmation or disconfirmation to theoretically based hypotheses concerning systems of genres and their linguistic manifestations.

3. Heterogeneous text

In trying to pin down variation and type in text, we come up against the problem that we do not seem to be dealing with a unidimensional phenomenon. Genre theorists usually appear to be happy to operate with a uni-level concept, namely genre itself, to account for typical discourse patterns, and some corpus linguists have explicitly adopted the same stance (e.g., Stubbs 1996). However, in other text analytical traditions the need has often been felt to include more than one dimension of text type classifications, for instance the German Texttyp and Textsorte, and similar distinctions, as for example that by Wikberg (1992). In systemic-functional linguistics the two main concepts are usually genre and register. There are two things that particularly motivate a dual view of text typology. One is the inherent heterogeneity of texts that many researchers have

observed, and the other is the wholeness of genres as complete discourse events, or the fact that they form meaningful and particular kinds of wholes.

There is an attested propensity of texts to be heterogeneous rather than homogeneous. It is often the case that a given discourse, an instance or a type, contains different stages or phases in which different kinds of texts are used, or that it draws on many sources in some other way, since it is multifunctional in its specific settings. For instance Ventola (1987) has shown how different stages in a spoken genre use different register parameters, and Virtanen (1988) speaks of "unitype" and "multitype" texts, etc. This heterogeneity can be related to one sense of intertextuality as well, as is done by Fairclough (1992), who in discussing the heterogeneity of discourses sees genres as essentially mixed. Genre-mixing and genre embedding are taken up also by Bhatia (1998), as an attempt to account for the observable heterogeneity without recourse to different levels or kinds of text typology.

However, if it is the case that heterogeneity can be observed in individual instances of discourses, and even types of discourses, then the implication is that there must be some relatively homogeneous types of discourse that are drawn on in producing and interpreting the heterogeneous ones. The homogeneity must at least cover enough ground so as to be recognisable. If a genre were not recognisable by a small fragment we would not be able to notice any heterogeneity. Genre-mixing and genre embedding would only be limited to the somewhat trivial case of marked quotations. Underlying the heterogeneity must therefore be an assumed homogeneity – a prototypical correlation of form and function, which is shared by members of the discourse communities using the genres.

Homogeneity of text type is assumed in register analysis, as well as in those typologies which distinguish such types as argumentative, narrative, expository etc. They tend to use key terms such as "formal" and "functional" in widely different ways – for example Biber (1998) uses the term "formal" of distinctions like expository and narrative, while e.g., Werlich (1976) and Wikberg (1992) talk about "functional" for the same distinction. Despite this woolly terminology, what Biber means by his formal typology seems to be very close to Halliday's notion of register. Halliday (1992: 68) characterises register as "a syndrome of lexicogrammatical probabilities", which sounds exactly like Biber's feature clusters. Typologies of these kinds seem to assume homogeneity within the types, but to adopt dual approaches to account for variability: they use at least two sets of criteria for categorising texts on different dimensions. Such a solution seems to capture linguistic reality better than unidimensional models. If we take, say, argumentative and expository text types, it is clear that they can be differentiated on a number of features. Yet neither can be called a genre in the sense of type of social behaviour linguistically performed – for instance news-

paper editorials and academic papers can both be argumentative, but not of the same genre.

Therefore, we need two dimensions (at least), to be able to account for heterogeneity. This need is reflected in pairs of concepts such as Texttyp and Textsorte, or genre and register; or we could simply distinguish between "type", recognisable by internal criteria, and "genre", based on social criteria (which is essentially in line with Biber, see next section).

4. Texts as wholes

In contrast to text typologies based on register, or the division of texts into types like argumentative, expository, etc., the concept of wholeness or completion has only been brought up in the context of genre. Thus the label "genre" is typically applied to entities which are recognisable as certain kinds of wholes. Barbara Couture (1986) talks about genres specifying the conditions for beginning, continuing and ending texts. A similar view has been adopted by Ventola (1987), and Jim Martin (1985) relates it to the function of genres in accomplishing particular rhetorical acts. As George Kamberelis (1995: 128) puts it, genres are organised according to a principle of closure. The principle of closure can be traced back to Bakhtin's (1986) "finalisation", which he defines as the process whereby a text becomes complete. This process takes place at different levels of discourse organisation, the highest of which is the text or discourse as a whole. Such ideas are also reflected in the way in which genre theory usually understands generic structure: organised at different levels, the top level being the text as a whole.

It seems, then, that in order to capture heterogeneity as well as closure in texts, and to handle typicality and variation, we do need two dimensions of categories.

Biber is one of those who recognises this need, linking it explicitly to external and internal criteria:

> "I use the term 'genre' to refer to categorizations assigned on the basis of external criteria. I use the term 'text type', on the other hand, to refer to groupings of texts that are similar with respect to their linguistic form, irrespective of genre categories ... In a fully developed typology of texts, genres and text types must be distinguished, and the relations among them identified and explained ..." (Biber 1988: 70)

I fully agree with this. Moreover, I believe good external criteria for corpus compilation ought to be informed by genre theory. Otherwise we remain in a circular position: we feed into our corpora intuitively categorised text grouping, and receive as output features of clusters describing what we had put in in the first place.

5. Mutual benefits?

Up to now, this discussion has focused on the benefits that a sound theory of genre might bestow on the selection of genres for corpora. Yet genre theory may also well be able to offer more specific advantages to corpus study, for example in the thorny problem of sense disambiguation. If it is the case that word senses vary with genre, that is, if the sense profile of a given expression is different in different genres, then it seems that in order to describe the use of this expression we need to take its generic distribution into account. Use is here understood closely intertwined with, even indistinguishable from meaning, in tune with Wittgenstein's and Firth's conceptions of meaning. We may even question the value of describing senses of lexical items as they result from a general corpus, i.e. one where genres are jumbled. Such senses may well represent some common core that a natural language can be said to have, but they result from highly abstract generalisations, and the diversified use in different genres are more likely to reflect senses that are meaningful to language users. Sense disambiguation for human users of natural language is quite likely to be heavily dependent on genre. For example readers bring top-down knowledge of this kind into the reading process from the start, and therefore need not engage in a complex disambiguation search, which for machines constitutes such a problem. If we had adequate genre information to feed into programs, we might be able to alleviate disambiguation difficulties.

Corpora compiled on a principled selection of genres can, in return, provide genre theorists with valuable descriptive tools and points of reference with which we can describe crucial features of genres. If text-internal parameters do not serve as viable criteria for determining genres, they are indispensable for testing hypotheses about genres defined by other criteria. The fundamental questions centre around the functional hypothesis – to what extent are the functions of language on the one hand, and linguistic features on the other mutually predictable? Moreover, corpus data can provide evidence for (or, as the case may be, against) our hypotheses concerning such questions as the heterogeneous nature of genres, and well as their typicality, stability and change.

One of the major strengths of corpus analysis is that it throws new light on linguistic patterning which has been hard to detect earlier. Insights from corpus research could therefore reorganise our notions of genre and the evidence supporting for instance functional assumptions of language.

However, current electronic corpora are mostly not yet appropriate for genre study. Most of the large ones aim at representativeness of a given language, and the way they try to achieve this is by gathering a little from all major categories of language as defined by the compilers. But the sampling that corpus compilers make within any given text category, be it genre or otherwise, is not even intended to be representative of that genre, as Nakamura and Sinclair (1995) observe. It is intended to make the corpus as a whole representative of the language as a whole.

Moreover, general corpora do not cover all aspects of language usage, or all thinkable genres. A case in point is translations: most general corpora reject them offhand, not considering the fact that some genres are heavily dominated by translated discourses while others are not. Stubbs (1996: 11) points out that genres of ordinary everyday discourse, such as diaries, personal letters, business correspondence, company reports, etc. are often missing from general corpora. Spoken language is also usually heavily underrepresented even in the largest corpora.

6. Conclusion

What I have been trying to put across in this paper is the idea that genres are of fundamental importance to language study. We cannot have corpora felicitously represent given languages without representing their genres adequately.

Genres should play an important part in discussions on what is shared and what is variable in human languages. Looking at genres as formations which organise language as social action in higher order structures and which act as resources for members of cultures and discourse communities should be a fundamental aspect of language study along with formal, semantic and pragmatic aspects of language. They are on the topmost layers of a Firth-inspired top-down approach to language description, beginning from sociological levels and going down to phonetics.

Genres are also important since they are types of language behaviour. Leontiev has suggested (1991) that all cultural phenomena involve "a dialectical unity of two flows: (a) overcoming of existing standards and stereotypes and (b) standardization and stereotypization of innovations" – this is where genres come in. In this, genres exhibit the same tensions as other levels of language use: languages tend towards convergence and stability on the one hand, and towards divergence and variation on the other.

Moreover, genres transcend cultural and linguistic boundaries, even if in somewhat transformed guises. They are thus an essential part of exploring language change and language contact.

What I would like to suggest as a way forward is that we start building corpora based on external criteria which are properly informed by genre theory, and conversely, start formulating our genre hypotheses in ways which allow some of the fundamental assumptions become relatable to, even testable with large amounts of data.

REFERENCES

Bakhtin, Mikhail
   1986   *Speech genres and other late essays*. Austin: University of Texas Press.
Bazell, C. E. – J. C. Catford – Michael A. K. Halliday – R. H. Robins (eds.)
   1966   *In memory of J. R. Firth* London: Longman.
Benson, James – William Greaves (eds.)
   1985   *Systemic perspectives on discourse*. Norwood, NJ: Ablex.
Berkenkotter, Carol – Thomas Huckin
   1995   *Genre knowledge in disciplinary communication*. Hillsdale, NJ: Lawrence Erlbaum.
Bhatia, Vijay K.
   1993   *Analysing genre*. London: Longman.
   1998   "Genre-mixing in academic introductions", *English for Specific Purposes* 16, 3: 181-195.
Biber, Douglas
   1988   *Variation across speech and writing*. Cambridge: CUP.
Couture, Barbara
   1986   "Effective ideation in written text: A functional approach to clarity and exigence", in: Barbara Couture (ed.), 69-88.
Couture, Barbara (ed.)
   1986   *Functional approaches to writing. Research perspectives*. London: Frances Pinter.
Fairclough, Norman
   1992   *Discourse and social change*. Cambridge: Polity Press.
Firth, J. R.
   1951   *Papers in linguistics 1934-1951*. London: OUP.
   1957   "A synopsis of linguistic theory, 1930-55", in: *Studies in Linguistic Analysis*. Special volume of the Philological Society, Oxford, 1-31.
   [1968]   [Reprinted in: Frank R. Palmer (ed.), *Selected papers of J. R. Firth 1952-59*. London: Longman, 1968, 168-205.]
Halliday, Michael A. K.
   1978   *Language as social semiotic*. London: Edward Arnold.
   1992   "Language as system and language as instance: The corpus as a theoretical construct", in: Jan Svartvik (ed.), 61-78.
Huckin, Tom
   1997   "Cultural aspects of genre knowledge", in: Anna Mauranen – Kari Sajavaara (eds.), 68-78.
Kamberelis, George
   1995   "Genre as institutionally informed social practice", *Journal Contemporary Legal Issues* 6, 115: 115-171.
Leitner, Gerhard (ed.)
   1992   *New directions in English language corpora*. Berlin: Mouton de Gruyter.
Leontiev, Aleksei A.
   1991   "Personality, culture, and language", (Paper presented at the AFinLA symposium, November 1991, Oulu, Finland.)
Mauranen, Anna
   forthcoming   "Kontrastiivinen retoriikka" (Contrastive rhetoric), in: Kari Sajavaara – Arja Piirainen-Marsh (eds.).
Mauranen, Anna – Kari Sajavaara (eds.)
   1997   *Applied linguistics across disciplines*. (AILA Review 12.). London: AILA.

Martin, James
   1985   "Process and text: Two aspects of human semiosis," in: James Benson – William Greaves (eds.), 248-274.
Miller, Carolyn
   1984   "Genre as social action", *Quarterly Journal of Speech* 70: 151-167.
McEnery, Tony – Andrew Wilson
   1996   *Corpus linguistics*. Edinburgh: Edinburgh University Press.
Nakamura, Junsaku – John M. Sinclair
   1995   "The world of woman in the Bank of English: Internal criteria for the classification of corpora", *Literary and Linguistic Computing* 10, 2: 99-110.
Rissanen, Matti
   1992   "The diachronic corpus as a window to the history of English", in: Jan Svartvik (ed.), 185-205.
Sajavaara, Kari – Arja Piirainen-Marsh (eds.)
   forthcoming *Soveltavan kielitieteen käsikirja*. (*Handbook of applied linguistics*). Jyväskylä: University of Jyväskylä, Centre for Applied Linguistics.
Sinclair, John M.
   1966   "Beginning the study of lexis", in: C. E. Bazell – J. C. Catford – Michael A. K. Halliday – R.H. Robins (eds.), 410-430.
   [1996]   [Reprinted in: Joseph A. Foley (ed.), *J. M. Sinclair on lexis and lexicography*. Singapore: UniPress, 1996, 1-20.]
   1992   "The automatic analysis of corpora", in: Jan Svartvik (ed.), 379-397.
   1996   "The search for units of meaning", *Textus* 9: 75-106.
Stubbs, Michael
   1996   *Text and corpus analysis*. Oxford: Blackwell.
Svartvik, Jan (ed.)
   1992   *Directions in corpus linguistics. Proceedings of Nobel Symposium 82, Stockholm, 4-8 August, 1991*. Berlin: Mouton de Gruyter.
Swales, John
   1990   *Genre analysis. English in academic and research settings*. Cambridge: CUP.
   1996   "Toward a textography of an academic site", (Paper presented at AILA 96, the 11th World Congress of Applied Linguistics.)
Teubert, Wolfgang
   1996   "Comparable or parallel corpora?", *International Journal of Lexicography* 9, 3: 238-264.
Tognini-Bonelli, Elena
   1996   Corpus theory and practice. [Unpublished Ph.D. dissertation, University of Birmingham.]
Ventola, Eija
   1987   *The structure of social interaction. A systemic approach to the semiotics of service encounters*. London: Frances Pinter.
Virtanen, Tuija
   1988   Discourse functions of adverbial placement in English: Clause-initial adverbials of time and place in narratives and procedural place descriptions. [Unpublished licenciate thesis, University of Turku.]
Werlich, Egon
   1976   *A text grammar of English*. Heidelberg: Quelle & Meyer.
Wikberg, Kaj
   1992   "Discourse category and text type classification: Procedural discourse in the Brown and LOB corpora", in: Gerhard Leitner (ed.), 247-261.