In praise of the dictionary

(c) John Sinclair 2004
Tuscan Word Centre

This paper deals with the theoretical and descriptive background to research that I am conducting into the design of language reference works. At the conference I hope to show aspects of it in practice with respect to English lexicography.

The reason that I sing the praises of the dictionary, dictionaries and the craft of lexicography is that among language reference tools they, and they alone, get right one very important matter – the priority that they accord to meaning. Meaning is the only thing that is ultimately worth bothering about in language and so a sustained focus on meaning is most laudable, and an example to other branches of linguistics.

This is in marked contrast to the attitude to meaning afforded by grammars. Some of them use meaning as a criterion, and talk of meaning and similarities and differences of meaning, but unfortunately the conventions of the grammars imprison them within a closed system that at its best only approximates to meaningful choices in the world outside.

The received model of meaning

I will return to the point about grammars, but meanwhile let me go over the most widespread model of how language makes meaning; I won't take long over it because it is so well-known, and it has stood the test of time and reached all corners of the globe because of its inherent simplicity.

In this model the primary units of meaning are the words; each word is a pointer to a meaning that is activated when the word is used in a sentence. Another kind of meaning is activated by the disposition of the words relative to each other in sentences, a kind of meaning that we call syntax. The two kinds of meaning are related by philosophers to "reference" and "truth"; the words are held to refer to entities or processes or attributes, and the syntactic arrangement of the words can be related to propositions, which characteristically have a truth value – that is, they can be right or wrong.

There are, of course, some complications, because language is never as tidy as those who describe it would like it to be. There are three variations of the basic position that must be taken into account:

1. Some words have more than one meaning. This does not invalidate the basic model, but to be adequate the theory should incorporate some means of deciding for each occurrence which meaning is activated.
2. Some meanings are activated by a group of two or more words rather than a single word. This is potentially destabilising unless it only affects a small number of easily recognisable items, like the *idioms* that we are familiar with.

3. A fairly small group of largely common words, often called "grammatical words", do not activate specific meanings when they occur, but are deployed internally in the management of the sentence. This also is potentially destabilising unless the grammatical words are easily recognisable and can safely be ignored in explicating the meaning.

The first variation, the apparent multiplicity of meanings of a word, has always been accepted as a normal fact of language, and is rarely questioned. This low profile may be possible because there is hardly ever any problem caused by the potential for ambiguity that a "one word, many meanings" model has. And there is hardly ever any problem because the other words around tend to focus the mind on only one of the available meanings.

From the point of view of an adequate theory of language, this is an unacceptable situation, since the only feature that prevents the "one word, many meanings" model leading to unintelligible sentences and unresolvable problems of interpretation is that speakers and writers harmonise their choices over short stretches of text. But this crucial feature is not normally considered part of language description at all. Neither dictionaries nor grammars make provision for it. The tail wags the dog with a vengeance.

The importance of the second variation depends on the number and frequency of multi-word units of meaning and we will deal with that below. The third variation, concerning the role of the "grammatical" words, contains a problem that is characteristically ignored. Grammarians do not agree on the set of grammatical words, but nevertheless rely on people being able to recognise them. "Everyone knows what they are" is a typical response to a request to list the grammatical words of a language, and this response contrasts sharply with the lack of an agreed list. It seems more likely that there is a continuum of "more grammatical – less grammatical" words, and different authorities (if they even bother) make a distinction at different places.

Illocution

In summary, the received model of how meaning is created in language has always had some unresolved problems, but they were not considered serious enough to threaten the widespread use of this model. It was pretty well universally accepted by linguists until the nineteen-sixties, when a book was published by a philosopher, and after that meaning was never going to be the same.

J.L.Austin (1962) pointed out that the application of truth-value criteria is not appropriate for large classes of sentences, those he called *performatives*. A promise, an apology, a greeting – these may be sincere or not, but the notion of "truth" does not apply to them. Questions and commands also have no prospect of a truth value. Austin went on to show that even sentences that notionally had a truth value also had an *illocutionary force*, by which he meant that they performed actions in the real world.

Remarkably, Austin lacked interest in language as interaction, although he made possible the serious study of that aspect of language structure. The transfer of his ideas into linguistics took place in the early seventies, and *discourse analysis* came into being. Also developing at that time was *conversation analysis,* an initiative from the branch of sociology called ethnomethodology, which studied the structure of conversations in terms of adjacent pairs of utterances, and the structure of the utterances from an interactive point of view.

The property of utterances that is relevant to linguistics is the illocutionary force, and not the truth value. The question of truth value just does not arise in the description of language as interaction, because there is no objective arena where it can be determined. Someone asserts something, and if the other participants do not challenge it then by default it acquires a provisional veracity.

So the traditional support for syntactic meaning, the connection with truth value, had to be abandoned. The hypothetical procedure that for every sentence it could be established whether or not it was true in the real world is irrelevant to language analysis. Participants in a conversation or in the reading process adopt an attitude that is similar to what Coleridge called "that willing suspension of disbelief for the moment, which constitutes poetic faith " (1817). We are not on such exalted ground as poetic faith, but merely engaged in the business of day-to-day conversation; we expect the attitude to be more a suspension of judgement than disbelief, and we anticipate that the suspension will last for more than a moment – perhaps until at least the end of the conversation. Essentially we do not assign truth values with any finality, but retain an open mind on the likely accuracy of what is said to us, along with its sincerity, plausibility and the intentions behind it.

It may not be so obvious that the other kind of meaning, usually called referential, is also undermined by Austin's work. The important point is not whether or not there are referential links between words and the world around – it is obvious that there are such links, at least for simple concrete objects. The question we must ask is whether or not these referential links are the means by which words acquire and maintain their meanings, or whether the meanings are a consequence of some other process. I will return to this discussion later, but at present it is only necessary to establish that the argument about truth value above has severed a spurious connection between utterances and the world. When we are engaged in discourse, if we do not look to the world for the truth of propositions then perhaps we do not need to look to the world for the meaning of words either.

Austin's work led to reconceptualisation of the nature of meaning when it was applied to language interaction; whereas the spoken language is quite clearly interactive, after a while people realised that the written language was also interactive at heart. Although both writing and reading can be done by an individual in isolation, and that is one of the great strengths of writing, the act of reading is interactive because the reader engages with the text, and the text has been prepared by the author for just this kind of occasion. The author, when composing the text, had in mind some target reader or readers, and wrote with them in mind; the most successful writings are often held to be those that anticipate the reaction of the target readers most accurately and take the reactions into account.

The lexical item

Gradually during the nineteen-eighties, another major development became a force to be reckoned with. The assembly of language corpora using computers had begun some twenty years before, but the early corpora were rather small to yield much of interest – tiny by today's standards – and were studied only by a few dedicated scholars who were willing to tolerate the shortcomings of the computers, the pathetic software and the assumptions of the time that computers were really intended for number-crunching.

No-one expected anything revolutionary from corpus research; the received models of language were largely trusted to have done their job – to have elucidated the main ways in which meaning was created. Researchers anticipated only clarifications, exhaustive enumeration of classes and details of changes over short periods of time. Indeed, quite a lot of corpus study proceeds on just this basis, preferring not to notice the more disturbing aspects of the early results of what is called *corpus-driven* research (Tognini Bonelli, 2001).

Let us return to the three variations on the basic received model of how language makes meaning. The first variation was the lack of any elucidation about how the apparent polysemy could be resolved, qualified by the observation that the nature of text made this a relatively minor matter, and further qualified by the observation that this valuable feature of text was not predicted by the theories. Corpus research offers a full explanation of this situation by pointing out that meaning is created, not over each single word, but over several words together. Polysemy is not an appropriate name for this phenomenon because meaning does not inhere in the word, and there is only very rarely ambiguity in a naturally-occurring text.

The second variation was just this point – that some meanings require several words. If there were just a few, they would not undermine the authority of the model. In English we can point to the usual few hundred idioms, and add from corpus research the useful point that they are not very frequent and cannot be used to fuel a strong counter-argument. Also there are the phrasal verbs, of which there a lot, and they are extremely frequent; however these are of such regular construction and pervasive occurrence that they can be seen more as an unusual grammatical feature rather than an exception to the lexical norms.

The received model would survive these depradations. The point where it becomes untenable is where it becomes clear that in a large proportion of text, over 50% and much higher in some texts, meaning is created by words in combination and not by single words. These are not idioms in the usual sense, but *coselections* that harmonise meanings across long stretches of text. There is nothing in received theory to prepare us for the structures described in Sinclair (1966, 1999a).

The third variation above concerned the grammatical words. Corpus research gives no support to those who would like a sharp dividing line between these and the other words, and it shows that they have a very strong lexical role in the multi-word lexical items that emerge as characteristic of the vocabulary of modern English. In fact preliminary studies of these words suggests that they are individualistic in their patterns of usage (Sinclair 1999b). So there is also no support for the idea that the "grammatical words" can be safely ignored in the explication of meaning – they have a central place in the making of meaning.

Gains for lexicography

The inevitable impact of the early investigation of corpora is that the received model of language is untenable. It is neither comprehensive enough nor flexible enough to cope with the new information that is coming in, and practically all of its assumptions are seriously challenged by the evidence that is accumulating. However, it is a much-loved and much-used model, and some time must pass before it is consigned to the archives.

When we abandon this model, three major gains for lexicographers become obvious.

1. The unit of description, the *definiens*, can be of indeterminate length. The idea of restricting it mainly to the word is no longer supported by theory. Until a lot of analysis has been done the identification of the units remains, as now, a matter for the professional discretion of the lexicographer, bearing in mind the following paragraphs. (It is true that multi-word units may still raise problems in practice – for indexing etc. – but that is a separate matter).
2. There is no support for the idea of words having more than one meaning. The task of the lexicographer becomes a systematic search for what I have called *canonical* forms – a unique expression that is reserved for each distinct lexical item.
3. Even after the canonical form of a lexical item is identified, the characteristic cotext is part of the meaning, and so is relevant to the definition of the item.

In order to profit from the second and third gains, and ultimately from the first as well, lexicographers need to get closer to corpora, and engage with them more systematically. The tradition of lexicography, which I support and defend in its determination to remain unrestricted in the expression of meaning, has inevitably stressed the subjective responses of the trained compiler. However, it is specifically beyond the personal resources of the individual to discover the canonical form of a lexical item, or to know that it is the canonical form if it is stumbled on accidentally. Personal intuitive judgement is called for eventually, but the sifting of the large amount of evidence required to present the strong recurrent patterns is the job of a powerful computer. Further, beyond the specification of a lexical item lie the cotextual links and harmonies that are so important in explaining how an item is used, and these also are inaccessible to the unaided intuition.

To get the best out of the present situation lexicographers need to rely more on corpus information, to take it more literally, and to focus subjective decisions within the guidelines of the corpora. These are proper responses to the challenge of lexicography today, and carry the pursuit of meaning to new standards of accuracy.

Full sentence definition

Here I am very pleased to point out that the defining technique called FSD, full-sentence-definition, is ideally suited to exploiting the gains just discussed. I developed this defining style for Cobuild some twenty years ago with no concern for theories of meaning – at that time it was the findings of spoken discourse analysis that convinced me of the ease and importance of the FSD, and they have made the Cobuild dictionaries distinctive ever since.

Early reactions to FSD were that it seemed to be an uneconomical technique, using many words which could be taken for granted. Since dictionaries are chronically short of space this argument had to be taken seriously; there are two main responses from the strict perspective of writing good dictionaries, and another from a more general point of view.

1. There is a lot more linguistic information in an FSD than in a traditional type of entry, and it does not all need to be explicit; all sorts of typical features of the cotext can be reproduced, suggested, alluded to, hinted at, without the necessity of a bald statement. At the present time this lack of precise accountability is a positive advantage. A traditional entry that attempted the same detail would have to commit itself at times when the evidence was diverse, or omit a potentially

important semantic observation. So Cobuild can say with impunity "If the police **arrest** you..." even though there exists the legal possibility of a citizen's arrest – Cobuild does not imply that no-one else can possibly arrest you; the traditional definitions are mealy-mouthed or vague in their attempts to wriggle out of commitment here[1].

2. An ordinary English sentence has a quality, akin to legibility, that we might call *interpretability*. It follows one of a familiar set of patterns and it is basically easy to understand. It has a modest lacing of redundancy to help it along; the typical definition sentence has a fairly elaborate *matching* grammar that highlights the superordinate, discriminators etc. (Barnbrook and Sinclair 2001). The same information in a set of notes is much less easy to handle, and will end up bulkier and more ungainly than a sentence.

The difficult side of FSD is not its supposed inefficient use of space, but the fact that without careful recording of corpus evidence the defining sentences cannot be constructed. A sentence commits its author to articulating certain elements whereas a set of notes is much less demanding. So the definition of *arrest* in the Longman Dictionary of Contemporary English (1987) is "to seize by the power of the law". This avoids specifying either who or what does the seizing or who or what gets seized – it is not even clear that it is people who get arrested. Nor is it clear how the power of the law is exercised, and to cap it all the meaning of *seize* is not any of those given in the same dictionary.

3. The third response to the charge of wasting space with FSDs is not directly relevant to lexicography as it is practised at the present time. The response draws attention to a feature of definitions that is so central to their nature that it may sometimes be overlooked. Definitions are composed in language; in monolingual dictionaries both definiens and definiendum are composed in the same language. No outside agency is involved – though non-formal, definitions form a closed set of expressions that perform their functions without the need to conjure up the world outside, ontologies or logical relations.

4. In the light of the third response above, it would be most unfortunate if definitions could not profit from being wholly within the language system. If they are sentences of English this allows them to be analysed by the rules of English, in a similar way to all the other sentences. Those definitions that fall short of sentence status lose access to the regularities of the language system and the essential connections with the other sentences of the language. While special grammars can be written for them, at a price, they still do not connect with the rest of the language[2].

The semi-formal nature of meaning

I promised to return to the difficulty of relating meaning and grammar. In general terms, grammars tend towards being formal systems, and some of them indeed are expressed in the terms of logical formalisms. In contrast, as I have argued above, meaning cannot be so confined, and lexicographers are right to refuse restrictions of this nature.

The grammars we know, whether formal or not, create meaning by paradigmatic choice. The meaning of each choice depends upon the totality of the choices available. In

systems of this kind, one single term has no exact meaning on its own, so the familiar terms like *singular*, *interrogative*, *second person* are only approximately related to the kind of meaning that the words chosen as terms seem to promise. Anyone who has tried to explain these terms to a lively undergraduate class will recall the hazards that emerge.

The methodology of lexicography is not fettered by these structural considerations; within the reasonable limits of editorial policy words can be defined in and for themselves, using whatever language is thought to be the most efficient and accurate for the task in hand. Each task is very small in scale, and the standards required are high, bordering on pedantry at times.

I am not sure how and when meaning went missing from systematic language description (other than lexicography). I imagine it was a slow process, starting perhaps with the mis-match of Latin terminology applied to English structure, and compounded more recently by the demands of a more formal approach to description. Further pressure is now being applied by the requirements of computational linguistics, which offers very considerable advances but which requires absolute precision of category definition, and this requirement makes it very difficult to accommodate meaning directly. Meaning gets simplified, systematised, tidied up - and risks losing its connection with our ordinary perception of it.

Computer Science and Information Science are quite unable to handle linguistic meaning, and have devised alternative strategies, of which I will mention just three of the more popular ones.

1. Words in isolation and combinations, without regard to meaning. The search engines rely on this technique, performing fuzzy searches purely on the character strings that are entered often with results that range from the alarming to the hilarious. Soon to be combined with (2) for added semantics.
2. Ontologies, organisation of words in relationships like hyponymy, antonymy, now called WordNets. Unfortunately the relationships are not comprehensive, and indeed seem to prioritise relationships that, though felt to be "core", are not often realised in texts. The ontologies that are popular today combine the worst features of grammars, being closed formal systems, with the appeal to referential explanations, whose relevance is questioned in this paper. It should however be said that there is nothing inherent in ontologies that precludes them from use in lexical analysis and description; it is the nature of the particular ontologies that have been developed and their unspoken assumptions that make them irrelevant to the study of linguistic meaning.
3. Statistical approximations, much used in automatic translation support. These take no account of meaning at all, and so they cannot be assessed for accuracy or utility value except by referring to external human adjudication.

Reflexivity, paraphrase and explicitness

We are left at this point with no alternative to the rather laissez faire world of the lexicographer, to the tenet that the best way of describing meaning is to use ordinary language in an unfettered way. There is no standard or best definition – all competent definition statements throw some light on the meaning of the definiens.

The lexicographer's definition skills have been concentrated on one particular area of meaning, what I will call the classificatory kind of meaning, where the characteristic technique is to associate the definiens with a superordinate and then qualify it with a discriminator. Other aspects of meaning and usage are normally relegated to appended notes. However, in recent years, another important area of meaning has become prominent, and this is non-classificatory – the emotional, connotative, attitudinal, pragmatic kind of meaning. Dictionaries rarely used to bother much about this kind of meaning, but the cotexts that come with citations from corpora make it clear that such meaning may well be the main reason for the choice of one lexical item rather than another. This kind of meaning – I use the term *semantic prosodies* to encompass the effects of all the adjectives above – can no longer be overlooked in reputable lexicography, and the job of including it will diversify the whole idea of a definition.

The act of definition makes use of two key properties of natural language, properties which enable our everyday language to be so flexibile and expressive. One is *reflexivity*, the ability of a user to talk about the language, and the other is *paraphrase*, the ability of a user to provide alternative phrasings of similar meanings. Any act of paraphrase which has as its aim the rephrasing, maintaining the meaning, of a word or phrase in the language, is definition, while *defining behaviour* is a broader category including any observations at all concerning another word or phrase in the language.

A third property needs to be added, because of the perceived purpose of definition, and that is *explicitness*. The definiendum should be more explicit than the definiens. So, in diagrammatic form,

definition = reflexivity + paraphrase + explicitness

Any sentence that co-ordinates all three of these properties should be some kind of definition.

This seems to be as close to formalism that we can get while keeping faith with meaning. The accuracy of paraphrase is a matter of subjective judgement and is likely to stay that way for the foreseeable future; the nature of explicitness is very complex although related to the size of the realisations in words. The number of possible definitions is unlimited, and they are all relevant to the meaning of the lexical items.

Theory of Meaning

This paper has taken shape principally as a critique of practical monolingual lexicography, but from the point of view of an emerging theory of meaning. Since we are obliged to abandon referential theories and keep a safe distance from ontologies, the dictionary comes

into focus as illustrating most of the properties that are required of the semantic component of a new theory. Of particular importance are:

1. The exclusive concentration on meaning. True, the focus has been somewhat narrowly centred on the classificatory side of meaning, with prosodics relegated to occasional notes, but there is perhaps some justification for that. Whereas a study of the occurrence of a lexical item in its cotext leads to the retrieval of its semantic prosodies, no such investigation can discover the strand of meaning that distinguishes the item from all others (Sinclair 2002).
2. The maintenance of an arena for processing meaning which is totally within the language system and not connected directly to any other system, organisation, algorithm or notation[3]. Following on the argument that meaning cannot be represented properly in a formal system, we have identified a semi-formal system with excellent properties of organisation, within which it meaning naturally resides. There now arises a need to make paraphrase a discipline in its own right, and to associate different paraphrases of the same original together, perhaps in the way that Harris (1952) envisaged using *equivalence classes*.

The theory of meaning takes as its starting point that meaning is holistic, unique to the text in which it appears and unique to each individual participant in the communicative process. The matter of individuality is confirmed by Weinrich's (2000) postulate of *discontinuity*; one of the most important characteristics of the language system is that people who are ultimately mysteries to each other can exchange ideas, stories and negotiate together without difficulty. A discourse participant has no knowledge of what a phrase means to another participant except operationally – that if it is used, the participants behave as if it had much the same meaning to all of them, and if they happen to explore the meaning by using the feature of reflexivity, then the results are not surprising. It is one of the main jobs of language description to pursue the common ground of meaning as far as it can be taken.

Another major job is to reconcile each unique cotext with the language system as a whole; to recognise those aspects of meaning which are arise from recurrent patterns like lexical items and to interpret the other patterns which are the result of the juxtapositions and fine tuning of the lexical items. Lexicographers do this all the time when they are working with concordances, just as they used to with the citation slips; they filter out the purely local, ad hoc meaning and concentrate on the meaning that is apparent in all or most of the instances.

Meaning is holistic – it does not survive an initial division into grammar and lexis. Winter (1977) pointed out that all the categories of grammatical meaning could be expressed without using the grammatical choices, but by making lexical choices with similar meanings. We said above that the categories of grammatical meaning were never quite what they said they were, but the correspondence of categories like *negative, modal, interrogative* are close enough.

This is not the occasion to expatiate on the relation between meaning and grammar; it is only necessary to reiterate that insofar as lexis and grammar can be separated, the provenance of meaning is the lexis, and that most of the contribution of grammar to meaning is of a secondary nature.

Future Prospects

In my talk at Euralex in July I will move on from this position statement to consider desirable developments. While I continue to praise the dictionary for keeping faith with meaning, I will summarise the reservations that have emerged so far, particularly the neglect of full sentence definitions. Then I will draw attention to more fundamental structural shortcomings, and go deeper into the nature of paraphrase and the notion of the dictionary as a text. From this I hope to propose some additional priorities for lexicography, and at the same time develop the present concept that we have of dictionaries into adequate vehicles for the semantics of a language. At present they appear to be too diverse, fragmented and idiosyncratic – but dictionaries are too successful to be taken for granted or criticised without good cause, and I hope to show that the cause is good.

Along the way I will pay some attention to that poor relation of the dictionary, the thesaurus. While thesauruses are well-intentioned, and – in abstract terms – have a clear place in the description of meaning, they seem unable to achieve their goals. In practical terms, the slowly growing acceptance of electronic publication will get over the massive problem of space that a really adequate thesaurus would face, but in theoretical terms a very interesting prospect arises of a thesaurus crafted out of corpus evidence.

To summarise: A dictionary is both a practical object on the bookshelf and an application of a theory of language to the description of meaning. A dictionary that is backed by a powerful theory is likely to be more authoritative than one that simply tries to attract customers, but there is no need for them to be forever at cross-purposes. While we have no reason to expect that the priorities of both will be the same, there is some value in trying to keep them aligned. A theory of meaning that is exemplified by a good dictionary can claim to have bridged the gap between the abstract world and the language as used in communication.

Endnotes

1. I have to record slightly raised eyebrows here in the use of *you*. In the first edition of Cobuild (1987) the pronoun was *someone*, following the conventions that I established for pronoun use. Since, unusually, the definitions in Cobuild address the reader directly as "you", compilers were urged not to ascribe socially undesirable activities to this pronoun. The impersonal pronoun distances the addressee from the events described, and we did not wish to imply that "you" were the kind of person who might do the sort of things for which one gets arrested. The second edition of 1995 altered the pronoun to *you* and subsequent editions have retained this phrasing, perhaps responding to the increasing lawlessness of society; I wonder when I will read in Cobuild "When you **assassinate** someone....."

2. For readers familiar with Barnbrook and Sinclair (2001) and other similar arguments over the years, this paragraph may seem to be at cross-purposes with the idea of *local grammars*. The paragraph is in fact carefully phrased to be consistent, but a word of explanation may assist. The principal point is that FSDs can be parsed by any adequate parser of English, even though a local grammar will be found to be more insightful. The property of "being a sentence of English" connects all such sentences, and the various subclassifications of sentences that arise in detailed description do not invalidate that connection

3. Some dictionaries, it is true, use diagrams, photographs and illustrations to support the text. This is a marketing matter, that takes into account the needs and abilities of certain classes of user, such as

small children or adult learners of the language in question. They should never take the place of the paraphrase, and should never be used to exonerate an inadequate definition.

References

Dictionaries

Collins Cobuild English Language Dictionary, 1987; second edition 1995; London, HarperCollins.

Longman Dictionary of Contemporary English, Second edition 1987: Harlow, Longman

Other works

Barnbrook, G. and J. Sinclair 2001 "Specialised corpus, local and functional Grammars" in M Ghadessy et al (eds) *Small Corpus Studies and ELT*: Amsterdam / Philadelphia; John Benjamins. Pages 237-276.

Coleridge, S. 1817 *Biographia Literaria*, chapter 14

Harris, Z. 1952 "Discourse Analysis" in *Language* 28, pages 1-30

Sinclair, J. 1966 "The search for units of meaning" in *TEXTUS* IX, 1 Special Volume eds Merlini L and J Sinclair: Lessico e Morfologia. pages 75-106. Reprinted in Corpas, G. (ed) (2000) *Las Lenguas de Europa: estudios de fraseología, Fraseografía y Traducción*; Editorial Comares, SL, Granada. Pages 7-38. Reprinted also in Sinclair 2004.

Sinclair, J. 1999a "The lexical item" in Weigand, E. (ed) *Contrastive Lexical Semantics* Amsterdam/Philadelphia, John Benjamins Volume 17 of series. urrent Issues in Linguistic Theory. Pages 1-24. Reprinted in Sinclair 2004.

Sinclair, J. 1999b "A way with common words" in Hasselgård, H. and S. Oksefjell (eds) *Out of Corpus* Studies in Honour of Stig Johansson; series Language and Computers: Studies in Practical Linguistics no 26 Amsterdam – Atlanta Rodopi. Pages 157 – 175.

Sinclair, J. 2001 "The Floating Dictionary" in Allén S., S. Berg, S. Malmgren, K. Norén and B. Ralph (eds) *Gäller Stam, Suffix och Ord* Meijerbergs Arkiv för Svensk Ordforskning 29; Göteborg, Meijerbergs Institut för Svensk Etymologisk Forskning; Göteborgs Universitet. Pages 393-422.

Sinclair, J. 2004 *Trust the text*. London, Routledge.

Tognini Bonelli, E. (2001) *Corpus Linguistics at Work* Amsterdam and Philadelphia; John Benjamins

Weinrich, H. 2000 "Von der Leiblichkeit der Sprache". Chapter 1 of *Sprache, das heisst Sprachen.* Tuebingen: Gunter Narr

Winter, E. 1977 "A clause-relational approach to English texts: a study of some predictive lexical items in written discourse" *Instructional Science* 6.1:1-92