

## Lexical signals of word relations

Antoinette Renouf  
University of Liverpool

### Editors' introduction

Renouf's contribution, like those of Scott and Berber Sardinha, uses Corpus Linguistics (CL) techniques, analysing large amounts of text by computer. The focus in her paper is on the identification of signals of semantic relations. Thus the frame "or, more exactly" might signal a general-particular relationship in "we stopped by the side of a lake, or, more exactly a loch". *Lake* is more general (a superordinate) than *loch* (a hyponym of *lake*). Renouf's endeavour was to trace signals like "more exactly" in large text databases, looking at and teasing out the meaning relations which crop up. Her work thus complements Jordan's analysis of Basis-Assessment in this volume, using quite different methods.

At the same time, she attempts to find out how good a match may be made between lexical signal and meaning - it is quite possible in language that one form might carry numerous functions (and conversely that the same meaning relation can be realised by numerous forms). This implies a need to pin down the meanings identified.

The context underlying this work is that of the influential work of Cruse (1986), which takes a non-CL view of meaning, heavily dependent on logic, and the notion of contextual normality by which Cruse means the test a near-native speaker or native speaker may make, as to whether a given string like "notable events such as a solar eclipse" seems normal. This is by no means a straightforward decision. Thus in

- (1) notable events such as meeting the President
- (2) notable events such as drinking tea

the degree of contextual normality would vary in rather unpredictable ways, depending on whether or not one regularly drank tea or worked in the President's office.

The approach Renouf takes is to examine large numbers of texts using a computer, to see instead what forms are actually attested. In so doing, she finds that Cruse's neat and logical patterns do not seem to be as neatly

reflected in the evidence of large numbers of examples from newspaper text. These two methods are not really a simple matter of pre-CL and post-CL, nor are they alternatives, in our view. For a start, the notion of "attested" examples in Linguistics is not new at all, and much early work before computers went into collecting slips of paper with heard or read examples of words. Perhaps the best known example is the many thousands of slips collected by innumerable contributors, for the construction of the Oxford English Dictionary. Second, we would argue that to use CL techniques does not constitute a new Linguistics, any more than using a spade constitutes "Spade Gardening". It is merely a matter of accessing resources. Third, it is not possible to use CL techniques without recourse to one's intuitions, e.g. as to what is contextually normal. A CL method is only able to identify positive hits: when one finds numerous examples of a given string of words, one may conclude that it must be contextually normal. If one finds one only, one does not know whether it is a joke or nonce construction or is just rather unusual; and if the string is not found even in a large database, that in itself does not guarantee that it would be contextually abnormal, since new contextually normal strings can be created at any time.<sup>2</sup> Renouf's paper thus shows that lexical semantics needs insights from logic and intuitions of contextual normality, together with CL methods enabling access to large numbers of examples.

## 1. Background

The role of words and phrases in the lexical signalling of word relations in text is an area which I turned to in 1991 in the search for a surface and thus automatable means of extracting sense-related word pairs from text. The automatic extraction of semantic pairs has two direct applications. One is in the field of text retrieval, where sense-related word pairs can function as alternative search terms; the other is in the field of linguistics, where they can inform a first description of the textual thesaurus.

The *ACRONYM* project was established in 1994-1997 to pursue this goal (Renouf 1996), and two strategies were adopted. The primary one was to use a collocational criterion for the identification of sense relatedness. Work on the earlier *AVIATOR* project (Renouf 1993) had shown that the collocational behaviour of a word is intrinsically related to (aspects of) its meaning, and in particular, that shifts in its collocational patterning can serve to identify a change in meaning. This encouraged me to believe that the phenomenon of

collocation could also be put to use in the identification of similarity in meaning; that two words (word-types) will be similar in meaning if they are found to occur to a significant degree in similar collocational environments. The matching methodology involves a statistical comparison of their individual "collocational profiles" (Collier and Pacey 1997), accumulated across many texts.

A second line of enquiry has been to investigate the extent to which individual cases of sense relationship are signalled by some kind of metalinguistic marker; that is, a word or grammatical string which links two words and signals a sense relationship between them. It is well-known that this type of correlation occurs to some degree; that a word like *especially*, or a lexico-grammatical string like *for example*, is used to introduce the superordinate-hyponym relation. In *ACRONYM*, I began to test the hypothesis that this metalinguistic signalling role was a systematic and all-pervasive feature of text, holding for all classes of sense relation; and further, that the association was sufficiently robust to be exploitable by automated means. Many candidate signals have since been tested in this light.

Within this second framework of investigation, I shall seek to answer two questions in this paper. The first concerns the closeness of match (or degree of "isomorphism"), if any, which is found between a sample selection of signals and the word pairs they link. The second concerns the nature of the sense relations, if any, which are signalled. I shall then summarise my findings about the role of lexical signals and word relations in text, with reference to their usefulness as aids in IT and language description.

## 2. Data

Two textual data bases were consulted at different points in this study: *The Independent* newspaper text from 1988-1998, numbering over 360 million words, from which I took a sample quarter year of text, of over eleven million words; and the *Financial Times* news text from 1988-1994, amounting to 198 million words. Although these obviously represent different textual domains, the *Financial Times* is unique among financial newspapers in its breadth of coverage, and in this study, the differences between them will not be the focus. Both corpora consist of unannotated text.

## 2.1 Lexico-Grammatical Signal: "Such as"

We begin by examining some potential signals of the relation of hyponymy. A promising sample set presents itself in the work of Hearst (1992). Coming from an NLP tradition, she takes an interest in the power of certain, what she terms "lexico-syntactic", patterns to identify the lexical relation of hyponymy in the electronic Grolier encyclopaedia text. She characterises these data as "unrestricted" (that is, unrestricted in comparison with Machine Readable Dictionaries), but they are nevertheless marked in style and content. The patterns she focusses on are shown in Table 1.

Table 1. Lexico-syntactic patterns of hyponymy (Hearst, 1992)

NP *such as* {NP, NP ..., ( and/or)} NP  
*such NP as* {NP,}\* { or/and)} NP  
 NP {,NP}\* {,I or other NP  
 NP {,NP}\*{,} and otherNP  
 NP {,} including {NP,}\* { or/and} NP  
 NP {,} especially {NP,}\*{ or/and} NP

Hearst adopts a simple method for the selection of these patterns, using intuition, together with observation of the contexts of certain term pairs for which the relation of hyponymy "is known to hold". She concludes that her chosen patterns "(almost) always indicate the relations of interest".

The patterns in Table 1 are indeed suggestive of hyponymy. Tested against our data, they are also found to map closely onto hyponymic pairings and groupings. However, there are some qualifications to make. I shall here restrict my observations to the first two patterns, referring to them for convenience as "NP *such as* NP" and "*such NP as* NP". Firstly, they do not match completely: to give just one example, the pattern "*such NP as* NP" allows 48 occurrences of:

### Example 1

there's no *such thing as* {a free mortgage/lunch, etc}.

Such oddities are sometimes amenable to correction: in this case, for instance, one could require the initial NP to be plural (although the singular NP, when uncountable, does contribute to this pattern, as in "*such investment/help as* NP"). Secondly, and more seriously, the realisations of the superordinate-hyponym relation which are automatically generated, whilst fascinating from a linguistic-descriptive point of view and useful for the purpose of "augment(ing)

and critiqu(ing) the structure of a large hand-built thesaurus" (Hearst 1992), are largely too unconventional and context -dependent for use in information retrieval, certainly without manual intervention. Hearst acknowledges this problem, citing the pairing "hyponym ('*steatomis*; '*species*')" as a case of "underspecification", and "hyponym ('*aircraft*; '*target*')" as "slightly askance of the norm". But we encounter a different order of unconventionality in our data. It is true that a single-word superordinate is likely to be under-specified; the top 10 superordinate noun heads for "*such NP as* NP" and "NP *such as* NP" are listed in Table 2 and the majority of their remaining superordinate NPs continue in a similarly abstract and general vein.

Table 2. 10 Most frequent noun heads for SUCH AS in *Independent* data

"SUCH NP AS NP"	"NP SUCH AS NP"
things	
companies	
films	
names	
countries	
people	
players	
issues	
classics	
areas	
events	
places	
phrases	
	countries
	people
	players
	issues
	items
	areas
	writers
	rivals

In fact, these general nouns are as often as not modified within the NP, though not necessarily very specifically, as shown in the selection below, for the superordinate NP *events*:

### Example 2

notable events *such as* a solar eclipse or the flooding of rivers  
 exciting events *such as* ekka-racing, kite-flying and cock-fighting

quirky events *such as* Mozart's murder by Beethoven during a game of leapfrog

innocuous events *such as* a full moon, Halloween, or Friday the 13th

with self-explanatory NPs such as:

### Example 3

international sporting events *such as* the 2006 football World Cup being the exception. Usually these nouns are only fully contextualised by the larger environment (which is even less manageable for an automated system).

One can count on one hand natural-world pairings of the kind:

**Example 4**

*such* toxins as arsenic and mercury

*such* diseases as bronchitis and emphysema

and even these are in fact incomplete sets, which are dependent on context for their interpretation, as can be seen in:

**Example 5**

"One of the NPIS's biggest recent investigations has been into the risk thrown up by the expanding 'alternative' medicines market. Many of these products include substances new to Europe and can contain such *toxins* as *arsenic* and *mercury*."

"Eric Clarke, MP for Midlothian, has pledged to make it easier for miners and their families to claim compensation for such *diseases* as *bronchitis* and *emphysema*."

where the contexts explain that the particular subset of the class of *toxins* are those which can be present in alternative medicines, and that the *diseases* are of the kind which miners contract.

But the real problem comes with the many quirky instances, exemplified here by:

**Example 6**

*such* howlers as the mispronunciation of "Missouri" and other American words

*such* blips as Blood Island

*such* endearments as Schwammi (little mushroom) and Mistblume (dung blossom)

*such* vehicles as shop catalogues

Our intention is not to make an exhaustive comparison with Hearst's findings, but to establish the degree of "isomorphism" between the patterns cited by Hearst and the superordinate-hyponym relation, and we agree that this holds, if in a somewhat idiosyncratic and indigestible fashion. The patternings do serve loosely as semantic markers of this relation.

In my previous investigation of such patternings, I was looking only for semantic relations, with the result that I only found those. However, in this current study, we are led to observe that such patterns serve a further signalling

purpose. The superordinate-hyponym relation might, all things being equal, be expected to be realised in the form of representative taxonomies of the superordinate NP, yet the lexico-grammatical signals cited in Table 1 are ushering in context-skewed and partial taxonomies. The logic of the selection has to be explained, and it is clear that the patterns are also serving a metalinguistic, discourse-organising function; the equivalent, perhaps, at word level, of the lexical signalling of logical relationships at clause level which has been described by Winter (1982) and Hoey (1983). Hearst's patterns play a focussing role, helping the reader by indicating that the hyponyms selected are to be understood as the particular members of the class of the superordinate which are relevant to and sufficient for the discourse purpose. Sometimes these partial "taxonomies" can be fairly full, but they are not signalled as being so, and thus exhaustive cataloguing is not generally the purpose here.

In the case of our *such as* signals, we further note that in both cases, there are in our data almost as many instances where a single NP hyponym occurs as there are of optional multiple NPs. In other words, the function of the *such as* patternings can often also be to single out one typical, or one specific, hyponym. In these many instances, of which we cite just three,

**Example 7**

"Here, like Anadyr, *there are more pressing matters*, such as *survival*. Any Chukchi born today is not likely to make it beyond four decades or so"

"Now it is the Left's turn and *former Communist functionaries* such as *Helena Brus*. The problem in this case is that the circumstances are more ambiguous. Mrs Brus is a Holocaust survivor who. . ."

"*Men* such as *Steve Knight*, whose style and youthful good looks make him widely tipped for stardom as the David Beckham of the new era. He is just 23. . ."

the single-word hyponymic NP is frequently taken up as the topic; thus the function of the *such as* pattern here seems to be to signal a new topic focus.

So our findings endorse those of Hearst in that these patterns are closely associated with hyponymic sense relations, but indicate that they are not all equally productive, and that there are variants on the patternings which also serve the purpose. As far as "isomorphism" goes, there is a very close match; the context-dependent nature of the semantic pairings is useful for manual augmentation of lexicons, but problematic for automated applications.

Some discourse-organising and discourse-structuring functions of the patternings have also been noted.

## 2.2 Lexico-Grammatical Signal: «More exactly»

Some lexico-grammatical signals have been identified by lexical semanticists as test-frames for isolating specific semantic relations. These test-frames are actually a routine element in non corpus-based semantic analysis, used as an "aide intuition", not because of the role they might actually play in naturally-occurring text. Cruse (1986), however, suggests that he comes close to our empirical approach when he talks of diagnostic frames which "collocate normally" and are «normally used». In the area of synonymy, he asserts that the frame *more exactly* "collocate(s) normally with pairs of lexical items whose semantic differences are relatively unimportant", citing the "dialectal synonyms":

### Example 8

"We stopped by the side of a lake - or, *more exactly*, a loch. . ."

He goes on to say that *more exactly* "is normally used to cancel a minor semantic trait and introduce a correction (between) two lexical items (which) should therefore be of the same level of specificity", citing non -cognitive synonyms:

### Example 9

"He was executed - or, *more exactly*, he was murdered."

So it was as a matter of interest that I decided to examine this framework. My own hypothesis was rather that this is a signal of the logical relation of general to particular, and thence of the sense relations of hyponymy (including taxonymy) or meronymy. For the sake of manageability, I took all nine instances of *more exactly* in 189 million words of Financial Times data, presented in Table 3.

In our data sample, the phrase *more exactly* is a signal of what it claims to be: of NP2 specifying more exactly what is meant by NP1. The question is whether this equates to the traditional relationship of synonymy. In fact, one would want to say that many word-pairs signalled here would pass as loose or marginal synonyms, but that they provide new insight into the nature of synonymy and invite finer sub-classification for its use in text. (1) evidences a lexical paraphrase with contrastive overtones. (2) represents partial synonymy, where *legitimate* has two semantic interpretations, one being highlighted here by the use of *lawful*. (3) is the rephrasing of a proper name as a specifically worded title, where only the determiner is changed in the phrase. (4) presents

Table 3. Contextualised instances of MORE EXACTLY in FT data

1. the so-called Big Seven of Spanish banking was reduced to the Big Five or, *more exactly*, to the Big Two, plus three others
2. swaps to cover actual borrowing are legitimate - or, *more exactly*, lawful
3. The Alpine Symphony, or *more exactly* "An Alpine Symphony"
4. Tuscany, or *more exactly* Medicean Florence
5. the fishermen of Wexford , or *more exactly* the about-to-be ex-fishermen of Wexford
6. Vivian Durante as the woman (and *more exactly*, as Woman)
7. what they believed was Nureyev , but which was *more exactly*, a shadow of the former star. . .
8. focusing on politics, or *more exactly*, on what the politicians... will do about interest rates
9. fable about colonialism: *more exactly* about Civilisation versus the Bush

*Tuscany* and *Medicean Florence* in a meronymous relationship, with an interesting additional element of time contrast, which is apparent elsewhere in our data, as in (5), which contrasts *fishermen* with the epithet *about-to-be-ex fishermen* along a time axis. (6) echoes (3) in presenting a grammatical shift of focus, but also exemplifies the way in which specification can be realised in movement not just from the general to the particular but from the particular to the generic. (7) echoes (5), in taking the time dimension in exploiting the contrast between *Nureyev* and the epithet which refers to the same real-world referent in a previous incarnation. (8) might loosely be described as a sub-class of hyponymy, where the topic or generic word *politics* includes the sub-topical phrase *what the politicians... will do about interest rates*. In (9), there appears to be a hyponymic phrase, in the form of an evaluative gloss on an aspect of the superordinate NP *colonialism*.

According to our data, then, Cruse is right in saying that *more exactly* can signal semantic relationships akin to synonymy, though not quite of the axiomatic variety that he might anticipate. But the match is not total; *more exactly* also signals kinds of hyponymy, antonymy (for contrast), meronymy, and a contrastive, time-sequencing relation, which can be realised in a variety of semantic relationships.

In fact, the instances of use shown in Table 3 incline me, even more than with the *such as* patterns, to see the lexico-grammatical signal in question rather as a signal of a functional relation, of specification or focussing, which cuts across established semantic classes. I observe also that it is, in contrast with *such as*, grammatically optional.

Similarly worded, and equally metalinguistic, signalling phrases such as *more precisely, to be {more} precise, to put it {more} simply, to put it another way* and *more accurately*, might be assumed to function similarly in text. Accordingly, I took the case of *more precisely*. The 22 examples in Table 4 are again all those which are found in the Financial Times 1988-1994.

Table 4. Contextualised instances of MORE PRECISELY in FT data

1. an opera, or *more precisely* a Singspiel
2. the democratic model or *more precisely*, the liberal democratic model
3. less government, and *more precisely*, less interventionist government
4. a single market or *more precisely* a single banking market
5. the machine - or *more precisely*... its central processor unit
6. The decade of the 1970's, or *more precisely* the period from 1973 to 1980.
7. animals or *more precisely*, those wonderful people who look after them when they fall ill
8. the accountancy profession, or *more precisely*, the Accountancy Standards Board
9. ADT, the vehicle auctions and electronic security group - or *more precisely* its chairman and president...
10. at the expense of business travellers or *more precisely*, their employers
11. it is accepted by the general public - or, *more precisely*, by the mass media
12. BTG's business is research, and *more precisely* the transfer of the fruits of research ... to industry and commerce
13. foreign policy issues, or *more precisely*... those issues which can most readily be translated into successes
14. a European standard... (or, *more precisely*, a scale of standards which embrace each country's own standards system)
15. due to the interest rate scenario - or, *more precisely*, the lack of it
16. Co-ordination within Europe Economic growth, or *more precisely* the lack of it
17. a balance (or *more precisely* a tiny deficit)
18. the work... delivers an impression of life, or *more precisely* imminent death, in the Warsaw Ghetto

Table 4 contains useful data from the linguistic point of view. The semantic relations signalled seem to be predominantly hyponymic (1-4); supplemented by meronymy (5-6); and reciprocity (7), in terms of "sick animals" and "animal carers". There is a kind of hyponymy, bordering on ranking, in (8-9), where the relation is between "professional group" and "a sub-group of representatives or controllers"; (10) and even (11) might also qualify as hyponymic, but without the overtone of ranking. There are more cases in (12-14) of the hyponymic evaluative paraphrasing of an aspect of the superordinate NP which was seen in Table 3.

There is also evidence (15-17) of an apparently paradoxical contrastive relation, whereby a common referent or idea is simultaneously referred to by a word and its opposite. This is expressed either paraphrasally:

**Example 10**

15. ... the interest rate scenario - or, *more precisely*, the lack of it

16. Co-ordination within Europe Economic growth, or *more precisely* the lack of it

or lexically:

**Example 11**

17. a balance (or *more precisely* a tiny deficit)

whilst (18) sets such a paradox in a time dimension, echoing examples (5) and (7) shown before in Table 3:

**Example 12**

18. ... an impression of life , or *more precisely* imminent death...

Our analysis of this phenomenon is that, in some cases, the native speaker is proposing the "positive" term in a semantic pairing; so *growth* will precede *the lack of it*; and, in others, making plain the general topic of what s/he is discoursing on, by fronting the topic or focal word that opens the particular lexico-semantic field. So *life* is the topic word for the field which includes *imminent death*. This analysis also accounts for most of the frameworks found earlier in Table 3.

Thus, the frameworks which I associated intuitively with hyponymy do signal that relationship, but more reliably in the case of Hearst's earlier set of signals: *such as, and other*, and so on. This is possibly because of their semantics and because they play a syntactic role of coordinating the word pair. Other frameworks, such as *more precisely*, play a more optional, interpolatory role, and seem not to signal semantics here so much as the discourse function of summarisation, modification, and specification. The apparently similar frameworks to these which I studied were also found to play a summarising role, to link longer stretches of text rather than individual words.

### 2.3 Lexico-Grammatical Signal: "Both X and Y"

I initially selected *both X and Y* as a potential signal of thesaural (synonymic) repetition for rhetorical purposes. The role of synonymy in text is well documented as being primarily one of discourse-organisation, where synonyms

more usually appear in successive clauses, for the purposes of reiteration or progression of the topic at hand, and so on. So my expectation was that this pattern, whilst more common in text than its predecessors, would yield relatively few instances, given the co-occurrence constraint.

The question concerning the extent to which this pattern is "isomorphic" with a particular sense relation is already answered by the small data extract in

Table 5, which shows that *both and* combines with more than one relation. The

question as to the precise nature of these sense relations is one that I should like now to pursue in somewhat more detail than for the previous signalling patterns, since *both and* is found to be a particularly good demonstrator of the functional modification that can be wrought on conventional sense relations.

For this pattern, all 832 instances of *Independent* data for the last quarter of 1998 have been retrieved. Here, X and Y each stand for a single word of any grammatical word class. We do not analyse fragments, i.e. where the first element only of a multi-word unit occupies the Y position. Of the X types resulting, we can see a ranked frequency list of the top 26 (non-hapax) in Table 5.

Table 5. Reverse-frequency ranked List of recurrent pairs for BOTH X AND Y

14 Both on and off	2 both women and men
7 Both men and women	2 both onshore and offshore
5 Both male and female	2 both on- and off-piste
5 Both left and right	2 both monetary and fiscal
5 Both club and country	2 both mentally and physically
4 Both physically and mentally	2 both males and females
4 Both inside and outside	2 both literally and metaphorically
4 Both good and bad	both for and against
4 Both black and white	2 both club and international
4 Both adults and children	2 both Scotland and Wales
3 Both compensatory and punitive	2 both Renault and Peugeot
3 Both boys and girls	2 both London and Manchester
3 Both before and after	2 both Exxon and Total
3 Both attack and defence	2 both 11- and 14-year-olds

### 2.3.1 BOTH X AND Y: Sense relation of antonymy or contrast

In our data, as shown in Table 5, it is clear that the particular recurrent word pairs do not generally represent synonyms, but what are conventionally called antonyms, if the conventional definition of the term is broadened to encompass

associated concepts such as *complementarity* (Lyons 1968), and to include all grammatical word classes.

In fact, *both X and Y* is a framework which Jones (1998) identified as a good signal of antonymy. Jones (2000), in a study of *Independent* news text, has selected a particular set of seed words, each from an established antonymic pair of the kind *good (bad)* which he calls "Umbrella" antonyms, and observed that they co-occur significantly with a number of lexico-grammatical patterns, of which *both and* is one.

Jones, in defining such uses as representative of the "Umbrella" type of antonymy, says that they "express inclusiveness or exhaustiveness" of scale; they "encompass all points on their given scale". In our data, it can indeed be seen that, in the context of *both and*, the "antonymic" pairs are not being used contrastively, but to express a kind of all-inclusiveness.

Our data contains a further 61 functionally non-contrastive, increasingly less conventional antonyms. Among the more straightforward are *town country* (which Cruse 1986, cites as "relatively weak opposites" because it is difficult to establish the relevant dimension on which they are opposed), and *twisted - straightened*. These last two are contextualised as follows:

**Example 13** "the Cosy Cuff, sheep skin on the inside, attractive trim on the outside, available in both *town* and *country* colours" "the hook, which is about one-eighth of an inch in diameter and is made of tempered steel, has been both *twisted* and *straightened*."

Given these more unorthodox antonymic pairings, it begins to be possible to see even wordpairs like *rapid - permanent* as having an element of antonymy too, as in:

#### **Example 14**

"The beneficial effects of this, he insists, were both *rapid* and *permanent*."

This pair could be seen simply as two collocating attributes of a shared NP, *beneficial effects*, serving to convey a positive evaluation. But the word *rapid* could also be understood to imply "immediate, though not necessarily long-lasting", and thus be contrastive with *permanent*.

Our growing awareness of the discursial function of *both and* as being to signal inclusiveness helps to explain the status of other unorthodox word pairs in Our data.

### 2.3.2 "Both X and y"; Sense *Relation of Inclusion or Inclusive Hyponymy*

There are five instances of a co-hyponymy relation shown in Table 6. Whilst not always being the sole pairings possible in their particular lexical area, they are presented by the signal as being inclusive in their coverage of a conceptual field in a given context.

Table 6. Inclusive Hyponymy signalled by BOTH X AND Y in *Independent data*

1. (she) must stand trial to determine whether she should pay both *compensatory* and *punitive* damages to the boy's family
2. he learnt at the coal face, both *literally* and *metaphorically*
3. It took a long time to get over the World Cup to be honest, both *mentally* and *physically*
4. The governments are...hedged in with both *monetary* and *fiscal rules*, which they accept as entry tickets to the world economic game.
5. We are very careful to ensure that soldiers are fully trained both *physically* and *mentally*

My interpretation of these contexts is that they indicate that there may well be further hyponyms of the superordinate NP, e.g. more than just *monetary* and *fiscal* for *rules*, (and certainly more hyponyms than just *physically* and *mentally* for "types of training for soldiers", such as *psychologically*), but the two hyponyms cited are to be understood as filling the relevant frame of reference.

### 2.3.3 "Both X and y"; Sense *Relation of Class Membership or Reduced Taxonymy*

The notion of "inclusiveness" which has just been identified in the kind of partial hyponymy we find signalled by *both X and Y* also applies to instances of the relation we call "reduced taxonomy". Here, two co-taxonyms linked by *both X and Y* represent the totality of their class, or at least the aspect of it which the writer wishes to focus on, as shown in Table 7. For instance, in the second example, Scotland and Wales are being cited as two devolved areas of the UK, which could not apply to England or Ireland. The wordpairs can also be seen to include proper-noun hyponyms.

### 2.3.4 "Both X and Y"; Sense *Relation of Mutual Entailment or Reciprocity*

A small group of pairings within the *both X and Y* framework are what we shall call "reciprocal" (or what Cruse 1986, has termed "relational converses"). These are shown in Table 8. These are word pairs which refer to two parties in more

Table 7. Reduced Taxonymy signalled by BOTH X AND Y in *Independent data*

1. It is indicative that the New Labour government has imposed new buildings on both *Scotland* and *Wales*
2. Labour's local by-election results in both *Scotland* and *Wales* (i.e. the two politically devolved areas of the UK) are worse than average
3. Bank shares (i.e. in France) fell sharply and both *Renault* and *Peugeot* were down sharply.
4. Nissan set up its own dealer network after sacking Octav Botnar's Nissan UK, and both *Renault* and *Peugeot* (i.e. the two French companies) own some of their UK dealerships.
5. Or Mash, in both *London* and *Manchester*, with its toasty Scotch Ale.
6. The Society of Black Lawyers, in its submissions to the Lawrence inquiry in both *London* and *Manchester* recently, advocated an overhaul for the whole criminal justice system.
7. However, shares in both *Exxon* and *Total*, the dominant partners in the respective mergers, fell on confirmation of the two deals
8. both *Exxon* and *Total*, the dominant partners in their respective mergers, may have fallen into just that trap
9. you should avoid sugar (both *white* and *brown*)
10. We conducted a blindfold test on carrots. Among our five testers, the organic carrot was unanimously the winner in terms of both *taste* and *texture*
11. the hottest are the small "bird" chillies, both *red* and *green*
12. In maths, results for both 11- and *14-year-olds* were depressed

Table 8. Reciprocal Pairs signalled by BOTH X AND Y in *Independent data*

bought - sold	client -lawyer
donors - recipients	employers _ employees
mentor - protege	owner - pet
patients - doctors	perpetrators _ victims
players - spectators	pursuer - pursued
readers - advertisers	rulers - ruled
sharer - receiver	speaker -listener
students - staff	teachers _ pupils

or less mutually defining social roles, where, in principle at least, the existence /of one entails the existence of the other. Thus, for there to be a *ruler*, there has to be someone who is *ruled*. Some word pairs are less 'exclusively' related in this way, as in the case of *mentor - protege* and *readers - advertisers* in Example 15:

**Example 15**

There must be a rapport in which both *mentor* and *protege* can bounce off each other. . . too many companies make the mistake of mismatching.

I fought... to make Punch a monthly... that would appeal to both *readers* and *advertisers*.

where one can envisage each word pairing with additional others, but nevertheless one feels that the pairings here do have something of the reciprocal about them. The situation is complicated by factors such as polysemy, and of Course in real life, the requirement for a reciprocal partner is not always met. For instance, one can be a "reader" without any advertisers being involved. Nevertheless, at least in the list of reciprocal pairings above, the two partners are in each case present and correct. This endorses the probability that the pairs are cited to encapsulate the totality of a particular domain.

2.3.5 "Both X and Y": Sense Relation of Individual to Institution or Meronymy  
 Meronymy is a term conventionally applied to the relationship of "part to whole", typically expressed by word-pairs referring to the parts and wholes of concrete objects, such as *toes - feet*. In our data for both Y and, however, we find only one such example, *sole - ball - heel*, in the context:

**Example 16**

Alexei Sayle, a man so impressively flat-footed that he seems unaware that the *sole* consists of both a *ball* and a *heel*...

Generally, the sort of meronymy we find is more abstract. It is primarily realised in terms of the individual to the institution, as in:

**Example 17**

buildings – people  
 club - player  
 firms - individuals  
 collective individual - team

individual- institutional  
 individuals - businesses  
 individuals - companies  
 player - club

Hearst actually refers to word pairs like these (her example is *group - member*) as "a subset" of the hyponymy relation. Out of context, they might equally be seen as contrastive rather than meronymous. In context, however, they are presented so as to indicate that, perhaps unexpectedly, they are to be regarded as the two parts making up a whole. This is shown in Table 9.

**Table 9.** Meronymy signalled by BOTH X AND Y in *Independent data*

1. Or Kale...where antiquities outnumber both *buildings* and *people*
2. Liverpool maintain that talks are still on going, but both *club* and *player* deny reports that he wants pounds 50,000
3. the FSA has the power to impose unlimited fines on both *firms* and *individuals*
4. Compensation cases – both *individual* and *team* titles now within their sights
7. The aim of Future Forests is for both *individuals* and *businesses* to reach a point at which they are "carbon-neutral"
8. the Japanese, both *individuals* and *companies*, are well aware of the need for change.
9. Vickery's condition, not yet career-threatening but of profound concern to both *player* and *club*...

2.3.6 "Both X and Y": Sense Relations of Scale and Rank

There are also in our "both X and Y" data a few instances of a kind of scaling and ranking in which two words represent two levels on a hierarchy of size or seniority. As for other sense relationships, these pairs are intended to indicate the totality of the class for the writer's purposes. We combine the two relations of scale and rank because they are small and notionally related. Scale is the primary one, and examples of this are given below:

**Example 18**

so persuasive was his advocacy both *locally* and *nationally*...

Yes, there is the opportunity, but taking advantage cannot be achieved by top-down planning. It has to come from bottom-up enterprise. There is a role for government, though. Both *local* and *national*.

The lobby, on 24 November, will involve both *national* and *regional* leasehold groups representing the majority of the UK's 2 million flat owners.

The ranking relation is traditionally associated with military and other hierarchies in public life. In our data, this relationship is realised rather in relation to levels of attainment, especially in the context of sport, as in *club - county level*. It has incidentally been noted that such usage occurs in multi-word NPs rather than in single word-pairs.

### 2.3.7 "Both X and Y": Sense Relation of Inclusiveness or Synonymy

Finally, the framework *both X and Y* does occur, though rarely, with loose synonymous pairs. However, the primary role of the *both X and Y* framework is interpreted in these cases to be a functional one, of indicating thoroughness, or exhaustiveness, of reference to the particular concept under discussion. This is achieved by exploiting the contrastive potential that is the property shared by all partial synonyms: the word *both* indicates that the synonymous pair are unexpectedly to be regarded as contrastive and thus encompassing the whole scope. This is demonstrated in Table 10.

Table 10. Loose synonymy signalled by BOTH X AND Y in *Independent data*

"Brown was both *cautious* and *prudent* in his June expenditure statement."  
 "these are the rights that have subsequently been both *emphasised* and *strengthened*." "The hinds are both *hunted* and *stalked* within the legal season..."  
 "detectives argue that he has no history of mental illness, and is both *intelligent* and *cunning*."  
 "It is both *logical* and *sensible* that the appeal be put back to a later date.

The framework in some cases also exploits the emphatic potential of the word *both* for rhetorical purposes. In this capacity, it joins a group of signalling devices, intensifiers such as *very*, *extremely*, *tremendously*, which I have observed to occur in the pattern "very X and very Y", ego where they create their rhetorical effect through grammatical and lexical parallelism, sometimes in conjunction with *both and*.

Summing up, I judge the *both and* signalling framework to join the previous frameworks in being a discourse organiser as well as, or even prior to, being a signal of sense relation. The word pairs signalled seem to represent a discourse relationship of inclusion or inclusivity, characterising and encompassing the full scope of the writer's concern at that point of the text.

## 3. Summary of findings

In this paper, we have examined the possibility that a certain lexical or lexico-grammatical item is in a close relationship with sense-related word pairs, such that an item like *in particular* functions as a reliable meta-linguistic signal of the nature of the sense relationship which the writer wishes to be understood to hold between the words in question. What we have in fact discovered is the following:

- I. Signals tend to be in more or less close but not "isomorphic" match with particular sense relations.
- II. There appears to be a closer relationship between a signal and a sense relation where the signal coordinates or connects the word pair syntactically, as in the case of *such as*;
- III. The lexical realisations of a signalled sense relation are generally rather unconventional and context-bound;
- IV. The sense related pairs found in text by this means are lexicologically interesting: including time sequence contrast, certain kinds of ranking, reciprocity, meronymy and "reduced taxonomy";
- V. The signals tailor the sense-related word pairs or groups to serve discourse functional requirements.

A basic duality in the role of these lexical signals, in signalling both semantics and discourse function, has been established. The intention is now to submit the whole area to more detailed study, among other things returning to signals previously analysed semantically, in order to uncover their discourse role. The outcome of such a study will hopefully be a set of classifications for use in corpus-based linguistics and in IT: of the nature of the semantic relationships associated with the individual lexico-grammatical signals, of the actual lexical realisations, of the lexico-grammatical signals themselves, and of the discourse functions imposed by these signals.

### Notes to Editors' Introduction

1. It seems that Cruse's more recent work also does not use computers for the purpose of analysing meaning. Thus Cruse (2000:43) discusses the "normality profile" of a linguistic item, that is "its pattern of normality and abnormality across the full range of descriptive contexts". What constitutes the full range is not made clear.
2. E.g. "notable events such as seeing Tony Blair drinking weedkiller", a string which has probably only just now come into existence.

### References

- Collier, A.J. and Pacey, M. 1997. "A Large-scale corpus system for identifying thesaural relations". In *Corpus-based Studies in English; Papers from the seventeenth International Conference on English Language Research on Computerized Corpora*, Stockholm, 15-19 May 1996, M. Ljung. (ed.), 87-100. Amsterdam: Rodopi.

- Cruse, D. A. 1986. *Lexical Semantics*. Cambridge: Cambridge University Press.
- Cruse, D.A. 2000. *Meaning in Language: An Introduction to Semantics and Pragmatics*. Oxford: Oxford University Press.
- Hearst, M. A. 1992. "Automatic acquisition of hyponyms from large text corpora". Proceedings of the ACL Conference, Nantes, France.
- Hoey, M. P. 1983. *On the Surface of Discourse*. London: George Allen and Unwin.
- Hunston, S. and Francis, G. 1999. *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins.
- Jones, S. 1998. "Approaching antonymy afresh". *Liverpool Working Papers in Applied Linguistics*, Vol. 4, No. 1. Department of English Language and Literature, University of Liverpool: 71-85.
- Jones, S. 2000. Investigating Antonymy in Text. Unpublished PhD thesis: University of Liverpool.
- Lyons, J. 1968. *Introduction to Theoretical Linguistics*. Cambridge: Cambridge University Press.
- Renouf, A.J. (with J McH. Sinclair) 1991. "Collocational frameworks in English". In *English Corpus Linguistics: Studies in Honour of Jan Svartvik*, K. Aijmer and B. Altenberg (eds), 128-143. Harlow: Longman.
- Renouf, A. J. 1993. "Making sense of text: automated approaches to meaning extraction". Proceedings of 17th International Online Information Meeting, 7-9 (December 1993): 77-86.
- Renouf, A.J. 1996. "The ACRONYM Project: Discovering the textual thesaurus". Papers from English Language Research on Computerized Corpora (ICAME 16), 1. Lancashire, C. Meyer and C. Percy (eds), 171-187. Amsterdam: Rodopi.
- Winter E.O. 1982. *Towards a Contextual Grammar of English*. London: George Allen and Unwin.