Corpus and Discourse

Corpus linguistics provides the methodology to extract meaning from texts. Taking as its starting point the fact that language is not a mirror of reality but lets us share what we know, believe and think about reality, it focuses on language as a social phenomenon, and makes visible the attitudes and beliefs expressed by the members of a discourse community.

Consisting of both spoken and written language, discourse always has historical, social, functional, and regional dimensions. Discourse can be monolingual or multilingual, interconnected by translations. Discourse is where language and social studies meet.

The *Corpus and Discourse* series consists of two strands. The first, *Research in Corpus and Discourse,* features innovative contributions to various aspects of corpus linguistics and a wide range of applications, from language technology via the teaching of a second language to a history of mentalities. The second strand, *Studies in Corpus and Discourse,* will be comprised of key texts bridging the gap between social studies and linguistics. Although equally academically rigorous, this strand will be aimed at a wider audience of academics and postgraduate students working in both disciplines.

Published and forthcoming titles in the series:

Studies in Corpus and Discourse
English Collocation Studies: The OSTI Report
John Sinclair, Susan Jones and Robert Daley
Edited by Ramesh Krishnamurthy, including a new interview with John Sinclair conducted by Wolfgang Teubert

Research in Corpus and Discourse
Meaningful Texts: The Extraction of Semantic Information from Monolingual and Multilingual Corpora
Edited by Geoff Barnbrook, Pernilla Danielsson and Michaela Mahlberg

Meaningful Texts

The Extraction of Semantic Information from Monolingual and Multilingual Corpora

Edited by Geoff Barnbrook, Pernilla Danielsson and Michaela Mahlberg

continuum
LONDON • NEW YORK

Continuum
The Tower Building
11 York Road
London SE1 7NX

15 East 26th Street
New York
NY 10010

First published 2005

www.continuumbooks.com

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

ISBN: 0-8264-7490-X (hardback)

Library of Congress Cataloguing-in-Publication Data
A catalogue record for this book is available from the Library of Congress

Typeset by RefineCatch Limited, Bungay, Suffolk
Printed and bound in Great Britain by
Cromwell Press Ltd, Trowbridge, Wilts

Contents

Part One: Monolingual Corpora

Part Two: Multilingual Corpora

9 Hidden culture: using the British National Corpus with language learners to investigate collocational behaviour, wordplay and culture-specific references

Dominic Stewart

Introduction

Preliminary remarks

In a previous article (Stewart 2000) I examined issues of conventionality and creativity in the area of Corpus Linguistics and Corpus Translation. Studies. Some of the topics discussed therein offer a useful preamble to the present work, and may be summarized as follows. It was suggested that arguably the most telling evidence electronic corpora have provided so far is that human beings are linguistic creatures of habit. While it is true that the importance of collocation and idiomaticity in language have frequently been emphasized in modern linguistics, it is only in recent times that corpora have become substantial enough to confirm the extraordinary pervasiveness of repeated patterns in language. Patrick Hanks (1996: 85) observes that 'the creative potential of language is undeniable, but the concordances to a corpus remind us forcibly that in most of our utterances we are creatures of habit, immensely predictable, rehearsing the same old platitudes and the same old clichés in almost everything we say'.

A number of translation scholars and corpus linguists, however, have assured us that, notwithstanding the prevalence of conventional patterns, this does not mean that creative flair and imaginative verve are completely swamped. As Dorothy Kenny notes (1998: 515), 'routine is not such a bad thing . . . It is what allows the creative use of language to be identified as such', while Mona Baker (1998: 483) offers the proviso that 'bien que les études basées sur le corpus s'intéressent d'abord aux régularités, elles ne s'intéressent pas moins à la créativité'. In the same vein John Sinclair (1996: 81) has underlined that the focus on recurring usage 'does not mean that unique, one-off events are necessarily ignored, but rather that they cannot be evaluated in the absence of an interpretative framework provided by the repeated events'.

It was further underlined (Stewart 2000) that the idea of identifying predictable language events in order to gain insights into the workings of unpredictable, imaginative usage lies at the heart of the notion of semantic prosody, defined by Bill Louw (1993: 157) as 'a consistent aura of meaning with which a form is imbued by its collocates'. In this connection it had previously been noted by Sinclair (1991) that almost all collocates of verbs such as *happen* and *set in* represent unpleasant things or events, a co-occurrence so powerful that any 'pleasant' collocates of these verbs are to be considered departures from recurrent patterns.

Wordplay in newspaper headlines

Departures from recurrent patterns are commonplace in newspaper headlines. The following headline from *The Times* newspaper introduces an article concerning the decision by the Ulster Unionist Council (the ruling body of the Ulster Unionist Party) to endorse George Mitchell's proposal for a power-sharing executive:

PEACE IS BREAKING OUT

The unusual co-occurrence of *peace* with the phrasal verb *break out*, which habitually collocates with unpleasant things and events such as *disease, riots* or *war*, not only attracts the reader's attention, but also serves to emphasize the difficulty in finding a peace agreement for Northern Ireland. Such collocational clashes, often adopted with ironic intent, are studied by Alan Partington (1995), who shows how linguistic creativity so often depends upon an upsetting of our collocational expectations. In a subsequent work, Partington (1998: 121–43) explores the way in which newspaper headlines not only play with semantic prosodies but also, and much more commonly, distort proverbs, quotations and idiomatic expressions, with journalists exploiting the 'framework of habit that collocation imposes on language' (ibid.: 121).

An interesting example of such phenomena is the following headline, particularly rich in wordplay:

THE MERCY BEAT FOR 'MAC DAD'

The article concerns a former policeman from Liverpool who at the time of writing was teaching computer literacy to Albanian children. There are fairly transparent references to the policeman's 'beat', to Mackintosh computers, and to the fact that the subject of the article is like a father ('Dad') to his pupils. Perhaps slightly less evident is the veiled allusion to the 'Mersey Beat', a term used to describe popular music in the Liverpool area (i.e. around the River Mersey) in the early 1960s, with 'mercy' (presumably to be interpreted as something to do with a mission of mercy) replacing 'Mersey'. This kind of substitution abounds in newspaper headlines, yet although it involves a capsizing of collocational expectations it is not so much an example of departure from semantic prosody as of a more general deviation from a norm. It is precisely this type of phenomenon that I intend to focus upon in the present paper.

Aims and methodology of the present work

As mentioned above, it has been claimed that corpora are particularly useful in providing a backdrop of repeated language events against which to identify and assess departures from conventional usage, providing insights which Louw (1993: 157) claims 'have been largely inaccessible to human intuition about language'. In the same way one wonders how accessible to intuition the 'Mercy Beat' type of newspaper headline might be, above all to non-native speaker intuition, and how effective or ineffective a corpus of English might be in helping to identify such deviations from conventional patterns. With the purpose of investigating precisely this question I decided to make it the theme of a final year module I taught recently in Linguistics at the School for Interpreters and Translators, University of Bologna. I selected around 60 headlines, with accompanying articles, from British and American newspapers and magazines, submitting these to groups of (Italian) students for analysis. The headlines were not chosen at random: I sifted out those which seemed to be of linguistic and cultural interest to the group. I devoted one lesson a week to this activity, and for each meeting two students were asked to prepare beforehand a number of headlines, usually four, and to submit their findings to the rest of the class.

Similar to the 'Mercy Beat' headline, all the headlines contained aspects of cultural and linguistic interest, usually involving variations upon idioms, culture-specific references, quotations, etc., the majority of which would be hidden to non-native speakers of English, even very proficient non-native speakers, without recourse to linguistic resources (though it should be emphasized from the outset that some of the more subtle and ingenious examples would also be hidden to the average native speaker of English). While a few of the headlines were fully understood by the students either with no resources at all or with the use of bilingual and monolingual dictionaries alone, other headlines remained quite beyond the scope of conventional learning aids. When this was the case, students were obligatorily required to consult for further information the British National Corpus (BNC),[1] which I had trained them to use and which is available on all the computers at our school. Naturally the students were free to consult other resources too, e.g. dictionaries of idioms, dictionaries of quotations and popular sayings, the World Wide Web, encyclopaedias, etc.

Types of deviation

Before reporting classroom feedback relating to the degree of usefulness of the BNC in attempts to 'solve' the kind of wordplay typical of newspaper headlines, it is perhaps worth pausing to examine some examples of the types of deviation featured (see also the classifications proposed by

Partington 1998: 125–8, and Moon 1998: 120–77). In each case the 'solution' is provided below the headline cited.

Inversion of key elements

NOT WITH A WHIMPER BUT A BANG

Original: 'This is the way the world ends
Not with a **bang** but a **whimper**'.

A line from T.S. Eliot's *The Hollow Men,* 1925.

Omission of original element

POLICEMAN'S LOT A HAPPY ONE

Original: 'When constabulary duty's to be done, A policeman's lot is **not** a happy one'. From the Gilbert & Sullivan opera *The Pirates of Penzance.*

Substitution of key element

SPIES IN THE WORKS

Original: the idiomatic expression *put/throw a* ***spanner*** *in the works.*

Insertion of new element (+ substitution)

A STAIRCASE TO INTERNET HEAVEN

Original: *Stairway to Heaven.* Title of a song by Led Zeppelin.

Orthographic alteration

DON'T LET THE BUGS BYTE

Original: 'Sleep tight, and don't let the bugs **bite**'. Line from a nursery rhyme.

Some headlines were slightly more complex:

HAYS' SHARES MAKE THE MOST OF THE SUNSHINE

Original: the proverbial expression *make hay while the sun shines.*

The article focuses on the fortunes of the financial group Hays.

REPAIRING JACK'S HOUSE

Original: *This is the House that Jack Built.* Title of a nursery rhyme.

The article discusses matters relating to the British Home Secretary Jack Straw.

A handful of the headlines examined were not deviations from a norm at all, appearing in their original, integral form, though in the following headline a bracketed exclamation is added:

LONG TO REIGN OVER US (SIGH!)

Original: 'Long to reign over us, God save our queen'. A line from the National Anthem.

Resources used

Obviously certain resources were more useful than others in individual cases, e.g. the *Cambridge International Dictionary of Idioms* for idiomatic expressions, the *Oxford Concise Dictionary of Quotations* for popular sayings. A pleasant surprise was that the *Cambridge International Dictionary of English*, a learner dictionary, was extremely helpful in tracing the source of not only idiomatic expressions but also popular sayings, e.g.

DON'T SHOOT THE PIANIST

In the entry for 'shoot' (and perhaps surprisingly not for 'pianist') is included:

> *Please don't shoot the pianist. He is doing his best* (Sign in a bar reported by Oscar Wilde in *Impressions of America*, Leadville, 1883) (*Cambridge International Dictionary of English*, p. 1319).

However, our main concern in the classroom was to test the usefulness of the BNC in revealing the original form of the type of headlines reproduced above. With this in mind, the following section reports student feedback, offering illustrations of how helpful or unhelpful the BNC proved to be in certain cases.

The BNC as an aid to comprehension

Cases where the BNC was unhelpful or not particularly helpful, with other resources proving more useful

PRIOR JOINS UP WITH THE ROYLE FAMILY

Original: *The Royle Family*. The title of a British TV soap.

The article in question discusses the transfer of the footballer Spencer Prior to Manchester City, whose manager is Joe Royle. Apart from the transparent punning on the Royal Family, there is also the said culture-specific reference to the hugely popular British TV series *The Royle Family*. The reason no trace is to be found of this show in the BNC is exclusively chronological: it first appeared on television in the mid/late 1990s and is therefore not captured by the BNC, whose most recent texts date back to 1994.

In the following case, on the other hand, it is regional factors which are decisive:

THE SECOND SHOE DROPS

Original: 'Wait for the other shoe to drop'. An idiom meaning 'wait for something bad to happen' (*Cambridge International Dictionary of Idioms*, p. 285).

This headline, introducing an article concerning political questions in Germany, represents a variation upon the entry in the *Cambridge International Dictionary of Idioms* inasmuch as it focuses upon the moment 'something bad' actually happens. The expression appears to be used more in American English, and has perhaps passed into popular British usage only recently. It is therefore no surprise that just one example is retrievable from the primarily British texts of the BNC. The query builder 'shoe|shoes # dropVERB', with a span of five, produced 18 concordance lines, of which the only relevant example was:

There's still another shoe to *drop* on whether they can survive the maelstrom

For the two headlines given above, the World Wide Web proved to be by far the most productive resource. The respective searches 'royle family' and 'shoe drops' produced a whole host of web pages with related titles, from which the 'puzzle' of the headline could be solved at once.

Cases where the BNC was just one of various resources able to reveal the hidden reference

VICAR WITH MOBILE PHONE DEFENDS BT IN THE BELFRY

Original: 'have bats in the belfry'. A dated idiomatic expression meaning 'be crazy'.

The article concerns the proposed installation of a British Telecom transmitter in a church belfry.

A BIT OF A CAMP SQUIB

Original: 'a damp squib'. An idiom used to describe something which is 'expected to be interesting, exciting or impressive, but fails to be any of these things' (*Collins Cobuild English Dictionary for Advanced Learners*, p. 1512). The review in question criticizes a disappointing musical featuring an abundance of camp characters.

In each of the two headlines above the words which turn out to be the unaltered nodes – *belfry* in the first case and *squib* in the second – do not belong to a high frequency lexical band, and do not appear in large numbers of idiomatic expressions. As a result the students were able to trace the original idiom in conventional resources (i.e. by looking up *belfry* and *squib* in a monolingual or bilingual dictionary or dictionary of idioms). In these cases the BNC was by no means crucial in identifying the source of the modified expression, though it was certainly useful for further exemplification and typical patterning. For instance, out of the 22 concordances featuring *damp squib*, seven contained the string *of a damp squib*:

But was it art, or just a bit *of a damp squib*? Video-taped reports
first week looks a bit *of a damp squib*, the full moon on
'Well, they're a bit *of a damp squib*!' We are really struggling
that it seemed more *of a damp squib* than a big band!
turned out to be something *of a damp squib* for the Slough, Berkshire
turns out to be something *of a damp squib*. I finished third in the
trading proved something *of a damp squib* as the stock added

Moreover a further five of the 22 concordances of *a damp squib* were immediately preceded by *like* or *as*, leaving the overriding impression that this expression prefers some sort of immediately preceding qualifier.

Cases where the BNC was crucial in revealing the hidden reference

Simple searches

SPECIAL QUEUE

Original: 'special brew'. A brand name of Carlsberg lager + title of 1980s pop song by Bad Manners.

The article reports that 50 people have applied for ten jobs as beer tasters at the Bass brewery in Staffordshire, England.

STARE CRAZY STARS SHOULD REFLECT ON FAME

Original: 'stir crazy'. An idiomatic expression meaning 'upset, angry and disappointed because you have been prevented from going somewhere or doing something for a long time' (*Cambridge International Dictionary of English*, p. 1427).

The article explains that the singer Madonna does not relish being stared at by flight attendants when she travels by aeroplane, and that as a result the cabin crew are under instructions to avoid eye contact with her.

CELL MATES

Original: 'soul mate'. An expression describing 'a person with whom one has a deep lasting friendship and understanding' (*Oxford Advanced Learner's Dictionary*, p. 1135).

The article concerns two men who fell in love after sharing a prison cell.

In each of the three cases above the unaltered nodes – *special, crazy* and *mates* respectively – belong to fairly or very high frequency lexical bands, and commonly occur with many different collocates. This renders the task of locating likely-looking 'original' collocates in the dictionary rather arduous. Under *mate*, for instance, in the *Cambridge International Dictionary of English*, the examples provided of collocates immediately preceding *mate* are *best, running, ship's, flat, team* and *work*, with no sign of *soul mate*. Looking up *special* in particular proved to be a futile enterprise, not only because it is an extremely common word with all manner of diverse collocates, but also because in the case in point the hidden expression is a brand name (*Special Brew*) and is therefore unlikely to feature in the dictionary.

The absence of tangible clues also made the concealed allusions impossible to find with search engines on the World Wide Web. Searches in the BNC, however, turned out to be more productive. A simple query for *special* retrieved 22,000 occurrences. These were too many to download in one go, but random searches of smaller numbers of concordance lines made visual scanning more manageable. After sorting alphabetically in descending order by the first word to the right of the node, and after exercising a good degree of patience, students finally hit upon an occurrence of *special brew*, which rhymed with *special queue* and seemed to match the context of breweries and beer-tasting. This was swiftly followed by a phrase query search 'special brew', of which there were nine occurrences:

A *Special Brew* was produced by Moor
loads of hippies going with cans of *Special Brew*, but that's not true,' he
a splashing can of Carlsberg *Special Brew* and asking me the time.
dogs on strings, and cans of *Special Brew*.'
launch of Gales Festival Mild, a *Special Brew* available for a limited period
violently sick. For the ten pints of *Special Brew* and vindaloo crowd only.
Centre all day with a can of *Special Brew* in your hand. Or perhaps
to women than the Carlsberg *Special Brew* brigade, something more
afterwards re-sold them as *Special Brew*. It was perhaps no wonder

Similarly the quick query 'crazy' produced 1760 concordance lines, which this time were sorted alphabetically to the left. After a number of wild-goose chases the students in question eventually hit upon two examples of *stir crazy*, for which there was no obvious connection with aeroplanes, Madonna or popular music, but whose phonetic similarity to *stare crazy* made it the likeliest candidate:

out of the house. You must be *stir crazy*.' She wandered over to the
some money out I'm gonna get *stir crazy* not being able to play badminton

The query 'mate' produced analogous results. There were 1877 concordances with *mate*, of which two featured a collocation with *soul* after sorting to the left. In this case the connection with *cell mate* was not only phonetic but also semantic, in that the two prisoners allegedly built up a very close relationship. Interestingly the students failed to notice that the *Oxford Advanced Learner's Dictionary* (p. 1135) is also of assistance, in that it provides a cross-reference to the expression *soul mate* under its entry for *mate* (i.e. 'See also SOUL MATE').

More complex searches

The BNC proved absolutely crucial when there was no key word with which to begin investigations, but only a structural pattern, for example:

OUT OF THE JUNGLE INTO THE POT

Original: 'out of the frying-pan into the fire'. An idiomatic expression meaning 'from a bad situation to one that is worse'.

The accompanying article deals with the question of meat for human consumption derived from wild animals, and in particular how some local economies in Central Africa rely for their income on the sale of bushmeat. Although one would certainly expect a native speaker of English to spot the disguised idiom, not one of my group of students was able to recognize it spontaneously. Nevertheless it was clear that some sort of wordplay was going on.

Unsurprisingly, dictionary and web searches for the keywords *jungle* and *pot* proved fruitless, as did those for the prepositions *out of* and *into,* both belonging to high frequency lexical bands. It was here, however, that the BNC came into its own, inasmuch as its 'query builder' option enables the user to look for key *patterns* rather than key terms. The following query builder,

('out of') ('the') (_) ('into') ('the')

with the so-called 'any node' in the middle representing any single word form, produced 71 concordances, of which three featured the hidden expression:

'So you're going *out of the frying-pan into the* fire?' Dr Abraham
perfect example of jumping *out of the frying-pan into the* fire. 'And I suppose
off.' That's it, then: *out of the frying-pan into the* fire; here's awful

Aside from the identical structural patterns, the semantic similarity between the *pot* of the headline and the *frying-pan* of the concordances also suggested that the correct solution had been found. A simple query 'frying-pan' then revealed further instances:

and said farewell. Out of the *frying-pan,* into the fire. Now all she
when they say, 'Out of the *frying-pan,* into the fire'? What do

In these two cases it would appear that the presence of the comma after *frying-pan* prevented the query builder from capturing them in the original search. Also worth noting in passing is that a greater number of relevant concordances would have been captured had it not been for the fact that *frying-pan* is often written as two words, i.e. *frying pan.*

The BNC: a backdrop of conventionality?

As mentioned in the Introduction above, corpora have been described as providing a backdrop of conventionality, of conventional language events, against which to measure creativity. Our investigations would suggest, however, that this is true only up to a point. Consider some further concordances generated by the query 'frying-pan':

jump from the frying-pan into the *frying-pan,* is there?' There you see
That would be jumping out of the *frying-pan* into a raging inferno.

Here the concordances in question already furnish evidence of creative variation upon an original form. Now while in the case in point the original form *out of the frying-pan into the fire* appeared in other concordance lines, and thus students were able to unearth the hidden phrase, in other cases the already modified usage in the BNC actually impeded the resolution of the problem. Consider the variation in the following headline, introducing an article from the beginning of 2000 concerning President Clinton's intention to step up his travelling plans:

COMING TO AN AIRSTRIP NEAR YOU

Original: 'Coming to a cinema near you'. A typical expression used to introduce trailers to forthcoming films.

Although the wordplay is perhaps fairly transparent – most of the students understood it more or less immediately – the source expression is actually quite difficult to trace in the BNC. The following query builder,
('coming to') ('a'|'an') (_) ('near you'),
produced 3 concordance lines:

> Variety Spectacular' is *coming to a college near you.* Forget your Roller
> by a band called Fuel – *coming to a shop near you* just about now
> is on tour and will be *coming to a town near you,* where you can

(There is in fact one example of the unmodified form in the BNC, though the concordance in question is interrupted after *near,* and is therefore not captured by the above search.)

Here of course it could be argued that the chances of finding a cinema trailer of this kind in the BNC are limited anyway. The following headline was more problematic in that its origin was not clear to the students at all, and once again, for the same reasons as above, the BNC did not prove helpful:

YOUNG, GIFTED AND BACKS

Original: 'Young, gifted and black'. A popular saying/The title of an album by Aretha Franklin/A film title.

Native speakers of English consulted were in agreement that this represented a variation upon *young, gifted and black*; *backs* is a technical term used in the game of rugby, to which the article in question refers. The phrase query 'young, gifted' produced two occurrences:

> *New Musical Express. Young, gifted* and plaque Fair and
> thing today. Er Let's have *young, gifted* and demanding it says.

It may be that in such cases the original saying, quotation or whatever has become engulfed by the deviation (there is currently a British television programme entitled 'Young, Gifted and Broke') and may in the course of time disappear altogether. This phenomenon also extends to common abbreviations such as *When in Rome* . . . in place of *When in Rome do as the Romans do* (explainable in terms of the Gricean maxim of quantity, i.e. 'do not make your contribution more informative than is required'), where the abbreviation would appear to be in the process of usurping the full form. What is particularly interesting is that if this phenomenon is proved to be widespread it could hold profound implications for language description, particularly lexicography. The author of the present article is currently conducting and supervising research in this area at the School for Translators and Interpreters, University of Bologna.

Conclusions

One of the principal objectives of the activities described above was to encourage use of the BNC as a source of both linguistic and cultural data, in that it contains a wealth of information which is beyond the remit of more conventional resources. Such resources, along with the World Wide Web, are already well exploited by students, but in my experience the extraordinary possibilities offered by electronic corpora still 'blush unseen', whether owing to lack of opportunity, to inadequate training or simply to natural reticence. However that may be, the study of newspaper headlines proved a good way of introducing learners to the idea of using corpora for problems of comprehension, in so far as headlines are particularly rich in wordplay, veiled references, departures from standard linguistic patterns, etc.

Further, the initiative was well-received in that the students were highly motivated to find a given solution. In my experience it can happen that students, even after receiving adequate training, remain somehow reluctant to exploit corpora, falling back a little too readily on more familiar resources. However, the fact of giving them a specific 'puzzle' to solve, combined with the fact that their usual resources were sometimes of little assistance, gave them a tangible, compelling reason to consult the corpus. Moreover, the fact that they were seeking not just linguistic but encyclopaedic information seemed to make the whole thing more challenging and stimulating.

Aside from the BNC's obvious merits as an important source of examples in context and of statistical, collocational, encyclopaedic, etc. information, in the investigations conducted the BNC tended to prove (i) extremely helpful in those instances where the unaltered constituent of the collocation sought belongs to a high frequency lexical band (*special queue*), and (ii) crucial where the *structural pattern* of the source expression had been preserved rather than the content words (*out of the jungle into the pot*). It was inevitably less useful in those cases where (iii) the source expression lies beyond the usual scope of the BNC (non-British usage; references to aspects and events subsequent to 1994), (iv) the source expression has apparently been ousted by variations upon it (*young, gifted and x*).

Notes

[1] The BNC is a 100-million-word general language monolingual corpus of contemporary, original (i.e. not translated) English, consisting of 90 per cent written texts and 10 per cent spoken. It was completed in 1994, and first released a year later, by an industrial/academic consortium led by Oxford University Press. It is encoded, and is a sample corpus, i.e. new texts are not added to it. See Aston and Burnard (1998: 28–40). Further information is available at the BNC website (http://info.ox.ac.uk/bnc).

References

Aston, Guy and Burnard, Lou (1998) *The BNC Handbook: Exploring the British National Corpus with Sara.* Edinburgh: Edinburgh University Press.

Baker, Mona (1998) 'Réexplorer la langue de la traduction: une approche par corpus', in Laviosa, Sara (1998), pp. 480–5.

Hanks, Patrick (1996) 'Contextual Dependency and Lexical Sets', *International Journal of Corpus Linguistics* 1(1): 75–98.

Kenny, Dorothy (1998) 'Creatures of Habit? What Translators Usually Do with Words', in Laviosa, Sara (1998), pp. 515–23.

Laviosa, Sara (ed.) (1998) *L'Approche Basée sur le Corpus/The Corpus-Based Approach.* Special edition of *Meta* 43: 4.

Louw, Bill (1993) 'Irony in the Text or Insincerity in the Writer? The Diagnostic Potential of Semantic Prosodies', in Baker, Mona, Francis, Gill and Tognini-Bonelli, Elena (eds) *Text and Technology: In Honour of John Sinclair.* Amsterdam/Philadelphia: John Benjamins, pp. 157–76.

Moon, Rosamund (1998) *Fixed Expressions and Idioms in English.* Oxford: Clarendon Press.

Partington, Alan (1995) 'Kicking the Habit: The Exploitation of Collocation in Literature and Humour', in Payne, J. (ed.) *Linguistic Approaches to Literature. English Language Research Journal.* Birmingham: University of Birmingham 17: 25–44.

Partington, Alan (1998) *Patterns and Meanings: Using Corpora for English Language Research and Teaching.* Amsterdam/Philadelphia: John Benjamins.

Sinclair, John McH. (1991) *Corpus Concordance Collocation.* Oxford: Oxford University Press.

Sinclair, John McH. (1996) 'The Search for Units of Meaning', *Textus* 9: 75–106.

Stewart, Dominic (2000) 'Conventionality, Creativity and Translated Text: the Implications of Electronic Corpora in Translation', in Olohan, Maeve (ed.) *Intercultural Faultlines. Research Models in Translation Studies 1: Textual and Cognitive Aspects.* Manchester: St Jerome Publishing, pp. 73–91.

Dictionaries consulted

Cambridge International Dictionary of English (1995). Cambridge: Cambridge University Press.

Oxford Advanced Learner's Dictionary (1995). Oxford: Oxford University Press.

Collins Cobuild English Dictionary for Advanced Learners (2001). Glasgow: HarperCollins.

Oxford Concise Dictionary of Quotations (1997). Oxford: Oxford University Press.

Cambridge International Dictionary of Idioms (1998). Cambridge: Cambridge University Press.