

## LEXICAL RESOLUTION

Antoinette Renouf, English Language Research, University of Birmingham, U.K.

The term 'resolution' which is used in the title of this paper is intended to draw an analogy between the fields of photography and image-processing, on the one hand, and the field of corpus linguistics, on the other. In photography, if a more detailed image is required, a higher resolution lens will supply it, provided that the original image is of sufficient depth and richness to withstand the greater magnification. In corpus linguistics, a similar situation obtains: in order to get a clearer or more differentiated picture of a word form, the data under scrutiny must be sufficient to be reliable. This paper will examine the way in which the character of a word form or lemma emerges more fully as the amount of corpus evidence available increases.

In any corpus of natural running text, approximately half the word types will occur once only, and the majority of the remainder will occur fewer than ten times. The 7.3 million word corpus of current English held at Birmingham, for example, which is part of the Birmingham Collection of English Text (Renouf, 1984), reveals the following (provisional) statistics:

no. of types	132,452
no. of types occurring once only	68,302
no. of types occurring 2- 9 times	42,255
no. of types occurring 10-19 times	8,151
no. of types occurring 20-29 times	3,450
no. of types occurring 30-39 times	1,839
no. of types occurring 40-49 times	1,251

So even in a relatively large corpus such as this, 5/6 of the types occur rarely. Or to put it another way, relatively little information on usage is yielded for about 5/6 of the types which make up the corpus.

For the purposes of grammatical research, with its broader focus, this phenomenon is less important. For lexicological purposes, however, it is crucial. The large-scale 'COBUIW' lexicographical research project at Birmingham, for example, has recognised the need for greater corpus evidence for many words, even those included in relatively small dictionary headword lists.

It is not the case, however, that all hapax legomena or other rare word forms in a text corpus are necessarily worthy of lexical analysis. Proper nouns can be one case in point. Errors, such as mis-spellings, are another. In the Birmingham Corpus, we had an early scanning problem with end-of-line hyphenation, so that a common collocate of FACTORY

appeared to be SATIS-. Other such oddities are introduced in the conversion of text to machine-readable form, but they are not of direct interest to the linguistic researcher.

But other word forms, which are not errors but which simply appear rarely in a smaller corpus, become linguistically more significant in a larger one. Words such as PANG, PAWN and PERTINENT, for instance, will carry a fuller range of uses. Referential words, such as OUTCROP, PEACOCK and CACTUS, will gain in linguistic interest and complexity in proportion to the amount of corpus data available.

In the case of items which, in traditional lexicographic terms, are regarded as inflexions and derived forms of 'bonafide' headwords, functioning solely as transformations of the base form according to their particular morphology, further instances of use will often reveal a character which is essentially distinct from that of the putative root word. The forms of a lemma occur in text with differing frequency, and the least frequent will not appear in a small corpus. But this does not mean that they do not have interesting and independent usage patterns.

A large text corpus, of say 20 million running words, will still yield little or no information on many word forms. However, it will provide more information, on a larger number of word forms, than can be found in a smaller corpus.

At Birmingham, we have had the opportunity of comparing the evidence offered by our smaller corpus, of 7.3 million words, with that of a larger one, of around 13 million words. These will be referred to in this paper as the 'Main Corpus' and the 'Reserve Corpus' respectively.

The patterns of word frequency in these two corpora do not entirely correspond. An extreme example of this is the case of MOORLAND, which occurs once in the Main Corpus, but 45 times in the Reserve Corpus. Such a discrepancy is often explicable in terms of the influence of a particular text, although not here. In most cases, however, the instances of occurrence of a word form in the larger corpus are found to be about three times those in the smaller one. This is rather more than expected, and a point which will be returned to at the conclusion of the paper.

Let us now look at the evidence offered by the two corpora for a range of words, and see the differing pictures which emerge, as the lexical resolution is enhanced. The citations are presented in a KWIC concordance format which has been truncated to fit the dimensions of this publication, so that less context is available to the reader than otherwise would be.

*Where no evidence exists in the Main Corpus*

Reserve Corpus

FADDISH  
un-clothes" of last year look faddish and freaky, and too much  
cessful that they cease to be faddish and become institutionalise  
vation which does not appear "faddish" because it is situated wi

FADDISM  
for the rampant intellectual faddism that already threatens to 0

FADDIST  
I think she should be a food faddist who drinks Tiger's Milk and

FADDISTS  
a household of food or bowel faddists, Mike would also have been  
our of advice from "experts", faddists and enthusiasts for honey

FADDY  
t is noisy, messy, and rather faddy about its food. A compounded  
ly if it helps to make people faddy, can actually do harm. Signs  
one else and not take up some faddy system that someone tells you  
e slight Above all, don't be faddy; there is no need. While so

These five word forms, which did not occur in 7.3 million words of text, are still relatively rare in the Reserve Corpus. But some modest observations can now be made on their use. The three instances of FADDISH are from two British English sources. In each case, FADDISH functions as a complement to a copula, and not as a modifier. The instances of FADDISM, FADDIST and FADDISTS operate as nouns, and are modified in all but one case. The exception is the last instance of FADDISTS, which is the only one in this group from a British English provenance. FADDY, in its turn, functions three times as a complement, and once as an attributive adjective; and all instances come from the same British English source.

Reserve Corpus

MOOT  
The court considers the point moot, and will reflect upon the mat  
erious he was about this is a moot point. On Saturday morning Pad  
six months. Although it was a moot point whether or not the Presi  
ices two-thirds bare. It is a moot point which issue most concern

mirable though they are, is a moot point. We had a most unnerving e was our centre, a crumbling moot tree; and few indeed of our 1

MOOTED

ial democratic party had been mooted. At the meeting they discuss the Severn Estuary was first mooted around the middle of the las edingly unlikely, it has been mooted by more than one worker in t Crab be, in Brittany, had been mooted earlier in the term, but obv h- potch of policies was being mooted in a desperate attempt to sa air environmental studies was mooted, we pointed out that there w nitions of planning are being mooted which could discard the emph

From the Reserve Corpus we learn that MOOT has no verbal function, and that it functions as a modifier in a complement, or as an object complement. The phrase A MOOT POINT occurs with some prominence.

In contrast, MOOTED is clearly verbal. It functions as the past participle in a passive construction. The evidence also shows that MOOTED means 'raised for discussion in a general and informal way'.

Reserve Corpus

OFF-KEY

t was my voice that was a bit off-key. If I want to be honest it fiddle played a mournful solo off-key. Yet the whole effect was r like this during a fiddler's off-key aria, and another to wake u, " while singing loudly in an off-key soprano. Now Mrs Cabel was

OFF-KEY occurs only in the larger corpus, where it is shown to be used in the semantic contexts of voice or musical instrument.

*Where little evidence exists in the Main Corpus*

ADVISEDLY

Main Corpus

s requested. I use the phrase advisedly, because it depends who i

Reserve Corpus

got to dance (I use the word advisedly) back to your general bas g. I say "frail Milanese" not advisedly but explicitly: having no y, where I'd deliberately and advisedly positioned myself as soon

Here, the Reserve Corpus endorses the indication in the Main Corpus that ADVISEDLY is used in an interpolated retort to the anticipated scepticism of the reader, as part of a rhetorical device to add weight to what is said. It occurs in metalanguage, on the 'interactional plane' (Sinclair, 1983), and the typical phraseology is probably something like:

'I USE/SAY the WORD/ PHRASE ADVISEDLY'.

ACCEDE

Main Corpus

with the law of contract. "To accede," he said, "to such a conten Mitterrand, if elected, would accede to Conununist Party demands fo ent that Spain was willing to accede to all American demands gran and upset by my declining to accede to his request He sent an e was his reluctant refusal to accede to the request of a group of

Reserve Corpus

he Book," the Yemenites would accede. The Hashed camp grew, and r the wisest course would be to accede gracefully. "Well, Mununy, I ed up mentally about sex. You accede or you don't, and you face t people felt it impossible to accede to the claims of this German from Miss Musson appeared to accede to this request; he went on: only to listen but finally to accede to their demands for a ban 0 will. The NF does not blindly accede to the imputation of meaning st consider how far he should accede to A's demands and thereby a y B in considering whether to accede to A's pressures. But there of Ignorance), then all will accede to two critical criteria. Th f the way she was prepared to accede to Machiavelli's wishes. Hav to induce the Florentines to accede to his demands, for he liste knowledge it is one thing. To accede to it, to shrug and clamber there may be trouble. If they accede to his demands, he will incr herself seemed glad enough to accede to this prohibition. The gir

The evidence from the Reserve Corpus brings into clearer focus the lexico-syntactic patterning of ACCEDE. Most typically, it is DEMANDS that one ACCEDEs to, whilst the collocational patterning to the left of ACCEDE centres round the semantic area of attitude, with WILLING TO or a modal verb figuring prominently. So we can say, on the basis of the accumulated evidence, that a typical pattern for ACCEDE is:

WILLING + TO + ACCEDE + TO + possessive pronoun + DEMANDS

and that the lexical items ACCEDE + TO + DEMANDS collocate in that particular grammatical relationship.

*Where the Reserve Corpus contributes to a more balanced view*

FAGGOT

Main Corpus

prick?" Oh, you shitty little faggot I'm going to get you right t

Reserve Corpus

bopped the sound man. Pompous faggot!" "I thought you wanted maxi  
e he'd see I was an old black faggot, and maybe he'd hate me for e  
calls himself. What kind of faggot name is that? But he's no fa "Go  
on, pi ss off, you misery faggot, or I'll turn you over to th agine! We  
even let that hairy faggot push us around. The Russians  
got name is that? But he's no faggot when it comes to weighing in

FAGGOT-WOOD

supplied, with no extra work, faggot-wood, used for heating bread

FAGGOTS

Main Corpus

beat them up for talking like faggots. Ab well, a mandarin's gott ace  
(because of "abortion and faggots and Mexicans"), and was con  
pect. Last August, a bunch of faggots staged a dance out by the c

Reserve Corpus

ions. And blood puddings. And faggots..... "Well," said *Mc* Endell  
ey could with stones, burning faggots of wood that sped like swar  
an, with possibly a bundle of faggots, so will you forgo being a

FAGGOTY

Reserve Corpus

te, to stage queer... a wacky faggoty voice full of innuendo: "Wa

In the case of FAGGOT above, the meaning and the pragmatic intent shown  
in the Main Corpus are also reflected in the Reserve Corpus. But the final  
instance in the larger corpus, ...but he's no faggot when it comes to  
weighing in..... introduces the additional meaning of 'weakness'.

In relation to FAGGOTS, the Reserve Corpus offsets the  
homogenous picture given by the Main Corpus by revealing the continued  
existence of other or earlier meanings of the word form.

*Where the Reserve Corpus highlights complexities of the language*

VAINLY

Main Corpus

I write that clinically, not vainly. I live in front of too many  
ronic independence, in which, vainly, I saw my own genes at work;  
ok off his glasses and looked vainly for something with which to  
style to which I have always vainly hoped to become accustomed.  
tion through groups of people vainly struggling to get clear, cha  
apanese - instead of striving vainly to beat them in industries i  
ob began to shout. Struggling vainly to understand what was being  
n inadequate pensions, trying vainly to catch up with the soaring  
d monetary commissions. Those vainly trying to keep a worn-out, I

Reserve Corpus

anity!"observed the holy man vainly. "Did he die in the church?"  
m which flamed on, endlessly, vainly. He spoke of this to no one,  
: "Having tried, honestly but vainly, all honourable means, a you after  
spending a lot of overs vainly attempting to find his touch  
rish landing at Gallipoli and vainly attempted to open the Dardan  
ceive the Count, let him hope vainly, building up dreams, which e s'  
cataclysmic downfall. They vainly defended themselves against  
rfor to find Fanny and Scylla vainly endeavouring to interest lit  
nt an awful long time waiting vainly for wickets to fall. I tried  
Praises on stones are titles vainly lent. A man's good name is h  
Shroud, which Fr Rinaldi had vainly made known to Cardinal Pelle

ller and Phil Edmonds toiling vainly on this ground six years ago or which the Labour Party had vainly pleaded during year after ye figures lurking in doorways...vainly searching for somewhere to s e Order of the Star, who also vainly stood their ground on the ba k, all he could do was to try vainly to get his thoughts into ord had an ejaculation, which he vainly tried to prevent. The dirtie possibility, while the rulers vainly try to keep things moving by cking on Mr Brooks's door and vainly trying to get a comment from hunched over their own laps, vainly trying to eat themselves. On ejudged nicely. Webster once vainly urged that the party should eory, it presumably was), and vainly urging the people, like bibl

In the Main Corpus, the adverbial VAINLY has the primary meaning of "in vain", or "unsuccessfully". This is clear from its association with the semantic areas of struggle and effort. The first two instances of VAINLY are less easy to categorise. There is some, albeit weak, contextual evidence to suggest that VAINLY there means "with vanity". In the first instance, VAINLY is set in contrast to "clinically", where both words can be interpreted as denoting attitudinal stances to the activity of writing which is mentioned; in the second instance, in so far as the limited context conveys any meaning at all, "with vanity" seems a marginally more likely interpretation of VAINLY.

The Reserve Corpus also shows "in vain" as the central meaning of VAINLY. But there are several concordance lines which are not readily interpretable. The Reserve Corpus evidence draws attention to the fact that the two principal meanings of VAINLY, while distinct, are not incompatible, and that several examples are ambiguous and may be deliberately so.

*Where the Reserve Corpus introduces the collocation of lemmas, and not merely of forms*

SORELY

Main Corpus

or Myra Breckinridge will be sorely disappointed. For lovers of Highness, who was said to be sorely missing his fiancée. The Pri of the NATO defences so long sorely needed was put in train at I Would Kuntajoin them? He was sorely tempted; this sounded even m earning such a lot that I was sorely tempted to stay for the rest

back at sea I shall miss you sorely..." She made an involuntary ir gardens. We shall miss her sorely. It has been decided by the iss Clare's tranquil presence sorely. Mrs Finch-Edwards could not April. We shall all miss her sorely. The vicar had already start haracters are four women, all sorely abused by life, all revealed the hectic helter-skelter of sorely afflicted commuters. Because d his wife, and he missed her sorely, and though he had no intent for Delhi, and they were both sorely in need of rest. And then, i , and I told him that we were sorely in need of good writers to g nd the like - which they miss sorely in the formal four lines for I am greatly relieved, he was sorely in need of companionship.' N es a sense of harmony that is sorely lacking. Such an approach is itten and spoken, which is so sorely lacking in this country scho methods of fighting Arabs were sorely lacking, and he felt Malcolm ear waste products, have been sorely led astray by the so-called of London that cannot but be sorely missed. If you start at St M ly accurate, this word too is sorely misused: it has the same Gre an say with vertainty that we sorely need legislation along the I kind of merchant venturer so sorely needed if Britain was ever t s over-confident little man a sorely needed increase in height. S at sense of humour, something sorely needed at Lord's in 1982 dur nagement. Such a programme is sorely needed. Professor Timothy O' irituallift at a time it was sorely needed. After the Jews had s nteers. The government, which sorely needed both the arms and the ounded leg would handicap him sorely on his long march home, woul coming interview, he was also sorely perplexed about a more perso though the moorland winds do sorely ruffle. Frederica crunched a n a crisis, but Churchill was sorely taxed by Beaverbrook's tantr breaks down. Maybe your son, sorely tempted, wonders whether you at these older people miss so sorely, the discussion of problems o have an ally to mollify the sorely tried Ousie Johanna, Francoi left alone. His patience was sorely tried. He fidgeted, he tosse n Different? Many parents are sorely troubled over the independen But one fact appears to have sorely troubled Margaret - the crue he cried himself to sleep, a sorely troubled Hintze from time to but unlike her he had been so sorely wounded that it was astonish where Francois had hidden the sorely wounded Xhabbo eighteen mont

From the instances of SORELY in the Main Corpus, we can already see that its meaning is clear, that this and its collocational patterning are essentially distinct from those of SORE, and that it is therefore arguably a different lemma; that it functions as an intensifier or adverb, preceded by a

form of BE, and followed by a participle, usually a past participle; and that it collocates twice with TEMPIED.

The Reserve Corpus, however, is sufficiently large to reveal the phenomenon of collocation by and between lemmas. It shows that the lemma SORELY collocates with more than one form of certain other lemmas. SORELY collocates with three forms of the lemma MISS: once with MISSING, 5 times with MISS, and once with MISSED. In each case, SORELY stands in the same abstract grammatical relationship with MISS (that of adjunct to the verb), although in several different syntactic realisations. Similarly, the larger corpus presents SORELY in collocation with various forms of the lemma NEED: once with NEED, 7 times with NEEDED (where NEEDED is a past participle, or a participial modifier) and, interestingly, 3 times with IN NEED OF. So this time, the Reserve Corpus demonstrates that the collocation of certain lemmas, here SORELY and NEED, can also survive a change in their grammatical relationship.

#### *Conclusion*

The question remains as to what the ideal size for a text corpus which is intended to serve lexicological needs might be. It is impossible to say. We have shown in this paper some of the comparative effects of interrogating a corpus of 7.3 million and then one of 13 million words. To the smaller data resource, the larger one variously brings evidence where there was none, or sharpens the focus where a vague picture has begun to emerge, or corrects an imbalance, or sometimes highlights an oddity where there appeared to be none. One must assume that a still larger corpus would continue this process of differentiation, of heightening the resolution for the word forms under scrutiny.

There are approximately 50,000 word types in the one-millionword LOB corpus, but only about 132,000 types in the 7.3-million-word Birmingham Main Corpus. The type count can be assumed to continue to decrease proportionately as the corpus grows, since we know that in 13 million words of text, types occur about three times as often as they do in a corpus approximately half the size. Eventually, a point may be reached in corpus development where all the word forms in which there is a lexicological interest are sufficiently exemplified. But we have yet to discover the precise statistical relationship between corpus size and the phenomenon of single occurrence. It seems unlikely, given the nature of text, that the number of hapax legomena, in a corpus of whatever type or magnitude, will dwindle to zero, or even deviate significantly from existing norms.

#### *Acknowledgements*

The author would like to express thanks for the help and support of colleagues in the COBUILD team. The data resources which inform this paper have been funded variously by the University of Birmingham and by William Collins Sons and Co Ltd, publishers.

#### *References*

- Renouf, A.J. (1984), 'Corpus Development at Birmingham University', in *Corpus Linguistics: Recent Developments in the Use of Computer Corpora in English Language Research*, eds. Aarts, J. and W.Meijs. Rodopi, Amsterdam.
- Sinclair, J. McH. (1983) 'Planes of Discourse', in *The Two-fold Voice: Essays in Honour of Ramesh Mohan* (ed. S.N.A. Rizvil), Salzburg Studies in English Literature, University of Salzburg.