On "not going there" in the Bilingual Dictionary: the Case of Grammatical-Word Idioms

Liam Rodger
Chambers Harrap Publishers Ltd
7 Hopetoun Crescent
Edinburgh EH7 4AZ
SCOTLAND
lrodger@chambersharrap.co.uk

Abstract
This paper considers the characteristics of a subclass of idiomatic English - namely, phrases constructed from strings of grammatical words (not including phrasal verbs) - and some of the issues these features raise for bilingual lexicography. There is a general discussion of the common features of such items from a lexicographical point of view, with particular reference to the production of general-purpose paper bilingual dictionaries. It is concluded that the problems associated with such linguistic items are likely to persist, recent powerful advances in language processing for lexicography notwithstanding, though ad hoc solutions may be reached.

1. Introduction

The advent of what Robert Burchfield called, not so very long ago, "the new technology of the green screen" (1989)[1] has already radically transformed the work of the lexicographer. The advances thereby made possible in corpus-based lexicography are acknowledged by all, and their full impact on lexicographers' day-to-day work has yet to be worked out. Similarly, the World-Wide-Web is an almost instantly accessible resource of a richness undreamt of by earlier generations of harmless drudges. However, there are some areas of language which may prove resistant to the transforming effects of these revolutionary changes. In this paper I will examine examples of idiomatic language which are tricky to tackle using standard corpus-query software, and which may also prove elusive in Internet searches. I will discuss possible methods which might help to remedy these problems.

2. Grammatical-Word Idioms

When the issue of "idiom" in the bilingual dictionary is to be discussed, we may most immediately think of the colourful and even bizarre expressions which typically feature on the covers of or the publicity material for the latest "Dictionary of English Idioms" - things like *to call a spade a spade*, *to get off on the wrong foot*, or the old favourite *to be raining cats and dogs* (which one doubts anyone ever really says, though we assume every native speaker knows the expression). However, I would like here to consider a rather different area of English idiom, which is (to put it no more strongly) at least as important a part of current usage as the flying pigs, blue moons and supernumerary kitchen personnel of the familiar colourful English idioms. The area in question is that of idiomatic phrases which are

structured from "grammatical" (or "function") words (mainly pronouns, prepositions, articles and modal verbs) rather than the "lexical" words which most strike the eye in the canonical idioms (*ill winds, roads to hell, scalded cats* and so on).

A large, and largely familiar, category of this kind of idiom comes under the heading of "phrasal verbs". Many of these, especially of the more colloquial sort, are formed with semantically lightweight verbs such as *be* (e.g. *to be up to* something - such as *a challenge*, say, or in a different usage, *no good*), *get* (e.g. *to get it on with someone* - or, as we would say in Britain, *to get off with someone*), or *do* (e.g. *to do for someone* - which could mean ruin them, or, in more dated English, be employed to clean their house). From this point we could proceed in one direction along a cline, as the main verbs in the phrasal combinations become semantically weightier (e.g. *make* or *take, come* or *go*), and we would eventually reach the indeterminate borderland between phrasal verbs proper, and verbs plus prepositional or adverbial phrases[2]. But we can also head, as it were, in the opposite direction - that is to say, to expressions where the prepositional or adverbial phrases have an autonomous semantic value, and continue to carry meaning when detached from a verb and transformed into sometimes bizarre variations. Take, for example, the idiom *to be with it*. Unlike what comes after the verb in a phrasal verb like *to be up to*[3], *with it* can be used with other verbs while retaining the same meaning (*to look/feel/seem with it*), or be used adjectivally (*our far from with it friends, a more with it style of dress*).

Some idioms of this kind do admittedly have a strong whiff of the 1960s about them - *(to be) where it's at* is another good example (again relating, perhaps significantly, to what is regarded as fashionable or trendy)[4]. However, there are plenty more in current use, some of which have only recently acquired currency: *(to be) out of it* (any of various states from unconscious, exhausted or dazed to drunk or drugged); *(to be) up for it* (eager and willing to participate - often, but by no means always, in a sexual sense); *to be up oneself* (conceited).

This type of idiom presents challenges at various stages of the lexicographer's task - for example in locating a range of relevant data to aid in identifying meaning or to provide exemplification, or in deciding where and how to present the end product of our analysis in the dictionary - in my case, general-purpose paper bilingual dictionaries. In what follows, the principal context of the discussion will be that of the revision of such dictionary text, although the issues are also relevant to the creation of new text.

2.1 New Idioms

One desirable feature of a good general-purpose bilingual dictionary is generous coverage of current idiomatic language, which of course means more than just the easily listable idioms of the canonical type, while certainly including them. It follows that revising an existing dictionary text will (or should) involve the inclusion of new idiomatic language, as well as, for example, new terms relating to innovative technology such as the Internet, mobile telephony, or biotechnology, or the new slang terms which excite so much comment in the press whenever a revision of a major dictionary is published.

There will certainly be additions to the list of idioms of the canonical type - recent additions to the time-hallowed stock of idioms in British English include *to be (as) sick as a parrot*, or the rather more recent *the dog's bollocks*[5]. Some of these may actually be recent coinages, while others may have already existed in regional or similarly restricted usage, but

have fortuitously come to prominence through use in the media - in a politician's or sports person's soundbite, in a comedy catch phrase or in a tabloid headline. However, there is also likely to be plenty of idiomatic innovation going on in the non-canonical class of idiom. Whether, ultimately, we can do equal justice to this area of innovation in our revision is another question, but I will take it for granted that we would certainly want to do so.

2.2 Sources for New Idioms
Where, then, do we find examples of these "new" non-canonical idioms? In this respect, a resource such as the British National Corpus (BNC) has the obvious disadvantage that its contents start to age as soon as the corpus has been established.

Ongoing resources, such as reading programmes like Chambers "WordTrack", manage to avoid this disadvantage, and are a rich source of data on language change. Teams of readers systematically search a range of text sources such as newspapers, magazines and novels to locate words which are not already in the dictionary. Examples, with citations, are collated and provide a bank of data which can be drawn upon in revising monolingual or bilingual dictionary texts. However, the most eye-catching items, and those which can most easily be checked against the current edition of the reference text, are inevitably "lexical" words. Innovations in the area of idiom we are concerned with here is likely to go largely unnoticed, or may in fact fall outside the readers' brief.

It is more likely that individual editors will start to register repeated instances of a new idiom of this kind in their general reading, TV viewing, or cinema-going, or from coming across its use in conversation. One-off apercus are all very well, and many an item has doubtless entered a dictionary by this means, but pooling of editors' pet new idioms is a much more valuable resource, as comparing impressions with others is a good preliminary way of sorting out the more marginal or ephemeral from items worth following up.

In the remainder of this paper I will discuss some of the distinctive characteristics of this class of idiom (from a lexicographer's point of view), and look in detail at some examples to consider how well our current resources equip us to incorporate this kind of language innovation in our dictionaries.

3. Some Characteristics of "Grammatical-Word Idioms"
The idioms we are discussing, which we might call "grammatical-word idioms"[6], share certain properties which raise particular challenges for the business of making bilingual dictionaries, and I will now consider two of the more salient of these problematic characteristics – accessibility (or rather the lack of it) and the problem of presentation.

3.1 Accessibility
One obvious difficulty is how to get hold of examples of this kind of idiom. If a phrase is made up of anonymous grammatical words, a search string (whether querying a formally organized corpus, or entered in an Internet search engine) is likely to return a mass of irrelevant data. If we are editing under the pressures of time and resources common in commercial dictionary production, this "noisy" data may well be all but useless.

Let's take as an example the relatively recent idiom "let's not go there". I choose this in particular because I recall about five years ago debating whether to include this in a

particular dictionary, and wondering how likely it was to last. Well I know now, and I am constantly being surprised by the new instances of its use that are to be encountered in the most unlikely places. As far as I can tell, the phrase probably originated in something like Californian Valley-Girl-speak, but quickly went bicoastal, and has been popularised on this side of the Atlantic through TV shows like "Friends", so it's decidedly odd to hear it in the mouth of a besuited MP, or a po-faced business spokesperson[7].

If we want to find examples of this phrase in context, the first problem is deciding what to look for, as there are many different instantiations of the same basic idiom: *We don't want to go there, You just don't want to go there, Let's not go there, shall we?* or the flat imperative *Don't go there!*. A search of the BNC[8] draws a blank on both *don't go there* and *don't want to go there. Not go there* produced 11 solutions, of which only three were not obviously literal:
a) We shall not go there.
b) Rather not go there!
c) Bah, do not go there!

On further investigation it quickly became clear that a) and b) were literal also - b) does come from a spoken text, but the *there* in fact refers to Liverpool. The last example did look the most promising of all. It has the right rhetorical ring - both contemptuous and peremptory - but turns out to be from a novel, and is spoken by a Frenchwoman fluent in the dialect of English spoken by Peter Seller's Inspector Clouseau:

"You seek lunch?" she said in English. "Bah, do not go there! The cooking, it is execrable".

We will obviously be more likely to find relevant examples in an Internet search, but a search for a fixed string using, say, Google, will throw up much else besides examples of the idiom desired - the humdrum components are likely to have many plausible literal uses: "Mario's? Their pizzas are stodgy and overpriced - you don't want to go there", for example. Similar problems can arise with searches for more canonical idioms, especially as punctuation and capitalization is generally ignored by Internet search engines like Google - "It was raining. Cats and dogs were roaming the streets" is a conceivable, perhaps even plausible find, but such cases would be relatively infrequent compared to the "noise" encountered in searching for grammatical-word idioms.

In context, of course, there may well be phonological cues, as many innovative usages of this kind originate in speech rather than writing. In the *go there* examples, there is likely to be a marked tonic stress, with falling intonation, on "go" rather than on "there" (the latter would tend to indicate a literal sense). Such information is not often readily accessible even in corpus data, however, let alone in cyberspace at large.

The example we have discussed at least allows us to grasp a range of possible instances for consideration, but if we are searching for cases of idiomatic uses of prepositional and adverbial phrases of the kind discussed above (e.g. *up himself/herself/themselves*, or - even worse! - *with it*), we are likely to be swamped by unmanageable amounts of largely irrelevant examples. The distorting effect of irrelevant data thrown up by Internet searches has been highlighted by Ross (2003), but here the problem is simply how to filter out enough material to provide a manageable set of data. Editors often develop a range of ad hoc strategies for coping with these problems when

searching for single lexical items - cases of homonymous proper names can be excluded by adding "the" to a search item; the query can be narrowed down by combining the search item proper with related words from the field desired (e.g. a search for *shipping* might include *invoice* and/or *delivery* when one sense is desired, or *tides* and/or *channel* when another sense is being targeted). Such techniques are far from infallible, but they can serve to quickly cut down options to a number it is feasible to examine while actively editing dictionary text. It is more difficult to find similarly quick-and-dirty fixes for grammatical-word idioms, not least because they are less likely to be restricted to specific contexts than straightforward (non-idiom) lexical items. One strategy worth exploring is that of generating plausible contextualised examples to test hypotheses about usage. An example follows which examines a possibly innovative use of the verb *do* - too short to be considered an idiom in itself, but presenting the same sort of problems of data-overload.

3.1.2 On "Doing God" (or not). Last year there was considerable comment in the British media about an incident which occurred during an interview given by Tony Blair to David Margolick, a reporter for the US magazine Vanity Fair. In the course of the interview the reporter asked Mr Blair about his Christian faith. As he was on the point of replying, he was interrupted by Alistair Campbell (an official whose position was variously described, but most frequently as "the Prime Minister's chief spin-doctor"), who said "I'm sorry, we don't do God." The Daily Telegraph felt strongly enough to devote an editorial to this remarkable interjection on Monday 5th May 2003, and the phrase has already become proverbial.

If we turn to our standard reference sources to see how they treat this use of *do*, we will not come away entirely satisfied, I think. To take an example where a fairly creditable job is done, there are 32 different uses or senses of the verb "do" in the third (2001) edition of the Collins COBUILD dictionary, and one comes pretty close to what we're after:

(2), [16] You can use do to say you are able or unable to behave in a particular way
□ *'Can't you be nicer to your sister?' - 'Nice? I don't do nice.'*

However, this use is listed as being a "V adj" construction, not "V noun". Furthermore, my feeling is that ability or inability is not the crux of the matter semantically, at least not in the sense used by Mr Campbell[9]. I feel the idea here is more "have in one's repertoire".

How do we seek more data to test such an intuition? Clearly, life (or at least one's current dictionary project) is simply not long enough to search through the 295,000 or so examples produced by a Google search for "we don't do". An alternative is to generate more limiting search strings by adding plausible nouns to *don't/doesn't do*. The following are some possible candidates (with hits for the *don't do* + [noun] given in brackets): sympathy (42), irony (227[10]), holidays (134) Mondays (74), apologies (34), goodbyes (21). How representative these examples are of the use of "do" we are trying to probe here is an open question, but at least by this method we can have a close enough look at citations to note, for example, that some writers place quotation marks round *do* in these collocations - perhaps indicating a feeling that this is a non-standard or at least rather colloquial usage, and perhaps also prompted by the marked stress the verb would have when spoken in such contexts.

3.2 Presentation

Assuming we've found our data, after wading through a lot of irrelevant evidence, we then want to put something in the dictionary with a neat idiomatic translation. With canonical idioms a citation form usually suggests itself easily enough (though whether the citation form is easy to translate is another matter) and the content words involved serve as useful salient hooks - spades, pigs, cooks, and so on - when deciding which headword to enter the item under. When all the elements are grammatical, or at best a verb like "go" or "get" which will have a vast entry, it is less likely that there will be an obvious citation form, or that any constituent word will be particularly salient. Turning to our example *don't go there*, the infinitive ("to not go there" or "not to go there") is likely to seem opaque even to native speakers familiar with the idiom. *Go* is the stressed word here, but the entry for such a word will be massive in any dictionary of the type in which this idiom might be considered for inclusion. *There* is the best of a not very satisfactory choice.

The difficulties of where and how to present such language items are particularly acute in the paper dictionary. Though random access to a text on CD-ROM does not automatically solve these sorts of problems, it certainly has the potential for greatly reducing them. Consideration of the case of grammatical-word idioms would, I think, be a useful component in the development of search tools for CD-ROM versions of dictionaries.

4. Conclusion

Some might argue that the sorts of idioms identified here are marginal, and unlikely to last. I would agree that many current phrases of this kind are likely to be ephemeral, but that is no reason for using this as an excuse to sweep the problems I have been indicating under the carpet. The same grounds could be given for ignoring any new words - something none of the major commercial producers of dictionaries in Britain (both bilingual and monolingual) seem inclined to do. I would go further, and argue that this area of language is far from peripheral, and that it provides a notable illustration of one of the fundamental characteristics of human language - the use of finite means to create innumerable, open ended encodings of meaning. One of the distinguishing features of "grammatical" words, after all, is that they form small closed sets (very small indeed in the case of the articles or pronouns), as opposed to the vast and non-enumerable open sets of the "lexical" classes of words such as noun or verbs. However, the class of idioms formed from strings of grammatical words is definitely not limited, and may in fact be a distinctively creative area of contemporary speech - witness the language (an early dialect of psychobabble) spoken by the characters in Cyra McFadden's 1978 novel "The Serial" (discussed illuminatingly by David Lodge (1980)), where, for example, the empathetic utterances "I can get behind that" and "I know where you're coming from" are blended into "I can get behind where you're coming from" for satirical comic effect. The years since McFadden's novel first appeared have shown little sign of a halt to the creation of new cases of such idioms.

The past two EURALEX conferences (Stuttgart 2000 and Copenhagen 2002) have recorded remarkable advances in corpus processing technology, and the "pay-off" of this research for practical lexicography promises to be revolutionary in its impact not just on the practice of lexicography (Atkins et al. 2003), but on the very nature of the dictionaries we

produce (Atkins 2002, Varantola 2002). I very much share the anticipation many lexicographers feel for the fruition of such promising advances in actual dictionary-making practice, just as I am sure that, as our tools and resources become more powerful, there will still be a need for a "good old fashioned lexicography" (Rundell 2002) to interpret the data thus generated and present it as usefully as possible for the target user. Still, there remains a doubt whether the nature of the new tools will determine the questions we ask, and the areas of language we focus on in the future. I think something of this is inevitable with any major technological and methodological advance, but the instruments of analysis should not in the end entirely define the field of study - this is the error of Eddington's apocryphal ichthyologist (Eddington 1939) who said "what my net can't catch isn't fish". As lexicography reaps the benefits of what shows signs of being a second corpus revolution, I hope we will also be able to pay more attention to areas of language of the kind discussed here, however recalcitrant they may often prove to be in practice, and I hope we can continue to develop strategies to tackle some of the characteristic problems I have indicated.

Acknowledgements

Endnotes

1. Quoted in Ross (2003), who comments how dated this description of computing already sounds.
2. A borderline which many bilingual and learners' reference texts seem to take a perverse delight in obscuring.
3. At least in the "activity" sense; the "capacity" sense seems more flexible – *he seems up to it* is acceptable, but not **the more up to it students*.
4. Though one noticeable feature of current colloquial British English (again, perhaps significantly) is its recycling of 1960s slang - most noticeably in the many uses of *cool*.
5. This idiom, meaning "something absolutely marvellous", has become current enough to have formed the basis of a punning TV series title "The Dog's Balearics", a Channel 4 series on uninhibited holiday hi-jinks among clubbers on Ibiza.
6. I avoid the less cumbersome term "grammatical idioms", as this might suggest affinity with the "grammatical metaphor" discussed by Halliday [1985].
7. It has just recently made its way into the headlines (11 February 2004) following its repeated use by US Secretary of State Colin Powell at a House of Representatives committee hearing. "There" in this case referred to the topic of President Bush's record of service in the National Guard.
8. Admittedly the BNC is showing its age, and it is by definition biased towards British sources.
9. An earlier instance of this particular collocation (from a 2002 interview with Richard Cadena of Australia's The Skeptic magazine) gives an explicit gloss: "A lot of Skeptics don't 'do' God. They don't choose to talk about that" (http://skepdic.com/cadenainterview.html).
10. Predictably, this is reduced to about 100 if we exclude hits with "Americans" (though by no means all the others appear to be actually endorsing this cliche).

References

Atkins, B.T.S., 2002. 'Bilingual Dictionaries: Past, Present and Future', in Corréard 2002.

Atkins, Sue, Michael Rundell, and Hiroaki Sato, 2003. 'The Contribution of FrameNet to Practical Lexicography', International Journal of Lexicography, 16/3, September 2003.

Burchfield, Robert W., 1989. Unlocking the English Language. London, Faber and Faber.

Corréard, Marie-Hélène, 2002. Lexicography and Natural Language Processing. A Festschrift in Honour of B.T.S. Atkins. EURALEX.

Eddington, Arthur, 1939. The Philosophy of Physical Science. Cambridge, Cambridge University Press.

Halliday, M.A.K., 1985. An Introduction to Functional Grammar. London, Edward Arnold.

Lodge, David, 1980. 'Where It's At: California Language', in: The State of the Language. ed. Leonard Michaels and Christopher Ricks, Berkeley and Los Angeles, University of California Press.

McFadden, Cyra, 1978. The Serial: a year in the life of Marin County. London, Picador.

Ross, Nigel J., 2003. 'Ensnared by the web of words', English Today, 19/4, October 2003.

Rundell, Michael, 2002. 'Good Old-fashioned Lexicography: Human Judgement and the Limits of Automation', in Corréard 2002.

Varantola, Krista, 2002. 'Use and Usability of Dictionaries: Common Sense and Context Sensibility?', in Corréard 2002.

Dictionaries consulted

Collins COBUILD English Dictionary for Advanced Learners (3rd edition), 2001. Glasgow, HarperCollins Publishers.

Collins Spanish Dictionary (6th edition), 2000. Glasgow, HarperCollins Publishers.

Harrap Spanish Dictionary, 2003. Edinburgh, Chambers Harrap Publishers.

Larousse Gran Diccionario Español-Inglés, 1993. Paris, Larousse.

Oxford Spanish Dictionary (2nd edition), 1998. Oxford, Oxford University Press.

Simon & Schuster's International Spanish Dictionary (2nd Ed.), 1997. New York, Macmillan.