

[Special Lecture]

Corpora, Dictionaries and Pragmatics: Challenges and Opportunities in the Age of E-lexicography

Michael Rundell

Lexicography MasterClass and Macmillan Dictionaries*

1. Introduction

Traditionally, dictionaries deal with semantics, not with pragmatics. In reality, the situation is a little more complex because the boundaries between these two aspects of language are not clear-cut. In describing the lexicon of a language, dictionaries have to account for common usages which instantiate features such as irony (*we were paid the princely sum of \$3 an hour*), euphemism (*learning about the birds and bees*), contempt (*another example of the bean counters taking charge*), and numerous other speech acts. While instances like these could be seen as incidental to the dictionary's primary purpose, many pedagogical dictionaries demonstrate a more explicit commitment to explaining pragmatic features. But it needs to be stated at the outset that the coverage of pragmatics in dictionaries designed for language-learners is at best patchy and incomplete. De Cock, for example, notes the under-representation of pragmatic features in learner's dictionaries. She compares the often extensive treatment of "classical" idioms (such as *spill the beans* and *kick the bucket*)—despite their relatively low frequency in real communication—with the generally weak coverage of what she calls "pragmatic prefabs" (expressions such as *I mean*, *in a way*, and *a bit of a*), which are far more frequent, yet either poorly explained or missing altogether (DeCock 2002: 471-2). Despite a number of initiatives (about which more will be said later), no current dictionary has a systematic, theoretically coherent approach to dealing with pragmatics. (A caveat: I refer throughout only to *English* dictionaries, though most of what is said here applies more generally.)

This matters. Effective communication in a second language depends on more than a mastery of denotative meanings. Without a good grasp of those pragmatic conventions which are well-established in the target language, the learner is seriously dis-

* This paper is a version of a talk given at the 15th Annual Conference of the Pragmatics Society of Japan, in Osaka in December 2012. I would like to thank the President and Committee of the PSJ for kindly inviting me to speak at the conference, and for their generous hospitality during my stay.

advantaged. In receptive mode, s/he risks misunderstanding the intended message of another's utterance. (At the most basic level, a learner of English needs to be aware that the greeting "How are you?" is not an invitation to provide an update on the state of one's health.) In productive mode, the learner risks being misunderstood or—perhaps worse—being perceived as impolite, sarcastic, or aggressive. There is a danger, in other words, of what Thomas calls "pragmatic failure". As she explains, "while grammatical error may reveal a speaker to be a less than proficient language-user, pragmatic failure reflects badly on him/her as a person" (Thomas 1983: 97).

For anyone in the pragmatics research community, it will be obvious that language-learners need reference resources which provide comprehensive coverage of the pragmatic features of their target language. For those of us in the lexicographic community, it is equally clear that the (thus far) sporadic treatment of pragmatics in dictionaries is inadequate for language-learners' needs. Against this background, there is a compelling case for pedagogical dictionaries to pay more attention to this aspect of language. The requirement, crucially, is to integrate its treatment into the structure of the dictionary, rather than (as is currently the case) seeing it as an occasional add-on or marketing-led feature.

2. Two revolutions in lexicography

What, then, is to be done? Recent developments in language technology and digital media provide grounds for optimism. In the first place, the "corpus revolution" which began in the 1980s (e.g. Hanks 2012) has not only transformed the way dictionaries are made, but has also provided empirical support for a view of language in which the role of context and co-text is central. Conventional monolingual dictionaries have been based on the implicit assumption that every individual word conveys one or more discrete meanings which we can confidently "define" on the basis of necessary and sufficient conditions (e.g. Atkins and Rundell 2008: 414–416). This context-free approach works well enough for certain classes of word: for items like *hedgehog*, *privatize*, or *equilateral*, conventional defining techniques can provide a description which is adequate for most users' needs. But what corpus linguistics has made clear is that, for huge and important areas of the lexicon, this traditional paradigm falls dramatically short.

The development of large language corpora (e.g. Atkins and Rundell 2008: 53ff), coupled with powerful corpus-querying software (e.g. Kilgariff and Rundell 2002) has provided linguists with the tools—and the evidence—to validate and take forward the ideas of pre-corpus scholars such as Harold Palmer, A. S. Hornby, and above all J. R. Firth. The work of John Sinclair demonstrates the importance of *recurrence* as a design feature of language: that is, speakers' observable preference—despite the theoretically infinite number of formulations available for encoding a particular concept—for defaulting to a relatively small number of "ways of saying" (Sinclair 1991: 110). According

to this analysis, meanings are seen not as autonomous entities, but as the product of individual communicative events, whose recurrence allows us to make generalizations which can be recorded in dictionaries. And in this enterprise, context has a central role. As Hanks points out, “in the overwhelming majority of cases, a correct meaning can be assigned to a keyword on the basis of clues in its immediate environment” (Hanks 2013: 81)—its context, in other words. “Context” here is an umbrella term describing a word’s observable (and quantifiable) tendency to behave and combine in particular ways, and these include its syntactic, collocational, and colligational preferences.

Two examples will clarify the position. First, the verb *treat*, whose meaning is almost entirely context-dependent. Syntactically, three distinct patterns instantiate distinct meanings:

1. V+OBJ+PP/*with* (e.g. *we treated the patient with antibiotics*): this encodes a meaning connected with medical treatment
2. V+OBJ+PP/*to* (e.g. *the time when David came in with the film star Rita Hayworth, and treated her to a mink coat*): this indicates a very different meaning, where the “treater” (in this case, David) provides something valuable or pleasurable for the “treated” (Rita Hayworth)
3. V+OBJ+ADVERB (e.g. *she married an older man who treated her very badly*): here the focus is on the manner (whether positive or negative) in which one person behaves towards another.

But *collocation* is important too. The meaning conveyed by the first pattern shown above depends on the category of nouns filling the object slot after the preposition: you can treat a patient (or an illness) with *penicillin*, *chemotherapy* and the like, but if a doctor treats a patient with *kindness*, *respect*, or *contempt*, a different meaning emerges (similar to that invoked by the third pattern above). It would be difficult to argue, therefore, that the verb *treat* conveys any meaning *on its own*: rather, it has what Hanks refers to as “meaning potentials”, and these are activated through context (Hanks 2013: 73–75).

Our second example, the noun *bunch*, illustrates the importance not only of collocation, but of colligation too (Colligation, as defined by Hoey, refers to a word’s preferences for appearing in a particular form, a particular position within a sentence, or a particular place in a sequence of words: Hoey 2005: 42–44.) *Bunch* here functions as a quantifier (*a bunch of x*), and corpus data shows that the *x* slot is filled by three frequent categories, words such as:

1. flowers, grapes, keys, bananas
2. people, guys, kids, mates
3. losers, idiots, hypocrites, crooks, thugs

When used with the first group, *bunch* is neutral with regard to the speaker’s attitude, while with the third group, the speaker’s choice of *bunch* clearly signals a dis-

paraging or contemptuous attitude. The second group is more complicated: like the third, it refers to people (though using very general terms), but in this case there are colligational constraints. Here *bunch* tends to be premodified by an adjective, and the prepositional phrase is optional: *My colleagues are a friendly bunch (of guys); they are a very nice bunch of girls who will do well.*

Corpus linguistics, as the cases of *treat* and *bunch* demonstrate, leads us to an understanding of how meanings are created which is determined by the context of an individual communicative event. This suggests a degree of convergence between the (fairly recent) findings of corpus linguistics and more established ideas in the field of pragmatics. Leech and Weisser (2003: 138) explain that “the idea behind a speech act is that meaning can be explained in terms of action, rather than in terms of concepts like reference and truth conditions”. Equally, lexicographers and lexicologists have begun to question the relevance of “concepts like reference and truth conditions” to the description of meaning. Thus the boundaries between semantics and pragmatics no longer seem so clear. And if an adequate dictionary entry for a word requires a full account of its typical contexts, this must include information which is traditionally categorized as belonging to pragmatics—as shown in the example of *bunch* (above), where some categories of context encode a distinct speaker attitude.

If the “corpus revolution” has given us the data we need for an improved account of meaning, a second lexicographic revolution—currently unfolding—provides opportunities for optimizing our presentation of this data for the end-user. The migration of dictionaries from print to digital media started as long ago as the early 1990s, with dictionaries published on CD-ROMs and handheld devices. But the process has accelerated in recent years, with some publishers abandoning printed dictionaries altogether in order to focus on the possibilities which the digital medium offers.

Freed from the space constraints which have traditionally limited our language descriptions, contemporary dictionaries are now able to exploit the possibilities of multimedia and hyperlinking, enabling us (for instance) to link dictionary entries to relevant examples in a corpus.

3. Pragmatics in dictionaries: the story so far

With abundant language data at our disposal, and unlimited space to play with, lexicographers are now well placed to provide a more satisfactory, and more systematic, account of the pragmatic features of the words that dictionaries describe. Before speculating on possible methods for achieving this, it would be worth reviewing the various strategies currently employed in dictionaries for conveying information of a pragmatic kind. We can divide these into three main types: labels, definitions, and supplementary material.

Dictionaries use labels to indicate any deviation from the “unmarked” norm. Assuming that most common words in the lexicon are unmarked and can be found in texts

of any type, a label will be applied to a word whose distribution across text-types is more limited. Thus some words are labelled *formal* if they show a clear tendency to occur in formal texts, and similar approaches are used to handle items which are domain-specific, characteristic of a particular regional variety, or no longer current. There is no generally accepted inventory of dictionary labels—every dictionary has its own set—but many dictionaries include labels which tell us something about the attitude of the speaker or the likely effect on a hearer. Labels such as *pejorative*, *approving*, *offensive*, *euphemistic*, or *humorous* are common. These may apply to a word or one of its meanings (as in Figure 1) or they may be attached to a specific example of usage (as in Figure 2):

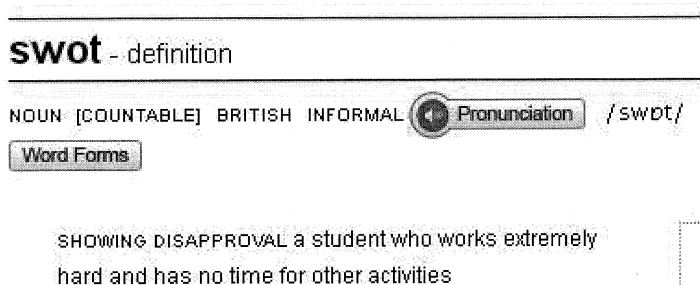


Figure 1: the label *SHOWING DISAPPROVAL*, used in the entry for *swot* (noun) in the *Macmillan English Dictionary*.

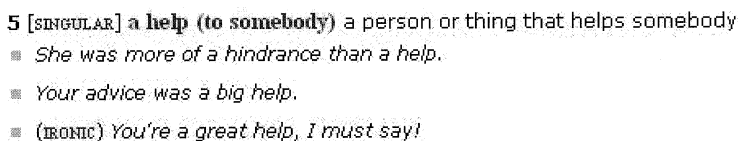


Figure 2: the *IRONIC* label, applied to an example sentence in the entry for *help* (noun) in the *Oxford Advanced Learner's Dictionary*.

The point of these labels is to indicate that there is more to the word than is signalled by the wording of the definition; in the case of *swot*, the definition alone could easily be interpreted as describing positive characteristics. Among the various learner's dictionaries developed in the UK, the COBUILD series (initiated by John Sinclair) have done more than most to take account of pragmatics, and many definitions are supplemented by a pragmatically-oriented label, such as “vagueness”, “politeness”, “emphasis” and “feelings”, as shown in Figure 3:

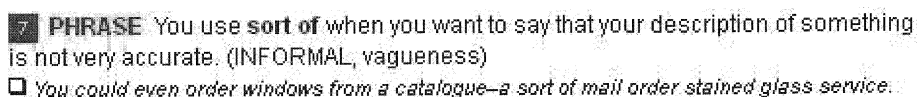


Figure 3: the label ‘vagueness’ applied to the entry for *sort of* in the COBUILD dictionary.

While conventional genus-and-differentia definitions convey nothing about the speaker's feelings or intentions, some dictionaries have developed strategies to extend the scope of what a definition can do. COBUILD's well-known "full-sentence definitions" sometimes employ what they refer to as a "displacement strategy" (Hanks 1987: 133–134) in order to indicate an extended use. Thus, while COBUILD's purely denotative definition for *wash* reads:

If you **wash** something, you clean it using water and soap...

the definition for *bourgeois* has a more complex structure, where the emphasis is on the speaker's motivation in selecting this word:

bourgeois If you describe people, their way of life, or their attitudes as **bourgeois**, you disapprove of them because you consider them typical of conventional middle-class people

The *Macmillan English Dictionary* has another approach, using a two-part definition structure: the first half is a straightforward explanation of the surface meaning, but a second sentence is used to indicate the speaker's attitude, as in this definition of *drama queen*:

someone who tends to treat situations as more serious or exciting than they really are. This word shows that you are annoyed by people like this.

The focus on function rather than denotation is also apparent in definitions which begin with formulae such as "used for showing..." or "used to indicate that...", as in this entry for the expression *forget it*:

- 1 used for telling someone that they should not worry about something because it is not important
'How much do I owe you?' 'Oh, forget it – it's nothing.'
- 1 Thesaurus entry for this meaning of forget
- 2 used for showing that you are annoyed because you think someone's comment or suggestion is completely unreasonable
'In the end I said to him, 'Look, forget it – I'm not paying you.'
- 'If you're just going to stand there and criticize, forget it.'

Figure 4: part of the entry for *forget it* in the *Macmillan English Dictionary*, showing a "functional" definition.

As well as using labels and definitions to convey pragmatic information, learner's dictionaries include a range of supplementary materials, such as usage notes and images at individual words, and dedicated sections dealing with speech acts, discourse conventions, and similar features. Thus, Figure 5 shows a usage note explaining the potential pitfalls of using the expression *of course*:

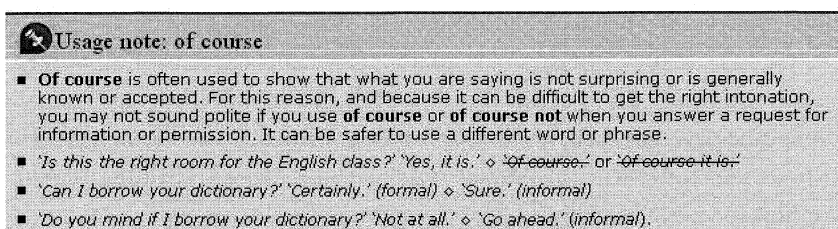


Figure 5: part of a usage note for the expression *of course* in the *Oxford Advanced Learner's Dictionary*.

Several dictionaries include separate sections (outside the A-Z text) which provide vocabulary choices for instantiating speech acts such as apologizing, giving your opinion, and making suggestions or requests. A central section in the print edition of the *Longman Dictionary of Contemporary English* has a four-page spread on signalling agreement or disagreement (with varying degrees of intensity), which includes the following information about using the expression *I know*:

used in spoken English when you have the same feeling or have had the same experience as someone: 'It's really hot today' 'I know—I wish I hadn't worn my sweater'

This phrase is commonly used in conversation, when sympathizing and agreeing with the other person.

The discourse conventions of academic writing are the focus of a section titled "Improve Your Writing Skills", currently in the centre of the print edition of the *Macmillan English Dictionary* and soon to be available online. This material deals with key rhetorical functions such as exemplifying, quoting from sources, and reformulating and paraphrasing, and it draws on data from learner corpora (Gilquin et al. 2007). It includes usage notes which explain common learner errors, as shown in Figure 6:

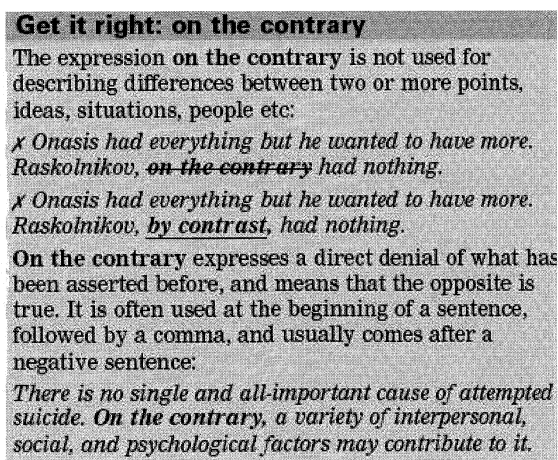


Figure 6: a usage note explaining the use of *on the contrary* in a section on "Comparing

and Contrasting” in the *Macmillan English Dictionary*.

These are welcome additions, no doubt, but what does it all amount to? While pragmatic features are not ignored in learner’s dictionaries, the approach tends to be *ad hoc* and sporadic, rather than systematic and comprehensive. Pragmatics often seems like an afterthought, and there is no sign yet of a truly coherent strategy.

As noted above, however, the conditions are now favourable for new ways of handling pragmatics which recognize its central role in communication. What is needed is, first, a realistic idea of what is possible (and what isn’t); second, a robust methodology for identifying features which need to be described; and finally, ideas about how to present pragmatic information to users in the most effective possible way.

4. Looking forward: some provisional proposals: (1) data analysis

The illocutionary force of many utterances depends on circumstances specific to a particular communicative event. Inevitably, therefore, a great deal of pragmatically interesting communication lies beyond the scope of even the most ambitious dictionary. Leech’s well-known example of a speaker asking someone to close a window or turn up the heating by saying “It’s cold in here” is the kind of indirect speech act which cannot be captured in a dictionary entry. It is possible, too, that many expressions of irony fall into the same category. In English, almost any utterance can be intended ironically, and in interpreting a speaker’s intended meaning, the listener has to take account of factors such as prosody and intonation, and what he or she knows about the speaker and the situation. When we look at corpus examples for a word such as *riveting*, we find cases where the adjective almost certainly *is* used ironically, such as this:

I excused myself from this riveting conversation and bolted for the bathroom.

(The broader context confirms this impression.) But it would only be legitimate to add this information at the dictionary entry for *riveting* if there was adequate evidence for the word being regularly used in an ironic way. (Some words do fulfill this criterion: corpus data shows that the expression *princely sum*—which literally means a great deal of money—has a strong preference for being used ironically: *for our two days’ work, we were paid the princely sum of \$12*.) This should be our starting point in determining which vocabulary items qualify for additional information describing pragmatic features: the same criteria apply as to any other information-type we describe in a dictionary entry. Whether we are recording information about word senses, register, syntactic or collocational behaviour, *or* about pragmatic usage, the corpus data must show that a given feature is both frequent and well-dispersed through a range of sources.

Before we proceed to outline some ideas for presenting pragmatic information in dictionaries, it will be helpful to summarize the current situation:

- most dictionaries do not explicitly describe the pragmatic features of a language,

and even those which do make some effort to cover pragmatics lack a systematic approach

- dictionary-makers now have the resources to do much better: abundant linguistic data in the form of large corpora; and the advantages that come with publishing in digital media: unlimited space, hyperlinking, sound, images and video, and so on.

There is, additionally, a great deal of corpus linguistic research which we can draw on. To give one example, Graeme Kennedy analyzed the use of “amplifiers” in the British National Corpus (words such as *utterly*, *completely*, and *totally*) and found that—so far from being interchangeable—each word has its own distinct characteristics and preferences, and thus has a particular illocutionary force. For example, “*perfectly* has exclusively positive associations”, whereas “*totally* tends to have mainly negative connotations (e.g. *unsuited*, *lacking*, *insane*)” (Kennedy 2003: 476). Material like this can guide our investigations and give a firmer shape to the messages we convey in the dictionary.

Fundamentally, however, our starting point for any dictionary entry is a detailed analysis of the corpus data. This process has two stages: first, we identify recurrent phenomena in the corpus (including those of a pragmatic nature), and then we record these in the database from which the dictionary entry is crafted. Any feature we identify should be accompanied by relevant contextual information. Thus in the case of *bunch* (above), when the reference is to humans rather than grapes or flowers, we noted that its disparaging use is signalled by one of a set of frequently-occurring *collocates*, while certain *colligational* features indicate a more neutral speaker attitude. All of this information is relevant to a dictionary description of the word, and the richer the underlying database, the better the eventual dictionary entry. This approach represents an established methodology for what we call the “analysis” stage of dictionary compilation (Atkins and Rundell 2008: 99–101), as distinct from the “synthesis” stage, where the information in the database is used for creating dictionary entries. So the only new requirements are to add database fields for optimally capturing pragmatically-interesting information, and to provide clear guidelines for the lexicographers tasked with finding and recording this type of data.

These goals suggest the need for a taxonomy of pragmatic types, such that information discovered in corpus analysis can be reliably assigned to a specific category. To do this properly, we would need to start by reviewing the relevant literature in pragmatics. This is not a trivial task, but in the meantime some provisional proposals will give an idea of how this might work. A possible approach is to tag corpus data according to two basic criteria: “function” and “mode”. “Function” refers to the speaker’s attitude or intention—their motivation for selecting a particular lexical item in order to encode a message. “Mode” indicates the lexical strategy employed in order to achieve the desired function. A few examples will clarify this.

Functions could include (among many others) the signalling of agreement or disagreement, irritation, disapproval or contempt, criticism, or scepticism. Mode would comprise features such as humour, emphasis, understatement, euphemism and exaggeration. Table 1 provides a few examples:

Table 1: some examples of a possible taxonomy of pragmatic functions and their lexical realizations

function	mode	examples
expressing disagreement	politeness	<i><u>I think you'll find</u> (that's my bag, etc); <u>I'm afraid</u> (that's nonsense); <u>with the greatest respect</u> (there is no evidence for this, etc)</i>
expressing criticism	archaism, humour	<i>his <u>amorous</u> advances; our <u>bibulous</u> vice-president</i>
	humour	<i><u>don't give up the day job</u>; who's been telling <u>porky-pies</u>?</i>
	understatement	<i><u>not exactly</u> (the best meal I ever had)</i>
expressing irritation	emphasis	<i><u>I can't begin to tell you</u> (how boring it was, etc)</i>
	irony	<i><u>thanks a bunch!</u>; he's a real <u>barrel of laughs</u>, isn't he?</i>
expressing uncertainty	hedging	<i><u>to the best of my recollection</u>; <u>as far as I know</u></i>

This is no more than a preliminary taxonomy, with a starter list of functions and modes. The goal would be to end up with a comprehensive but finite set of items in each of the first two columns which would, collectively, allow lexicographers to categorize most instances of pragmatically-interesting features attached to specific lexical items. One question to consider is whether such a taxonomy should include discourse functions: thus a function could be “introducing a summary” and the mode could be the use of discourse organizers (such as *In conclusion*, or *To summarize*, ...). A further addition might be a field for showing neutral alternatives to pragmatically-loaded ways of expressing an idea: the difference, for instance, between *peddle* ideas (negative) and *promote* ideas (neutral), or between *swallow* (negative) and *believe* or *accept* (neutral).

While this is all quite provisional, the point is to indicate the need for a set of clear categories for tagging individual words and phrases in the database, as a first step in describing their pragmatic features in a dictionary entry.

5. Looking forward: some provisional proposals: (2) presentation in the dictionary

Let us assume that a detailed corpus analysis process has provided our lexicographers with a rich database of information about the headwords which the dictionary will describe. The next requirement is a set of strategies for presenting this information to the dictionary user. (Different approaches might be used according to the pragmatic function being explained.) As discussed earlier, the digital medium offers exceptional opportunities. The absence of space constraints means that fuller explanations of meaning and function can be provided, while the risk of “information overload” can be averted through the use of on/off toggles.

With regard to definitions, the two-sentence approach (illustrated in section 3, above) would work well in this model: here, the first half of the definition is (more or less) neutrally denotative, and this would appear by default. Users would then have the option of activating the second half of the definition (by “toggling” it on), where the pragmatic force of the word is explained:

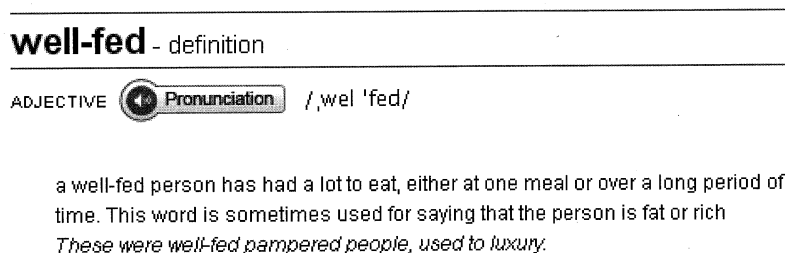


Figure 7: the entry for *well-fed* in the *Macmillan English Dictionary*, illustrating the two-sentence approach to defining.

There is also scope for providing multiple example sentences (here again, users could choose to show or suppress these). A link from the entry would give access to examples from the corpus which illustrate the full range of typical contexts. The same approach can be used for describing collocational preferences. We saw in section 2 that, in the case of the word *bunch*, the speaker’s attitude is encoded through its combination with any of a set of negatively-charged nouns (*idiots, losers, crooks* etc): the digital medium (and abundant language data) now make it feasible to provide lists of the most typical collocates. Multimedia features could also come into play: audio files can be used to indicate the (sometimes unpredictable) intonation patterns typical of pragmatically-loaded expressions such as *tell me about it!* or the sceptical *yeah, right!*; and images or video clips could illustrate features which are difficult to describe verbally.

Some concrete examples will help to show the potential. First, we will look at the expressions *I must say* and *I have to say*, which are used (typically in spoken discourse) to introduce an opinion about something. A good starting point will be the entries for

these items in two well-known learner's dictionaries:

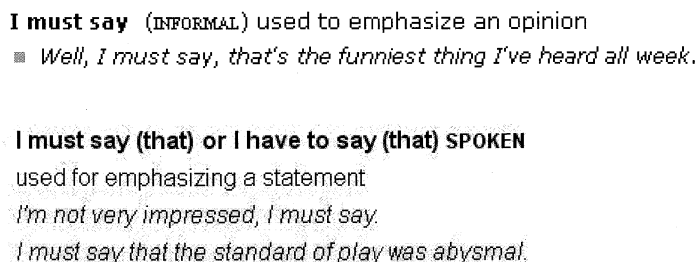


Figure 8: entries for *I must say/I have to say* in (first) the *Oxford Advanced Learner's Dictionary* and (second) the *Macmillan English Dictionary*.

Both entries use a label to indicate the text-type (“informal” or “spoken”) in which these phrases typically occur, and both definitions show that these items are used for adding emphasis. The second entry also provides *colligational* information: the two examples show (implicitly, at least) that these expressions can be either sentence-initial or sentence-final. But it should be clear that the current lexicographic record is underspecified in pragmatic terms. Entries like these are barely adequate as guides to “decoding” a text, and of very little use if the user wishes to “encode” them in his or her own text. (On this distinction between a dictionary’s “receptive” and “productive” functions, see Atkins and Rundell 2008: 407–411.) The corpus data broadly supports the idea—shown in these definitions—that *I must say* and *I have to say* are used for adding emphasis. It also confirms the (implicitly made) colligational point that these phrases have a marked preference for appearing at either the beginning or end of a sentence. What the data also shows—but the dictionaries fail to record—is that speakers usually select this device for one of two purposes: to express a negative opinion, or to concede, somewhat reluctantly, that something is better than they had expected.

To reflect what the linguistic data tells us, a productively-useful dictionary entry would make all this information explicit, and complement it by making available a user-specified number of authentic examples from the corpus. As a minimum, the entry might look something like this:

I must say OR I have to say

definition used for adding emphasis when giving your opinion about something

syntax followed by a *that*-clause (with optional *that*), unless at the end of the sentence

pragmatic features mainly used for expressing a negative opinion; sometimes used when admitting that something is better than you expected it to be

colligation usually used at the beginning or end of a sentence or clause

examples (first five: click for more)

I must say I doubt the value of employing these overpriced consultants.

One thing he does have is character, sadly missing in so many younger politicians, I must say.

'So you've come back,' he grunted in a tone that held little welcome. 'I must say you've taken your time about it.'

I must say I'm rather enjoying our little outing.

I must say I was quite impressed by his determination.

A second case is the use of the noun *shred* (in expressions such as “a shred of evidence”).

➤ a very small amount of something:

There's still a shred of hope that a peace agreement can be reached.

There isn't a shred of evidence to support her accusation.

2 [OFTEN SINGULAR] a very small amount of something

(not) a shred of something: *There's not a shred of evidence to support his claim.*

Figure 9: entries for the noun *shred* in (first) the *Cambridge Advanced Learner's Dictionary* and (second) the *Macmillan English Dictionary*.

Again, the current dictionary account works well enough for the user in decoding (or receptive) mode, but provides too little information to be a reliable guide to encoding (or language production). An improved entry would complement the denotative definition with the information that this device is used for adding emphasis and expressing the speaker's certainty. From a colligational point of view, it is important for users to know that (in this meaning) *shred* is almost always used in the singular, and almost always in a negative or “broad negative” formulation; these could be listed, the most common being: *not a shred*, *not one shred*, *not a single shred*, *hardly a shred*, and *without a shred*. The most frequent collocates of *shred* (the nouns following *of*) divide roughly into two sets, the most frequent members of which are:

evidence, truth, proof, credibility, justification

self-respect, dignity, decency, humanity

All of this information should be provided, backed up by corpus examples, e.g. five examples for each node + collocate pair, as here:

shred+evidence:

At no time was there a shred of evidence to link any Irish group to the incident.

'There's not one shred of evidence that these tests benefit human health', he added.

Both Blair and Bush have made allegations against Saddam, but these have not been supported by a single shred of solid evidence.

In short there is not a shred of objective evidence to support this hypothesis

It's probably just someone's biased opinion, without a shred of evidence to back

it up.

These are no more than initial suggestions, and they leave plenty of issues unresolved. The metalanguage of dictionary entries needs to be carefully chosen, and in the case of the first entry above, we may decide to refer to “speaker’s attitude or intentions” (rather than “pragmatic features”), or to substitute “context” for the less familiar “colligation”. The question of which information-types appear on the screen by default is another design decision, but ideally the user would be free open up or “collapse” the entry as s/he sees fit, so as to see as much or as little of this information as needed for a given look-up. There are further resources which could be used to enhance the dictionary’s coverage of pragmatically-interesting or productively-useful information. Data-visualization techniques could be used for presenting quantitative information. In the case of *shred*, for instance, a simple histogram would be a good way of showing the ratio between the frequency of negative formulations as against positive ones (well over 90% of uses are in negative expressions like *not a shred of*). Another possible approach is to link a dictionary entry to headwords which have similar characteristics. Assuming the kind of taxonomy described in Table 1 above, search mechanisms could be provided, enabling users to find all examples of a particular feature (such as euphemism or ways of expressing annoyance). Equally, hyperlinks could be made from dictionary entries to relevant material in external sources. Many dictionary publishers (including Oxford and Macmillan) have language blogs which regularly feature posts on specific pragmatic issues, while online forums (such as those at WordReference.com or the British Council’s “LearnEnglish” forum) include discussions on similar topics. Links could be set up to any material which would deepen the user’s understanding of how specific vocabulary items are used, and, more generally, of the key role of pragmatics in communication.

Not all of this is new. The novelty of these proposals lies in the goal of replacing the current approach to pragmatics in dictionaries (which is random and incomplete) with a data-driven, and productively-useful programme which makes systematic use of existing strategies, and exploits the opportunities of digital media to add new features and thus further enhance our coverage. Although we have the data we need to underpin such an enterprise, this is a far from trivial task, and will require significant editorial resources. But the reward will be a dictionary which takes pragmatics seriously—and that would be a first.

References

- Atkins, B. T. S. and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- De Cock, S. 2002. “Pragmatic Prefabs in Learners’ Dictionaries.” In A. Braasch and C. Povlsen (eds.) *Proceedings of the 10th EURALEX International Congress*, 471–481. Copenhagen: Center for Sprogteknologi.

- Gilquin, G., S. Granger and M. Paquot. 2007. "Learner Corpora: The Missing Link in EAP Pedagogy." *Journal of English for Academic Purposes* 6, 319-335.
- Hanks, P. W. 1987. "Definitions and Explanations." In J. Sinclair (ed.) *Looking Up: The COBUILD Project in Lexical Computing*, 116-136. London: Collins.
- Hanks, P. W. 2012. "The Corpus Revolution in Lexicography." *International Journal of Lexicography* 25:4, 398-436.
- Hanks, P. W. 2013. *Lexical Analysis: Norms and Exploitations*. Cambridge, Mass.: MIT Press.
- Hoey, M. 2005. *Lexical Priming: A New Theory of Words and Language*. London: Routledge.
- Kennedy, G. 2003. "Amplifier Collocations in the British National Corpus: Implications for English Language Teaching." *TESOL Quarterly* 37:3, 467-488.
- Kilgarriff, A. and M. Rundell. 2002. "Lexical Profiling Software and its Lexicographic Applications: A Case Study." In A. Braasch and C. Povlsen (eds.) *Proceedings of the Tenth Euralex Congress*, 807-818. University of Copenhagen, Denmark.
- Leech, G. and M. Weisser. 2003. "Pragmatics and Dialogue." In R. Mitkov (ed.) *The Oxford Handbook of Computational Linguistics*, 136-156. Oxford: Oxford University Press.
- Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Thomas, J. 1983. "Cross-cultural Pragmatic Failure." *Applied Linguistics* 4:2, 91-112.