Colin Yallop, Macquarie University

Grammatical Information in Dictionaries: How Categorical should it be?

Abstract

Dictionaries vary in their presentation of grammatical information but agree in assigning words to discrete grammatical categories, or at least in implying such assignment. But linguists like Halliday and Sinclair have argued that grammar is probabilistic rather than categorical. Corpus evidence shows that it is indeed more realistic to describe typical patterns than to insist on categorisation. Carefully chosen examples can signal typical usage without the need for rigid categorisation.

1. Introduction

This paper draws on research into grammar and lexis carried out at Macquarie University by Emeritus Professor Ruqaiya Hasan, Dr Carmel Cloran and myself, supported by the Australian Research Council and by the Macquarie University Research Grant scheme. The paper also makes use of data from the Macquarie Dictionary corpus of Australian English, known as Ozcorp, made available by Macquarie Library Pty Ltd, the publishers of the dictionary.

Most of the examples are verbs of saying: these verbs illustrate complex patterning of a kind that is elusive, to lexicographers and to users and learners of the language.

2. Current practice in presenting grammatical information

Dictionaries vary in their presentation of grammatical information but agree in assigning words to discrete grammatical categories, or at least in implying such assignment. The Macquarie, for example, is relatively conservative, confining itself to category labels such as *n.*, *v.i.* and *v.t.* Thus, *say* has entries not only as *v.i.* and *v.t.* but also as *n.*, because of uses like *it is now my say.*

Some dictionaries, especially those intended for advanced learners, give more elaborate grammatical information. LDOCE gives us *say* not just as a noun, but with the codes [S; U], telling us that the noun *say* is a special singular noun, uncountable but nevertheless useable in a phrase

such as *a say.* OALD signals similar details by somewhat different means, marking this nominal meaning of *say* as [sing, U]. COBUILD's "extra column" gives comparable information.

Collins Today's English Dictionary omits grammatical notes and labels. As Sinclair puts it in the preface, "we do not include any grammatical parts of speech, because our full-sentence definitions make it clear whether a word is a noun or a verb, an adjective or an adverb, and so on." And indeed, the definition "When you *say* something, you make a comment or statement or ask a question" clearly differentiates a verbal use from the nominal use of "If you *have your say*, you give your opinion on something."

3. Practice and theory

Despite differences in presentation, the dictionaries mentioned here all imply categorical classification: a word may be categorised as a verb, or as an intransitive verb, or as an intransitive verb and a noun, and so on; but there is no suggestion that a word might be almost always a noun but occasionally an intransitive verb, or usually a transitive verb and very rarely intransitive. The more elaborate labelling of some dictionaries, such as LDOCE and OALD, does offer refinement, but this finer categorisation rarely acknowledges frequency or probability. Even in Today's English Dictionary, where labelling is abandoned, the approach remains essentially categorical. The definitions imply that a word is a verb or a noun or an adjective.

Notwithstanding this lexicographical leaning towards discrete categories, linguists like Halliday, and to some extent Sinclair himself, have argued that grammar is probabilistic rather than categorical. Halliday makes the point that frequency of occurrence (as evident in a corpus) is the "instantiation" of probability in the grammar (1992, p.66). Sinclair talks of using corpus evidence to establish "norms and regularities" (1991, p.61) and notes that many observations about words are "probabilistic – they show strong tendencies in the behaviour of words rather than clear-cut alternatives" (p.78).

Of course, neither Halliday or Sinclair is arguing that grammars and dictionaries can, or even should, be immediately and easily rewritten in some non-categorical probabilistic way. (Halliday mentions in his 1992 paper the difficulty of incorporating probability into grammatical description, which he describes as a major theoretical problem.) But it should be clear that frequency or probability is a significant part of linguistic description.

4. Telling examples

If we take as illustration one of the most common verbs of saying, namely *tell*, it is clear that there are several uses or senses which a lexicographer wants to distinguish, for example *I told him the store was shut, I told him to shut the store, I could tell the store was shut*. But, as the third of these examples already indicates, these are not just semantic differences independent of grammatical patterning. The third example shows that *tell* in the sense of judging or knowing is usually accompanied by a modal. Ozcorp gives us examples such as

- you could tell what everyone was thinking
- I can tell who's been talking to you
- I can't tell why I never became a gun

It seems reasonable to say that this is useful or important information. It is nevertheless an overgeneralisation to say that *tell* in this sense is always accompanied by *can* or *could*. Sometimes this modal effect is achieved by other wording, for example

- it was impossible to tell whether any hits or near misses
- it is hard to tell what human beings may have done
- it's hard to tell when he's telling the truth

The most accurate report of *tell*, at this level of detail, would be to say that it is usually preceded by a modal *can* or *could*, and if not, by an expression such as *it's hard to* or *it's impossible to*.

Ozcorp's data on *scold*, a rather less common verb, show comparable significant patterning. For example, with *scold*, the target and receiver are one and the same: you can *criticise* me to my boss or my friends, but you can only *scold* me to my face. In general, this target/receiver is expressed, as in:

- his mother always scolded him when she found out
- but Hugh himself scolded the child roundly

But occasionally, it isn't immediately present:

- Coonardoo scolded angrily and sent them away again
- "don't scold, there's a darling"
- say you are sorry and she won't scold any more

And, rarely, there is some hint that the target and receiver aren't quite identical, perhaps because of the possibility of scolding in print rather than orally:

– has scolded this column for advertising
– she's quick to scold John McEnroe for the antics

5. Discussion

The evidence of a large corpus shows the dangers of making simple categorical generalisations on mechanical criteria. For example, it would be misleading to say that *scold* must have a human direct object . What is true about the object (target/receiver) of scolding is that it is often present; that, if not, it is usually recoverable from the context; and that we find only very rarely patterns like *scold at each other.*

It is tempting to exclude some of the rarer patterns as "metaphorical" or "extended" use. But this only raises the troublesome question of what exactly constitutes a metaphor: if it is somehow metaphorical to speak of a bird scolding or of scolding a column, is it perhaps equally metaphorical to speak of an adult scolding an adult? or of anyone other than a mother scolding anyone other than her child? It is not clear where to set the boundary between a supposedly canonical literal use and other uses.

If there is a useful criterion here it is precisely frequency of use. We can say that certain patterns are, by their predominance in large quantities of evidence, prototypical, while others are relatively rare and therefore marginal. It may be that categorisation is warranted as a means of simplification for the benefit of, say, children and adult learners. But the danger is that the categorisation not only fails to do justice to the true nature of linguistic behaviour but also dulls the lexicographer's and the learner's sensitivity to change and flexibility. It is, for example, a reasonable oversimplification to say that the most common verbs of saying, *ask, say, speak, tell*, in their unaffixed forms, cannot be used as nouns. But this generalisation not only overlooks (as most dictionaries do not, of course) the use of *a say* or *my say*; it also marginalises *speak* in patterns like *Treasury-speak* and *Keating-speak* (which most dictionaries treat as a suffix); and it overlooks the incipient use of *ask* as a noun, as in the sporting commentators' reference to *a big ask*. In short, our categorical bulkheads tend to leak.

One can see the case for abandoning grammatical apparatus, as Collins Today's English Dictionary has done, particularly if senses are exemplified and presented in order of frequency. Even in those diction-

aries that are reluctant to go that far, the increasing use of corpora and citations shows that lexicographers know the value of trying to inform by example. This strategy of using carefully chosen examples does at least offer a prospect of signalling prototypical uses without being rigidly categorical or asking the reader to cope with a complex grammatical apparatus.

References

COBUILD = *Collins Cobuild English Language Dictionary*, 1987
Collins Today's English Dictionary, 1995
Halliday, M.A.K. 1992. "Language as system and language as instance: The corpus as a theoretical construct." In J. Svartvik (ed.), *Directions in Corpus Linguistics*, Mouton de Gruyter, Berlin & New York, pp.61–77
LDOCE = *Longman Dictionary of Contemporary English*, 1987 (2nd ed.)
The Macquarie Dictionary, 1991 (2nd ed.)
OALD = *Oxford Advanced Learner's Dictionary of Current English*, 1989 (4th ed.)
Sinclair, J. 1991. *Corpus, Concordance, Collocation*, Oxford University Press, Oxford.