Elizabeth Walter, Cambridge University Press

Semantic set-defining: benefits to the lexicographer and the user

ABSTRACT: In this paper, I discuss the insights which can be gained from defining groups of near-synonyms in relation to one another. In particular, I discuss the tendency of many dictionaries to define by synonym, and point out that this is often inaccurate and misleading. I examine some methods of identifying the distinguishing characteristics of near-synonyms, and discuss how information on near-synonyms is best presented to users. This paper is written primarily with the needs of learners of English in mind.

1. Introduction

One of the main aims of a lexical reference work must be to help a user of the language to choose appropriate words for specific contexts. When learners of a language progress beyond an elementary level, they begin to acquire a larger and potentially more subtle vocabulary, but they need help in learning how to distinguish between semantically related words, or near-synonyms. There is a great deal of potential for confusion, offence, or hilarity if an inappropriate word is chosen, and language users may rely on dictionaries to guide them in their choice.

One of the main dangers language learners and users encounter is the tendency of lexicographers to define by synonym, often with little regard for the fact that, as D. A. Cruse points out: "One thing becomes clear when we begin a serious quest for absolute synonyms, and that is that if they exist at all, they are extremely uncommon." (1986, 270)[1]. In failing to identify the distinguishing features of near-synonyms, important information about words is often missed, and it is easy to see how learners can be led to use words inappropriately.

In what follows, I discuss the various ways in which near-synonyms differ from one another, and how these ways can be recognized. I also suggest some ways in which additional distinguishing information could be incorporated into dictionary entries.

2. Substitutability

One of the easiest ways to test for synonymy is to take several citations of the words in question and to try to substitute each word in each sentence. This technique usually gives clues as to the distinguishing features of the words if they are discovered not to be synonyms.

This is the simplest way of discovering whether or not words are true synonyms, and examples of the technique are given throughout this paper to illustrate various types of distinguishing features of near-synonyms.

3. Types of disambiguating information

Near-synonyms can differ from one another in several ways. What follows is a checklist of the most important of these. I have included some discussion of syntax as a distinguishing feature between words, because although semanticists might argue that this is not relevant to a discussion of synonymy[3], it is certainly relevant to the information needs of learners of English. Of course, many near-synonyms differ in several of the ways described below, and the differences interrelate.

3.1 Semantic nuance

By this, I mean a specific nuance of a word which means that its range of use is not exactly the same as that of a near-synonym. For instance, the word 'exhilaration' can only be used in relation to events in progress or recently ended, not events in the future. However, the idea of 'exhilaration' tends to be defined by lexicographers in terms of 'excitement' and 'happiness':

> COBUILD: **exhilaration** a strong feeling of excitement and happiness
>
> LDOCE: **exhilarate** to make (someone) cheerful and excited **-ration**

'Excitement' often does exhibit the feature of anticipation, and trying to substitute 'exhilaration' in the following authentic citations would have made this difference clear:

> *The next high tide at dawn was awaited with much more excitement than December 1992.*
>
> *As the sun went down over Nicaragua's volcanic horizon, the excitement steadily mounted in expectation of Presidente Daniel.*

Another difference in semantic nuance which exists between the notion of 'excitement' and 'exhilaration' is that 'exhilaration' is always a positive feeling, whereas 'excitement' is often used in contexts indicating erratic behaviour or lack of self-control:

> *Will they trigger clashes with excited crowds?*
>
> *His father was told by doctors not to get too excited.*

For this reason 'exhilarated' would not be an adequate substitute in the above sentences. It could be argued that the positive aspect of 'exhilaration' is implied in the words 'happiness' and 'cheerful' in the definitions above, but we need to ask ourselves whether it is reasonable to expect a user to intuit which of the features of any defining word are applicable to the headword.

3.2 Context

The definitions of 'exhilarate' and 'exhilaration' quoted above do not mention that 'exhilaration' tends to be used in two particular contexts:

1) contexts involving a thrill of real or perceived danger

2) contexts involving a feeling of physical well-being, especially invigoration, e.g. in cold weather, wind, or cold water.

The following authentic citations of 'exhilaration' demonstrate this point:

You are soon shrieking with terror and exhilaration as you surf over 10ft waterfalls.

After the grind of the ground school and the exhilaration of the flying, what remains is a warm glow of self-satisfaction at having overcome fundamental fears.

'Excitement' is a much more general word, and although it may be used in similar contexts to 'exhilaration', implying fright or physical well-being, it can equally be used in contexts where neither is implied:

Compound Q, a substance which caused excitement in recent laboratory studies, has shown problems of toxicity in patients.

In dictionaries, especially learners' dictionaries, typical context is often shown in examples. The LDOCE examples for 'exhilarated' are as follows:

I was exhilarated by my ride in the sports car.

This sea air is most exhilarating.

The issue of relevance to this paper is the question of how explicit an entry can or should be in talking about typical context. The LDOCE examples are clearly attempting to show the contexts I identified earlier, and of course, this is one of the main functions of example sentences. However, the fact that these contexts are typical for 'exhilaration' while the contexts shown above for 'excitement' would not be, is not made *explicit*. The learner does not know if the headword is always, usually, or only sometimes used in these contexts.

3.3 Collocation

This is a very important disambiguating feature, especially because the identification of collocation patterns can often lead to insights about semantic nuances.

For instance, looking at citations for the words 'humid' and 'muggy', it can be observed that 'humid' often collocates with words indicating a place: *humid desert zone, humid equatorial regions, a humid outpost.* On the other hand, 'muggy' rarely collocates with places, and is more likely to collocate with words indicating a period of time: *a muggy morning/evening.* This leads to the observation that 'muggy' is far less likely to be chosen than 'humid' when describing *permanent* states, a feature which might not have been noticed without a comparison between the words.

To return to 'excited' and 'exhilarated', the following citation illustrates another difference between them, to do with collocation:

Unfortunately one dog began to bite a child and the dogs became excited.

Although it is demonstrated here that it is both possible and natural-sounding to describe an animal as 'excited', 'exhilarated' could not collocate in the same way, but seems to be restricted to humans.

3.4 Fixed Phrases

Much has been written about the fact that a great deal of the language we use is made up of 'clusters' or 'chunks' of words which seem to us to belong naturally together[4]. This is another distinguishing feature of near-synonyms, since in fixed phrases the substitution of another word, however close in meaning, is usually not possible. For instance, definitions of the word 'dread' usually imply that it is a synonym of 'fear'. However, in the following examples, the substitution of 'fear' by 'dread' would be completely inappropriate because 'fear' is embedded in very fixed constructions:

The government had previously rejected his application for fear of upsetting the US during negotiations over the exchange of nuclear information.

Captain Coetzee, who has apparently joined the ANC after fleeing South Africa in fear of his life ...

3.5 Grammar

One very difficult problem for learners of English is when semantically similar words follow different grammatical patterns, as the following authentic examples of learners' text demonstrate:

* *They have had the possibility to vote for a long time.*

* *People want that we support the development of solar energy.*

It is easy to see how these mistakes could be made: 'possibility' cannot be followed by the infinitive, but 'opportunity' and 'chance' can; 'want' cannot be followed by a that-clause, but 'desire' can. Although learners' dictionaries generally show this information at the words concerned, it is likely that it would be more effective presented in a contrastive way.

3.6 Register

A major distinguishing feature of near-synonyms is register, such as level of formality. As Frank Palmer vividly puts it: "A nasty smell might be, in the appropriate setting, an *obnoxious effluvium* or an *'orrible stink.*" (1981, 89). In general, level labelling in dictionaries is not subtle enough to describe narrow gradations of register, so we often get cases where an unlabelled word is defined by a so-called synonym, but where in fact there is a difference of level:

OALD: **require** ... depend on (sb/sth) for success, fulfilment, etc; need

There is nothing here to tell the user that 'require' is more formal than 'need'. Another example is the phrasal verb 'dispose of'. LDOCE, OALD, and COBUILD all use 'get rid of' as the core of their definitions, with no indication that 'dispose of' is the more formal phrase[5].

3.7 Attitude of user

This is what semanticists often refer to as the 'emotive' meaning of a word, as opposed to the 'cognitive' meaning. Many dictionaries have labels such as 'derog' or 'apprec' for 'derogatory' and 'appreciative', but there are often cases where the nuance is felt to be too subtle for a label.

e.g. LDOCE: **be-all and end-all** the most important thing; the whole purpose of something

However, compare:

a) *As far as my boss is concerned, product design is the most important thing.*

b) *As far as my boss is concerned, product design is the be-all and end-all.*

The second example indicates definite disapproval by the speaker of the importance attached to product design, implying that other issues are being disregarded because of it, whereas the former example does not carry the same disapproval.

A similar case is the difference in intensity between 'waste' and 'squander'.

OALD: **squander** ... waste (time, money, etc); use sth wastefully

The use of 'squander' in the following citations suggests an accusation of irresponsible recklessness, a nuance which would be far less strongly expressed were 'waste' to be substituted.

Political leaders on the other hand fear that impatient youths will squander the gains of the protests.

A lot of big stores squander vast budgets every year on new mannequins.

4. Implications for dictionary entries

Having established that dictionaries often fail to provide all the necessary information to enable users to choose between near-synonyms, it is necessary to ask how, or indeed if, it is possible to do so. Clearly, different user groups have different needs, and it would be ridiculous to say that a dictionary entry 'should' be structured in a certain way or 'should' always contain all the information resulting from the preceding checklist. However, I will briefly discuss a few possibilities.

4.1 Presentation in semantic groups

Treatment in groups, as opposed to alphabetical order, can lead to more subtilety of information, e.g. level labelling. It is easier for a user to contrast entries which are physically close. However, at an advanced level, presenting a large number of near-synonyms together may overwhelm the user with information. In an A-Z dictionary, cross referencing and usage notes contrasting certain words can help, as could the kind of extra-definitional material suggested in 4.2, below. As more and more work is done on the analysis of learners' texts, there will be better information on which words actually cause this kind of usage problem, so information can be targeted at relevant words.

4.2 Extra-definitional material

There are many possible ways of introducing material not traditionally included in a definition. Learners' dictionaries often have usage notes, for example, or short 'synonym essays'. It would be possible to include comparative information within an entry, perhaps separated in some way from the definition proper, especially when contrasting with a more common word:

> e.g. **handy** *adj* [more informal than **useful**]
>
> **idle** *adj* [less common, and often used more pejoratively than **lazy**]

This method could potentially cover a fairly sophisticated comparison, which may be particularly appropriate when words are presented in semantic groups rather than alphabetically:

> e.g. **futile** *adj* [implies more contempt than **pointless**]
>
> **reluctant** *adj* [suggests a lesser degree of unwillingness than **unwilling**]

Another type of extra-definitional material which is sometimes shown in dictionaries is the typical object or subject collocating with the headword:

> e.g. LDOCE: **incubate** ... (of eggs) to be kept warm until the young birds come out

However, a far greater level of subtlety than this could be possible. Consider, for instance, the information conveyed in the following:

> **parade** *vt* [obj: own body or sth that can be carried]
>
> **flaunt** *vt* [obj: esp. sth which may cause resentment or disapproval in others]

4.3 Bilingual formats

One problem for lexicographers working on learners' dictionaries is that the use of a restricted defining vocabulary may not allow for subtle distinctions. Information on the distinguishing features of near-synonyms given in the language of the user would make quite sophisticated detail accessible to a wider ability range. In a bilingual format, the necessity to distinguish adequately between near synonyms is perhaps even greater than

with monolinguals - if more than one English word is translated by the same word in the user's language, that translation needs to be qualified by an explanation of the differences in the English in order to be accurate.

> e.g. **sort** *vt* [ordonner selon le type, la taille, etc. Obj: ex. des lettres, des vêtements, des fruits] classer, trier
>
> **classify** *vt* [implique un système plus formel que **sort** ou **order**] classer, classifier

With bilingual dictionaries there is of course the added complexity that even when a single word in one language is translated by a single word in another language, there may not be exact equivalence between them. Many bilingual dictionaries would be more accurate if the translations they gave were qualified by a description of the way they differed from the headwords.

4.4 Negative information

Negative information is rarely shown in English dictionaries, but could be used in some cases, for instance to warn that a word does not have all of the same properties as a near synonym. An example of this would be the collocational possibilities of 'flexible' rather than 'supple':

> **flexible** *adj* [describes: materials, *not* usu people]

4.5 Qualifying definitions

It may be that definitions would be more useful if lexicographers were less afraid of using the word 'often' and even 'sometimes'. There is no reason to exclude information on a word simply because it does not apply in every case.

For instance, in the following example, the extra-definitional text helps to distinguish between the words 'crisp' and 'brittle', even though the nuances described do not apply in every single instance of use of the words:

> **crisp** *adj* [usu. appreciative, implying freshness. Describes: esp. food]
>
> **brittle** *adj* [a negative word. Often implies that the thing described is weak]

One way of including information on a way a word is sometimes or often used, is to attach it to a particular example demonstrating the point. For instance, when thinking about the difference between the word 'cheeky' and the word 'insolent', it is apparent that 'cheeky' often implies an element of humour which 'insolent' does not. One way of covering this might be as follows:

> **cheeky** *adj* disrespectful, esp. towards someone older; rude *Don't be cheeky to your mother* [often implies use of humour] *a cheeky allusion to the minister's private life*

Thus the user is given explicit information which would either not be found at all in a traditional dictionary definition, or at best would only be implied in the choice of examples.

5. Conclusion

Semantic set-defining is a useful technique for identifying important features of words, even if those words are then presented in an alphabetical list. The question of the definition of semantic groups themselves is beyond the scope of this paper, but existing thesauruses provide a useful basis to work from. At a simpler level, I believe definitions could be much improved if at least the words used in the definitions were compared in ways described in this paper to the headword they are defining. It is important for learners of a language that they are given all the information they need to avoid using words inappropriately, and I believe that semantic set-defining in all dictionary compilation would help lexicographers achieve this.

Endnotes

1 See also Lehrer (1974), Lyons (1977), McCarthy (1990), Palmer (1981)

2 Dictionaries quoted in this paper are identified by the following abbreviations:
OALD: Oxford Advanced Learner's Dictionary (4th edition, 1989)
COBUILD: Collins Cobuild Dictionary of the English Language
LDOCE: Longman Dictionary of Contemporary English (2nd edition, 1987)

3 e.g. Cruse (1986, 269)

4 e.g. Bolinger (1976), Cowie (1981), McCarthy (1990)

5 In fact, taking the first 25 instances of 'get rid of' picked randomly from our corpus, 'dispose of' could not have been substituted in a single case

Bibliography

BOLINGER, D. (1976): "Meaning and Memory". In: Forum Linguisticum 1.1:1-14

Collins COBUILD English Language Dictionary (1987) Collins, London and Glasgow.

COWIE, A. P. (1981): "The treatment of collocations and idioms in learners' dictionaries". In: Applied Linguistics 2:203-206

CRUSE, D.A. (1986): Lexical Semantics. Cambridge University Press, Cambridge.

LEHRER, A. (1974): Semantic Fields and Lexical Structure. North Holland Publishing Company, Amsterdam.

Longman Dictionary of Contemporary English. Second edition. (1987) Longman, Harlow.

LYONS, J. (1977): Semantics, Volumes I and II. Cambridge University Press, Cambridge.

Mc CARTHY, M.J. (1990): Vocabulary. Oxford University Press, Oxford.

Oxford Advanced Learner's Dictionary. Fourth edition. (1989) Oxford University Press, Oxford.

PALMER, F.R. (1981): Semantics. Second edition. Cambridge University Press, Cambridge.

KEYWORDS: semantic sets, synonymy, near-synonyms