TOWARD A USABLE DEFINITION OF SYNONYMY

(POUR UNE DEFINITION FONCTIONNELLE DE LA SYNONYMIE)

Gillian Michell
School of Library and Information Science
The University of Western Ontario
London, Ontario
N6A 5B9

ABSTRACT

The lack of a usable definition of synonymy calls into question the results of retrieval experiments designed to test the effectiveness of indexing with controlled vocabularies. This paper describes the nature of the problems that arise in defining synonymy in order to suggest where an adequate definition might be found. (L'absence d'une définition fonctionnelle de la synonymie remet en question les résultats des tests de repérage ayant pour but de mesurer l'efficacité de l'indexation avec vocabulaires contrôlés. L'auteur analyse la nature des problèmes associés à la définition de la synonymie et propose quelques suggestions pour en arriver à une définition adéquate.)

A USABLE DEFINITION OF SYNONYMY

Synonymy, or sameness of meaning, is a property of lexical items or words which has received considerable attention in certain aspects of information science. This is because of the problems it raises with respect to maximizing the effectiveness of information systems. To improve recall, the information scientist wants to exploit synonymy relations among lexical items so as to retrieve all the information relevant to any query. To improve recall without loss in precision, he wants to be assured that the mapping between synonymous terms is as accurate as possible. The importance of synonymy, then, is unquestionable; however, the lack of a usable definition continues to plague the field. Some information scientists have turned to the work of philosophers and linguists for answers, but a coherent and relevant view of the nature and dimensions of synonymy has not yet been produced in the field of information science. This paper will attempt to suggest an approach to such a view.

Previous definitions of synonymy in information science have been unworkable, it seems to me, because they have attempted to distinguish different kinds of synonymy at a fine level of detail without first making the right general distinctions which must precede such lower level distinctions, if they are to be valid. Synonymy has been used to cover a range of sameness relations which includes at least those illustrated by the pairs of terms in (1).

(1) a. lawyer/attorney
     b. IR/information retrieval
     c. program/programme
     d. aspirin/A.S.A.
     e. the Morning Star/the Evening Star
     f. lift/elevator
     g. data bank/data base
     h. marine propellers/propellers, marine
     i. computer/computers
     j. computer/computing
     k. sand blasting/abrasive blasting
     l. roughness/smoothness
     m. flammable/inflammable

Undeniably each term is closely and systematically related to the other member of its pair; however, the nature of the sameness relations involved varies enormously. Some of the relations do not depend on meaning at all, such as that between computer and computers; while others depend on one or another of several distinct kinds of meaning relation. These meaning relations differ also in the ease with which they can be explicated and applied.

The major distinctions which have not been recognized are two: first, that some sameness relations are not meaning relations at all and ought therefore to be defined and handled differently from meaning relations;

second, that there is an important difference between items related by sameness of sense and those related by sameness of reference, a difference which has implications for information retrieval purposes. A third distinction which needs to be recognized is between 'cognitive' and 'emotive' synonymy, and a fourth and final point concerns the validity of the distinction which is often made between 'true' and 'contingent', or context-dependent, synonymy. It is hoped that an examination of these considerations will produce a better understanding of what synonymy is. This in turn should lead to more fruitful attempts to handle synonymy for the purposes of information retrieval.

We should recognize one interesting point to begin with. This is that it seems to be the case that synonymy in common parlance is no problem at all. It is a simple notion, easily grasped and easily illustrated. It can come as a surprise, then, to find it is a problematic issue in philosophy, linguistics, and information science. Why is there this discrepancy? I think we can find its source by considering the kinds of examples of synonymy that are generally proffered and accepted in non-technical discussions. Typical synonymous pairs are (2) and (3):

(2) <u>oculist</u> / <u>eye-doctor</u>

(3) <u>aspirin</u> / <u>A.S.A.</u>

For practical purposes the first member of each pair has the same meaning as the second member, and it is this sameness of meaning which is the relation that we use the term synonymy to label. Where are the problems then? The pairs in (2) and (3) illustrate two of the considerations mentioned above that complicate the definition of synonymy.

The first of these considerations has to do with the distinction between 'cognitive meaning' and 'emotive meaning' (Ullmann 1966). Cognitive meaning refers to the primary, intellectual meaning of lexical items, while emotive meaning refers to those aspects of meaning which have to do with emotional or creative or affective uses of language. An extreme example of a pair of terms that is cognitively but not emotively synonymous would be <u>doctor</u> and <u>sawbones</u>. In the full range of actual language use it is clear that the choice of words will on occasion depend on distinctions in emotive meaning; however, I think it is safe to say that for many information retrieval tasks no importance will attach to meaning differences which can be attributed to properties of lexical items like style, social status, or emotional weight. We will follow John Lyon's (1968) lead, then, in restricting the use of the term 'synonymy' to cognitive synonymy. By excluding non-cognitive meaning differences from consideration, we will take care of comments to the effect that <u>oculist</u> and <u>eye-doctor</u> are not really synonymous because they differ in their stylisitic connotations in certain contexts. What is usually of prime interest to us in information science is not those

situations where the overlaid connotations of a word distinguish it from other words; rather, we are concerned with situations where distinctions—cognitive distinctions—between words do not exist. This matter we will return to in treating the distinction between true and contingent synonymy.

A second consideration that raises problems with synonymy is exemplified by the aspirin / A. S. A. example and has to do with the distinction philosophers have called the sense - reference distinction. Reference is a relation between words and entities or events in the world. The name Gill Michell, for example, is a lexical item that picks out or refers to me as a physical entity in the real world. The reference relation thus relates language to reality. Sense, in contrast with reference, has to do with the meaning of words as they relate to other words in the vocabulary. Sense has nothing to do with entities in the real world. Sense-relations of lexical items in a vocabulary are often characterized in terms of semantic elements or components. Boy, for example, can be said to have semantic components like [male], [human] and [non-adult], and values of these components can be used to distinguish the sense of boy from that of girl or woman as in (3).

(3)

boy	girl	woman
[+human]	[+human]	[+human]
[+male]	[-male]	[-male]
[-adult]	[-adult]	[+adult]

Words like boy, girl, and woman differ in sense relations as characterized above. They also differ in reference, of course, in that they can be used to refer to different kinds of entities in the real world. There are other words, however, that do not seem to have both sense relations and reference relations. When we look at a pair of words like aspirin and A. S. A., for instance, it is clear that they are identical in reference because they refer to the same objects in the world, but it is not clear that they are identical in sense; as we will see it is not even clear that they have sense relations with each other or other lexical items. It seems to be a characteristic of names, in fact, that they have no other significant meaning beyond their function of giving linguistic labels to real world entities.

Whether synonymy should be confined to identity of sense or identity of reference or the combination of the two is a major problem in defining synonymy. Lyons (1968, p. 427) maintains that synonymy is a matter of sense, not reference. On this account, aspirin and A.S.A. would not be related by the synonymy relation, but only by identity of reference. (We note that of course names with identical referents may differ in emotive

meaning--a difference that advertising tries to foster: "only aspirin has the Bayer cross".) For information retrieval purposes, one might choose to have names like aspirin and A.S.A. linked together but it should be recognized that they are linked by a relation distinct from that linking terms with the same senses.

It may seem intuitively that this distinction is unimportant, that the sameness relations of sense and reference are of the same order. This misconception arises from the fact that many word pairs can be found that have both identity of sense and identity of reference. It can be shown, however, that if we do not define synonymy as identity of sense and instead take identity of reference as a sufficient criterion for synonymy, systematic problems can arise. We would, for example, have to consider the term the leader of the Liberal party and the term the Prime Minister synonymous. This would lead to errors, because, although at this time these two terms have only one possible referent, Pierre Trudeau, the phrase the Prime Minister may have a number of different real world referents (at different times and in different places) because of the sense of the phrase. These need not coincide with the referents of the leader of the Liberal party, because of its different sense. We must concur with Lyons, then, in saying that, for lexical items that have both sense and reference, identity of reference is a necessary but not sufficient condition for synonymy. Identity of sense is the only sufficient condition.

We turn next to the question of true synonymy as opposed to contingent or context-dependent synonymy. If we look at the different treatments of synonymy in the literature, we find that in most cases they are concerned only with true synonymy; that is, synonymy that holds in all contexts. Linsky's (1968) account of the influential treatments of synonymy in recent philosophy discusses the approaches of Quine, Carnap, Goodman, and the ordinary language philosophers. Quine's and Carnap's arguments depend on the view that synonymy means identity of sense in all contexts, while Goodman tries to give an account of true synonymy also, but in terms of identity of reference. An examination of the criteria for synonymy discussed by these philosophers leads to the radical conclusion that no two expressions are exactly synonymous. Counter to these views are those of ordinary language philosophers, like Rollins (1950) and Shwayder (1954). They believe that a pair of terms should be considered synonymous if they are interchangeable not everywhere but only in contexts relevant to a given discussion. Once context is admitted to have an influence on the definition of synonymy, then synonymy changes from a semantic question, in which matters of usage are irrelevant, to a pragmatic one, where usage and other, non-linguistic considerations may play a role. The importance of context has been recognized by others, including Quine (1953, p. 24-5) and, particulary, Lyons (1968, p. 428, 452). In view of the fact, then, that there are no lexical items which in the

final analysis can be proved to be truly synonymous in all possible
contexts it seems that the only acceptable definition of synonymy must
resemble that of Katz:

> A constituent C_i is synonymous with another
> constituent C_j on a sense just in case they
> have a reading in common (1972, p. 48),

where a constituent can be understood for our purposes as equivalent to
a lexical item, and where by "having a reading in common" is meant that
one possible interpretation, or meaning, of C_i is the same as one possible
interpretation of C_j. Recognition of the role played by context is
implicit in this definition, since only context can determine which read-
ing of a lexical item is the relevant one on any given occasion. The
notion of true synonymy, then, can be abandoned as untenable and unsup-
ported, and context dependent synonymy accepted in its place.

It is noteworthy that attempts to identify synonymous terms auto-
matically have had some success, but this success has been achieved only
when terms are grouped by their common contextual properties, and when
the range of synonymy is limited to specialized subjects, rather than
extending over language in general. As Sparck Jones and Kay say, in
describing research on synonymy in automatic classification:

> Although computers cannot recognize sameness of meaning
> directly, they can recognize sameness of contextual
> behavior, which is just what we expect synonyms to
> exhibit (1973, p. 162).

Accounts of synonymy in information science which do not recognize con-
textual boundaries in their definitions are of questionable value and
serve only to confuse an already complex issue.

Typical of such accounts are those which distinguish synonymous
terms from 'quasi-synonymous' or 'equivalent' terms (e.g., Hutchins 1975,
p. 39; Soergel 1974, p. 110). An example of quasi-synonymy is the pair
of terms library science and information science. These terms fit Soergel's
requirement of having meanings which "overlap widely". (Note that no
attempt is made to indicate what "widely" means in extent.) He says later:

> Synonymous and equivalent terms should be lumped together
> in the user version [of the thesaurus]. The distinction is
> difficult and of limited operational significance (p. 232).

The distinction that quasi-synonym is supposed to label seems in fact not
to be a clear cut distinction but rather to be only a matter of degree.
That is, quasi-synonyms are used interchangeably in a somewhat smaller
range of contexts than are synonyms. But since there is no context-free

156

A USABLE DEFINITION OF SYNONYMY

synonymy, the distinction has little content as stated. What would be more meaningful is the claim that with respect to a certain context-- say, a subject area for which a thesaurus is being developed--specified terms are used completely interchangeably, while in certain more restricted aspects or subparts of that subject area there are other sets of terms that are used interchangeably. Thus in a thesaurus for library and information science the terms library science and information science might be considered to have distinct senses with respect to a broad view of the subject area, but in many contexts within this area the distinctions are neutralized.

The converse situation to that in which contextual considerations neutralize distinctions is that in which contexts introduce distinctions. Limiting a thesaurus to a particular time context, for example, will introduce distinctions among terms that have changed over time. If the term electrical condenser is not a currently used term, having been super-seded by capacitor, these two terms will not be treated as synonymous in a system which has reason to maintain time distinctions. Similar contrast-introducing contexts are regional and social variations in usage.

We have one point left to deal with: relations of sameness that do not have to do with meaning. These include spelling variants and morphologically variant forms. It has been suggested that these relations account for the major proportion of sameness relations which must be dealt with for information retrieval purposes. If this is so, the problem of dealing with relations of sameness is much simplified, since spelling and morphological variations are much more transparent and predictable aspects of language than are semantic considerations. Spelling variants merely represent minor differences in the conventions by which speech is reduced to written form. Morphological variations include inflectional differences indicating tense, aspect, or number, as in computer / computers, and derivational differences marking grammatical distinctions such as that between noun and verb as in computer / compute. Some of these variants may reflect differences in grammatical form that have become significant, as the standard example program / programming illustrates. Only the purpose a system is designed to answer will determine whether these differences are significant in practice. It is also the case that on some occasions morphological variants are tied to differences in regional or time contexts. Bar these last two considerations spelling and morphology produce only super-ficial and insignificant differences among lexical items which will not cause interesting problems.

One further question might be raised about context-dependent synonymy. This question is the following: if synonymy cannot be defined without reference to context, how is it that synonym dictionaries work? The answer to this depends crucially on context too. Synonym dictionaries, of which Roget's Thesaurus is the prototype, group together terms which

are synonymous in one context or another and leave it up to the user's knowledge of his language to decide which terms are synonymous for the particular context he is concerned with. Roget, for example, offers as synonyms for <u>nice</u> the items <u>savoury</u>, <u>discriminative</u>, <u>exact</u>, <u>good</u>, <u>pleasing</u>, <u>fastidious</u>, and <u>honourable</u>. Only previous knowledge of the appropriate uses of these words make such a list useful. One piece of evidence for this is the fact that non-native speakers cannot use such dictionaries, and another is that standard bilingual dictionaries that specify contexts of use for the different words offered as translations of particular words into the target language are much more useful to the translater than simply a list of words without comment. The question for the information scientist is to decide when the system can leave specification of context up to the user and when and for what purposes contextual restrictions must be made explicit.

I have in this paper treated four considerations which are important in defining and categorizing types of synonymy and sameness relation: (1) the cognitive / emotive distinction; (2) the sense / reference distinction; (3) the putative existence of true synonymy, as opposed to contingent synonymy; and (4) sameness relations which are not semantic in character. If account is taken of these considerations, the problem of synonymy becomes a much more manageable one.

REFERENCES

BLOOM, Lois 1970 Language development: form and function in emerging grammars. Cambridge, Mass.: M.I.T. Press.

HUTCHINS, W.J. 1975 Languages of indexing and classification. Stevenage, Herts.: Peter Peregrinus.

KATZ, Jerrold J. 1972 Semantic theory. New York: Harper & Row.

LEECH, Geoffrey 1974 Semantics. Harmondsworth, England: Penguin.

LINSKY, Leonard 1968 Synonymity. In Encyclopedia of philosophy, v. 8, pp. 54-57. Edited by Paul Edwards. New York: Macmillan.

LYONS, John 1968 Introduction to theoretical linguistics. Cambridge: Cambridge University Press.

MANDERSLOOT, William G.B. et al. 1970 Thesaurus control -- the selection, grouping, and cross-referencing of terms for inclusion in a coordinate index word list. Journal of the American Society for Information Science 21: 29-57.

QUINE, W.V. Two dogmas of empiricism. In From a logical point of view, ch. 2. Cambridge, Mass.: Harvard University Press.

ROLLINS, C.D. 1950 The philosophical denial of sameness of meaning. Analysis 11: 38-45.

SCHWAYDER, David 1954 Some remarks on 'synonymity' and the language of semanticists. Philosophical Studies 5: 1-5.

SOERGEL, DAGOBERT 1974 Indexing languages and thesauri: construction and maintenance. Los Angeles: Melville.

SPARCK Jones, Karen, and Kay, Martin 1973 Linguistics and information science. New York: Academic Press.

ULLMANN, Stephen 1966 Semantic universals. In Universals of language, pp. 217-62. 2d ed. Edited by Joseph H. Greenberg. Cambridge, Mass.: M.I.T. Press.