

*Dušan Gabrovšek*  
*University of Ljubljana, Ljubljana*

## **The general thesaurus: the what and the who of it**

The general thesaurus is a semantics-based onomasiological reference tool listing undefined sets of semantically related lexical items. Traditionally, it has been associated primarily with the provision of synonyms and the contrast between conceptual and alphabetical macrostructural arrangement.

---

Most thesauruses are built on the basis of either sense relations or semantic fields, both including rather vague lexical relations and associative fields. The key organizing principle of thesauruses remains non-subjective word association, yet we know of few successful non-alphabetical arrangements of lexical information. Also, we know very little about how the thesaurus is actually used and by whom. Thesauruses should be constructed on the basis of surveys into users' needs, which would help us frame answers to important questions concerning their form, their uses and their users.

### **Introduction**

If asked to define the dictionary, we are likely to find it easy, because alphabetization and the provision of word meaning are virtually synonymous with dictionariness, so to speak. On the other hand, defining the thesaurus is not so easy; traditionally, the provision of synonyms and conceptual vs. alphabetical arrangement seem to have been its most frequently quoted features. However, I believe the truth is far more complex than this; there are in fact several problems involved: those of definition, those of organization, and those of the user. My basic point in this paper, which focuses on the general monolingual thesaurus of English — thus leaving out such related genres of word-books as reverse dictionaries, pictorial dictionaries, and a variety of other

wordfinders –, is that we have to look hard at the basics so as to gain a global and realistic view of the thesaurus, its forms, uses, and users.

### **Definition**

The general monolingual thesaurus is a semantics-based reference tool listing undefined sets of semantically related lexical items. Specifically, it is onomasiological, that is, it is characterized by the concepts-to-terms-designating-them approach. In this, it is often contrasted with the dictionary which is semasiological, that is, characterized by the term-to-meaning approach. Nevertheless, it is easy to find the thesaurus described as a kind of a dictionary (Hartman 1994, 138; Busmann 1996, 484). Also, it follows from this characterization that the dictionary of synonym discriminations is not a thesaurus either. The macrostructure of a thesaurus is an array of lexical items listed either alphabetically or conceptually. In broadest possible terms the term thesaurus refers to a thematic organization of lexis (Kay and Roberts 1994, 4603); this type of wordbook basically attempts to display some of the semantic relations of words by arranging them according to their similarity of meaning (Ball 1987, 167). In most discussions, the problem of arrangement is restricted to the macrostructure, while the microstructural considerations (selection and presentation of the matter given within the entry article) are largely ignored.

### **Organization**

To discuss the organization of a thesaurus in terms of the dichotomy of conceptual vs. alphabetical arrangement is not really fruitful, because it is simply too general and vague. What we must do instead is to scrutinize the overall organization and the entry articles themselves in order to identify the organizational criterion or criteria. However, this is not to deny the rich tradition of thematic lexicography going back at least to the Scholastics that today comprises much more than thesauruses: there are, for example, travelers' phrase books, specialized topical vocabulary lists, classified glossaries, and a host of others (cf. McArthur 1986, 161). Let us note in passing that today's success of alphabetization is the result of the convenience of the alphabet as an invariant series universally learned and used in more or less the same sequence, whereas thematic ordering is chiefly regarded as arbitrary and necessarily idiosyncratic, themes or topics being seen as a variant series.

I would suggest that most thesauruses are built on the basis of

a) sense relations, including not only synonymy, antonymy, hyponymy, meronymy etc. that are discussed in most textbooks on semantics (e. g. Lyons 1995), but also — and this is something of a problem — a number of more or less vague lexical relations that cannot be captured and described by purely linguistic methods, based either on association and intuition or on extralinguistic relationship captured by encyclopedic knowledge (Lipka 1990, 157–58). In

actual fact many thesauruses and writers on the topic favor synonymy as the most important sense relation (e. g. Bussmann 1996, 484), even though it is not at all clear whether this is so from the users' viewpoint. Opinions differ, however: some authors regard hyponymy as the most important semantic relation underlying the thesaurus (Hudson 1995, 15), while for others, it is the semantic or associative field (McCarthy 1990, 138). Moreover, some linguists seem to be in two minds about this (McCarthy 1990, 19 vs. 138).

b) semantic fields, usually defined as conceptually related vocabulary areas, the words within each field forming an interdependent whole. In thesauruses they are often coupled with rather vague associative fields. Here, ontological relationships typical of field thesauruses are often taken into account; terminographers call them associative relationships (e. g. PROCESS — STEP: operation — incision; cf. Aitchison and Gilchrist 1987).

It follows from the above that the semantic-field principle will mostly result in lengthy and rather heterogeneous entry articles. Indeed, this is one of the major characteristics of today's conceptual variety of the thesaurus: it offers a very broad picture of a given vocabulary, with several word classes being represented in one place and semantically closely related categories following each other. There are, however, exceptions: the comprehensive *Macquarie Thesaurus* (Bernard ed. 1990) arranges its 800-odd »keywords« alphabetically. This type of thesaurus is associated with extreme inclusiveness (Landau 1984, 108). Importantly, its use calls for a two-stage lookup process, which takes more time<sup>1</sup> (e. g. Chapman ed. 1992, Daintith et al. eds. 1993).

By contrast, the alphabetical variety has a narrower and tighter entry structure because its A-Z ordering imposes heavy restrictions on the macrostructure: each headword is a veritable »world of its own.« It is associated with more selectivity — not only are there hardly any proper names, specialized or obsolete terms, proverbs, and word lists, but also such a work in essence almost always provides mere alternatives to a lexical item (i. e. synonyms or near-synonyms listed directly after each alphabetized headword). Thus it is easy to consult but necessitates a great deal of repetition. In any case, this is the dominant type of the general thesaurus today (e. g. Urdang 1991, Gilmour ed. 1995).

Whatever the type, it is probably true to say that the key organizing principle of the monolingual general-language thesaurus is word association (Piotrowski 1994, 124); it goes without saying that since non-general word associations are likely to vary from person to person, one can realistically expect that the thesaurus can only incorporate general word associations. But I hasten to add that despite the importance attributed to the thematic arrangement in thesauruses (McArthur 1994, 3–4)<sup>2</sup>, there have been so far very few successful

---

1 This is only true of printed versions; the electronic thesaurus obviates the need to first consult the A-Z index of terms typical of the conceptual variety ever since it was first published (Roget 1852).

2 McArthur (1986, 158–59) suggests discussing reference works of all kinds in terms of a »reference rectangle« comprising four broad options: encyclopedic, wordbook, thematic and alphabetic. Roget's *Thesaurus*, for example, is an instance of a thematic wordbook.

arrangements of lexical information organized non-alphabetically (Kipfer 1992, x), Roget's, own thesaurus of 1852 being no exception: he included a sizable alphabetical index, even though as a secondary means of accessing words. Nevertheless, many compilers and users of today seem to be unaware that from the viewpoint of Roget's, sameness or similarity of meaning is not the primary key to compiling a thesaurus — rather the key is membership in the cluster of linguistic signs that go to make up some very large and general concept (Chapman 1977, xvi).

Let us not forget that historically, there has been so far little real progress in general thesaurus-making; however, one must not ignore the *Thesaurus Dictionary* (1902) by the American philologist Francis A. March, an A–Z wordbook combining dictionary and thesaural entries, and the thesaurus in dictionary form, first prepared in 1931 by the British-born American orientalist C. O. Sylvester Mawson, basically Roget's *Thesaurus* recast in dictionary form, which changed Roget's concept of a broad semantics-based wordbook created as a means toward a clear and unambiguous expression of ideas, into a much more narrowly conceived collection of undiscriminated synonyms.

### Uses and users

The number of thorough analyses of the thesaurus is small, and the existing surveys are brief (e. g. Marelllo 1991, Kay and Roberts 1994). Moreover, there are surprisingly few empirical studies of the typical users and uses to which thesauruses are put. Of course, most thesaurus blurbs tell us that no serious student, writer, and translator should be without one, but no one seems to care to explain why this should be so. To begin with, the needs of the three classes of users are far from being identical: students probably consult a thesaurus to improve their prose writing; writers are likely to draw on it chiefly for the sake of stylistic variation; as to translators, the situation is quite complex. It has been observed, for instance, that thesauruses do not help at all in the translation process, because the synonyms listed occur in various unpredictable collocations that are not provided (Heliel 1990). It seems fair to say that a translator will or will not need to use a thesaurus depending on the specific type(s) of information needed in the process. Second, most translation is done from a foreign language into a mother tongue, where thesauruses are simply not needed. At any rate, blurbs such as the one just mentioned indicate that in reality little is known of how the thesaurus is actually used and by whom. This view is supported by most writings on the subject. To say, for instance, that you go to a thesaurus when you already know the MEANING but do not have the exact word you need to express that meaning accurately and colorfully, that the thesaurus gives you a choice of words with SIMILAR meanings, one of which MAY meet your needs (Kingsley and Heffner 1984, 71–72), or that it is to help us think of NEEDED words that do not readily come to mind (Riggs 1989, 147), in reality does little more than beg the question: how do we know a meanings, if not by having a word or a brief para-

phrase for it? When are meanings similar? What if the words do not meet our needs? Which are the needed words and who needs them?

The lack of studies of the user perspective basically demonstrates that thesaurus makers largely rely on themselves, assuming the role of the user, which they should not do because their knowledge of language and their profession make them far different from most of their users in terms of lexical competence and lexical needs. While one can think of important research within linguistics that thesaurus makers might profitably draw on (the possibility of paying special attention to semantic primitives <Wierzbicka 1992>, or of organizing the thesaurus by word–association responses <Miller 1991, 155–58>), the most urgent task seems to be to undertake a wide–ranging examination of users’ needs, whether for composition, translation, or even setting or solving crossword puzzles.

## Conclusions

Thesauruses — whether printed volumes or electronic versions — should be constructed on the basis of surveys into users’ needs that could go a long way toward helping us frame at least tentative answers to several important questions, notably

1. Are there several identifiable groups of thesaurus users with identifiable reference needs? The common view that the search for variety of expression is still the prime function of a thesaurus (Kay and Roberts 1994, 4604) doubtless implies a strong native–speaker orientation. This is hardly to be disputed: who, aside from the educated native speaker can be realistically expected to successfully handle lists and lists of semantically related but largely or entirely indiscriminated lexical items? It is the awareness of this orientation that seems to be responsible for the idea that thesauruses are mere memory joggers.
2. What do users need thesauruses for? Do they typically need synonyms, hyponyms, etc., or rather items belonging to a given semantic field, if we choose to disregard here the notoriously vague associative fields?
3. Are word lists (kinds of diseases, largest lakes, architects, branches of psychology etc.) useful at all, and if so, are they really to be given in separate lists? These lists are mostly restricted to the conceptual variety of the thesaurus (but cf. Kirkpatrick comp. 1995, an A–Z thesaurus with over 400 such lists).
4. Should the entry article list only insertable alternatives and/or related items, or should lengthy explanatory phrases and glosses be given as well?
5. Given that in recent years several wordbooks have appeared that combine thesaural features with those of a dictionary (McArthur 1981, Summers ed. 1993), is it advisable to make the thesaurus richer, that is, more varied in terms of types of information provided, or should one settle for a diversified

thesaurus (cf. e. g. McCutcheon 1995)? Most importantly, are these developments a reflection of actual users' needs or mere ingenuity of thesaurus makers?

6. Can we make principled decisions about what is to be the headword (the item being looked up) and what is to be the entry (the item needed, listed within each entry article)? It is essential to draw and maintain a clearcut distinction between the two. It is evident that the former is to be based on the general vocabulary, and that the latter is to contain thousands of rarer, restricted and specialized items.

7. Is the idea of a thesaurus that ignores word-class divisions worthy of serious consideration?

8. Why is there so little common ground when it comes to comparing similar thesauruses? For instance, checking a common item such as *dictionary* in a number of thesauruses of a similar size, one finds a maze of entry types, starting with the non-existent one (Stein and Flexner eds. 1984), continuing with those listing only three of four entries but also adding an example of use (Urdang 1991) and those that have ten or more entries as well as brief definitions (Kipfer 1992). These are all alphabetized thesauruses; the conceptual variety likewise shows little consensus: the headword may be given in only one or two categories but be granted a special word list (Chapman ed. 1992), or be listed as many as six times or more in as many categories but have no special list of its own (Daintith et al. 1993). I very much fear that this lack of even the most rudimentary agreement is at least in part due to the indifference to users' needs, and that it is a reflection of the performance of different minds.

## References

- Aitchison, J. and Gilchrist, A. 1987. *Thesaurus Construction*. 2nd edn. London: Aslib.
- Ball, C. 1987. Lexis: The Vocabulary of English. In *The English Language*, rev. edn., ed. by W. F. Bolton and D. Crystal, 165–97. London: Sphere.
- Bernard, J. R. L. ed. 1990. *The Macquarie Encyclopedic Thesaurus*. McMahons Point, NSW: Macquarie Library.
- Bussmann, H. 1996. *Routledge Dictionary of Language and Linguistics*. Trans. and ed. by G. Trauth and K. Kazzazi. London: Routledge.
- Chapman, R. L. ed. 1977. *Roget's International Thesaurus*™. 4th edn. New York: T. Y. Crowell.
- Chapman, R. L. ed. 1992. *Roget's International Thesaurus*®R. 5th edn. New York: HarperCollins.
- Daintith, J, Martin, E. A, Alexander, F, and Wright, J. eds. 1993. *Bloomsbury Thesaurus*. London: Bloomsbury.
- Gilmour, L. ed. 1995. *Collins Thesaurus*. Glasgow: HarperCollins.
- Hartmann, R. R. K. 1994. The Onomasiological Dictionary in English and German. In Hüllen ed. 1994, 137–49.
- Heliel, M. H. 1990. Lexical Collocations and Translation. In *Translation and Meaning, Part 1*, ed. by M. M. Thelen and B. Lewandowska-Tomaszczyk, 129–39. Maastricht: Euroterm.
- Hudson, R. 1995. *Word Meaning*. London: Routledge.

- Hüllen, W. ed. 1994. *The World in a List of Words*. Tübingen: Max Niemeyer.
- Kay, C. J. and Roberts, J. 1994. Thesaurus. In *The Encyclopedia of Language and Linguistics* (10 vols.), ed. by R. E. Asher, vol. 9, 4603–5. Oxford: Pergamon Press.
- Kingsley, H. H. and Heffner, E. D. 1984. *Mastering Word Skills: Dictionary, Thesaurus, Vocabulary*. Columbus, OH: Charles E. Merrill.
- Kipfer, B. A. 1992. *Roget's 21st Century Thesaurus in Dictionary Form*. New York: Dell; London: Robert Hale, 1993.
- Kirkpatrick, B. comp. 1995. *The Concise Oxford Thesaurus*. Oxford: Oxford University Press.
- Lipka, L. 1990. *An Outline of English Lexicology*. Tübingen: Max Niemeyer.
- Lyons, J. 1995. *Linguistic Semantics*. Cambridge: Cambridge University Press.
- Marello, C. 1990. The Thesaurus. In *Wörterbücher/Dictionaries/Dictionnaires* (3 vols.), ed. by F. J. Hausmann, O. Reichmann, H. E. Wiegand, and L. Zgusta (1989–91), vol. 2, 1083–94. Berlin: Walter de Gruyter.
- McArthur, T. 1981. *Longman Lexicon of Contemporary English*. Harlow, Essex: Longman.
- McArthur, T. 1986. Thematic Lexicography. In *The History of Lexicography*, ed. by R. R. K. Hartmann, 157–66. Amsterdam: John Benjamins.
- McArthur, T. 1994. Reference Materials and Their Formats. In Hüllen ed. 1994, 1–16.
- McCutcheon, M. 1995. *Roget's Superthesaurus*. Cincinnati: Writer's Digest Books.
- Miller, G. A. 1991. *The Science of Words*. New York: Scientific American Library.
- Piotrowski, T. 1994. British and American Roget. In Hüllen ed. 1994, 123–36.
- Riggs, F. W. 1989. Review of *The Wordtree*, by H. G. Burger. *International Journal of Lexicography* 2, no. 2 (Summer): 147–51.
- Roget, P. M. 1852. *Thesaurus of English Words and Phrases*. London: Longman, Brown, Green, and Longmans.
- Stein, J. M. and Flexner, S. B. eds. 1984. *The Random House Thesaurus. College Edition*. New York: Random House.
- Summers, D. ed. 1993. *Longman Language Activator*™. Harlow, Essex: Longman.
- Urdang, L. 1991. *The Oxford Thesaurus*. Oxford: Clarendon Press.
- Wierzbicka, A. 1992. The Search for Universal Semantic Primitives. In *Thirty Years of Linguistic Evolution*, ed. by M. Pütz, 215–42. Amsterdam: John Benjamins.

### *Opći tezaurus: njegovu tko i što*

Opći tezaurus jest posebna vrsta priručnika koji karakteriziraju semantička organizacija, onomaziologija i popisi nedefiniranih semantički organiziranih leksičkih jedinica. Ipak, u literaturi se najčešće napominju princip sinonimije i kontrast između abecednog i konceptualnog makrostrukturnog načela.

Tezaursi se uglavnom temelje na semantičkim odnosima ili semantičkim poljima, te prilično nejasnim leksičkim odnosima i asocijativnim poljima. Njihovo osnovno načelo jesu općenite nesubjektivne asocijacije između riječi, no broj uspješnih neabecednih leksičkih radova veoma je malen. Problem je osobito u malom broju studija o porabama i porabnicima tezaursa. U svakom slučaju potrebno je sastavljati tezaurse na temelju raščlambe potreba korisnika, čime bi se mogli dobiti odgovori na pitanja u svezi s oblikom i tipičnom namjenom tezaursa te profilom njegovih porabnika.