Meaning Sort
— Three Examples: Dictionary Construction, Tagged Corpus Construction, and Information Presentation System —

Masaki Murata, Kyoko Kanzaki, Kiyotaka Uchimoto, Qing Ma, and Hitoshi Isahara

Communication Research Laboratory, MPT, 2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0289, Japan, {murata,kanzaki,uchimoto,qma,isahara}@crl.go.jp, http://www-karc.crl.go.jp/ips/murata

Abstract. It is often useful to sort words into an order that reflects relations among their meanings as obtained by using a thesaurus. In this paper, we introduce a method of arranging words semantically by using several types of 'is-a' thesauri and a multi-dimensional thesaurus. We also describe three major applications where a meaning sort is useful and show the effectiveness of a meaning sort. Since there is no doubt that a word list in meaning-order is easier to use than a word list in some random order, a meaning sort, which can easily produce a word list in meaning-order, must be useful and effective.

1 Using Msort

Arranging words in an order that is based on their meanings is called a meaning sort (Msort). The Msort is a method of arranging words by their meanings rather than alphabetically. The method used to list the meanings is described in the next section.

For example, suppose we obtain the following data in a research project:

| an event |
| --- |
| a temple, a formal style, an alma mater, to take up one’s post, the Imperial Household, a campus, Japan, the Soviet Union, the whole country, an agricultural village, a prefecture, a school, a festival, the head of a school, an established custom, a government official, a celebration, a Royal family |

This is a list of noun phrases (NPs), each followed by the word gyoji (an event) in the form NP X no gyoji (an event of NP X) in Japanese. To find the most useful way to examine the list, we first arrange the NPs alphabetically:

\[ \text{We actually obtained this data from the EDR co-occurrence dictionary.} \]
an agricultural village, an alma mater, a campus, a celebration, an established custom, a festival, a formal style, a government official, the head of a school, the Imperial Household, Japan, a prefecture, a Royal family, a school, the Soviet Union, to take up one’s post, a temple, the whole country

This list is not easy to use, so we next arrange the NPs by frequency of appearance:

an established custom, a school, a formal style, Japan, a prefecture, the whole country, a temple, an agricultural village, a Royal family, the Soviet Union, a festival, a campus, to take up one’s post, a celebration, an alma mater, the Imperial Household, a government official, the head of a school

Yet, even arranged this way, it is too difficult to use the list.

We then use an Msort to arrange the NPs semantically, by using following categories: Human, Organization, and Action:

(Human) the Imperial Household, a Royal family, a government official, the head of a school

(Organization) the whole country, an agricultural village, a prefecture, Japan, the Soviet Union, a temple, a school, a campus, an alma mater

(Action) a celebration, an established custom, a formal style, to take up one’s post, a festival

This list is much easier to use than a listing in alphabetical or frequency order. Note that the words in each line are also arranged in an order that reflects relations among their meanings. For example, Japan and the Soviet Union are listed side by side, as are a school, a campus, and an alma mater.

Although the list shows a variety of events, we can see at a glance that some are events related to certain special persons, and some are events related to a certain organization, and the others are miscellaneous forms of events.

The Msort is also applicable to other situations as described in later sections. The Msort enables users to more easily and efficiently recognize and examine various types of problems.

2 Implementing Msort

To sort words in an order that reflects relations among their meanings, we first need to determine an order for the meanings. The Japanese thesaurus Banrui Goi Hyou [10], an ‘is-a’ hierarchical thesaurus, is useful for this. We refer to it as BGH. In BGH, each word has a category number. In the electronic version of BGH, each word has a 10-digit category number that indicates seven levels of