Håvard Hjulstad

TERMINOLOGY WORK IN NORWAY AND THE NORWEGIAN TERM BANK

The Norwegian Centre for Technical Terminology[1] has been organizing work on technical terminology in Norway since the 1940s. It has published 45-50 dictionaries and given advice on terminological matters to industry and the authorities. The Norwegian Standards Association (Norges Standardiseringsforbund) has published a number of standards containing technical terminology; and the Norwegian Language Council (Norsk språkråd) have participated in some terminology projects.

Most published dictionaries cover a specific subject field. Most contain less than 2,000 entries. Norwegian is the primary language, and there may be equivalents in English, German, French, Swedish, and sometimes other languages. Dictionary entries are written by subject specialists in collaboration with terminologists. There is, however, a shortage of trained terminologists, and a relatively large portion of the work is being done by the subject specialists.

In 1979 the Norwegian Term Bank (NoTe)[2] was established at the University of Bergen. Its aim is to coordinate, or at least record, terminological projects going on in Norway. It also develops computational methods in terminology and lexicography, and maintains a national data-base on terminology. Most terminological dictionaries published in Norway since 1979 have been edited and printed with the assistance of the NoTe computer. Some terminological research and teaching at university level is also being promoted. NoTe is cooperating with all the other terminological institutions in Norway.

When NoTe was set up, the cooperating institutions were working on a number of dictionary projects. These dictionaries were to be printed using the NoTe computer. In order not to delay production, a working program system had to be established that would pay special attention to the problems of producing a printed book, such as printing styles, codes for special characters, margins and indentation. It was also to produce indexes in different languages. Direct access to data by other users was not needed at that time, but would be developed at a later stage.

The first task was to work out a format for terminological records. Each record should contain information about one concept. In this way the record is identical with the entry in a printed book. It is, however, essential to let the record format differ from the book entry in at least two respects: (1) division into explicit fields, and (2) inclusion of all relevant conceptual information.

A field usually starts off with a field code. It is this code, and not the position within the record, that indicates what kind of information the field contains. The NoTe field codes are multi-dimensional. The first portion consists of a language

identification tag, using the ISO two-letter symbols for languages. The second portion is a field label which shows whether the field in question is a primary term field, a synonym field, a definition field, a context field, etc. It is also possible to add extra fields or information on terms that need special treatment in order to be correctly sorted in alphabetical order (γ-ray, 0,0-dimethyl-S-ditio-phosphate).

In a printed dictionary there are usually many cross-references, in particular synonymous terms and deprecated terms. A computerized dictionary should contain the cross-reference information in the record referred to. This complies with the idea of a conceptual structure and enables us to keep track of references while editing and updating the dictionary. The NoTe system keeps explicit cross-reference information in the target record, while implicit references in definition texts are marked for easy checking.

So far the NoTe data-base has been composed of several sequential files. A large number of programs have been written to handle data in these files. Now we are ready to develop a direct-access data-base.

The NoTe data-base is not just an automatic dictionary. We are thinking in terms of a 'knowledge bank', containing four categories of information: (1) data on single concepts; (2) texts; (3) bibliographical data; (4) graphic data. Traditional term banks usually include information of the first kind only. Each concept is defined, the relevant classification codes and terms are added, and the position within a structured system of concepts is indicated. Texts may be relevant manuals and textbooks of various kinds entered in full form, while the bibliographical data also include references to sources not found in the data-base itself. The graphic data to be included are in the main relatively simple line drawings. Later it may be possible to enter more complicated figures or even moving pictures.

The term/concept relationship forms the centre-piece of the data-base structure. It is necessary to link the information between the terminological and conceptual records. A particular field within the terminological record may relate to the record as a whole, or to another field. Giving information on gender or pronunciation, for example, it is obviously essential that it is connected with the correct term in the appropriate language. It may be practical to separate the processing of terms (and formal information relating to single terms) from the processing of concepts. A separate conceptual record would contain definitions, classification codes, and pointers to structured systems of concepts. There may also be direct pointers between the conceptual records. The terminological records would each contain one term with formal information such as gender, inflection and pronunciation. Relevant conceptual and terminological records would be inter-connected both ways by means of pointers. Several terminological records may have pointers to and from the same conceptual record; in cases where one term is used for different concepts, that terminological record is entered only once, with pointers to and from several conceptual records.

The various parts of the data-base must be connected with the term/concept system. A search may start by accessing a term; the relevant term/concept records then appear, together with information on occurrences in other parts of the data-base. The user may then read the texts referred to, look at the bibliographical information given, or have figures supplied. The connections will usually be one-way, from the term/concept records to the other parts.

Searches using display terminals locate the information from the data-base by one of five main output methods: (1) elaborate paper edition, book; (2) simple paper edition, print-out; (3) microfilm; (4) display terminal, tele-data; (5) machine-readable edition, customer data-base. Each of these methods may have several variations, and different sets of equipment require different treatments. New equipment and output techniques are also constantly being developed. The internal representation of the different types of data must therefore be of a general nature and independent of output equipment.

Commercial data-base systems are not primarily made to meet the needs of linguistic data banks. It is important to develop the appropriate data-base structure for these data banks, to allow the improved processing of their complex data.

Notes

[1]   The Norwegian Centre for Technical Terminology: Rådet for teknisk terminologi (RTT), Riddervolds gate 3, Oslo 2, Norway.

[2]   The Norwegian Term Bank: Norsk termbank, Strømgaten 53, Bergen, Norway.