# A taxonomy for dictionary data and its use for a scholarly historical dictionary[1]

## Jane Rosenkilde Jacobsen and Ebba Hjorth

In a small linguistic area like Denmark, the growing internationalization of society is constantly increasing the need for foreign-language dictionaries. To meet this need it will be essential that we can recycle lexicographical data from new and existing dictionaries. This is one of the reasons why an inter-institutional group under the auspices of the Danish Research Council for the Humanities[2] has worked out a proposal for a general taxonomy for lexicographical information, primarily with a view to data-processing.

By a **taxonomy** we understand a classification of the type of information that can appear in dictionaries, in contrast to a *format*, meaning the editorial conventions of the individual dictionary formulated as a structured subset of the categories of the taxonomy required by the dictionary in question. We understand *dictionary* in the broadest sense of the term, as any body of lexicographical data, irrespective of organizing criteria.

Classifying the information for data-processing is nothing new. Classifications have been suggested for various types of dictionary — for example, for monolingual dictionaries (e.g. Hess/Brustkern/Lenders 1983: 33 ff., Schaeder 1981: 101 ff., LDOCE 1978: VIII ff., Michiels/Noël 1984: 386), bilingual dictionaries (e.g. Wilton 1978: 408, Grundt 1983, Al 1987) and term banks (e.g. DANTERM record 1984, 1986).

Unlike such classifications, our proposal for a taxonomy has a broader aim. The information categories are generalized so that data from many different types of dictionary can be covered.

We are presenting a descriptive apparatus that is independent of the issue of what a *word*, a *lexeme* or a *term* is. For the meaning of these expressions is closely connected with one's intentions with the lexicographical material. One

---

[1]  This paper is based on:
Jane Rosenkilde Jacobsen: A Taxonomy for Lexicographical Data & Ebba Hjorth: The Use of a General Taxonomy for the Processing of a Historical Dictionary.

[2]  The research group consists of the following scholars:
Gert Engel, The Southern Jutland School of Economics and Business Administration; Ebba Hjorth, The Society for Danish Language and Literature, Copenhagen; Jane Rosenkilde Jacobsen, University of Copenhagen; Bodil Nistrup Madsen, The Copenhagen School of Economics and Business Administration; Ole Norling-Christensen, Gyldendal Publishers, Copenhagen; Hanne Ruus, University of Copenhagen. A fuller account of the work of the research group will be available in "Lexicographica. Series Maior". Tübingen, Autumn 1987.

chooses one's point of departure for each different work or project; that is, one selects the units one wishes to give information about.

The choice of approach, onomasiological or semasiological, and the precise nature of the units to be dealt with must, then, be decided for every work or project. These factors vary in ways that are not amenable to treatment in a general taxonomy.

There are two other elements in the description of lexicographical works and projects that belong to the characterization of works as a whole rather than to that of the categories of information they contain: information about the lan-

**Project-specific**

| Main groups | Categories |
|---|---|
| Language | - language stage |
| | - dialect |
| | - sociolect |
| | - national language |
| Approach | - semasiological |
| | - onomasiological |
| Medium of expression | - writing |
| | - picture |
| | - speech |

**General**

| Main groups | Categories |
|---|---|
| Etymological information | - origin |
| | - parallel |
| Grammatical information | - part of speech |
| | - flexional information |
| | - word formation |
| | - syntactic information |
| Graphic information | - orthographic information |
| | - graphic symbol |
| Phonetic/phonological information | - suprasegmental features |
| | - segmental features |
| Pragmatic information | - textual/contextual information |
| | - usage |
| | - external reference |
| | - evaluative information |
| | - administrative information |
| Semantic information | - subject classification |
| | - semantic relations |
| | - content specification |
| | - equivalence |

Fig. 1: Proposed taxonomy for lexicographical data.

guage or languages dealt with in the work, and information about the medium of expression used for the work.

Fig. 1 gives an overview of the proposed taxonomy. It is divided into three project-specific classes of information and six general ones. Each class of information constitutes a main group that is further sub-classified into a number of categories.

The taxonomy builds on comparative studies of a number of different types of lexical work where the Danish language is involved (Hjorth 1984, 1985); on a review of various proposed formats for lexical data (Jacobsen 1984); and on lexicographical scholarship in general. The fineness of the distinctions in the main groups and categories is related to needs as we perceive them and the consensus on subclassification registered in our surveys.

We shall now describe the various main groups and categories and exemplify their use by relating them to information in a comprehensive Danish-English dictionary (VB) (1976); some examples will be taken from THE CONCISE OXFORD DICTIONARY (COD) (1972) and from the LONGMAN DICTIONARY OF CONTEMPORARY ENGLISH (LDOCE) (1978). The examples are given in the figures.

The three project-specific types of information, *language, approach* and *medium of expression*, determine the nature of the lexical units one deals with in the specific work, and assign the work to a particular lexicographical discipline with its traditions.

Every lexicographical work deals with one or more languages. The definition of **language** depends on the nature of the lexicographical work; as shown in Fig. 2 it may be a matter of stages of a language, as in historical dictionaries; of the languages of different regions, as in dialect dictionaries; of different sociolects, as in slang dictionaries, for example; of different national languages, as in all translation dictionaries.

As far as **approaches** (Fig. 2) are concerned, the basic dividing-line goes between two types: general lexicography that takes *words* for granted and aims to comment on their form, structure, meaning and usage, the semasiological approach; this approach involves a definition of headword, lexeme and lemma, for example; on the other hand terminography and technical or specialized lexicography take concepts for granted and describe these by means of definitions and the terms and words used to express these concepts (e.g. Felber 1984: 116, 189, 225). This onomasiological approach is committed to the definition of concepts and terms.

Until recently lexicographical works were presented in the form of printed books. The written word was the predominant **medium of expression** (Fig. 2), with occasional examples of the use of pictures. But with the modern media of expression the written word, pictures and sound can all be combined. For example, one could have a translation dictionary available in an information retrieval system where one could learn the pronunciation of the word one "looked up".

| Main groups | Categories | Examples |
|---|---|---|
| - language | - language stage | |
| | - dialect | |
| | - sociolect | |
| | - national language | Oxford English Dictionary |
| | | VB: Dansk-engelsk ordbog |
| | | (Danish-English Dictionary) |
| - approach | - semasiological | VB: headword/lemma |
| | | VB: sublemma } Danish lexical units |
| | | VB: equivalent } English |
| | - onomasiological | concept |
| | | term |
| - medium of expression | - writing | Grolier's Encyclopedia |
| | | LDOCE |
| | | Van Dale |
| | | VB |
| | - picture | Grolier's Encyclopedia |
| | | LDOCE |
| | - speech | Grolier's Encyclopedia |
| | | Van Dale |

Fig. 2: The main groups: language, approach and medium of expression.

So the authors of a lexicographical work must now consider which medium of expression is best suited to their data and users.

The central feature of the general component of the proposed taxonomy (see Fig. 1) is its nineteen mutually exclusive categories of information. Thus any given item of information can be assigned to one and only one category.

On the basis of their dominant features the categories are arranged in main groups corresponding to the various linguistic disciplines: etymology, grammar, orthography, phonetics and/or phonology, pragmatics and semantics. This grouping is based on one overall view of the interrelations among the different linguistic disciplines. There is less general agreement, however, on the exact placing of some of the linguistic categories in the main groups. For example, information about word-formation is sometimes considered to be etymological, like information about cognate words etc., and sometimes, as here, grammatical, like flexional information. The fact that it is possible to classify some categories in one or the other main group does not detract, however, from the mutually exclusive nature of the categories. It is rather a symptom of a certain amount of competition between the linguistic disciplines.

In the overview the main groups are ordered alphabetically to bring out the mutually exclusive nature of the categories. The individual categories themselves are quite comprehensive and can be subdivided in different lexicographical works according to the work-specific degree of specialization in particular types of information. Subdivisions, too, should be mutually exclusive. In the following we show some subcategories which will be sufficient for the great majority of dictionaries, but they should not be considered as exhaustive. Their realization in actual formats may require further subdivisions.

**Etymological information** (Fig. 3) tells us about the *origins* of words and the forms and meaning of *parallel* or cognate words in other languages.

**Grammatical information** includes morphological and syntactic categories: *part of speech, flexional information, word-formation* and *syntactic information.* Fig. 3 shows one application of the categories — with the subcategories relevant to the Danish-English Dictionary (VB 1976) and some proposals for other subcagories.

**Graphic information** (Fig. 4) tells us how words and concepts appear in writing. *Orthographic information* provides us with the *alphabetical form* of a given word or expression. Other items of orthographic information are *variant spellings* and *abbreviated forms of words. Graphic symbol* specifies the symbol used for a given word or concept, for example the plus sign (+) for *plus* or *plus sign.*

**Phonetic** and/or **phonological information** (Fig. 4) consists of two categories: *suprasegmental features* informs us about stress, intonation and accent. This is the type of phonetic information most often found in lexicographical works. Very few dictionaries can manage without indicating the stress in some words. *Segmental features* indicate the pronunciation of segments using a phonetic notation like the IPA or an orthographical notation that approximates the pronunciation.

**Pragmatic information** (Fig. 5 & 6) collects together the information categories that indicate use in the broadest sense — both the use of the words described and the use of the lexical information.

*Textual/contextual information* (Fig. 5) comprises the subcategories *quotation* and *possible context. Quotations* are authentic examples of the use of words and expressions, with the actual wording of the source. *Possible context* gives examples of the use of a given word or expression.

*Usage* (Fig. 5) places the use of a given word or expression in a *temporal, spatial* or *social dimension,* or tells us about how often a word or expression is used — its *frequency.* In the temporal dimension we can be informed, for example, that a word is old-fashioned or obsolete. As far as the spatial dimension is concerned, we can be told that a word or expression is used in a particular geographical area. In the social dimension we register use in specific social strata or contexts, genres, for example slang or colloquialisms. One type of frequency information can be provided in the Danish-English Dictionary (VB 1976) by *(ofte =).*

| Main groups | Categories | Subcategories | Examples |
|---|---|---|---|
| - etymological information | - origin | | f. Gmc *horsam, -az |
| | - parallel | | OE hors / OS, OHG hros / ON hross |
| | | | COD: [OE hors, OS, OHG hros, ON hross |
| | | | f. Gmc *horsam, -az] |
| - grammatical information | - part of speech | | adj |
| | | | VB: pæn adj ... nice ... |
| | | - subclassification (gender) | et |
| | | | VB: æble et -r apple ... |
| | - flexional information | - flexional ending | -r |
| | | | VB: æble et -r apple ... |
| | | - flexional form | bedre, bedst |
| | | | VB: god adj (bedre, bedst) ... good ... |
| | - word formation | | compound |
| | | | VB: folkevandring migration. |
| | | | VB: vandring en -er ... |
| | | - elements of compound word | |
| | | - compositional element | |
| | | - root | |
| | | - derivational affix | |
| | - syntactic information | - valency | i |
| | | | VB: gnistre vi sparkle ... |
| | | - syntactic function | alm kun prædikativt |
| | | | VB: glad adj ...; (alm kun prædikativt) |
| | | | glad ... |

Fig. 3: The main groups: etymological information and grammatical information.

| Main groups | Categories | Subcategories | Examples |
|---|---|---|---|
| - graphic information | - orthographic information | - alphabetical form | bister / fierce / grim / gruff<br>se bister ud / frown<br>VB: II. **bister** <u>adj</u> fierce, grim gruff; <u>se ~ ud</u> (<u>ogs</u>) frown. |
| | | - orthographic variant/ variant spelling | zibeth<br>VB: **zibet** <u>en zo</u> (Viverra zibetha) zibet(h), Indian civet.<br>Diesel<br>VB: **diesel**- diesel (<u>el.</u> Diesel).<br>Diesel<br>VB: **diesel**- diesel, Diesel. |
| | | - abbreviated form of word | fk kompagni<br>VB: **co.** (<u>fk kompagni</u>) Co.<br>fk CPI<br>VB: **forbrugerprisindeks** consumer price index (<u>fk</u> CPI). |
| | - graphic symbol | | +<br>VB: **plustegn** plus (sign) (+). |
| - phonetic/ phonological information | - suprasegmental features | | 'fortegning<br>VB: I. **'fortegning** <u>en -er</u> (<u>model</u>) model.<br>for'tegning<br>VB: II. **for'tegning** <u>en -er</u> (<u>forkert tegning</u>) incorrect drawing, (<u>ogs fig</u>) distortion. |
| | - segmental features | | haʋs<br>LDOCE: **house**[1] /haʋs/ <u>n</u> ... |

Fig. 4: The main groups: graphic information and phonetic/phonological information.

| Main groups | Categories | Subcategories | Examples |
|---|---|---|---|
| - pragmatic information | - textual/ contextual information | - quotation | "We can send you a map if you wish" LDOCE: if[1]... conj ... "We can send you a map if you wish" (SEU S.) |
| | | - possible context | fx the devil's advocate VB: advokat en -er ... 5 (fig) advocate (fx the devil's advocate) ... fx money, wages VB: fortjent adj ...; surt ~ hard-earned (fx money, wages); ... |
| | - usage | - temporal dimension | glds VB: fæstemand (glds) betrothed. |
| | | - spatial dimension | amr VB: efterår autumn; (amr) fall; ... |
| | | - social dimension | S / T VB: III. fyr (mandsperson) fellow, chap; T feller; S bloke; ... |
| | | - frequency | ofte = VB: fuldkornsbrød wholemeal bread; (ofte =) brown bread. |
| | - external reference | - literature reference | |
| | | - source reference | SEU S. LDOCE: if[1]... conj ... "We can send you a map if you wish" (SEU S)... |

Fig. 5: Main group: pragmatic information (1).

| Main groups | Categories | Subcategories | Examples |
|---|---|---|---|
| - pragmatic information | - evaluative information | | NB firm of barristers bruges ikke<br>VB: **advokatfirma** firm of lawyers; (2) firm of solicitors; (3: NB firm of barristers bruges ikke); ... |
| | - administrative information | - collector<br>- editor<br>- other processor<br>- compilation date<br>- data entering date | VB: front-matter/slips |
| | | - internal reference | se empiriker<br>VB: **empirist** en -er se empiriker.<br>se A: han har gode forbindelser<br>VB: **forbindelse** en -r ...; B: ...; han har -r, se A: han har gode -r; ... |
| | | - homograph number | I. / II.<br>VB: I. **dukke** en -r ... doll; ...<br>II. **dukke** vt duck, ... |
| | | - technical information | VB: lemma: semi-bold<br>VB: sublemma: italics<br>VB: equivalent: roman<br>VB: end of entry: full stop<br>VB: usage: in brackets |

Fig. 6: Main group: pragmatic information (2).

*External reference* is reference to the literature and source references. A *literature reference* recommends finding more information in another work. A *source reference* also points outside the given lexicographical context. *Evaluative information* (Fig. 6) is especially found in normative lexicographical works, in the form of recommendations of or warning against particular words or phrases.

*Administrative information* (Fig. 6) indicates the *collector, editor* or *other processor* of the data described, and the *compilation* and *data entering dates. Internal references* (i.e. references to other information in the same collection of data) and *homograph numbers* are also administrative information. In addition there is *technical information,* which can for example be about the use of type-faces like semi-bold for the lemma, italics for the sublemma and Roman for the equivalents, etc., or the use of punctuation marks like the comma, semicolon, full stop and various kinds of brackets: for example, full stop for the end of an entry, or brackets around usage information. Information like this is normally stated explicitly in the list of editorial conventions and front matter, and of course manifests itself throughout the work in the individual entries.

**Semantic information** (Fig. 7 & 8) serves to identify and describe the content of a linguistic sign – its meaning – and corresponds to what in lexicography is normally called the description of meaning in the broadest sense. *Semantic information* can be subdivided into the categories *subject classification, semantic relations, content specification* and *equivalence.* Most of the subcategories to be shown are used in specialized or technical lexicography; they can also be found in general lexicography, but are rarely used systematically. The examples are a taxonomic characterization of some types of information that do in fact exist in the Danish-English Dictionary (VB 1976), but which this dictionary only in rare cases explicitly marks as this type of information. In other words, this dictionary would often not be able to take part in any systematic exchange of data within the main group, semantic information.

*Subject classification* (Fig. 7) provides information about fields or subjects. Field information is regarded as a characteristic similar to other characteristics describing properties (for example, function, position, construction, purpose, components) and can thus be included in the definition itself, that is, can be a part of *content specification* (cf. Fig. 8). But this information can also be given separately, for example as a notation or code belonging to a more or less detailed *classification system*, whether general or worked out for a given collection of data. In many dictionaries there is no more than a rough subject classification in the form of an alphabetical list of *subject abbreviations,* for example *med.* for medicine, *tekn* for technical. The Danish-English Dictionary (VB 1976) also uses *unabbreviated subject information* for meaning discrimination. Subject information may also be given by the title of a specialized dictionary or data collection.

By information about *semantic relations* (Fig. 7) we understand descriptions of the relationships between concepts – for example hyponomy, hyperonymy and co-hyponymy. Usually this is a matter of the internal systematization of a

Main group: semantic information

| Categories | Subcategories | Examples |
|---|---|---|
| - subject classification | - classification system | EURODICAUTOM: Lenoch |
|  |  | Scandinavian Term Banks: NORDTERM |
|  |  | Some terminological data collections: UDC |
|  | - subject abbreviation | med. |
|  |  | VB: **hjerteforkalkning** (med.) coronary sclerosis. |
|  |  | tekn |
|  |  | VB: **hjulnøgle** (tekn) wheel wrench |
|  | - unabbreviated subject information | på hjuldamper |
|  |  | VB: **hjulaksel** ...; (på hjuldamper) |
|  |  | paddle shaft, wheel shaft. |
| - semantic relations | - concept system |  |
|  | - generic super-/subordinate relationship | legems- / maskin- |
|  |  | VB: III. **skade** en -r ...; (legems-) injury; |
|  |  | (maskin-) (engine) breakdown, ... |
|  |  | vandløb |
|  |  | VB: I. **flod** en -r 1 (vandløb) river; |
|  |  | 2 (mods lavvande) high tide ... |
|  | - partitive super-/subordinate relationship |  |
|  | - successive relationsship |  |
|  | - causal relationship |  |
|  | - antonymy | mods lavvande |
|  |  | VB: I. **flod** ... 2 (mods lavvande) high tide ... |
|  | - thesaurus |  |
|  | - descriptor |  |
|  | - non-descriptor |  |
|  | - numbering of meanings |  |

Fig. 7: Main group: semantic information (1).

Main group: semantic information

| Categories | Subcategories | Examples |
|---|---|---|
| - content specification | - analytical definition (intensional/extensional) — | forkert tegning<br>VB: II. **for'tegning** <u>en —er</u> <u>(forkert tegning)</u> incorrect drawing, <u>(ogs fig)</u> distortion.<br>LDOCE |
| | - synthetic definition (relational) | |
| | - denotative definition (by exemplification) | |
| | - ostensive definition (e.g. illustrations) | LDOCE: pictures |
| - equivalence | <u>within a single language</u> | |
| | - synonyms/quasisynonyms (of lemma/sublemma) | = tage fast form<br>VB: **fæstne** vt ...; ~ sig (= tage fast form) assume definite form; ... |
| | - synonyms/quasisynonyms (of equivalent) | [,] fierce / grim / gruff<br>VB: II. **bister** <u>adj</u> fierce, grim, gruff; ... ǀ<br>[()] engine breakdown / breakdown<br>VB: III. **skade** ... <u>(maskin-)</u> (engine) breakdown, ...<br>[(el.)] by a river / near a river<br>VB: I. **flod** ...; <u>ved en ~</u> by (<u>el.</u> near) a river; ...<br>[(, )] a packet of pins / a packet of needles<br>VB: **nål** <u>en —e</u> ...; <u>et brev ~e</u> a packet of pins (, needles); ... |

between two languages

- full equivalence                    type-face change to roman

                                      VB: II. **bister** adj fierce, ...

- partial equivalence                 omtr = / type-face change to roman

                                      VB: **frugtgrød** (omtr =) stewed fruit.

                                      svarer til / type-face change to roman

                                      VB: **folketingstidende** [official report of parliamentary

                                          records]; (i Eng. svarer til) Hansard.

- zero equivalence                    kan gengives / type-face change to roman

                                      VB: **frugtfromage** (kan gengives) fruit cream,

                                          fruit mousse.

                                      [] / type-face change to roman

                                      VB: **folketingstidende** [official report of

                                          parliamentary records]; ...

Fig. 8: Main group: semantic information (2).

particular dictionary or data collection: for example a *concept system* worked out on the basis of a terminological analysis of a restricted field. The systematization may be realized as a hierarchy of *generic* and/or *partitive superordinate* and *subordinate relationships*. In many dictionaries information about semantic relations is used as a means of discriminating between meanings. The system of concepts may also be based on a *successive relationship* (according to the order of the concepts in a process); a *causal relationship* (based on cause and effect); or may feature many other more or less well-defined relationships, including *antonymy*. A special type of lexicographical data collection is the *thesaurus* for information and documentation retrieval. Such a thesaurus lists the permissible expressions, the *descriptors* for indexing and accessing information in databases. The descriptors are ordered systematically to show the relations between them. In addition, non-admissible expressions, synonyms or *non-descriptors*, are listed. The *numbering of meanings* in dictionary entries can also be regarded as a form of semantic structuring. Often a special significance is attached to the order of the meanings.

By *content specification* (Fig. 8) we understand the description of meaning in the narrow sense (definitions). Definitions can be *analytical* (intensional or extensional), *synthetic* (relational definitions), *denotative* (by exemplification) or *ostensive* (for example, an illustration) (e.g. Frandsen 1982). Primarily analytical definitions may be formulated in a formalized defining vocabulary as in the LDOCE (1978) or may make use of formalized semantic features. Information on conceptual content also includes information on subjects and semantic relations. These types of information are therefore often included in definitions, cf. what was said under subject classification and semantic relations.

*Equivalence* (Fig. 8) informs us about similarities and differences between concepts, i.e., common and distinctive semantic features. Equivalence relations *within one language* are characteristic of *synonyms* and *quasi-synonyms*. In many dictionaries synonyms are only indicated typographically, for example by means of changes in type-face or equal-signs. Synonyms and quasi-synonyms in the target language of a bilingual dictionary are often listed separated only by a comma or semicolon. Translation equivalents indicate relations of equivalence *between two languages. Full equivalence* is often signified by zero marking, frequently a change in type-face. Information on *partial equivalence* is found in traditional lexicographical works in the form of various comments before the type-face change. The information that a concept in one language has no direct correspondence in another, i.e. information on *zero equivalence*, and that the item listed is a proposed translation, is also equivalence information.

The justification for this taxonomic proposal (see Fig. 1) is that dictionary data should be capable of electronic processing, and that individual dictionaries should be able to take advantage of one another's data. It is a necessary condition for such cooperation that data are classified in accordance with a common taxonomy. This gives us the following advantages:

— we achieve a non-ambiguous description of the content of the dictionaries that will be available for comparative work;

— the totally or partially automatic exchange of lexicographical data between different dictionaries becomes possible;

— lexicographers will be enabled to compile dictionary entries with the greatest possible consistency and least possible waste of time;

— it will be possible to produce dictionaries with different contents on the basis of the same lexicographical data;

— the same dictionary can be presented in different ways, both in print and in other media;

— it will become easier to consult, revise and update dictionary data.

We shall now describe why it is necessary to employ a general taxonomy for the processing of a historical dictionary and the advantages to be gained by so doing. The dictionary in question is the Old Danish Dictionary (Gammeldansk Ordbog).

The aims of the Old Danish Dictionary are:

— to present a linguistic description in lexical form of the Danish language in the period between about 1100 and 1515;

— to provide a philological reference work that will be of assistance to the reader of texts from the Old Danish period.

The basic material for the Old Danish Dictionary has been handed down directly from the medieval period, either in manuscript form or as a printed text. The work of assembling a collection of slips upon which the editorial process could be based began in 1956. The collection of slips was built up according to familiar and well-tried principles: the transcription of source material, duplication of the transcripts, alphabetization etc. All the different phases in the production of the slips have been carefully controlled to avoid errors. The collection of material, which is now more or less complete, consists of approximately 1 million slips (Fig. 9). According to the original plan, this collection of slips should result in a nine-volume dictionary. It is intended that the dictionary, like other scholarly, historical dictionaries, should contain:

— information that serves to identify the headword;

— information about the inflexion of the headword;

— information about the history and etymology of the headword;

— information about the semantics of the headword.

These four main types of information will be documented in a scholarly dictionary in various ways:

— by quotations from, or references to, the sources upon which the dictionary is dependent;

— by references to relevant specialist literature;
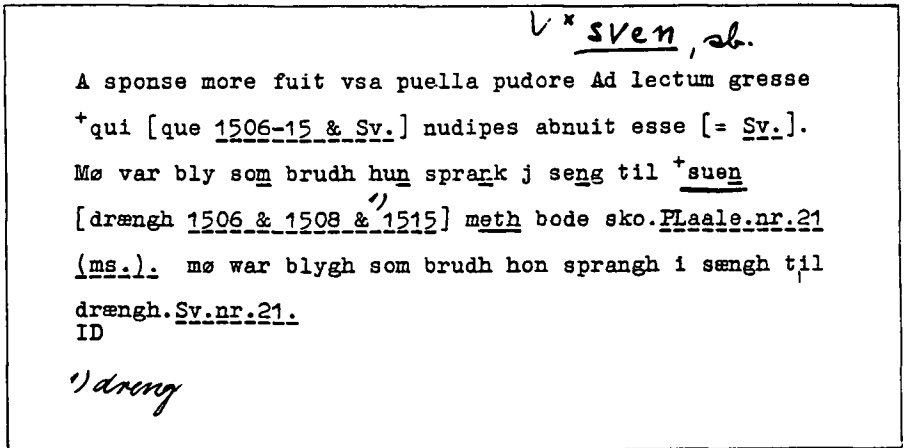
— by references to other entries in the same dictionary.

$\vee$ ˣ $\underline{sven}$ , ɔb.

A sponse more fuit vsa puella pudore Ad lectum gresse

⁺qui [que 1506-15 & Sv.] nudipes abnuit esse [= Sv.].

Mø var bly som brudh hun sprank j seng til ⁺suen

[drængh 1506 & 1508 & 1515] meth bode sko.PLaale.nr.21

(ms.). mø war blygh som brudh hon sprangh i sængh til

drængh.Sv.nr.21.
ID

¹) dreng

Fig. 9: This slip from the collection of the Old Danish Dictionary contains:
— a headword and indication of word-class;
— the Old Danish context in which the headword appears;
— the original source — here in Latin — from which the Old Danish text is translated;
— younger — but still Old Danish — parallels to the Old Danish text;
— a younger, Old Swedish parallel to the Danish text;
— a source reference.

In 1975, towards the end of the collection phase, a number of specimen articles were compiled and provisional editorial principles established. Now the collection of material is practically completed, and in working out the definitive editorial principles for the Old Danish Dictionary, it has been natural for us to take our starting-point in the general taxonomy and benefit from its advantages for editorial work.

The existing specimen entries and the provisional editorial principles have thus formed the basis for experiments with establishing a definitive format (cf. above). These entries and the provisional editorial principles had been carefully and thoroughly enough prepared to serve as a starting-point for deliberations and decisions on the types of information that can or should be included in an entry in the Old Danish Dictionary. At the same time, the specimen entries were not too numerous to preclude any changes in their structure made necessary by the wish to fix the structure of the entries on the basis of the general taxonomy.

### The project-specific types of information and the Old Danish Dictionary

**Language**. As mentioned earlier, the aims of the Old Danish Dictionary are two-fold. On the one hand the dictionary is to present a linguistic description of the Danish language in the medieval period, i.e. a description of the language stage Old Danish. On the other hand it is to be a work of reference for reading texts dating

from the Old Danish period. The semantic content of the Old Danish words is thus described with the aid of the language stage Modern Danish. In other words, the dictionary translates from *the language stage Old Danish* to *the language stage Modern Danish*. The Old Danish Dictionary is thus at once a native-language dictionary and a translation dictionary.

**Approach.** The lexical unit in the Old Danish Dictionary is any word which can have an independent or superordinate function in the context. Prefixes and suffixes, however, cannot be headwords; nor can word-combinations. The approach in the entries is *semasiological,* and they are to be arranged in strict alphabetical order.

**Medium of expression.** The medium of expression for the Old Danish Dictionary is *print.* There are no immediate plans for replacing any definitions by illustrations.

## The general types of information

We will now exemplify the *two different kinds of work-process* which have to be carried out, and describe the two different kinds of problem that must be solved for each individual item of information in a dictionary entry whose format is based on the principles of the taxonomy.

The first work-process involved is ensuring that each of the types of information in the dictionary entry can be assigned to one and only one category in the taxonomy. This work-process can be exemplified by the information-type *word-class* or *part of speech.*

As far as the indication of word-class is concerned, the editorial principles from 1975 read as follows, in extract:

> The word-class is to be indicated by such abbreviations as *sb., adj., part. adj., adv., vb.,* etc. In the case of substantives the definite singular form and the indefinite plural are to be cited whenever the recorded material permits, and the abbreviation *sb.* is then to be omitted. If a substantive is not recorded in definite form, the abbreviation *sb.* is not to be included. If attributive or predicative words reveal the gender of the substantive, then this is to be indicated by *fk.* (common gender) or *n.* (neuter).

This quotation illustrates how many years' refinement of lexicographical practice have meant that lexical information is often only implicitly present in an entry. If the editorial principles from 1975 were to be followed when establishing the definitive editorial principles, there would only be complete agreement between the information types of the general taxonomy and the information categories of the Old Danish Dictionary in cases where a substantive is not recorded in definite form. Whenever a substantive is recorded in definite form, the abbreviation *sb.* would be omitted from the space for the indication of word-class, and the information that the headword is a substantive would be implicit in the

quoted definite singular form and (where recorded) indefinite plural form. In such cases the lexical information in the Old Danish Dictionary could not be assigned to one and only one of the types of information in the general taxonomy, but could be classified both as *part-of-speech information* and as *flexional information*.

Since the expediency of the implicit form of communication is very open to question, and explicitness is a prerequisite for data-processing and data exchange, word-class should always be indicated by the abbreviations *sb., adj., vb., adv.,* etc., while flexional information should appear in the section on the inflexion of the headword (the B-section in the outline of the Old Danish Dictionary format in Fig. 10).

There will then be complete agreement between the information type in the general taxonomy and the information category in the Old Danish Dictionary, and the indication of word-class will be classified uniquely as *information on part of speech.*

One of the characteristics of a so-called scholarly dictionary is, as mentioned above, its comprehensive documentation of the information supplied in the dictionary entries. The usual method of documenting a definition is by providing a quotation from one of the sources upon which the dictionary is based. If such a quotation, in the form in which it has traditionally appeared in dictionaries, is analysed in relation to the general taxonomy, it will be revealed that it contains information from two categories, namely an element of *textual information,* the sub-category *quotation* (cf. Fig. 5), and an element of *external reference,* the sub-category *source reference* (cf. Fig. 5). In order to obtain absolute agreement between the types of information in the dictionary and the categories of the taxonomy, it is necessary to treat the documentation of a definition as consisting of two elements, namely textual information in the form of a quotation, and external reference in the form of source reference.

If we take a closer look at the sub-category *quotation* in a dictionary such as the Old Danish Dictionary, we shall find that the sub-categorization proposed by the authors of the taxonomy is not sufficient to satisfy the needs of a dictionary with the complexity of the Old Danish Dictionary. Quotations from the sources on which the dictionary is based are used as documentation for

— either a flexional form, or
— a definition,

i.e. there are two different kinds of quotations, which might be referred to as

— flexional-form quotation, and
— definition quotation.

Over and above these primary quotations, the lexicographer needs to use quotations from the foreign texts from which some of the dictionary's sources are translated or from other texts upon which the sources are dependent in one way or another. Such secondary quotations can often explain or throw light upon remark-

able word-forms in a primary quotation. It is thus necessary to operate with two additional kinds of quotations. These can be referred to as

— flexional-form quotation from original text, and
— definition quotation from original text.

Finally, it can sometimes be expedient for the lexicographer to cite a reading from a younger or contemporary parallel text, where this can throw light on a primary definition quotation. So there is yet another type of secondary quotation, which can be referred to as

— definition quotation from parallel text.

The specific aims and structure of the Old Danish Dictionary thus require the replacement of the single sub-category *quotation* in the format by these five different kinds of quotation.

This is an example of the second kind of work process which has to be carried out when establishing the format, namely the definition of the sub-categories required by the individual dictionary.

Summing up, we have thus far demonstrated the working method to be used when a format is to be worked out on the basis of the general taxonomy:

1. Provisional editorial principles are to be worked out on the basis of the general lexicographical experience of the lexicographer.
2. The types of information in the provisional format are to be confronted with the general taxonomy to determine whether the types of information of the format can be assigned to one and only one category in the taxonomy. The taxonomy thus orders lexicographical thinking, and increases the lexicographer's awareness.
3. Now the lexicographer is able to revise the structure of the dictionary entries and the revised editorial principles can in turn be confronted with the general taxonomy with a view to
   — checking that the types of information are mutually exclusive, and
   — defining sub-categories.
   For all the types of information in the dictionary processes 2 and 3 are to be repeated as often as necessary to fulfil
   — the uniqueness requirement,
   — the sub-categorization requirement.
4. Finally the definitive format can be worked out.

The next process in the establishment of the format is the fixing of rules for the order in which these types of information should appear and for whether they should be omitted or repeated. In this way we arrive at the definitive editorial principles or the format for the Old Danish Dictionary, as shown in Fig. 10.

| Field code | Type of information |
|---|---|
| (IHOM) | Indication of homograph |
| IOPO | Headword |
| (IUDN) | Indication of coefficient of utilization |
| IOKA | Indication of word-class |
| [ IOPV ]* | Variant |
| (IOVL) | Reference relating to variant |
| [ IOVK ]* | Constant part of source reference relating to variant |
| [IOVV]+ | Variable part of source reference relating to variant |
| [ IHEO ]* | Referring word |
| [IHEN]+ | Cross-reference |

I – section

| | |
|---|---|
| [ (BSUB) ]* | Sub-classification |
| BFLE | Indication of inflexion |
| (BØKO) | Comment on inflexion |
| (BØFO) | Inflexional form |
| [ BØCI ]* | Quotation showing inflected form |
| BØKK | Constant part of source reference for inflected form |
| [BØKV]+ | Variable part of source reference for inflected form |
| (BFLC) | Quotation from original text |
| (BFCS) | Language of original text |

B – section

| | |
|---|---|
| [ SOBN ]* | Overwiev definition number |
| (SOGK) | Overwiev syntactical information |
| SOBB | Overwiev definition |
| (SBEN) ]* | Definition number |
| (SSGK) | Syntactical information |
| (SORF) | Word combination |

S – section

$$\left\{\begin{array}{l}\left(\begin{array}{l}\text{SIHO}\\ \text{[SIHE]}^{+}\end{array}\right)\\ \text{(SBCI)}\\ \text{SBCP}\\ \text{SBCK}\\ \text{[SBCV]}^{+}\\ \left[\begin{array}{l}\text{SCKF}\\ \text{[SKFO]}^{+}\end{array}\right]^{*}\end{array}\right.$$

SIHO    Internal referring word

[SIHE]⁺    Internal reference

(SBCI)    Definition quotation

SBCP    Parallel of definition quotation

SBCK    Constant part of source reference for definition

[SBCV]⁺    Variable part of source reference for definition

SCKF    Original for source for definition quotation

[SKFO]⁺    Information about original for source for definition quotation

(ELDK) *    Information about loan or formation

ELPS *    Indication of language of loan/parallel

ELPO    Loan-/parallel word

(ELPI)    Information about loan-/parallel word

(ELPB)    Definition of loan-/parallel word

(ELPG)    Basic meaning of loan-/parallel word

EJSA *    Comparison

(EJSS)    Language of comparison

EJSO    Comparing word

(EJSB)    Definition of comparing word

(EEHL)    Reference to secondary literature

(EBRO)    Usage information

unmarked field code = the type of information occurs just once

( )      = the type of information occurs once or no times

[ ]⁺      = the type of information occurs once or several times

[ ]*      = the type of information occurs no times, once or several times

Fig. 10:Outline of the Old Danish Dictionary format.

# References

## Cited dictionaries

THE CONCISE OXFORD DICTIONARY OF CURRENT ENGLISH (COD)
H.W. Fowler et al., Oxford: Clarendon (1972).

DANSK-ENGELSK ORDBOG (VB)
H. Vinterberg and C.A. Bodelsen, Copenhagen: Gyldendal (1976).

GROLIER'S ENCYCLOPEDIA
Grolier Inc.: Banbury, Conn. (1986).

LONGMAN DICTIONARY OF CONTEMPORARY ENGLISH (LDOCE)
P. Procter et al., London: Longman (1978).

WEBSTER'S NEW WORLD DICTIONARY OF COMPUTER TERMS
L. Darcy and L. Boston, New York: Simon & Schuster (1984).

## Other literature

Al, B.P.F. (1988) "Ordinateur et Lexicographie. Université Libre d'Amsterdam, CNRS, UA Silex 382. Sept. 1985", in: A. Zampolli (ed.): *Scritti in Onore di Roberto Busa S.J.*

DANTERM record (1984) and (1986), Handelshøjskolen i København, Terminologiafdelingen.

Felber, H. (1984), *Terminology Manual. General Information Programme and UNISIST*, International Information Centre for Terminology, Paris: UNESCO: Infoterm.

Frandsen, L. (1982), "Definition. Objekter, metoder og regler, med terminologisk inventar for ventilation og luftkonditionering", ARK 15. Maj 1982, Det erhvervssproglige Fakultet, Handelshøjskolen i København.

Grundt, L.O. (1983), "En større norsk-fransk ordbog", in: Henrik Holmboe (ed.), *Sprog Og Mennesker*, Nordisk forskerkursus om datamatunderstøttet leksikografi, august 1982, Institut for Lingvistik, Aarhus Universitet. Århus.

Hess, K./Brustkern, J./Lenders, W. (1983), *Maschinenlesbare deutsche Wörterbücher. Dokumentation, Vergleich, Integration,* Tübingen: Niemeyer.

Hjorth, E. (1984), Rapport til Statens Humanistiske Forskningsråd over projektet "Analyse af informationerne indeholdt i danske ordbogsartikler og udarbejdelse af et standardformat egnet til elektronisk databehandling" (Journal nr. 5.25.35.00, januar 1984).

Hjorth, E. (1985), "Ordbogsarbejde og EDB", in: *Humaniora* 6, 1983—84, Forskningssekretariatet, København.

Jacobsen, J.R. (1984), Rapport til Statens Humanistiske Forskningsråd over projektet "Analyse af informationerne indeholdt i danske ordbogsartikler og udarbejdelse af et standardformat egnet til elektronisk databehandling" (Journal nr. 5.25.35.00, september 1984).

Lenoch, H. (1971—1972), *Klassifikationssystem. Verzeichnis der Haupt- und Untergruppen, erstellt für das Terminologiebüro der Kommission der europäischen Gemeinschaften,* Luxemburg.

Michiels, A./Noël, J. (1984), "The pro's and con's of a controlled defining vocabulary in a learner's dictionary", in: Hartmann, R.R.K. (ed.), *LEXeter '83 Proceedings,* Tübingen: Niemeyer, 385—394.

NORDTERM (unpublished) NORDTERM-arbejdsgruppe 2.

Schaeder, B. (1981), *Lexikographie als Theorie und Praxis,* Tübingen: Niemeyer.

Wilton, M.T. (1978), "The Canadian Dictionary: An Experiment in Automated Lexicography", in: *Vingt-cinq ans de linguistique au Canada: Hommage à Jean-Paul Vinay,* Montréal: Centre Educatif et Culturel.