

# The “Thesaurus of Old English” database: a research tool for historians of language and culture

*Lynne Grundy and Harold Short*

---

*Research Unit in Humanities Computing, King’s College London, Strand, London WC2R 2LS, England*

KEYWORDS: Old English, database, thesaurus

AFFILIATION: King’s College London

E-MAIL: L.Grundy@kcl.ac.uk

H.Short@kcl.ac.uk

FAX NUMBER: 44 171 836 1799

PHONE NUMBER: 44 171 873 2684

## Abstract

The “Thesaurus of Old English” is the not inconsiderable offshoot of the Historical Thesaurus of English, a project based at Glasgow University. Although it is a research tool of considerable interest in its own right, it is also intended that the TOE will serve as a “pilot”, in which the classification structures and working practices designed for the Historical Thesaurus may be tested. The printed “Thesaurus of Old English” (TOE), recently published by King’s College London Medieval Series, is drawn from an INGRES database, and other published resources, paper and electronic, are planned. The proposed presentation will explore some facets of the corpus of interest to historians of language and culture, and display the database in practical use. Some technical aspects of the database design and management will be addressed, and some lessons learned by us in the process of managing it, and in generating the printed work with its index, will be discussed. These lessons will prove invaluable in feeding into the work of the Historical Thesaurus, and in preparing the later TOE publications.

The “Thesaurus of Old English” has been created as a database, using the INGRES relational database software. Although the relational capabilities of the software have not so far been exploited, in the future we may want to use them as we move away from the original tasks of preparing both the Old English constituent of the Historical Thesaurus and the independent volumes that comprise the published, paper, edition of the “Thesaurus of Old English”. The database contains (Modern English) subject headings, and for each subject heading one or more Old English words. In total there are at present 72,612 database records, of which

50,502 represent Old English words and 22,110 Modern English headings. The subject headings are grouped by category and sub-category. Each group and sub-group has a classification number associated with it for identification purposes. Each database record can be uniquely identified by its category and sub-category references and a serial number. Other data stored for the Old English words include part of speech, a primary alphabetical sort key, Roget categorization (dating from the time before the new classification was devised by Michael Samuels and Christian Kay, but still of use in placing slips as a clue to where errant ones should belong) and cross-references; also we record a modern or earlier English etymological tie-up. A further field affords a space for comments: for example, a location reference for each rare word, meaning or usage; this field also serves as a convenient space for recording observations of the sort that would normally go in a footnote. Finally, we have a field in which we put a code if the lexical item is a hapax legomenon – that is to say, it occurs once only in the corpus of Old English texts – (o), is recorded only in poetry (p), is recorded in glosses (g) or in some way has a query attached to it (q).

Text is recorded in the database in ASCII, so the Anglo-Saxon characters *thorn* and *ash* have to be entered as symbols (we cheat slightly by having *thorn* stand for both *eth* and *thorn*). Thus, in the database *thorn* is represented by “}” and *ash* by “{”, and these conventions, together with an underscore to denote the macron of a long vowel, serve to present the language adequately. At present, using the database requires an effort of transliteration, one which is acceptable to its primogenitors, but not, probably, to the researchers we hope will want to use the database materials for ongoing research. We are therefore determined that any electronic version should carry with it an acceptable Anglo-Saxon font (in particular, it must offer macrons to mark long vowels).

The database classification starts within each new category with the most general concepts or ideas, and descends to more and more refined headings. Here, for example, are the categories at the beginning of the classification of words to do with the activity of the mind:

06 Spirit, soul, heart

06.01 The head (as seat of thought)

06.01.01 Thought, the faculty of thinking, mind

06.01.01.01 Thinking about, minding, heeding

06.01.01.01.01 Thought, cogitation, meditation

06.01.01.01.01.01 Consideration, rumination

06.01.01.01.01.02 Forethought, consideration

It is often said that modern human beings have lost the faculty their ancestors had of being able to

memorize very long stories or histories. Feats of memory (such as the recitation of a saga many hours long) have certainly been recorded in the twentieth century, but in general this facility of the mind is not something of which we boast. For the Anglo-Saxons the mind, the active seat of thought, is conceived of simply as within the body, which itself may be viewed as a dwelling place ("sawolhus") or a container, a coffer ("hordcofa") for the inner life. The mind, the heart, the spirit and the soul are all manifestations of that inner life, and the thesaurus illustrates the range of ideas recorded for the way in which the mind works. That the mind may be regarded as a treasury kept within the body is illustrated by the concepts of "sawolhord" or "lichord", the metaphysical hoard corresponding in value to a hoard of gold and jewels. It is here that the intelligence is locked away for security ("gewitloca"). More than fifty nouns attempt to give some concrete account of the inner life of the body. Over forty verbs express the ideas of thinking, considering, reflecting, pondering, illustrating a way of looking at the mind that recognizes the hugely wide-ranging capacity of this most mysterious of human functions.

Searching the database at present is by one of two methods: the Query-by-forms procedure provided within INGRES, or SQL. However, making effective and fruitful searches is at present very much a matter of competent use of SQL. In the future, the database will have a more user-friendly front-end which will allow ready access to the data (possibly with an option of adding comments to the discussion of a particular word). The results of a sequence of searches may perhaps be placed in adjacent windows for comparison, or further refinements of a search already made may be used to bring the researcher ever closer to the fields required. Progress with these developments will be described, and demonstrated.

The paper will describe briefly the technical problems encountered in producing the paper publication of the TOE, and in particular those related to the generation of the word-subject index, including issues in the secondary and tertiary sort sequences which were needed to accommodate both words and phrases. The automated procedures developed for producing camera-ready copy from the database using LaTeX will also be briefly described.

However, the main emphasis in the technical part of the paper will be on the issues related to the electronic publication of the thesaurus material. The project is addressing questions of medium (CD-ROM and/or on-line database access and/or Internet access), format ("mainframe" database, personal computer database(s), SGML mark-up) and user interfaces and manipulation tools. The project is not only taking note of work done else-

where, but is also carrying out small-scale comparative studies with sample materials drawn from the database. The key objective is to find an appropriate balance between effectiveness of use and technical complexity. The paper will describe the progress of these developments, and the conclusions drawn.

With the ongoing publication of fascicles of the Dictionary of Old English (by the University of Toronto) there is a great deal of interest in the vocabulary of Old English at the moment. The thesaurus offers a unique way of exploring what words were available to the Anglo-Saxons when they wanted to call a spade a spade or when they wanted to consider the most remote elements of the nature of God (although naturally this record must be recognized as incomplete because of the haphazard survival of Anglo-Saxon texts). The thesaurus documents, using all the sources available to us, the way the Anglo-Saxons denoted their world.