Christian J. Kay

THE HISTORICAL THESAURUS OF ENGLISH

Introduction

THE HISTORICAL THESAURUS OF ENGLISH was begun by Professor M.L. Samuels in 1964, and I myself have been associated with it since 1969. Our target date for completion is 1987, so in academic dictionary terms it is a relatively speedy project. One reason for this is that the THESAURUS draws its data from existing dictionaries, notably the OXFORD ENGLISH DICTIONARY (OED), the ANGLO-SAXON DICTIONARY of Bosworth and Toller, and the CONCISE ANGLO-SAXON DICTIONARY by Clark Hall. Another is the large number of people - academics at Glasgow and elsewhere, postgraduate students and research assistants - who have been associated with the project over the years.[1]

The main purpose of the THESAURUS is to provide a research tool for people working in semantics and historical linguistics. Essentially, what we are doing is grouping words according to the concepts they express and then presenting them in chronological order within these groupings. The sample given in the Appendix shows part of the section in the THESAURUS dealing with water in movement. Within each sub-division, words are arranged chronologically, beginning with the Old English, which is underlined to distinguish it from the rest. The first section deals with words for 'river', all of which are considered approximately synonymous as exponents of the general concept. Perhaps the most immediately striking thing about this section is that what we might consider the most basic general word for the concept, river itself, is a relatively late accession, coming in from French and first recorded in 1297. Other things which might strike one are the small number of words under 'Large river' - only one at the time this sample was put together - and the considerable number of first recorded uses under 'Small river' for the 16th and early 17th centuries. This period is well-known as a time of great lexical innovation in English, and the material in our sample bears out the general observation for one particular semantic area. Later, comparison with other semantic areas will be possible, and we will be able to show how the rate of accession of vocabulary varied according to the preoccupations of people in different periods.

Historical information about individual words has, of course, been available to scholars ever since the OED was published. There are also conceptual classifications available in thesauri such as Roget's THESAURUS OF ENGLISH WORDS AND PHRASES. What is unique about the HISTORICAL THESAURUS is that for the first time these two major approaches to lexicography are combined, allowing us to see how the structure of the vocabulary has developed and changed.

The THESAURUS has grown considerably since its early days. At first it was envisaged that only abstract vocabulary would be

included, as the most linguistically interesting changes could
be expected to occur there. However, at an early stage, the
decision was taken to include vocabulary for concrete areas,
thus ensuring the comprehensiveness of the project, but at the
same time vastly increasing the amount of work involved. As a
result of this decision, large parts of our archive now offer
an intriguing record of the impedimenta of everyday life over
the past 1000 years. If you would like to know what people were
eating or wearing in the 14th century, for instance, the THESAURUS
is one place to look. A second reason for growth was the decision
to incorporate material from the OED supplements. This too has
increased the work of the project, but means that the archive
is comprehensive and up-to-date for the modern period. Compilation
of data from the OED was completed in December 1982. All the Old
English material has also been compiled, a task carried out single-
handed by Dr. Jane Roberts of King's College, London (cf. Roberts
1978). We have also worked our collective way through the first
two OED supplements, have nearly completed the third and are
eagerly awaiting the fourth.

Editorial procedures

     Present interests, especially my own, have moved away from
slip-making and centre on two quite different areas, firstly the
refinement of editorial style, and secondly the development of the
classification. Inevitably, the finished THESAURUS will be a
compromise between linguistic theory and the practicalities of
the printed page. Our first grand idea was that the numbering
system should represent the hierarchies of the classification, so
that the logical place of each category could be deduced from its
numerical head. Various attempts were made to devise a suitable
notation, but these were largely defeated by their own complexity.
The present sample represents this compromise in that the main
categories have numerical headings, but the smaller sub-divisions
have only verbal heads. We hope to introduce a greater degree
of systematic subordination into the smaller divisions, and prom-
ising work has recently been done on a layout based on systematic
indentation (cf. Chase 1983). In certain areas of vocabulary,
notably the more abstract ones, we have worked with componential
analysis in an attempt to produce a more systematic meta-language
of definition (cf. Kay and Samuels 1975).

     Within the lists of words, we have tried to combine visual
clarity with maximum historical information about individual forms.
Certain conventions have been adopted to achieve this, for example
the use of the plus sign +, which indicates a break in the currency
of a word, and the use of parentheses to indicate uncertainty
about the continued currency of a word. These conventions can be
seen in some of the more complex entries under 'Small river',
for example rillet. The string of dates here is interpreted thus:
the word is first recorded in 1538 and was apparently in continuous
use until 1678, then comes a break, then another period of use
from 1830-90, with the possibility left open that the word con-
tinued in use beyond then. Many of the uncertainties about closing
dates will of course disappear once the material from the
supplements is collated with data from the main volumes. One
of our present jobs is matching up supplement slips with main
volume slips which may have been completed and had a classificatory

number assigned five or ten years ago. This can be a tedious
process, but can also throw interesting light on how our collective
view of classification has developed over the years.

The classification

Since the project began, we have been using the 990 heads
of the revised edition of Roget's THESAURUS as a storage system
for our data, always with the assumption that we would eventually
move to a new system of classification. My own research interest
in the project has lain largely -in the development of the classi-
fication, with considerable help from other members of the project,
notably Wotherspoon (1969). The classification thus arrived at can
perhaps best be described as a modified folk taxonomy, and is
based on the assumption that we begin by classifying the immedi-
ately observable phenomena in our environment. The first section
thus consists of the vocabulary of the material universe – the
earth, the seas and the heavens. Such an approach contrasts with
that of a thesaurus like Roget's, which begins with abstract
notions; in our system the abstract notions are taken to be
deduced from the physical phenomena and are therefore placed after
them. In all there are some 60 major fields within the new classi-
fication, some matching those in Roget, others where there is
no equivalent.

Our experience has been that only the broadest general prin-
ciples of classification can be laid down in advance. Ideally,
perhaps, we should lay down no guidelines, but simply put all the
material into a large pile and start sorting it out; however, with
700,000 slips in the archive, such a pile is beyond contemplation.
The classification has now reached an interesting stage where
approximately one-half of the material has been sorted into the
new categories, and several major sections have been worked on
in detail.[2] One thing that has emerged very clearly is the extent
to which categories vary in their internal semantic organization,
just as they vary in their chronological profile. One of the
last general principles to which I clung was the attempt to impose
a fixed order of parts of speech within categories: noun, verb,
adjective, adverb. This was successful for the concrete categories,
but work on abstract categories has shown that some categories
are predominantly verbal or adjectival, and that greater flexi-
bility is needed. One of the interesting things to emerge from the
project as a whole may be the relative predominance of parts
of speech across categories.

Another interesting development of the past year has been the
reorganization of the Old English material into the major
categories of the new classification. Once this material has
been worked through in detail, we will have a better idea of
the success or otherwise of the new classification as a whole;
one problem of compiling a thesaurus, as opposed to an alpha-
betical dictionary, is that one is constantly having to revise
one's ideas in the light of data from other sections, so that the
structure cannot be regarded as complete until the work is
finished. The Old English material is by no means representative
of the vocabulary as a whole, but it provides us with a manageable
corpus which can serve as a pilot study.

Conclusion

Two questions are often put to people working on the HISTORICAL THESAURUS. The first is, "Why didn't you get a computer to do all this?" - a singularly irritating question if one has spent the best years of one's life plodding along writing out slips. The answer is simply that, in the state of knowledge in the mid-60's, neither the hardware nor the software was available to us. Nowadays, things might be different, and we intend to make full use of computer facilities in the final stages for storage, retrieval and printing. The second question, often delivered in tones of utter disbelief, is "What are you doing it for?" The potential uses in linguistic and historical research have already been mentioned. The linguistic implications are considerable, extending from historical linguistics through sociolinguistics to psychologically oriented research into conceptual systems and cognitive processes. The THESAURUS will provide a data-base for research for many years to come, joining other data-bases which are being accumulated in the Faculty of Arts at Glasgow. Once the project is complete there will also be the possibility of retrieving specialized thesauri, such as a complete listing of terms in warfare from the 15th century onwards, or of games and pastimes in the 19th century, or of new words in areas of interest to the 20th century. At the moment, however, our main task is getting the project finished, at which point the real work can begin.

Notes

1
  We are also indebted to the British Academy, the Leverhulme Trust, the Axe-Houghton Foundation, the Vogelstein Foundation and the Manpower Services Commission for financial support.

2
  In addition to the topics covered by Chase (1983) and Wotherspoon (1969), these include The Animal Kingdom (Lorna Knight), Meteorology (Günter Kotzor), Astrology and Astronomy (Angus Somerville), Authority (Hannah Stone), Good and Evil (Freda Thornton).

References

Chase, T.J. (1983) A Diachronic Semantic Classification of the English Religious Lexis. Ph.D. thesis, University of Glasgow
Kay, C.J. and Samuels, M.L. (1975) "Componential analysis in semantics: its validity and applications" Transactions of the Philological Society 49-81
Roberts, J. (1978) "Towards an Old English Thesaurus" Poetica 9: 56-72
Wotherspoon, I.A.W. (1969) A Notional Classification of Two Parts of English Lexis. M.Litt. thesis, University of Glasgow

Appendix

Historical Thesaurus of English: Sample of Classification

1.5 FLOWING WATER, RIVER

N. gytestream, lacu, lagu-
stream, flood<flod OE-1814 obs
or poet, stream OE-, water-
stream<waeter-stream OE-,
flume c1175-1652, river 1297-,
fleam c1300-1516, current
c1380-1863, watercourse 1510-,
nymph 1591-1889, fluent 1598-
1705, lymph a1630- poet, flux
1637-1769, currency 1758.

A. fluvial 1398-, fluviatile
1599, fluminal 1633, riverine
1871-1898, potamic 1883-.
LACKING RIVERS: riverless
1860-.

1.5.1 KINDS OF RIVER

  Large river

N. main flood 1549/62-1596.

  Small river, stream

N. riþ, rindle<rynel OE-1686+
c1746-1861 dial or poet, lake
<lacu OE-1630+1842/71-dial,
sitch<sic OE-1684+1826-dial,
sichet c1133-1319, siket
1300-1550+1894 dial, syke
c1330- N&Sc, channel 1483-
1705, riveret 1538-1670+1807,
rillet 1538-1678+1830-(1890),
rill 1538-, purl a1552-1651,
runnel 1577-, rundle 1587-1651+
1877/86 dial, rivulet 1587-,
riverling 1591, vein 1600-
1613, prill 1603-1614+1862-
dial, vein-riveret 1656,
riverlet 1674+1883, creek
1674-US, skerth 1851 dial,
rillock 1893+1903.

  Torrent

N. hlimme, hlyde, raes,
streamfaru, þrymm, waelstream,
waeterþeote, flood-gate 1388-
1533, race 1480-a1649, flush
1529-, flood 1531-, torrent
1601-, cataract 1634-(1860),
flash 1677-, lasher 1677-,
torrent-stream 1728, lavant
1774-, torrent-flood 1825.

A. swiþstream, sharp 1665-
1875, torrential 1849-.

ADV. torrentially 1882.

  River system

N. myþe, confluence 1615-
(1742), water-system 1833-.
FLOWING TOGETHER: confluence
1538-, confluxion 1599-1610,
conflux 1606-1781,
corrivation 1621+1635,
confluvium 1670.
BRANCH, TRIBUTARY: creek
1577-, marigot 1759- W.Afr,
sprout 1794, feeder 1795-,
tribute-river 1819, affluent
1833, tributary 1836-,
confluent 1850-(1861), tine
1875, off-stream 1876.

V. intr. conflow 1609+1872,
vent 1784 US, put 1755-1807
US, tail 1899-.

A. tributary 1611-, confluent
1612-, multifluvian 1807.

NOTE: This sample was put
  together before the data
  were complete.