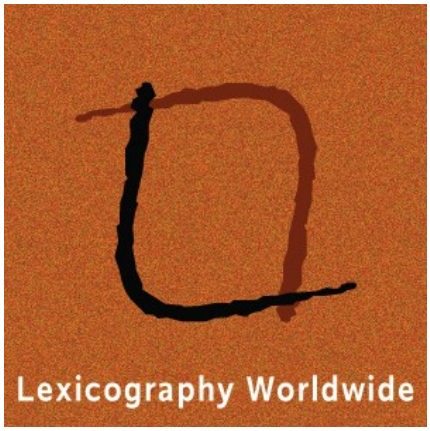
Michael Adams
(ed.)

"Cunning passages, contrived corridors": Unexpected Essays in the History of Lexicography



Polimetrica
International Scientific Publisher

2010 Polimetrica ® S.a.s.
Corso Milano, 26
20052 Monza – Italy
Tel. ++39.039.2301829
Web site: www.polimetrica.com

ISBN 978-88-7699-207-0 Printed edition
ISBN 978-88-7699-208-7 Electronic edition

The electronic edition of this book is not sold and is made available in free access.
Every contribution is published according to the terms of "Polimetrica B License". "Polimetrica B License" gives anyone the possibility to distribute the contents of the work, provided that the authors of the work and the publisher are always recognised and mentioned as such. It does not allow use of the contents of the work for commercial purposes or for profit. Polimetrica Publisher has the exclusive right to publish and sell the contents of the work in paper and electronic format and by any other means of publication. Additional rights on the contents of the work are the author's property.

Contents

The *Historical Thesaurus of the Oxford English Dictionary*

The *Historical Thesaurus of the Oxford English Dictionary*

Introduction

The *Historical Thesaurus of the Oxford English Dictionary* (*HTOED*) was published by Oxford University Press on October 22, 2009. It contains the contents of the second edition of the *Oxford English Dictionary (OED)*, supplemented by Old English vocabulary from *A Thesaurus of Old English*, arranged in hierarchically structured semantic fields. Each level of the hierarchy displays lists of synonyms for the concepts in question, with their dates of use, under brief explanatory headings. The work thus forms a semantic map of the development of the English vocabulary over its 1300-year recorded history.

The four papers presented here deal with various aspects of the production and potential of the work. They can be read in sequence or independently. Wotherspoon's paper describes the progress and methodology of the project from its inception in 1965 as a pencil-and-paper operation to the technology-dependent final product. Kay discusses *HTOED*'s unique system of classification, designed to accommodate historical data, and its theoretical basis in further detail. Wild takes the particular example of the conceptualization of childhood, and shows how a thesaurus presentation sheds new light on words and their cultural context. Alexander concludes by demonstrating how *HTOED*'s rich conceptual hierarchy can be used to produce a novel visual display of the shifts and balances in the history of the English lexicon.

A bibliography of works deriving from *HTOED* can be found in Kay *et al*. 2009, vol. 1, pp. xxxiii-xxxv, and on http://libra.englang.arts.gla.ac.uk/WebThesHTML/homepage.html

Michael Adams (ed.), *"Cunning passages, contrived corridors": Unexpected Essays in the History of Lexicography.* ©2010 Polimetrica International Scientific Publisher Monza/Italy

Michael Adams (ed.), *"Cunning passages, contrived corridors": Unexpected Essays in the History of Lexicography*, 255-270 ©2010 Polimetrica International Scientific Publisher Monza/Italy

Classification: Principles and Practice

Christian Kay - University of Glasgow
c.kay@englang.arts.gla.ac.uk

1. Introduction

Conceptually organised thesauri, such as the *Historical Thesaurus of the Oxford English Dictionary* (*HTOED*, Kay et al. 2009) are one of the oldest forms of lexicography in the English-speaking tradition, yet they are also one of the rarest. Hüllen, indeed, notes evidence of topical word-lists, covering such things as plants, foodstuffs, tools, and musical instruments, dating back to early civilizations such as Egypt and China (1999: 30-31). The earliest English glosses, usually of Latin texts, date from the 8th century; thereafter, lists from different texts were gradually collected together for convenience of reference, and thus divorced from their original contexts. The words in such amalgamated lists were arranged according to subject matter, often as a pedagogical aid, concerned as much with imparting information about the natural and social worlds as with teaching vocabulary. Hüllen (1999: 62) attributes an important role to these word-lists and suggests that:

> The most noteworthy example of an Old English universal topical glossary […], representing the beginnings of English onomasiological lexicography, is the one attached to Ælfric's (c.955-1020) grammar.

Despite this auspicious start, and sporadic attempts at thesauri and synonym dictionaries in later periods, the market in the English-speaking world has been dominated by alphabetically-organized dictionaries, sometimes with a nod in the direction of thesauri through lists of synonyms and antonyms attached to individual headwords. In the modern period, the thesaurus tradition has been largely sustained through the work of Peter Mark Roget, whose

ground-breaking *Thesaurus of English Words and Phrases* was first published in 1852.[1] Since then many versions of his work have appeared, in both English and other languages, often with little change in the overall design. One which departs significantly from Roget's framework is Chapman's (1992) *Roget's International Thesaurus*, which reorganizes the overall structure of categories while remaining true to Roget's general principles (Fischer 2004: 43). More radically different structures are proposed in Hallig and von Wartburg (1952) and McArthur (1986). No one, as far as the *HTOED* team was able to determine, has ever attempted a thesaurus of either the size or the diachronic perspective of *HTOED*.

2. Planning the *HTOED* Classification

For most of the period of *HTOED* data collection, starting in 1965, the 990 heads of Dutch's edition of Roget's Thesaurus (Dutch 1962), plus an extra head under which to assemble the body and its parts (see Wotherspoon, this volume), were used as a temporary filing system, but the intention of the *HTOED* team was always to devise a new system of classification which would do justice to both the nature and the quantity of our material. The sheer weight of vocabulary involved – almost 800,000 meanings taken from the *Oxford English Dictionary* (*OED*) and *A Thesaurus of Old English* (*TOE*, Roberts and Kay 2000) – was one issue. Another was how to organize all this material, representing many changes in culture and attitudes, within a coherent conceptual framework. We were also keen to develop a hierarchy of headings and subheadings which would offer a much greater degree of semantic discrimination than is found in most thesauri. While the resultant framework inevitably coincides with those of other thesauri at certain points, as a whole it offers a new and uniquely detailed system for semantic classification.

[1] References to *Roget's Thesaurus* or simply *Roget* throughout this paper are to Dutch (1962), the edition used in the preparation of *HTOED*.

Michael Adams (ed.), *"Cunning passages, contrived corridors": Unexpected Essays in the History of Lexicography*. ©2010 Polimetrica International Scientific Publisher Monza/Italy

Starting Points: HTOED and Roget

Thesaurus-makers who depart from Roget's overall plan are united in at least one respect: rejection of his starting point. Roget's six major classes are as follows:

1. Abstract relations
2. Space
3. Matter
4. Intellect: the exercise of the mind
5. Volition: the exercise of the will
6. Emotion, religion and morality

Class 1 includes the following sections: Existence, Relation, Quantity, Order, Number, Time, Change, and Causation. In his introduction, Roget says of this scheme:

> In constructing the following system of classification of the ideas which are expressible by language, my chief aim has been to obtain the greatest amount of practical utility. I have accordingly adopted such principles of arrangement as appeared to me to be the simplest and most natural, and which would not require, either for their comprehension or application, any disciplined acumen, or depth of metaphysical or antiquarian lore. (Dutch xxvi)

Hüllen (2009: 43) points out that Roget's original macrostructure survived in all six main British editions of his work after the first, writing:

> All of them retained this 'Plan of Classification' with minimal changes; it was obviously felicitously designed for the absorption of new vocabulary and so enhanced the general success of the book.

Earlier in the same article he had claimed that "…this macrostructure was able to combine the impression of scientific universality with easy usability" (2009: 40). Neither of these claims, nor Roget's claim above that his scheme was "simple and natural," was borne out by the experience of *HTOED* slip-makers, who often found Roget's system counterintuitive and difficult to operate. Admittedly, we were using it in a somewhat unusual way, trying to find places in his classification system for the entire vocabulary of English past and present, rather than looking for better ways to express ourselves or answers to crossword puzzle clues. We were

tempted, however, to suppose that subsequent *Roget* editors had experienced similar difficulties when trying to expand the coverage of the volume, especially when it came to including more of the vocabulary of the material universe, the positioning of which is sometimes hard to predict.

While there is a certain logic to beginning the classification with abstract ideas which are implicit in later sections, it is odd to find sections like 11 **Consanguinity: relations of kindred** so near the beginning of the book, especially when the fairly comprehensive set of kinship terms listed there is separated from other categories such as 169 **Parentage** and 170 **Posterity** (subsets of Causation), the latter also separated from 132 **Young person. Young animal. Young plant** (a subset of Time).[2] Without wishing to belittle Roget's considerable achievement, it has to be said that the practical problems arising from interpreting sections such as these drove the *HTOED* team inexorably in the direction of a classificatory system based on the conceptual or semantic field (see Section 3), that is the domain of experience or activity in which a word is likely to be used.

Our theoretical position on *HTOED* has always been that the classification at whatever level should develop from the data rather than be imposed upon it using some predetermined schema. A preliminary survey of the data (Kay and Samuels 1975), where recurring keywords in *OED* definitions were treated as meaning components, led to the establishment of three major divisions: I The External World, II The Mental World, and III The Social World, each with a single one-place category at level 1 in the classification system. The first of these had seven two-place categories at level 2:

[2] When I was teaching semantics, I used to illustrate this point by asking students to find words in *Roget* without using the index. Two which particularly frustrated them were *DOOR*, which they expected to find along with buildings rather than in 263 Opening, and *NECK*, which they expected to find with other body parts rather than in 218 Support. Even though opening and supporting are attributes of doors and necks, the students' instinctive response was to look in the obvious field of meaning.

Michael Adams (ed.), *"Cunning passages, contrived corridors": Unexpected Essays in the History of Lexicography.* ©2010 Polimetrica International Scientific Publisher Monza/Italy

I. The External World

01 The world
01.01 The earth
01.02 Life
01.03 Physical sensibility
01.04 Matter
01.05 Existence in time and space
01.06 Relative properties
01.07 The supernatural

Figure 1. Section I, levels 1 and 2

These seven categories in turn served as superordinates for between four and eleven three-place level 3 categories, as in:

I. The External World

01 The world
01.01 The earth
01.01.01 Region of the earth
01.01.02 Geodetic references
01.01.03 Direction
01.01.04 Land
01.01.05 Water
01.01.06 Named regions of earth
01.01.07 Structure of the earth
01.01.08 Minerals
01.01.09 Earth science
01.01.10 The universe
01.01.11 Atmosphere, weather

Figure 2. Section I, levels 1, 2 and 3

Here, the seven level 2 categories and the lower level categories dependent upon them cover much of the material in Roget's first three major classes, but with a very different emphasis. Instead of starting with "abstract relations" we put them near the end in 01.06 **Relative properties**, choosing instead to begin with the immediately observable physical universe, followed by living beings in 01.02, and

then by more abstract concepts such as "Matter" and "Existence." Thus the very first words in *HTOED* denote the world itself:

01 (***n.***) *The world*

brytengrund OE · brytenwangas OE · eardgeard OE · eormengrund OE · eorþstede OE · eorþweg OE · eþel OE · foldwang OE · foldweg OE · grund OE · grundwæg OE · grundwang OE · gumrice OE · hruse OE · leoht OE · middan(g)eard OE · moldweg OE · þeodeorþe OE · wundorworuld OE · ymbhwyrft OE · earth-ric<eorþrice OE–*c*1200 · world-riche<woruldrice OE–1390 · fold<folde OE–*c*1440 · mould<molde OE–1614; 1810 (*poet.*) · earth<eorþe OE– · mother<modor OE– · world<woruld OE– · all the world *c*1175– · wone *c*1205–1748 · mound *a*1290–*c*1320 · wilderness *a*1340– · mappemonde *a*1390–1560 · the whole world 1534– · the globe of the world 1553 · the/this globe 1575– · mother earth *c*1586– · orb 1601–1667 · glimpses of the moon 1602; 1828– · the earthly globe *c*1630 · universe 1630–1820 · terrene 1667–1830 · a world 1676– · Orphic egg 1684–1789 · all outdoors 1830– (*US colloq.*) · outer world 1868– · Midgard 1882 · overworld 1911– · spaceship earth 1966– **01** *as God's creation* ealdgeweorc OE · frumgesceap OE · frumheowung OE · frumsceaft OE · frumsetnung OE · frumweorc OE · fyrngesceap OE · fyrnweorc OE · landgesceaft OE · sæ and eorþe OE · (ge)sceaft OE · woruld OE · woruldgesceaft OE · creature *a*1300–1611 · creation 1611–1783/94 **01.01** *before Christ's redemption* gehæftworld OE **02** *as capable of growth* vegetable *a*1676

Figure 3. Category 01, noun, The world, with subcategories.
© University of Glasgow 2009

Other Starting Points

Although the large number of words from Old English in Figure 3 (labelled "OE" in the text) can be at least partly attributed to the survival of many religious texts from this period, contrasting the earthly and the heavenly life, the substantial early lexicalization exhibited by the category helps to justify its place in the overall framework of a historical thesaurus. In thus positioning the category, we are at odds with McArthur, whose first category is "A. Life and Living Things" (1981: v) and Chapman, who begins with "1. The Body and the Senses" (1992; quoted in Fischer 2004: 43). McArthur has the universe much lower down his structure in "L. Space and Time" (1981: v), while Chapman has it in "3. Place and Change of Place" (1992; quoted in Fischer 2004: 43). Of those conceptual thesauri we were able to consult, only Dornseiff (1959 [1933]; quoted in Fischer 2004: 45) begins with "1. The Inorganic World, Matter", followed by "2. Plants, Animals, Humans".

Comparing classifications is an interesting exercise, but must take into account such factors as the purpose and intended readership of a work. Chapman, with a 20th-century audience very firmly in mind, claims that Roget's scheme

> …does not coincide with the way most people now apprehend the universe. *Casting about for a more fitting arrangement* [italics added] I chose what I call a 'developmental-existential' scheme […]. The notion has been to make the arrangement analogous with the development of the human individual and the human race. […] This seems to me 'the simplest and most natural' array in *the mind of our own time* [italics added]. (1992; quoted in Fischer 2004: 43)

Twenty years earlier, the *HTOED* team would have agreed heartily with Chapman's first point, but, bearing in mind not only "the mind of our own time" but those of previous generations represented in the diachronic scope of our work, we chose to begin with the natural world, i.e. the context which predates human beings, in which they find themselves and on which they depend. As Chapman says above, when making such decisions one is indeed casting about for the most fitting arrangement, that is the arrangement which best serves one's data and one's audience. A case in point for *HTOED* is the positioning of 01.01.10 **The universe** near the end of the list of co-

ordinate categories at level 3 (see Fig. 2). To a modern mind, it would seem more logical to make **The universe** the level 2 category, with **The earth** subordinate to it, but the weight of the evidence suggested that the immediately present world was more salient in earlier periods.

A more extreme case of a classification adapted to historical data is that of Eugene Nida for New Testament Greek, which makes an initial division between "I. Entities, II. Events, III. Abstracts, and IV. Relationals" (1975: 178-186). Such a system, although interesting, often has the effect of splitting up semantically related parts of speech, so that, for example, foodstuffs are in I but verbs like "to cook" are in II. Another system which departs from most others from the beginning is that of Hallig and von Wartburg (1952) for French, which makes a broad tripartite division between "A. The Universe (*L'univers*), B. Man (*L'homme*), and C. Man and the Universe (*L'homme et l'univers*)" (Hüllen 1999: 19-20; Fischer 2004: 46-49).[3] Class A. contains plants and animals as well as the inanimate world, B. contains intellectual and social aspects of human life as well as man's physical existence, and C. contains abstract terms, science and technology. Although this classification was "meant as a foundation for onomasiological dictionaries of all languages" (Hüllen 1999: 19), it is difficult to apply to the lexicon, since, at least for English, the postulated distinction between human and other animates is not in evidence, for example in body parts or in basic verbs such as *TO LIVE*, *TO BREATHE*, *TO EAT*, and *TO MOVE*, which can have non-human animals and in some cases plants as their subject. A classification based on such a system would result in a good deal of uninformative repetition.

Section II

Section II, The Mental World, is the smallest of the three main divisions in *HTOED*, and presented fewest problems in its overall structure. It includes much of Roget's Class 4. Intellect, and Class 5.

[3] I gratefully acknowledge my use of Fischer's translations from the original French and German for quotations from the thesauri discussed in his article.

Michael Adams (ed.), *"Cunning passages, contrived corridors": Unexpected Essays in the History of Lexicography.* ©2010 Polimetrica International Scientific Publisher Monza/Italy

Volition, as well as Emotion from Class 6, on the grounds that emotions are mental in origin even though they may have physical manifestations. One of the most problematic categories was 02.07 Having/possession. H-J Diller (2008: 125) queries why it is here rather than in III The Social World, since many aspects of possession, such as materials, trade and commerce, are to be found in 03.10 **Occupation/work**. Although to some extent we shared his doubts, our decision was influenced by a comment under *HAVE* in *OED*2, describing *HAVE*, alongside *BE* and *DO*, which occur basically in *HTOED* section I, as "the most generalized representatives of the verbal classes", predicating, in its weakened senses, "merely a static relation between the subject and object." The presence of this relationship suggested a mental process. We therefore decided that this more abstract notion of possession, along with associated concepts such as giving, taking, wealth and poverty, should be separated from the huge body of material in Section III. Somewhat similar issues were raised by the split between Language in Section II as an intellectual activity, and Communication as a social activity, involving information transmitted in books, journalism, correspondence, broadcasting, telecommunications, etc., in Section III.

Section III

The changes in Section III overall were more radical. Initially we tried to make a distinction between Section I as the wholly natural world, and Section III for man's social activities, including his operations upon the world, but this proved impractical. In categories such as 01.01.05.00.01.03 **Channel of water** and its subcategories, there was often no way of telling whether the item in question was manmade or not. Given this, it seemed unhelpful to separate channels from clearly man-made things in the same domain of meaning, such as locks and dams. It was therefore decided to leave activities connected with physical existence, such as **Farming** and **Food**, in Section I, while moving those with a more clearly social dimension, such as **Kinship** and **Inhabiting / dwelling**, including buildings, to Section III. In some cases, there was a clear distinction between the physical and the social, leading to a category in each section, as in 01.05.08 **Movement** and 03.09 **Travel**. A similar kind of separation

was made between 01.07 **The supernatural**, which occurs at the end of Section I as a way of explaining the universe, and covers supernatural creatures and practices of all kinds, from angels and witches to spiritualism, and 03.07 **Faith** in Section III, which covers all aspects of organized religion. Such a division seemed best to encapsulate how the lexis of this area, and the attitudes it reflects, has developed over the years.

Sometimes, a concept might end up with a place in all three sections. One which got moved around a lot as classifiers claimed or rejected it was gem-stones, which can be minerals (Section I), a means of adornment (Section II Beauty), or an industrial material (Section III), and in fact were put in all these categories along with associated vocabulary.

3. Semantic Fields and Folk Categories

The need to make decisions about the nature and placing of categories may be one reason for the relative scarcity of conceptual thesauri (see Section I). The compilers of a dictionary at least have the alphabet as a structural starting-point, and, in the case of a major dictionary produced over many years, such as *OED*1, could generate income by publishing in fascicles. Thesaurus-makers are in a different position, since they are dealing with structures which are endlessly flexible. For both categories and individual words, the placing is often a choice between better or worse rather than right or wrong, and the work as a whole cannot be considered complete and offered to the public until every last meaning is slotted into place. Although online publication can allow interim publication of alphabetical dictionaries provided the status of each entry is clear, it does not help in the case of an incomplete thesaurus, where words may be moved or categories added right up until the end. Throughout compilation of *HTOED*, we were happy to supply sections of data to scholars writing articles or theses related to particular fields of meaning, but always with the proviso that such data were inevitably incomplete. Usually such incompleteness was part of the master plan, but sometimes there were nasty surprises, as when I received an email from a colleague on 5 February 1998 reading: "Also, I ought to

mention that Wit, which C. did right at the start when she did Ridicule, is in cupboard 6 and was intended to go somewhere in Mental Capacity." I carried this note around until 2004, when I started the final editing of Section II and duly found a place for Wit in 02.01.08 **Understanding**. One advantage of using paper slips throughout the classification process is that one has the reassurance of knowing that the project is finished when all the cupboards are empty.

Semantic Fields

While the term "semantic field" is used freely in linguistics, its definition is problematic, especially in relation to the associated term "conceptual field." One of the earliest and most influential scholars in this area of semantics was Jost Trier, who did not use the term "semantic field" at all. Instead, in his pioneering work of 1934, he used the terms *SINNFELD* and *WORTFELD*, usually translated as 'conceptual field' and 'lexical field'. The term *BEDEUTUNGSFELD* 'semantic field' was used by other German scholars, notably Walter Porzig, who, in another pioneering work of 1934, proposed a system of classification radically different from that of Trier (Lyons 1977: 251).

Strictly speaking, the conceptual field lies outside linguistics, referring to the domain in the external world in which an object, event, relationship, etc., occurs. The analyst identifies the field and then supplies the words which express the concepts within the field. This is the procedure advocated by Hallig and von Wartburg, who impose a set of conditions for developing conceptual systems, of which the first is:

> A scheme should be structured according to concepts (*Begriffe*), not words. Words are added or "filled in" when the scheme is used to study the lexicon of a particular language or dialect. (1952: ix, quoted in Fischer 2004: 46)

As Fischer goes on to say, this starting point may be one reason why Hallig and von Wartburg's scheme of classification has had little impact among either lexicographers or semanticists.

By contrast, a lexical or semantic field starts from the words of the language, and, by examining how they are related, the lexicographer establishes a set of fields. This was essentially the

procedure adopted in *HTOED*, and is what we mean by saying that the classification should "emerge" from the data. Since our data are words, our fields are by definition lexical fields, but we have tended to follow common usage and refer to them as semantic fields.

In practice it is difficult to separate conceptual and semantic fields. Even if the conceptual field is thought somehow to have an existence outside of language, it can only be described by using language; in other words, we are faced with the lexicographer's familiar problem of being unable to separate the object of description from the language of description. Equally, the meaning of words cannot always be elucidated by analysis of semantic relationships alone: the analyst often has to call on the encyclopaedic knowledge of the world which human beings accumulate from their culture, education and life experience. Thus, although *GIRL* may be defined as an antonym of *BOY*, and both *GIRL* and *BOY* are hyponyms of *CHILD*, identifying these relationships does not solve the problem of who may appropriately be referred to as *GIRL* in any particular context.

Folk categories

At every level of the *HTOED* system of classification lies the contention that, within certain limits, such as a fixed order of parts of speech, each category should be allowed to develop its own semantic profile. Since much of our data predates the development of expert scientific categories, this decision led to the development of a taxonomy based on natural or folk categories, which are "…grounded in the way people normally perceive and interact with the things in their environment" (Taylor 2003: 75). A similar stance was taken by Hallig and von Wartburg, whose second condition reads:

> The scheme should represent a "linguistically conditioned pre-scientific mode of cognition." Scientific categories should be added only when folk categories are missing or insufficient. (Fischer 2004: 47)

Folk categories are available to all members of a linguistic community, but communities are influenced by factors such as time, place and occupation. *HTOED*, for example, contains categories for creatures such as salmon or sheep according to their ages, but such

categories are likely to be beyond the knowledge of urban dwellers nowadays. Folk categories are also notoriously fuzzy at their edges. This applies to individual words – how old is a person referred to as a girl or a boy, for example? – and to classes of meanings. Wild (this volume) notes the problem of deciding whether to classify *TODDLER* as a baby or a child, thus drawing attention to the fuzziness of both these categories – when does someone stop being a baby and become a child?

Expert categories, on the other hand, are the work of people of acknowledged competence in a field, and do not admit of fuzziness. Items such as plants, animals or minerals either belong to a particular category within an established scientific taxonomy or they do not, in which case either the boundaries have to be redrawn or a new category set up. In other words, whereas folk categories cluster round a prototype, admitting both good and less good examples of it, expert categories operate on an either/or basis.

In *HTOED*, both folk and expert categories are used, with the former predominating. In major scientific categories, such as **Plants** and **Animals**, expert categories take over; the folk categories simply do not operate at the level of detail required by the vast accumulation of scientific knowledge of recent centuries. In some cases, a major category may contain both folk and expert categories at lower levels. **Animals**, for example, contains categories for wild animals, pets, strays, and a cow with a bell round its neck as well as the scientific taxonomy. **Plants** has categories for ornamental plants, poisonous plants, and fruits regarded as vegetables – this last reminding us that educated people can operate both kinds of categories, knowing that tomatoes are fruits, but treating them as vegetables for all practical purposes (see further O'Hare 2004).

4. Internal Structures

Semantic field theory is essentially structuralist in nature. As Lehrer says (1974: 7):

> The lexicon of a language is a network in which each item is related to other lexical items in a variety of ways. There is no single semantic description which shows all of these relationships, and different approaches highlight different aspects of the network.

Trier and Porzig differed in what they considered to be the primary relationships which structure a field.[4] Trier advocated a paradigmatic model, in which relationships such as synonymy, antonymy, hyponymy, and meronymy are used to group words on the basis of whether they are alike in meaning, unalike, instances of a kind, or parts of a whole. Porzig, on the other hand, took a syntagmatic approach, based on collocation, that is, words which consistently appear in proximity to one another, as *WHITE* does with *WINE* or *COFFEE* as well as with a host of other nouns.[5] In the early days of field theory, these two approaches were seen to be in competition, but nowadays, as Lyons says (1977: 261):

> There can no longer be any doubt that both Trier's paradigmatic relations and Porzig's syntagmatic relations must be incorporated in any satisfactory theory of lexical structure […].

This has certainly been our experience on *HTOED*. Our primary approach has been closer to Trier's in that initial classification was always done on the basis of sorting according to likeness, so that piles of synonyms, antonyms, etc. emerged at the first stage. Sorting on the basis of collocational relations would have led to the same sort of repetition as Hallig and von Wartburg's attempts to separate animal and human attributes (see Section 2), since many words have a wide collocational range; some lead in the direction of a semantic category, as in the relationship between *dog* and *bark*, others in no particular direction, as in *BRIGHT LIGHT*, *SUN*, *STAR*, *EYES*, *TEETH*, and *COLOUR*… Such a procedure is too open-ended to form a basis for classification, but where collocations are frequent enough for the *OED* editors to allocate them a separate sense, they are of course included in the appropriate category of *HTOED*. Thus a place could be found in 01.02.08.04 **Hunting** for OED2 *PAD* verb 2 "To place or pack (big game, etc.) on the pad of an elephant," but a word of such broad extension as *CASE* noun 2, sense 2a, "The outer protective or

[4] A detailed description and critique of both Porzig and Trier can be found in Lyons (1977: 256-266).

[5] Such an approach has become more powerful in recent years with the advent of massive online corpora which enable collocations to be identified and quantified.

Michael Adams (ed.), *"Cunning passages, contrived corridors": Unexpected Essays in the History of Lexicography*. ©2010 Polimetrica International Scientific Publisher Monza/Italy

covering part of anything, as the case of a watch, a fire-work, a sausage; a natural outer covering, sheath or receptacle, e.g. a seed-vessel, the 'case' of a pupa or chrysalis, of a case-worm, etc." could only be put in a general category of coverings.

5. Conclusion

Compiling a thesaurus is an art rather than a science. No single set of principles is sufficient to encompass the whole, especially when the whole has the scope of *HTOED* and has been edited by many hands over a long period of time. Our approach to linguistic theory has necessarily been somewhat eclectic, but it is has nevertheless been invigorating to watch the project develop alongside theories of semantics in the twentieth century.

References

Chapman, R. L. (ed.) (1992) *Roget's International Thesaurus*, 5th ed., HarperCollins, New York.

Diller, H.-J. (2008) "A Lexical Field Takes Shape: the Use of Corpora and Thesauri in Historical Semantics", *Anglistik* 19 (1), pp. 123-140.

Dornseiff, F. (1959 [1933]) *Der deutsche Wortschatz nach Sachgruppen*, Walter de Gruyter, Berlin.

Dutch, R. A. (ed.) 1962 *Roget's Thesaurus of English Words and Phrases*, Longman, Harlow.

Fischer, A. (2004) "The notional structure of thesauruses", in Kay and Smith (eds), pp. 41-58.

Hallig, R. & W. von Wartburg (1952) *Begriffssystem als Grundlage für die Lexikographie: Versuch eines Ordnungsschemas*, Akademie-Verlag, Berlin.

Hüllen, W. (1999) *English Dictionaries 800-1700. The Topical Tradition*, Clarendon Press, Oxford.

Hüllen, W. (2009) "Dictionaries of Synonyms and Thesauri," in A. P. Cowie (ed.) *The Oxford History of English Lexicography, vol. II: Specialized Dictionaries*, Clarendon Press, Oxford, pp. 25-46.

Kay, C. J. & M. L. Samuels (1975) "Componential Analysis in semantics: Its validity and applications", *Transactions of the Philological Society*, pp. 49-81.

Kay, C. J. & J. J. Smith (eds) (2004) *Categorization in the History of English,* John Benjamins, Amsterdam and Philadelphia.

Kay, C., J. Roberts, M. Samuels, & I. Wotherspoon (2009) *Historical Thesaurus of the Oxford English Dictionary,* Oxford University Press, Oxford.

Lehrer, A. (1974) *Semantic Fields and Lexical Structure*, North-Holland Publishing, Amsterdam.

Lyons, J. (1977) *Semantics* (2 vols.), Cambridge University Press, Cambridge.

McArthur, T. (1986) *The Longman Lexicon of Contemporary English*, Longman, Harlow.

Nida, E. A. (1975), *Componential Analysis of Meaning: an introduction to semantic structures*, Mouton, The Hague and Paris.

OED: The Oxford English Dictionary. 1884-1933, ed. by J. A. H. Murray, H. Bradley, W. A. Craigie, & C. T. Onions; *Supplement*, 1972-1986, ed. by R. W. Burchfield; 2nd edn, 1989, ed. by J. A. Simpson & E. S. C. Weiner; *Additions Series*, 1993-1997, ed. by J. A. Simpson, E. S. C. Weiner, & M. Proffitt; 3rd ed. (in progress) *OED Online*, March 2000-, ed. by J. A. Simpson, Oxford University Press, Oxford.

O'Hare, C. (2004) "Folk Classification in the HTE Plants Category", in Kay and Smith (eds), pp. 179-191.

Porzig. W. (1934) "Wesenhafte Bedeutungsbeziehungen", *Beiträge zur deutschen Sprache und Literatur* 58, pp. 70-97.

Roberts, J. & C. Kay with L. Grundy (1995) *A Thesaurus of Old English*, King's College London, London; 2nd ed., 2000, Rodopi, Amsterdam.

Trier, J. (1934) "Das sprachliche Feld. Eine Auseinandersetzung", *Neve Jahrbücher für Wissenschaft und Jugendbildung* 10, pp. 428-449.

Taylor, J. (2003) *Linguistic Categorization*, 3rd ed., Oxford University Press, Oxford.

Michael Adams (ed.), *"Cunning passages, contrived corridors": Unexpected Essays in the History of Lexicography*, 271-287 ©2010 Polimetrica International Scientific Publisher Monza/Italy

The Making of *The Historical Thesaurus of the Oxford English Dictionary*

Irené Wotherspoon – University of Glasgow
Irene.Wotherspoon@arts.gla.ac.uk

1. Introduction

The Historical Thesaurus of the Oxford English Dictionary (*HTOED*), published in October 2009 (Kay, Roberts, Samuels and Wotherspoon 2009), is a unique resource in which users can see the history of all the words used for a given concept or object, and also the range of words available to express any concept or object at any given time in the history of the English language. This was made possible by using the data from the *Oxford English Dictionary* (*OED*), which gives dated citations indicating the period(s) when the words were used in written English.

The idea of making a thesaurus from the *OED* was publicly proposed by M. L. Samuels in 1964 at a lecture given to the Philological Society (Samuels 1965). He saw it as a way of revealing the information on semantic and social change which was locked up in the alphabetical arrangement of the dictionary (Samuels 1972: 180).

The structure of *HTOED* is a modified folk taxonomy; it is based on the world view revealed by the pre-scientific terminology of English, but modified by reference to more scientific taxonomies for certain domains where this was appropriate, e.g., plants and animals. It is organised within each major section according to the principle of hyponymy (i.e., from the most general terms to the most specific), making extensive use of the concepts "types of" and, for terms for concrete things, "parts of." The outline structure was devised by Christian Kay and M. L. Samuels (Kay and Samuels 1975), mostly on the basis of recurrent key terms in the definitions in the *OED*, but at lower levels the nature, quantity and distribution of the terms themselves dictated how they were organised. *HTOED* is divided at the most general level for practical convenience into three main

sections: the external world, the mental world, and the social world. These in turn are divided into the 26 major categories devised by Kay and Samuels, reflecting the main activities and preoccupations of users of the language.

01 The world	**02 The mind**	**03 Society**
01.01 The earth	02.01 Mental capacity	03.01 Society/the community
01.02 Life	02.02 Emotion	03.02 Inhabiting/dwelling
01.03 Physical sensibility	02.03 Philosophy	03.03 Armed hostility
01.04 Matter	02.04 Aesthetics	03.04 Authority
01.05 Existence	02.05 Will	03.05 Morality
01.06 Relative properties	02.06 Refusal/denial	03.06 Education
01.07 Supernatural	02.07 Having/ possession	03.07 Faith
	02.08 Language	03.08 Communication
		03.09 Travel/travelling
		03.10 Occupation/work
		03.11 Leisure

Figure 1. The structure of *HTOED*, first and second level categories.

Further sub-categories follow, moving from the most general ideas to the most specific. Headings leading to 'musical instrument' and a selection of its subcategories illustrate this:

03 Society
03.11 Leisure
03.11.03 The Arts/art
03.11.03.01 Music
03.11.03.01.08 Musical instrument
03.11.03.01.08.03 Keyboard instrument
03.11.03.01.08.03.01 Stringed keyboards
03.11.03.01.08.03.01/06 pianoforte
03.11.03.01.08.03.01/06.03 sound-board
03.11.03.01.08.03.01/06.03.01 parts of

Figure 2. Sample subcategories of Society

Within each sub-category the items are organized by grammatical category and then in chronological order according to the citation dates in *OED*, as can be seen in the example at the end of this paper.

2. Sources of the Data

HTOED includes data from OED 1 and 2 and the supplements and additions, and therefore reflects both the strengths and weaknesses of *OED*, on one hand unparalleled scope and detail, but on the other bias towards the literary, and towards periods for which most reading for *OED* was done. It is hoped that revision of *HTOED* will take place in line with the third edition of *OED* as that progresses.

As a matter of policy, *OED* includes no word whose history ended by 1150. *HTOED*, however, was conceived so far as possible as a complete classification of the senses of words in English from the earliest written records onwards. *A Thesaurus of Old English* (*TOE*), compiled by Kay and Roberts and published in 1995, was in effect a pilot project for *HTOED*, with only 48,000 senses as opposed to more than 790,000 in *HTOED*. It uses the same principles of classification, but, reflecting the society in which Old English was used, has fewer categories than *HTOED*; e.g., for "The Arts" *HTOED* has 5599 headings for nouns, compared with 137 in TOE, for "Work" 9021 compared with 338, and for "Communication" 4106 compared with 397. Note that these are the numbers of headings, i.e., categories and subcategories, not lexical items. Nevertheless these figures indicate the extent to which these semantic areas have expanded. *TOE* was compiled as a separate operation at King's College, London. Roberts gathered the data from the standard Anglo-Saxon dictionaries of Clark Hall and Bosworth Toller, plus the material then available from the Toronto Dictionary of Old English and its corpus of texts. The data from *TOE* was incorporated in *HTOED*, along with some further additions from the advancing work at Toronto. Not only does the inclusion of this Old English material make the historical data in *HTOED* more complete than in *OED*, but it also reveals connections between some Old English words and their descendants which are not evident from *OED*. For instance OE *becca*, 'a pick or mattock', appears in *OED* as *beck*, dated 1875. Concepts which were lexicalised in Old English appear relexicalised much later: for instance OE *hygegebend*, 'heartstrings', a term which appears in *OED* with first citation 1596.

3. Gathering the Data

When the Historical Thesaurus project was proposed by M. L. Samuels to the staff of the Department of English Language at the University of Glasgow it was envisaged that each member would take part in collecting the data from *OED* and would classify and research the terms for at least one semantic field. It was hoped that post-graduate students would also be attracted by this kind of research, and to some extent this was so. At the beginning staff members worked on one *OED* volume each. Extracting the data consisted of filling out by hand a paper slip for each sense in the dictionary. Senses were ticked off in the volumes as a record of what had been done. This was similar to, but in a way the converse of, the method used for compiling *OED* from paper slips containing extracted quotations. It was laborious and time-consuming. Of course once *OED* was digitised, attempts were made to gather the data for the thesaurus electronically, by using searches for specific words in definitions, or by the subject labels used by *OED*, but this was just not possible because the *OED* data is not consistent enough in these respects, especially the latter. This is just one of the many problems of adapting the dictionary data to the purpose of a thesaurus. Nevertheless when *OED* became available in electronic form it was an invaluable time-saver when definitions or dates had to be checked during the routine work of the thesaurus compilers.

Dictionary Senses versus Thesaurus Senses

Generally, the senses for any given word in *OED* will be paralleled in *HTOED*, facilitating reference from one work to the other. If a word has more than one sense in *OED* it will appear in more than one place in the *HTOED*. The policies adopted by *HTOED* for dealing with *OED* definitions and sense divisions are the subject of another paper (Kay and Wotherspoon 2002), but may be briefly outlined here. The different purposes of dictionary and thesaurus meant that occasionally the *OED* sense divisions were too broad, obscuring nuances that might be significant in a thesaurus. An example of this is *OED HOST* verb 2, 1a, where the definition includes being the host at a dinner party and being the compere of a

show. Here the compiler's awareness of separate well-established lexical domains would indicate that separate senses could be recorded, provided there are citations to support them. (In fact this item appears in the section for **Hospitality** and in the section for **Performing arts** in *HTOED*.) More often, but still rare as a proportion of senses, the dictionary definition is too narrow for *HTOED*, resulting in the potential for two identical forms to occur in the same semantic category of the thesaurus. Examples of this are where the dictionary divides senses by context, such as '(a) of persons, (b) of things', or where the distinction is grammatical only, such as whether a verb is followed by an infinitive or a noun.

Splitting up one *OED* sense into several thesaurus senses can occur where the main grammatical category for a lexeme is finely divided in the dictionary, but others are more generalised: *HARSH*, adjective, has many senses, but *HARSHNESS*, noun, has one general definition. In such cases the thesaurus compiler had to split up the noun citations into the same thesaurus categories as the adjectives, where possible. Whenever dictionary senses were split up for the thesaurus care was taken to select the appropriate citations and hence dates.

The *OED* treatment of transferred and figurative meanings is rather inconsistent, sometimes giving them separate sense numbers and definitions, and sometimes merely marking them as 'also transf.' or 'also fig.'. In the thesaurus a separate sense was made if the importance of the concept or number of citations justified it. Otherwise figurative usage is indicated in the field for style and status labels.

4. The Slips

To render indubitable, support by proof or testimony 478

stablish n
a.1325 – 1561
+ 1885/94 ai

Figure 3. A sample *HTOED* slip.

The diagram above illustrates what a typical slip looked like.

The slips formed the raw material which went through several processes to become *HTOED*.

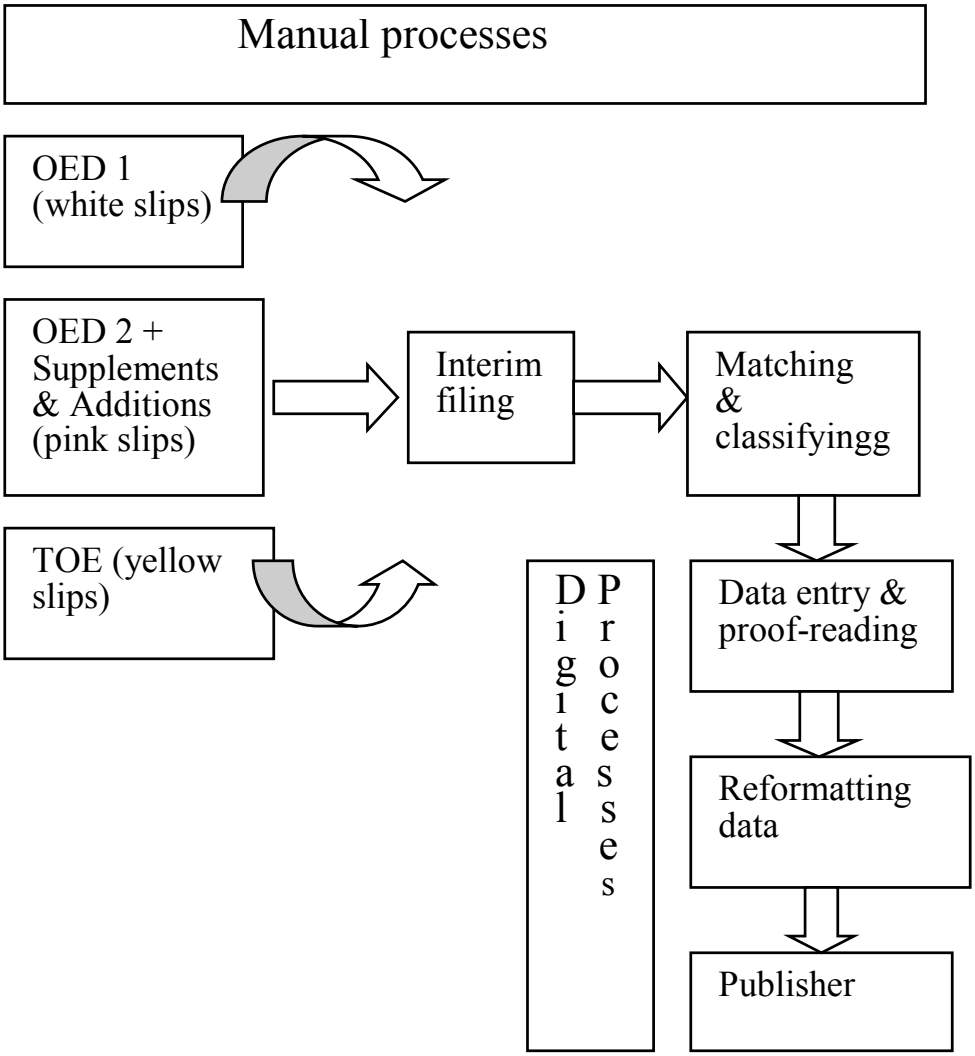


Figure 4. Processes underlying *HTOED*

Writing out the thesaurus slips was by no means a mechanical task. Decisions had to be taken about every part of any entry. Information from every part of an *OED* entry except etymology, orthographic variants and pronunciation, which were considered the province of a dictionary, was included on a thesaurus slip.

The Headwords

A very few *OED* headwords did not find their way into the thesaurus. Words with a single quotation were usually omitted if they were nonce words, or found only in a dictionary, or were

undefined. In the early stages of compiling *HTOED* it was the policy to omit very specialized obsolete technical and scientific terms. However as data collection progressed this policy changed, mainly because it became apparent that there would be a large number of modern technical terms from the supplements which would have to be included in any case. The same kind of creeping inclusion happened with words which *OED* recorded as dialectal only. It gradually became apparent that it would be safer to include them in case further evidence emerged. Some obvious compounds were omitted, especially those not individually defined in *OED*, and many negatives such as those in *UN-* and *-LESS*. On the other hand, phrases treated as lexical items by the dictionary were included, and also some specialized collocations which citational evidence and parallel formations indicated should be treated as phrases in the thesaurus. An example of this is *POSITIVE LAW*, treated as a phrase to balance *NATURAL LAW*. Headwords on thesaurus slips were in lower case, with initial capital only where this was the case with the *OED* headword.

Grammatical Categories

Part of speech appears on the slip as an abbreviation of the usual name of the grammatical category, taken from *OED*, except that "sb" was never used, always "n", and verbal nouns were entered as simply "n." Important attributive uses of nouns, marked 'attrib' in *OED* were called "aj" (for adjective) on slips if a separate slip was made for them on grounds of extensive citational evidence. This was all straightforward except for the issue of the transitivity of verbs. *OED* has a variety of ways of indicating this, and is not consistent when it comes to verbs followed by infinitives or clauses. The OE examples often do not have enough citations to judge transitivity (exclusively transitive use in a limited record does not guarantee exclusively transitive use in speech), and many OE verbs have an object in a case other than accusative. Impersonal and absolute uses are usually included with intransitive verbs in the *HTOED*. The verbal categories are "v" (where transitivity is not determined), and "vi", "vt" and "vr", with occasional uses of "v.pass" and "v. impers."

Definition

The full *OED* definition was written on the slips. In the early stages of slip-making the definition was not written on cross-reference slips, but this soon proved to be inefficient, as the definition had to be checked in *OED*.

Dates

HTOED does not include the dates of all the *OED* citations for any given word. The intention is to indicate the period or periods when the word was in use in written English. So the date of the first citation is used as a starting date. If the word is in use in the present day, the starting date is followed by a dash; in *HTOED* any *OED* sense with a date later than 1870 is considered to be still in use or to have the potential of being used. If the word is obsolete, the starting date is followed by a dash and the date of the last *OED* citation. However, if there is a gap of 150 years or more between *OED* citations this is reflected in the thesaurus entry, which indicates separate periods of usage with a ";". Although *HTOED* uses the ante and circa dates as in *OED*, the vaguer dates have been simplified, so that 14.. becomes a1500 and ?1500 becomes c1500. OE is used to indicate any date before 1150, dates in this period being particularly problematic. Of course some of the later dates given in *OED* are also notoriously controversial, such as those for Shakespeare's plays, and some are inconsistent for the same work, for instance those given for *Sir Gawain*. *HTOED* has to live with this, as does *OED*, until the third edition of *OED* is completed. However this does mean that senses traced by a user between the thesaurus and the dictionary can be confirmed by correspondence of dates.

Style and Status Labels

The labels used in *OED* to indicate region, domain, or status of a word, e.g., "colloq.", "US", "Law", etc., constitute valuable information about restrictions on usage and are included in *HTOED*. They are added to the appropriate date-range: a word might start life in a dictionary, and be labelled as such for the opening date only, then become a well-used term; or it might have been used in standard English and then become restricted to dialectal use, so having a later

date range labelled as dialectal or as specific to a region. These labels, inconsistent in form in *OED*, were rationalised in *HTOED*. The labels are not used in *HTOED* where they would be redundant; e.g., items labelled *Law* in *OED* are not so labelled in the Law category of *HTOED*, only where they appear in other categories.

The above descriptions of the handling of the grammatical categories, dates, and style and status labels from *OED* can be seen exemplified in the extended example from *HTOED* at the end of this paper.

Roget Numbers

At top right on every paper slip a category number (or where more than one was needed to represent the sense, numbers) from Roget's Thesaurus was written. Before they were classified the slips were temporarily filed according to these numbers as a handy way of roughly sorting them. The *HTOED* classification was evolved from the data as it was collected, so there was no detailed pre-conceived scheme. It was never intended that *HTOED* would be organised according to Roget's scheme. The basis of Roget's classification, abstract, through concrete, then mental then emotional categories, and a heavy reliance upon antonymy, is quite different from that of *HTOED*. Even before classification began it was obvious that some Roget categories would not be useful, and that adjustments would need to be made to the filing scheme. For example because Roget's system is based on abstractions, terms for the parts of the body are distributed among dozens of numbers; *FOOT* is in the right number for lowness, *HEAD* in the number for height, etc. For thesaurus filing an extra number was added at the end to collect the body parts. Like any thesaurus, including inevitably *HTOED*, Roget is very much of its time, e.g., the listings in Roget of almost any kind of sexual activity under the heading "Impurity."

5. Archive Management

The paper slips were stored in metal filing boxes, inside metal cabinets. The building in which the project was housed suffered a serious fire in 1978. Some of the marked volumes of *OED* were

badly singed, but the slips themselves were all safe. After this, slips were made in triplicate, with the extra copies stored at King's College, London, and at the main archive at the University of Glasgow. The English Language Department moved to a different building, where a room was converted to be a specially secure fire- and water-proof environment for the *HTOED* archive of slips.

6. Classification of the Slips

Classification of the slips was a labour-intensive and lengthy process for which no technological aid was conceivable. It began in the late 1970s, before slip-making was completed, and by the late 1980s was the main focus of work on *HTOED*. It was carried out by research assistants funded at different times by the Leverhulme Trust, the British Academy, the Arts and Humanities Research Council, other foundations, and a small but continuous grant from the University of Glasgow. For a short time the project also benefited from a government training scheme for unemployed graduates. Those who came to work at the thesaurus project helped with rough classifying, sorting into grammatical categories and editing the slips to ensure consistency for data entry.

The method of classification meant that the structure of each major category was determined by its content, rather than being imposed in advance. Generally, a large section of several thousands of slips corresponding to one of the second-level categories of the thesaurus (e.g., Armed Hostility or Music) was taken by a classifier. Usually this was the equivalent of several Roget numbers, i.e., several archive drawers; for example, Armed Hostility ran from 718 to 723, Music from 410-414. The classifier sorted the slips roughly by sense, according to recurrent words in the definitions, then sorted progressively more finely, still by sense, until semantic groups of near-synonyms were arrived at. The words within each group would not necessarily or even usually be exact synonyms, but would have most of their semantic content overlapping, more or less in the manner now reflected in prototype theory. In some areas this resulted in very small groups, and in others in groups of more than a hundred words, reflecting the extent to which different types of

concept have been lexicalised. The terms within the semantic groups were then sorted by grammatical category, and finally, by date of first citation.

Ordering of the Grammatical Categories

In *HTOED* the number of terms in each of the grammatical categories is very uneven: nouns, verbs and adjectives are numerous, and adverbs relatively few. This reflects the *OED* data and the fact that the potential for forming an adverb from an adjective nearly always exists. Much consideration was given to whether the order in which the grammatical categories are presented in *HTOED* should vary with the type of concept being categorised. Some concepts are inherently more adjectival than nominal, or more verbal than nominal. In her thesis dealing with the classification of terms for good and evil, Freda Thornton cogently argued for putting adjectives first, as being the most basic grammatical category for these concepts (Thornton 1988). However the editors of *HTOED* came to the conclusion that, because of the size and varied nature of *HTOED*, the least confusing arrangement would be to follow a canonical order: noun, adjective, adverb, verb, then the more minor parts of speech as detailed in *HTOED* (1.xxii).

Matching the Slips

Concurrent with the process of classification, the (pink) slips from the *OED* supplements and the (yellow) slips duplicated from the *TOE* archive were matched to their equivalent (white) *OED* slips.
In the case of slips made from the supplements this was usually straightforward, as notes were made on the pink slips to ensure a correct match. The Old English slips were a very different matter, except where there was an obvious modern English descendant of an Old English form. As can be seen in *TOE*, the distribution of lexis and the mapping of semantic fields is different from that of modern English, and of course many now familiar concepts simply did not exist. Also figurative senses were not usually recorded separately for *TOE*, due to the overall small numbers of extant examples. In

HTOED modern English forms are shown linked to their OE antecedents by "<", e.g., *slackness<sleacnes OE-*.

Headings

The classifiers wrote out a heading slip on green paper for each grammatical category of each semantic group which (s)he established. This slip went at the beginning of the series of slips to which it belonged and bore a verbal heading of no more than fifty characters. Repetitive headings were abbreviated, e.g., "one who", "quality of." *HTOED* does not use antonymy as a feature for distinguishing categories unless the number of lexical items justifies it. Usually negative terms such as those in *UN-* or *-LESS* will be in a sub-category with an abbreviated heading such as 'not' or 'without', where the meaning of the heading is understood from the next above category heading. The practice of *HTOED* is to use in the headings only words which are in current use if at all possible.

The heading slip also bore a category or sub-category number. These were decimal numbers of the type 01.01.01.01.01 to seven places for categories and to five places for sub-categories. The verbal headings for sub-categories were preceded by a number of dots appropriate to the level, from one to five (not replicated in published version).

7. Data Entry and Storage

Almost from the start it was known that *HTOED* would be printed electronically from a database. Consultations with technical staff at OUP in 1981 meant that we began with a robust database package which may seem ridiculously archaic today, but has in fact been able to support every form we subsequently wanted to put the data into. This database was dBaseII.4. The structure of the thesaurus database was devised by Alastair Forsythe. There are 29 fields, shown below in three columns for compactness; the first column has the data for the word and heading, the second the date and style data, and the third temporary classification data.

Group 01.05.05.09.04.01	**oe** OE	**Roget 1**	679
Sub-group	**Dsh1** -	**Roget 2**	
Number 10 (not shown)		**Ac1Roget 3**	
Word sweer<swær	**Dt1** 1668	**Newx 1**	
P o s aj	**Dt1a**	**Newx 2**	
	Abb1 Sc&nd	**Newx3**	
	Dsh2 +		
	Ac2		
	Dt2 1865		
	Dt2a		
	Abb2 Sc&nd		
	Dsh3		
	Ac3		
	Dt3		
	Dt3a		
	Abb3		
	Bkt		
	Current		

Figure 5. Structure of the data entry database with a sample entry

Note "+" for discontinuity between dates, and two-letter codes for grammatical category and style and status labels. In print "+" becomes ";" and the two-letter codes are expanded.

The fields provided for new cross-reference numbers were in fact not required. Fields were provided for four possible date ranges, Old English and three others, though it was rare for all of these fields to be populated. The combination of the group number, subgroup number, and number in lists of words provided a unique identifier for each entry. dBase11.4 was used for data entry until almost the very end of the work, though the data has been stored in many other successively more up-to-date forms over the years, e.g., Access, Ingres, Mysql. Modern databases were necessary to support complex data retrieval, which became ever more important as the size of the database increased and classifiers had to be able to find whether an item was already in the database and if so where.

Data entry was done by undergraduate and post-graduate students and by people on a government training scheme for the unemployed, under the supervision of Flora Edmonds, the *HTOED* database officer. They worked from bundles of classified slips and headings,

made up into files of about fifty items each. After the data for each thesaurus file had been entered it was proof-read twice.

Eventually an electronic version of *HTOED* will provide links with *OED* and offer the possibility of refined searches not only by semantic area but also by period of currency, status label, and grammatical category, or any combination of these. A prototype for this was devised in the early 1990s (Wotherspoon 1992), and has been substantially developed by Flora Edmonds for internet access.

8. Final Insertion Stage

During classification of the slips it was inevitable that some would be reassigned to different numbers. In fact some difficult-to-place items were still going the rounds until the very end. So, unlike a dictionary, a thesaurus cannot be published a bit at a time – it is not finished until the last item is placed. Also, since classifying began before all the slips were made, extra slips would arrive for sections for which the data had already been entered. This meant that after reaching 'the end' of classifying first time round, many sections of the thesaurus had large numbers of additional items to be inserted. This task was mostly done by post-graduate students, who slotted the additional items into the already established categories or made additional subcategories as appropriate.

9. Index

A thesaurus should to an extent be its own index, since the terms for any given concept should have the terms closest in meaning nearby and those more distant in meaning proportionately more distant in the thesaurus. However, the scale of *HTOED* on the one hand, and the level of detail on the other, is such that an alphabetical index was considered necessary to provide an easy way into the enormous numbers of lexical items. The index was generated electronically by the publishers, and forms volume II of *HTOED*. In order to keep the volume similar in physical size to the thesaurus, the index does not include Old English words, nor words which had no continuation after the 14th century, nor phrases of more than four words. As well

as convenient access to the word-lists, the index provides cross references between categories, as of course every category and sub-category in which a form appears in *HTOED* is listed under that form. The highly polysemous character of much of the lexis of English is evident at a glance, even more so than in a dictionary, though of course all the meanings of any word are not necessarily in use at any one period, and, differing from the arrangement in the dictionary, homonyms are under the same index heading. Only the category numbers are given in the index, not the verbal headings. However some semantic information can be deduced from the numbers themselves. Widely different category numbers, for instance if they show the occurrence of the item in section 1 of the thesaurus (numbers beginning 01) and also in section 2 (numbers beginning 02) or 3 (numbers beginning 03), sometimes indicate homonyms, or more often, the operation of metaphor. Concrete and abstract categories linked by metaphor are especially found where a form has categories in section 1, The External World, and section 2, The Mental World. Examples of this are the intransitive verbs *SWELL* and *BOIL*, which have 01 categories and also appear in 02.02.28 'Be proud' and 02.02.15.01 'Become ardent/fervent' respectively, and the adjective *THICK*, which has 01 categories and 02.01.09.03 'Dull, stupid'.

10. Conclusion

More than 40 scholars world wide have already made use of parts of the *HTOED* data in published research, on topics as varied as religion, death, expectation, and sex, love, and marriage.[1] However publication of *HTOED* is in a way the real beginning of a new phase of identifying and analysing the interesting things about language and society which can only be revealed by a thesaurus which shows terms in both their semantic and their historical context.

[1] For a list of topics and authors see the bibliography in Kay et al. 2009, vol. 1, xxxiii-xxxv.

References

Bosworth, J & N. T. Toller (1898) *An Anglo-Saxon Dictionary*, London.

Cameron, A. F., A. C. Amos, A. di P. Healey, S. Butler, J. Holland, D. McDougall & I. McDougall (eds.) (1986-) *Dictionary of Old English*, PIMS, Toronto.

Clark Hall, J. R. & H. D. Meritt (1960) *A Concise Anglo-Saxon Dictionary*, 4th ed., Cambridge University Press, Cambridge.

Kay, C., J. Roberts, M. Samuels & I. Wotherspoon (2009) *Historical Thesaurus of the Oxford English Dictionary*, 2 volumes, Oxford University Press, Oxford.

Kay, C. J. & M. L. Samuels (1975) "Componential Analysis in semantics: Its validity and applications", *Transactions of the Philological Society*, pp. 49-81.

Kay, C. & I. Wotherspoon (2002) "Turning the Dictionary Inside Out: Some Issues in the Compilation of a Historical Thesaurus" in J. E. Diaz Vera (ed.) *A Changing World of Words*, Rodopi, Amsterdam/New York, pp. 109-135.

Roberts, J. & C. Kay with L. Grundy (1995) *A Thesaurus of Old English*, King's College London, London; 2nd edition, 2000, Rodopi, Amsterdam.

Roget, P. Mark (1962 [1852]) *Thesaurus of English Words and Phrases*, ed. Robert A. Dutch, Longman, Harlow.

Samuels. M. L. (1965) "The Role of Functional Selection in the History of English", *Transactions of the Philological Society*, 1965, pp. 15-40.

Samuels. M. L. (1972) *Linguistic Evolution*, Cambridge University Press, Cambridge.

Thornton, Freda J. (1988) *A Classification of the Semantic Field 'Good and Evil' in the Vocabulary of English*, Ph.D. Thesis, University of Glasgow.

Wotherspoon, I. A. (1992) "Historical Thesaurus Database Using Ingres", *Literary and Linguistic Computing*, 7 (4), pp. 218-225.

The Oxford English Dictionary. 1884-1933, ed. by J. A. H. Murray, H. Bradley, W. A. Craigie & C. T. Onions; *Supplement*, 1972-1986, ed. by Robert W. Burchfield; 2nd ed., 1989, ed. by J. A. Simpson & E. S. C. Weiner; *Additions Series*, 1993-1997, ed. by J. A. Simpson, E. S. C. Weiner & M. Proffitt; 3rd ed. (in progress) *OED Online*, March 2000-, ed. by J. A. Simpson, Oxford University Press, Oxford.

01.05.05.09.04.01 (*n.*) *Sloth/laziness*
aswundennes OE · solcennes OE · swancornes OE · uncafscipe OE · aswolkness<asolcennes OE–*c*1175 · unlust<unlust OE–*a*1470 · sleuth<slæwþ OE–1629 · sweerness<swærnes OE–1676 (*Scots & northern dial.*) · slackness<sleacnes OE– · slauth *c*1175–*a*1400 /50 · slougth *a*1225–*c*1489 · sweering *a*1300 · sloth-head 1303 · unlusthead 1340(2) · slouth 1340/70–1597 · sweerdom *c*1375 (*Scots*) · slogardy *c*1386–1616 · nicety 1387–*c*1440 · sluggardy 1390–1606 · sluggardness 1398– · musardry *a*1400 /50 · nicehead *c*1440 · slugness *c*1440–*a*1500 · sleuthfulness *c*1470–1520 · lashness *c*1477–1673 · truantness 1483 · sluggardry 1513 · passibility 1526 · slothfulness 1526– · sluggardise 1532–1606 · luskishness 1538–1642 · desidiousness *a*1540–1651 · restiness *c*1540–1571 · ocivity 1550 · niceness *a*1557 · sloth 1575/85– · laziness 1580– · easiness 1581–1825 · poltroonery 1590 · facility 1615 · sloath *a*1618–1697 · otiosity 1632 · easefulness *a*1639 · indolence 1710– · dronishness 1731–1753 · indolency 1741 · do-nothingness 1814; 1879 · far niente 1819–1894 · oisivity 1830 · dronage 1846–1875 · faineance 1853– · faineancy 1854– · lazyhood 1866– · sleevelessness 1882– · bone-laziness 1925 · sluggardliness 1977– **01** *personified* Sloth 1362–1769 · slug *c*1425 **02** *as a disease* fever-lurden *c*1500–1808 · litherlurden *a*1590 **03** *lazy person* caynard 1303–*c*1386 · slogard 1398–*a*1548 · lusk *c*1420–1694 · sluggard 1423– · truant *c*1449–1596 · daw *c*1460; 1500/20–1862 (*Scots*) · slug *a*1500– · hummel 1500/20 (*Scots*) · sloven 1523–1576 · drone *a*1529– · draw-latch 1538–*a*1610 · slim 1548–1611 · slouk 1570 (*dial.*); 1867 (*dial.*) · do-nothing 1579–1887 · do-little 1586–*a*1654 · lazy-bones 1592– · luskin 1593 · loiter-sack 1594 · do-nought 1594–1870 · bed-presser 1598–1601 · lazy lizard 1600–1629 · lazy-back 1611 · trivant 1624; 1851/85 (*dial. Dict.*) · nothing-do 1629–1633 · slothful 1648 · lolpoop 1661; *a*1700 (*cant Dict.*); *a*1825 (*dial.*) · fainéant 1819– · haggis 1822 · lazy-boots 1831–1863 · lazy-legs 1838 · never-sweat 1851 · poke 1860– (*US*) · lollop 1896–1919 · trouble-shirker 1908– · warb 1933– · poop-butt 1973 **03.01** *group of* idle bellies 1530–1634 **03.02** *loutish/stupid* lubber 1362– (*now arch. & dial.*) · slouch 1515– · lubberwort 1575 · litherback 1577 · lubbard 1586–1712; 1724– (*Scots & northern dial.*) · litherby 1598 · lither lurden 1615 **03.03** *fat* panguts 1617–1775 **03.04** *an idler/loafer* lurdan *a*1300–1865 · player 1382– · loon *c*1450–1851 (*chiefly Scots & northern dial.*) · lounger 1508–1513 (*Scots*); 1711– · idler 1534– · rest man 1542 · holidayman/holidaywoman 1548–1600 · friar-fly 1577 · lingerer 1579– · idleby 1589–1708 · shite-rags 1598 · loll 1600; *a*1807 (*dial.*) · idlesby 1611–1681 · idle 1633–1709 · lollard 1635–1659 · loiterer 1684– · saunterer 1688–1798 · scobberlotcher *a*1697; 1933– · shoat 1800– (*dial. & US*) · waffler 1803– · ruminator 1827 · loafer 1830– · quisby 1837–1851 (*slang*) · bummer 1855– (*US slang*) · moocher/moucher 1862– (*slang*) · dead-beat 1863– (*slang, orig. US*) · scowbanker 1864–1898 (*slang*) · bum 1864– (*slang, orig. & chiefly US*) · scowbank 1881 (*slang*) · schnorrer 1892– · ikey 1906–1941 (*slang & dial.*) · layabout 1932– · lie-about 1937– · spine-basher 1946– (*Austral. slang*) · limer 1970– (*West Indies*) **03.04.01** *mock title* idleship 1860–1865 **03.04.02** *a gathering of* lounge 1798–*a*1865 **03.04.03** *in specific place* market-beater *c*1380–1483 · market-dasher *c*1440 · market-runner 1486 · bencher 1534–1860 · bench-whistler 1542–1618 · bench-babbler 1549 · Paul's man 1616; 1816 · Paul's walker 1658 · wharf-rat 1860– (*slang*) · boulevardier/boulevardeer 1879–1916 · bar-loafer 1889 · bench warmer 1892– (*US slang*) · stoepsitter 1934– (*S. Afr.*) · sidewalk superintendent 1940–1970 (*joc., chiefly US*) · beach bum 1962– (*slang*) **03.04.04** *watching fire* wait-gleed *c*1310 **03.04.05** *watching television* couch potato 1979– (*slang, orig. US*) **03.04.06** *fashionable* flâneur 1854– · lounge lizard 1918– (*slang, orig. US*) · saga boy 1949– (*West Indies*) **03.05** *lying late in bed* slug-a-bed 1592– · lier on/in bed 1657–1844 · morning-killer 1753 · lie-abed 1764– **03.06** *depending on others for support* smell-feast 1519–1884 · sponger 1677– · useless mouth 1722– · parasite 1883– · bot 1919–1960 (*Austral. & NZ slang*) · free-loader 1947– (*slang, orig. US*) **03.07** *in luxurious ease* lotus-eater 1847– (*transf.*) **03.07.01** *in Greek legend* lote-eater 1587–1638 · lotus-eater 1832– · lotophagist 1839 **03.07.01.01** *collectively* lotophagi 1601–1725 **04** *behaviour* truandise *a*1300–*c*1440 · lompering *c*1315 · sleuthing *c*1450–*a*1585 · truanting 1532 · slugging 1532–1633 · lusking 1579 · lazing *a*1626– · lounging 1793– · loafing 1838– · loaf 1855–1897 (*US*); 1900 · bumming 1857– (*US slang*) · sluggarding 1864 · mooch 1867– (*slang & dial.*) · flânerie 1873– · spine-bashing 1941– (*Austral. slang*) **04.01** *specific* bar-loafing 1889 **05** *period of lazing/lounging* lounge 1806– · laze 1862– · thoke 1891 (*dial. & slang*) · bludge 1969 (*Austral. & NZ slang*) **06** *place for lounging/loafing* lounge 1775– · loafery 1898–1903 **07** *imaginary land of laziness* lubberland 1598– · lotus-land 1842–

Figure 6. An illustrative example from *HTOED*

Michael Adams (ed.), *"Cunning passages, contrived corridors": Unexpected Essays in the History of Lexicography*, 289-308 ©2010 Polimetrica International Scientific Publisher Monza/Italy

Angelets, Trudgeons, and Bratlings: The Lexicalization of Childhood in the *Historical Thesaurus of the Oxford English Dictionary*

Kate Wild - University of Glasgow
k.wild@englang.arts.gla.ac.uk

1. Introduction

The recently published *Historical Thesaurus of the Oxford English Dictionary* (*HTOED*) promises to be an invaluable resource for both linguists and historians. Already, material from the *HTOED* has been used both to shed light on the history of subjects as diverse as horses and seaweed, and to investigate linguistic processes such as metaphor, metonymy and phonaesthesia. The purpose of this paper is to examine the English lexis for 'child' from both a historical and a linguistic perspective, and to show how data from the *HTOED* can inform us of the history of the conceptualization of childhood.

After a survey of previous research in this area, and some brief remarks on the concept of childhood, the lexemes for 'young person', 'child', and 'baby' in the *HTOED* are described and discussed from two angles: firstly, the extent to which they follow metaphorical and metonymical patterns; and secondly, what they suggest about the history of childhood and the conceptualization of childhood in English-speaking countries. It will be shown that the *HTOED* offers previously unavailable insights into the relationship between lexical change and cultural change.

2. Previous Research on the Lexicalization of Childhood in English

There has been significant research on aspects of the lexical field 'child' in English. An early article by Taylor (1929) examines words

for the concept 'child' in a variety of languages (English, Latin, Greek, Arabic, Celtic, and Native American languages, among others), giving examples of lexemes based on notions such as 'small', 'foolish', and 'dirty'. Bäck's (1934) study of the OE vocabulary for 'child' discovers that most words for 'child' in OE are based metonymically on etymons meaning 'birth' or 'womb', and that other metaphorical lexicalizations are not found.[1] Diensberg (1985) examines the restructuring, due to Romance loanwords, of the fields 'child' and 'servant' in ME. The words *BOY*, *GIRL*, and *BABY*, the etymologies of which are still uncertain, have also been examined in some detail (see e.g., Bammesberger and Grzega [2001] on *GIRL*; Breeze [2002] on *BABY*; Liberman [2008] on *BOY* and *GIRL*). Furthermore, there has been extensive research on the semantic structuring of the field 'human being' (e.g., Kleparski [1997 and 2004] on 'girl/young woman'; Grygiel [2006] on 'man/male human being'), some of which touches on the related area of children/young people. As yet, though, there has not been a study of words for 'child' across all periods of English, and this is something that the *HTOED* allows us to do with ease.

3. Some Remarks on Childhood as a Historical, Cultural, and Conceptual Phenomenon

While it would be beyond the scope of this paper to survey the history of childhood in the English-speaking world, it is worth making a few points about the development of the concept of childhood as a distinct period of life, and children as a distinct section of society. It has been argued that the idea of childhood is a modern phenomenon; that, in the famous (translated) phrase of Philippe Ariès, in the Middle Ages "the idea of childhood did not exist."[2] Recent research has gone a long way to disprove this claim,

[1] The following abbreviations for periods of English are used throughout: OE = Old English; ME = Middle English; EModE = Early Modern English; LModE = Late Modern English; PDE = Present Day English.

[2] Ariès (1960) *L'Enfant et la Vie Familiale sous l'Ancien Régime*, translated by Robert Baldick as *Centuries of Childhood* (1963), quoted in Heywood (2001: 5). As

and to show the extent to which children in western cultures have always had a separate identity. Crawford, for example, uses archaeological and documentary evidence to argue that Anglo-Saxon children "had their own characteristics and their own games" (1999: 174). There is also substantial evidence of medieval awareness of childhood in discussions (inherited from the Graeco-Roman tradition) of the stages of life, which included the stage before walking, the stage of dentition, the stage of growing facial hair, and so on (Heywood 2001: 14).

However, research has also explored the extent to which childhood is a socially constructed phenomenon. As Mintz (2008: 93) points out, "[a]ge categories are not natural; rather, they are imbued with cultural assumptions, meaning, and values." Thus, the transition point to adulthood at the age of eighteen in modern Britain and America is arbitrary and culturally variable; in other societies and other periods the transition might be at the age of, say, ten or sixteen, or it might not be at a fixed age. For example, Crawford (1999: 52), using evidence from law codes, shows that Anglo-Saxons became responsible for their actions – and thus, arguably, classed as 'adults' – at either ten or twelve years old. Another transition point from childhood to adulthood might be the age of starting work, but as Heywood (2001: 123) shows, this was highly variable before the modern period:

> Not until the late nineteenth and early twentieth centuries is there evidence of a crisp transition from childhood into the adult world of work, marked by the ritual of leaving school at the minimum age required by the state.

Thus, in examining the history of words for 'young person', 'child' and 'baby' in English, the fuzziness and culturally-dependent nature of these categories will be considered.

Heywood (2001: 19) points out, the original phrase *sentiment d'enfance* is ambiguous, and although it is usually translated as 'idea of childhood', it could equally mean 'feeling for childhood'.

4. The Lexicalization of Childhood in the *HTOED*

Lexemes related to childhood appear in the *HTOED* within the section 'people', which is a subcategory of 'life'. Relevant parts of the hierarchy are as follows:

01. The World
01.02 **Life**
01.02.07 **People**
01.02.07.04 **Person**
01.02.07.04.01 **Man**
01.02.07.04.02 **Woman**
01.02.07.04.03 **Young person**
01.02.07.04.03.01 **Youth/young man**
01.02.07.04.03.02 **Young woman**
01.02.07.04.04 **Child**
01.02.07.04.04.01 **Boy**
01.02.07.04.04.02 **Girl**
01.02.07.04.05 **Baby/infant**
01.02.07.04.06 **Adul**t
01.02.07.04.07 **Middle-aged person**
01.02.07.04.08 **Old person**
01.02.07.04.08.01 **Old Man**
01.02.07.04.08.02 **Old Woman**

Thus, while the category 'person' is first divided by gender ('man' and 'woman'), it is thereafter divided by age groups ('young person', 'child', 'baby/infant', 'adult', 'middle-aged person', and 'old person'). In turn, these categories are sometimes further divided by gender ('old person' into 'old man' and 'old woman', for example), but sometimes not (there are no separate categories for 'boy baby' and 'girl baby'). Since the categorization of the *HTOED* was data-driven (see Wotherspoon, this volume), these categories follow the evidence of salience in the senses and quotations in the *Oxford English Dictionary* (*OED*). That is, while there is sufficient evidence that 'old man' is or has been a salient category in English, with separate lexemes from the more general category 'old person', there is no evidence that would justify separate categories for 'boy baby' or 'girl baby'.

Michael Adams (ed.), *"Cunning passages, contrived corridors": Unexpected Essays in the History of Lexicography.* ©2010 Polimetrica International Scientific Publisher Monza/Italy

In this paper, only the three broad categories related to childhood are analyzed: 'young person', 'child', and 'baby/infant'. The gender divisions 'young man', 'young woman', 'boy' and 'girl' are not considered here, partly because these have already been researched elsewhere (see the summary of research in section 2 above), and partly because the focus of this paper is on age, rather than gender, as a historical and conceptual category. It should also be noted that there is a separate section in the *HTOED* for 'child' in the sense 'offspring': this appears within 03.01.01 'kinship' along with sections for 'father', 'mother', 'sibling', and so on, and is not included in this analysis.

Each category in the *HTOED* is divided into parts of speech, and each part of speech into subcategories (see Wotherspoon, this volume). Like many concrete categories, 'young person', 'child' and 'baby/infant' are predominantly nominal, with fewer adjectives and hardly any verbs.[3] Because of this, the focus of this paper is on nouns. In addition, more attention will be paid to the main categories, with selected examples from the subcategories.

The approach in analyzing these sections was to examine the words in the *HTOED*, and to discern salient types of lexicalization (such as prominent metaphors and metonyms), using the data in the *OED* to determine etymologies and sense developments. Data from the *OED* are from the second edition, as this is the edition on which the *HTOED* is based.

When referring to salient notions in different periods, I refer to the date at which a lexeme is first recorded, rather than the dates of its continued use. For example, *kid* is first recorded as meaning 'child' in 1599, and at this date its basis on the animal metaphor 'young goat' would have been alive; when *kid* is used today to mean 'child', however, this animal metaphor is arguably dead. Thus it is the date of coinage which is evidence of the salience of the metaphor.

[3] For example, in 'young person' there are 91 nouns, 33 adjectives, 4 adverbs, and only 2 verbs.

Young Person

Figure 1 shows the nouns for 'young person' in the *HTOED*. In OE and ME, most of the coinages are based on the form 'young': *YOUNGLING* (from *GEONGLING*), *YOUTH*, *YOUNG* and *YOUNG PERSON*. The only exception is *GIRL,* which could originally refer to a young person of either sex. Other lexemes based on *YOUNG* (or the Latinate form *JUVENILE*) continue to be coined in subsequent centuries: *JUVENAL*, *NONAGE YOUTH*, *JUVENILE*, *YOUNG 'UN*, *YOUNG BLOOD*, *JUVENILE ADULT*, *YOUNG ADULT*, *YA* ('young adult'), and *YOUF* (a jocular form of 'youth').

It is not until the 16th century that the first metaphorical lexicalizations are attested: the plant metaphors *PRIMROSE*, *SAPLING*, and *SLIP* (originally 'twig, sprig'), followed in the 17th century by *A SLIP OF A BOY/GIRL*. Animal metaphors first appear in the 17th century: *CHIT* (originally 'whelp; pup; kitten') and *EGG.* These are followed by *KID* in the 19th century and *SPRING CHICKEN* (usually in the phrase 'no spring chicken') in the 20th. The contemptuous *QUAT*, which originally meant 'pustule, boil', is the only metaphor relating to the supposed dirtiness of young people. *BOOTS* is an interesting example of the overlap between the concepts 'servant' and 'young person' (as discussed in Diensberg 1985; see section 2 above): it originally referred to a servant who cleans boots, and later extended to any young member of a club or regiment. *YOB* is backslang for *BOY*, and was used in the 19th century to denote any young person; it has since specialized in meaning to refer specifically to hooligans.

> **01.02.07.04.03 (*n.*)** *Young person* youngling<geongling OE- · youth c1250- · girl c1290-c1450 · young a1300-a1500 · young person 1535- · primrose a1568 · slip 1582- · juvenal 1588-1664; 1820 · sapling 1588-1905 · quat 1604-1623 (*contemp.*) · egg 1605 (*contemp.*); 1835 (*contemp.*) · airling 1611-1775 · nonage youth 1628 · chit 1649- (*contemp.*) · a slip of a boy/girl a1660- · juvenile 1733- · boots 1806 (*slang*) · young (')un c1810- · snip 1838- · yob 1859-1886/96 (*slang*) · young blood 1862; 1885 · kid 1884- (*colloq., orig. US*) · spring chicken 1910- · chiseller 1922- (*Irish*) · juvenile adult 1926; 1959 · YA 1974- (*US*) · yoof/youf(f) 1989 (*chiefly colloq. joc. & iron.*)

Figure 1. Nouns in the main section of 'young person' in the *HTOED*

It is of further interest to look at the subcategory 'adolescent', shown in figure 2. The concept was not lexicalized until the 14th century, a fact that could not easily have been ascertained without the *HTOED*. Furthermore, the first recorded lexeme, *PUBER* – from the Latin *PUBERTAS* 'age of maturity'– was restricted to legal and official usage. *ADOLESCENT* itself – from the Latin *ADOLESCERE* 'to grow up' – was first used in English in the 15th century. Metaphors and metonyms for this concept (the animal metaphor *FARROW* and the clothing metonym *JEAN-AGER*) do not appear until the 19th and 20th centuries.

A similar pattern can be seen in the subcategory 'adolescence/puberty' (see Fig. 3). The pre-19th-century coinages are Latinate and technical (*PUBERTY*, *ADOLESCENCY*, and so on), whereas the 19th- and 20th-century coinages give more of a sense of the character of adolescence: *HOG-AGE*, *THE AWKWARD AGE*, and *JEAN-AGE*.

> **05** *adolescent* puber c1315 (*Scots*); 1545 (*Scots*) · adolescent 1482-1495; 1815 · farrow 1820 · teen 1820- · teener 1894- · teen(-)ager 1941- · jean-ager 1959; 1961 · teen(e)y 1969- (*colloq.*) · skin 1970- (*slang*)

Figure 2. Nouns in the subcategory 'adolescent' in the *HTOED*

> **05.03** *adolescence/puberty* puberty 1382- · adolescency 1398-1719 · adolescence 1430- · pubes 1637 · pubescency 1658-1684 · teens 1673- · pubescence 1822/34 · hog-age 1848-1893 (*US*) · teenhood 1893 · the awkward age 1895- · prepubescence 1916 · prepuberty 1922- · teenagedness 1952 · jean-age 1960 · teenager 1960-

Figure 3. Nouns in the subcategory 'adolescence/puberty' in the *HTOED*

Child

> **01.02.07.04.04. (*n.*)** *Child* bearn OE · gebyrd OE · magotimber OE · umbor OE · wuscbearn OE · wenchel<wencel OE-a1225 · burden<gebyrþen OE;1489-1667 · littling<lytling OE;1721- (*Dict. dial.*) · child<cild OE- · hired-child c1205 · wee a1300 · chick c1320- · broll a1325-1575; 1865- (*obs. exc. dial.*) · innocent c1325-1641 · congeon/conjon c1330-a1400 (*contemp.*) · imp c1380-1809 · young one 1382- · faunt a1400-a1400/50 · chicken

a1400-1809 · enfaunt c1450 · schild c1450 · whelp 1483-c1591;1852 (*joc.*) · brat c1505- (*contemp.*) · little one 1526- · younkerkin a1529 · kitling 1541-a1745 · urchin 1556 · budling 1577 · breed 1586(2) · tadpole 1588; 1881 · pledge 1590- · bantling 1593-1831 (*derog.*) · two-year-old 1594/5- · bud 1595- · lambkin 1597- · eyas-musket 1598 · kid 1599- (*orig. slang*) · bratchet a1600- (*contemp.& joc.*) · younker 1601-1866 · dandling 1611-1695 · butter-print 1616-1709 (*slang*) · doveling a1618;1888 · whelpling a1618- · chit c1624-1864 (*contemp.*) · ninny c1626 · lap-child 1655 · chitterling 1675-1826 · nyny 1687 · wee ane 1692-a1774 (*Scots & dial.*) · kinchin a1700;1838- (*cant*) · cherub 1705- · charity child 1714-1861 · wean 1725- (*Scots dial.*) · youngster 1732- (*colloq.*) · poult/pout 1739- (*colloq. & dial.*) · elfin 1741-1804 · two-shoes 1766-1870 · piccaninny/pickaninny 1785- (*humorous*) · small boy/girl 1786- · piggy/piggie 1799- (*joc.*) · totum a1800-1898 (*Scots*) · suck-(a)-thumb c1800-1890 · olive 1803- · piece of goods 1809; 1895 (*humorous & contemp. dial.*) · stumpie 1820- (*Scots*) · sexennarian 1821 · totty 1821; 1849 · poppet 1830- · brancher 1833 · toad 1836- · bosom-child 1838 · young 'un 1838- (*colloq.*) · two-penny 1844 · weeny 1844 (*colloq.*); 1973; 1977 · tottykins 1849 · brattock 1858 (*local*) · ninepins 1862 (*humorous*) · piggy-wiggy 1862- (*joc.*) · ragazzo 1862- (*Ital.*) · angelet 1868 · tenas (man) 1870 · nipper 1875- (*slang*) · pee-wee 1877- (*N.Amer & dial.*) · tad 1877- (*colloq.*) · tacker 1885; 1893 · joey 1887 · simplicity 1887 · thumb-sucker 1891- · littly 1893/4 - · (little) tyke 1894 - · che-ild/chee-(y)ild 1896- (*iron.*) · kipper 1905- (*slang*) · small 1907- · God forbid 1909- (*slang*) · toto 1916 – · snookums 1919-1928 · problem child 1920- · trottie 1924 · squirt 1924 – (*colloq.*) · tiddler 1927- · pre-adolescent 1930- · teeny 1931- (*colloq.*) · accident 1932- (*colloq.*) · sprout 1934- (*US slang & colloq.*) · juvie 1941- (*US* slang) · sprog 1945- (*slang, orig. Nautical*) · subteen 1952- (*orig. US*) · pre-schooler 1954- · subadolescent 1957 · pint-size 1959- · subteenager 1959- · saucepan lid 1960- (*slang*) · pre-teenager 1966- (*orig. US*) · pre-teen 1967- (*orig. US*) · rug-rat 1968/70- (*US slang*) · preteener 1969

Figure 4. Nouns in the main section of 'child' in the *HTOED*

Figure 4 shows the nouns in the main section of 'child'. As discussed in Bäck (1934), OE words for 'child' are mainly based on the concept of childbearing: *BEARN*, *GEBYRD,* and *GEBURÞEN* from 'that which is born';[4] and *UMBOR* and *CILD*, which are thought to derive from etymons meaning 'womb'.[5] After the OE period, words with similar etymologies are coined, but only for 'child' in the sense 'offspring', for example, *BIRTH*, *CONCEPTION*, *SPAWN*, and *BEGET*. As noted above, these are all found in the *HTOED* in section 03.01.01 'kinship'. It is only in OE that words for 'child' meaning 'young person' are based on the concept of birth.[6]

There is also one OE word, *WENCEL*, which is thought to relate to *WANCOL*, 'totter', and is perhaps a precursor of the 19th-century words for *TODDLER* (Bäck 1934: 80). One OE word, *LYTLING*, denotes small size, and this concept is the basis of several subsequent lexemes, such as *WEE*, *LITTLE ONE*, *WEAN*, and *PINT-SIZE*.

From the 14th century, animal metaphors are frequently employed, often with the diminutive suffixes -*LING* and -*KIN*: children are referred to, for example, as *CHICKENS*, *WHELPS*, *KITLINGS*, *URCHINS*, *CHITTERLINGS*, *TADPOLES*, *LAMBKINS*, *KIDS*, *DOVELINGS*, *WHELPLINGS*, *PIGGIES*, *POULTS*, *TIDDLERS*, *KIPPERS*, and *TOADS*. Plant metaphors are, perhaps surprisingly, less productive, and there are only a few coinages in each period: *IMP* in the 14th century, *BUDLING* and *BUD* in the 17th, *OLIVE* in the 19th and *SPROUT* in the 20th.

In the 18th and 19th centuries, two words are coined on a metaphorical extension of 'angel'– *CHERUB* and *ANGELET* – which reflect the Romanticization of childhood in this period (for example, the image of 'heaven [lying] about us in our infancy' in Wordsworth's "Intimations of Immortality", 1807). Related to these

[4] Or perhaps 'that which is carried' (Bäck 1934: 66).
[5] See Bäck (1934: 32; 78).
[6] The 16th-century coinage *BREED*, which is listed under 'child' (as young person), appears to have been miscategorized: both the *OED* quotations indicate that it was only used in the sense 'offspring':

> **1586** WARNER *Alb. Eng.* I. ii. 4 Cybell [had] brought to light Her second breede, a smiling boy. *Ibid.* X. lv. 253 When Junos Breed on farther bankes his passenger had set.

are the coinages based on the metaphor 'fairy/magical being': *CONGEON/CONJON* (from 'changeling') in the 14th century and *ELFIN* in the 18th. One further lexeme is based on the metaphor 'doll': *POPPET*, 1830. Two metonyms are based on the perceived qualities of children: *INNOCENT* in the 14th century and *SIMPLICITY* in the 19th. There is only one lexeme related to the notion of dirtiness: *SQUIRT*, 1924, with the original meaning 'diarrhoea'.

Other types include foreign loans (e.g., *NYNY* and *NINNY* from Spanish *NIÑO*, *KINCHIN* from German *KIND*, and *TOTO* from Swahili *MTOTO*) and rhyming slang in the 20th century (*GOD FORBID* and *SAUCEPAN LID*).

The 19th century sees a spate of coinages based on children's actions or activities: *SUCK-(A)-THUMB*, *NIPPER* ('one who nips'), *TACKER* ('one who attaches'), and *THUMB-SUCKER*. The most notable development, though, is the number of words showing finer distinctions in age in the 20th century. There were a few words of this type coined before this period, such as *SEXENNARIAN*, but far more in the 20th century, when nearly a third of the coinages are age-related. One refers to children before school-age – *PRE-SCHOOLER* – and there are a further seven words referring to children before they become teenagers: *PRE-ADOLESCENT*, *SUBTEEN*, *SUBADOLESCENT*, *SUBTEENAGER*, *PRE-TEENAGER*, *PRE-TEEN*, and *PRETEENER*. As noted above, *TEENAGER* itself is in the category 'adolescent' (within 'young person') and this raises issues in classification: where does one draw the line between a child and a young person? Regardless of how they are classified, though, the fact that in the 20th century a large number of words are coined which finely discriminate the category 'child' suggests the increased attention paid to children as a section of society.

Slightly different patterns of lexicalization emerge in the subcategory 'childhood', shown in Figure 5. Until the 17th century, all the lexemes are based on the form *CHILD* or Latinate or French versions such as *ENFAUNCE* and *PUERICE*. In later centuries, though, we find clothing metonyms – *LEADING-STRINGS* in the 17th century and *CAP AND FEATHER DAYS* and *TUNICHOOD* in the 19th – which are

not attested in the nouns for 'child' itself.[7] There is also one 19th-century metonym based on what children tend to eat: *BREAD-AND-BUTTERHOOD*. It is notable that, as was the case with the concept 'adolescence', discussed above, there is no evidence of the metaphorization of childhood until the 19th century. This is the kind of information which would be very difficult to access by looking at individual words; by presenting the words conceptually and chronologically, the *HTOED* allows such insights.

> **07** *childhood* cildgeogoþ OE · cildyldu OE · childhood<cildhad OE- · bairnheid a1300-1588 · childhead a1300-1588 · enfaunce c1400 · puerice 1481; 1660 · puerility 1512-1849 · child-age 1548-1638 · childishness 1660 · leading-string(s) 1677- · impuberty 1785 · cap and feather days 1821 · bairnhood a1835 (*Scots*); 1894 (*Scots*) · pupilarity 1846 · tunichood 1859 · child-life 1865- · bread-and-butterhood 1884 · latency 1910- (*Psychology*) · preadolescence 1930- · puerilism 1940 (*Psychology*)

Figure 5. Nouns in the subcategory 'childhood' in the *HTOED*

Baby/Infant

Figure 6 shows the section 'baby/infant' in the *HTOED*. Again, only the OE words are based on the concept 'womb': *UMBOR* and *CILD*. Many of the other early words are variants of *BABY* and *INFANT*, such as *INFANT*, *FAUNTEKIN*, *FAUNTELET*, *BABY*, and *BABE* (all first recorded in the 14th century). Even though the word *INFANT* derives from the Latin 'unable to speak', there is little evidence that this meaning was still alive in the English derivatives, and 'speechlessness' does not seem to be a salient notion in the lexicalization of this concept. *BABY* is probably onomatopoeic (although see Breeze 2002 for the argument that it is a Celtic loanword).

[7] *BRAT* is a possible exception: it has been argued that it is related to *BRAT* meaning 'cloth, rag' (see e.g., Bammesberger and Grzega 2001), but this etymology is uncertain.

01.02.07.04.05 (*n.*) *Baby/infant* cradolcild OE · umbor OE · child<cild OE- · baban c1230 · chrisom-child c1275-1680 · fauntekin 1377-c1446 · baby 1377- · infant 1382- · fauntelet 1393 · babe 1393- (*now chiefly poet.*) · lakin c1440-1570· mop c1460-1598 · tenderling 1587-a1900 · chrisom 1596-a1667; 1755 (*Dict.*) · chrisomer 1605 · flosculet 1648 · childling 1648; 1855 (*arch.*) · bratling 1652; 1796 · lullaby-cheat 1671 (*cant*) · stranger a1674-1896 (*joc.*) · hoppet 1695 (*northern English dial.*) · tot 1725- (*colloq.& dial.*) · bambino 1761; 1863 · weanie/weany 1785- (*dial.*) · nursling babe 1793· dab 1797; 1833- (*dial.*) · tiny 1797- · toddler 1812- · trudgeon 1814 · nursling child 1818 · child/baby in arms 1819- · toddle 1825-1882 · toddles 1828 (*dial.*); 1854 · chrisom-babe 1829 · yearnling 1829 · toddlekins 1852-1904 · mite 1852- · trot 1854-1905 · dot 1859- · nurseling 1860 · tiny 1863-1883 · nestler 1866 · babelet 1867 · spoon-child 1868 · toddlekin 1879-1904 · the new-born 1879 · kiddy 1889- (*colloq.*) · toddleskin 1890 · kidling 1899 (*slang*) · kidlet 1899- · babe in arms 1912- · liddly 1929- · maumet 1932 (*dial.*); 1949 (*dial.*) · neonate 1932- · snork 1941- (*Austral. & NZ slang*)

Figure 6. Nouns in the main section of 'baby/infant' in the *HTOED*

There is a smattering of metonyms based on clothing, derived from the chrisom-robe that babies wore at baptism: *CHRISOM-CHILD*, *CHRISOM*, *CHRISOMER*, and *CHRISOM-BABE*, all of which are now obsolete. A chrisom-robe was also used as a shroud if a child died within a month of baptism, so *CHRISOM* and *CHRISOMER* came to refer to those who died in infancy, and the general sense 'baby' became obsolete, as it was presumably felt to be incompatible. There are also a few animal metaphors, but not until the 19th and 20th century: *NESTLER*, *KIDDY*, *KIDLING*, *KIDLET*, and *SNORK* (which has the semantic progression 'snort' – 'young pig' – 'baby'). There is only one plant metaphor: *FLOSCULET* in the 17th century. Two lexemes are based on the metaphor 'toy, doll' – *LAKIN* in the 15th century and *MAUMET* in the 20th – and one derives from the sense 'fool': *MOP*, c1460-1598.

Lexemes referring to size appear from the 18th century, for example, *WEENIE*, *DAB*, *MITE*, *DOT*, *TINY*, and *LIDDLY*. The most significant development, though, is the group of lexemes referring to

how babies move. Before the 19th century there is only one lexeme of this type: *HOPPET*, 1695, glossed as 'a young child danced in the arms' in the *OED* quotation. In the 19th century, the coinages refer to babies beginning to walk, and most of these are variants of *TODDLER*. *TODDLER* itself is first recorded in 1812 to refer to a 'child that toddles', and it is highly productive in the 19th century: there follows *TODDLE*, *TODDLES*, *TODDLEKIN*, *TODDLEKINS*, and *TODDLESKIN*. There are two other words based on the way that babies move: *TRUDGEON* and *TROT*, coined in the 19th and 20th century. Like *TEENAGER*, these raise problematic issues in classification: is a toddler a kind of baby, or would it be better classified under 'child', or in a category of its own? No matter which section these words are grouped in, though, it remains that all the words for babies beginning to move around and find their feet are coined from the 19th century onwards, even though babies would of course have toddled before that time.[8] These coinages for more finely discriminated stages of infancy suggest an increased attention to and interest in this concept, and account in part for the vast increase in 19th-century coinages for 'baby/infant'.

As further evidence of this, the subcategories of 'baby/infant' are presented in Figure 7. Over half of the sub-categories are first lexicalized in the 19th or 20th century. For example, the subcategory 'premature baby' is first lexicalized (with the phrase *SEVEN MONTHS CHILD*) in 1859; the first lexeme in the category 'baby reared by specific method' (*BOTTLE-BABY*) is first recorded in 1893; and the first lexicalization of 'presence/expectation of [babies]' is *PATTERING (OF LITTLE FEET)*, attested in 1849. These reflect in part the history of raising babies – bottle-feeding only became a viable option in the mid-19th century with the invention of rubber teats, while premature babies were unlikely to survive before the development of incubators in the late-19th century – but they also show the increased salience of infancy.

[8] See Heywood (2001: 89) for a summary of evidence of the way babies first learned to walk in earlier periods.

Subcategory	Date of lexicalization
babyhood/infancy	OE
suckling	OE
weanling	1532/3
caesarean baby	1540
baby girl	1611
personality of	1648
presence/expectation of	1849
premature baby	1859
babies collectively	1860
baby boy	1881
baby reared by specific method	1893
baby conceived in specific method	1935
baby exhibiting signs of abuse	1963

Figure 7. Subcategories of 'baby/infant' (noun) in the *HTOED*, arranged by date of lexicalization

By way of contrast, Figure 8 shows one subcategory which was more highly lexicalized in earlier periods of English: *SUCKLING*.

> **06** *suckling* deonde OE- · diend OE · meolcsucend OE · sucking<sucende OE-1382 · sucker 1382-c1440 · suckerel c1440; 1813 · suckling c1440- · pap-hawk c1450 · milksop c1460· nursling/nurseling 1607- · teatling 1631 · nuzzling 1638 · nursery 1642-1650

Figure 8. Nouns in the subcategory 'suckling' in the *HTOED*

All thirteen words in this section were first recorded in OE, ME or EModE; none were coined in the LModE period. Furthermore, most of the words are now obsolete in this sense. In some cases their figurative senses now predominate: *SUCKER* has been taken over in modern use by the extended sense 'simpleton', while *MILKSOP* has always been more common in the sense of a feeble person or effeminate man. There are now only two words available for the concept 'suckling': *NURS(E)LING* and *SUCKLING*. *SUCKLING* is now usually applied to animals, which perhaps limits its acceptability as a word for human babies. *NURS(E)LING* is quite

rare: it occurs only twice in the 100-million word British National Corpus, and ten times in the 400-million word Corpus of Contemporary American English.[9] In PDE a phrase such as *BREAST-FED BABY* is generally preferred. The decline in words in this section is partially a reflection of the history of feeding babies: more babies are bottle-fed now, and babies tend to be weaned earlier.[10] However, it is also suggestive of the way that babies are conceptualized: more as acting and doing, and less as attached to their mothers' (or wet-nurses') breasts.

Summary: Salient Notions in the Lexicalization of Childhood

The following is a summary of the most prominent underlying notions in the lexicalization of the categories 'young person', 'child', and 'baby/infant', considered as a whole. Figure 9 presents the number of nouns in the main sections (not subsections) of these categories, divided into the most salient notions identified: birth/womb; small; young; animal metaphors; plant metaphors; age-distinction; and 'other'. Figure 10 illustrates a selection of these types.[11]

In OE, the most salient notion in the lexicalization of childhood is the concept of childbearing: over half of the nouns for 'young person', 'child', and 'baby' derive from etymons meaning 'birth' or 'womb'. After OE, this notion becomes much less salient as the basis of words for 'child as young person', although it continues to be salient for 'child as offspring'. The concepts 'small' and 'young' are fairly prominent underlying notions in all periods, with no clear diachronic pattern.

[9] Both available at Davies' (2004-, 2008-) online corpus sites.

[10] Crawford (1999: 71) suggests that Anglo-Saxon babies were breastfed until they were two or three: dental evidence shows decay after this time, indicating a sudden movement onto adult food.

[11] The higher number of coinages in the 16th and 19th centuries is partially due to the fact that the *HTOED* is based on the *OED*, which overrepresents these periods in its quotations (see e.g., Brewer 2000). For this reason, I have focused on types and proportions of coinages rather than overall figures.

		birth/ womb	**small**	**young**	**animal**	**plant**	**age/stage - distinction**	**other**	**total**
OE	coinages	7	1	1	0	0	0	4	13
	% of total	**54**	**8**	**8**	**0**	**0**	**0**	**31**	
13th c.	coinages	0	1	2	0	0	0	4	7
	% of total	**0**	**14**	**29**	**0**	**0**	**0**	**57**	
14th c.	coinages	0	0	1	2	1	0	9	13
	% of total	**0**	**0**	**8**	**15**	**8**	**0**	**69**	
15th c.	coinages	0	0	0	1	0	0	4	5
	% of total	**0**	**0**	**0**	**20**	**0**	**0**	**80**	
16th c.	coinages	0	1	3	7	5	1	6	23
	% of total	**0**	**4**	**13**	**30**	**22**	**4**	**26**	
17th c.	coinages	0	1	2	6	2	1	13	25
	% of total	**0**	**4**	**7**	**24**	**8**	**4**	**52**	
18th c.	coinages	0	5	2	2	0	0	9	18
	% of total	**0**	**28**	**11**	**11**	**0**	**0**	**50**	
19th c.	coinages	0	7	3	11	1	9	28	59
	% of total	**0**	**12**	**5**	**19**	**2**	**15**	**47**	
20th c.	coinages	0	4	4	5	1	8	13	35
	% of total	**0**	**11**	**11**	**14**	**3**	**23**	**37**	
total	coinages	7	20	18	34	10	19	90	198
	% of total	**4**	**10**	**9**	**17**	**5**	**10**	**45**	

Figure 9. Types of lexeme for nouns in the main sections of 'young person', 'child' and 'baby/infant', by date of coinage

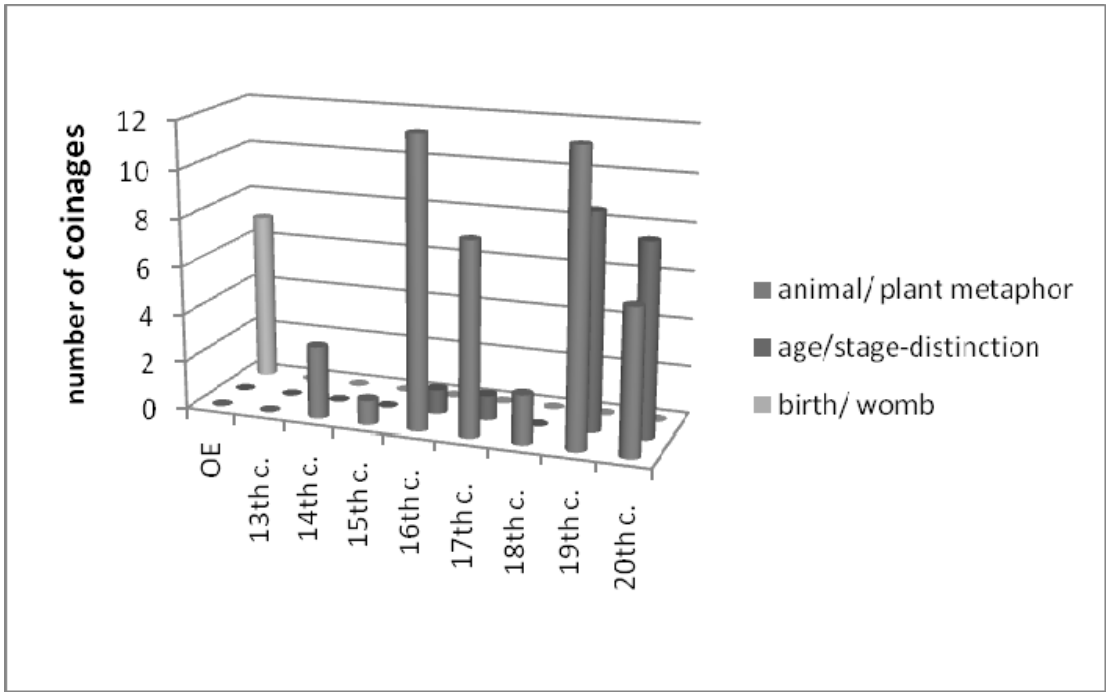


Figure 10. Number of coinages for three types: birth/womb; animal and plant metaphors; and age/stage-distinction

The single most frequent notion is the animal metaphor: thirty-four lexemes, or 17% of the total, are of this type. This follows the pattern of metaphorization for human beings generally. Kleparski (1997: 230-41) shows that the concept 'girl/young woman' is frequently lexicalized by animal metaphors, especially birds (e.g., *BIRD*, *HEN*, *CHICK*) and mammals (*JADE*, *KITTEN*, *COW*). Similarly, the concept 'man' has often been lexicalized by mammal metaphors (*HOUND*, *RAT*), and especially horses (*STUD*, *STALLION*) (Grygiel 2006: 61). Neither 'man' nor 'woman' tends to be lexicalized by fish or amphibian metaphors (Kleparski 1997: 232 gives only *BACKFISH*, recorded as meaning 'young woman' in the 19th century). The animal metaphors used for 'man' and 'woman' perspectivize particular characteristics or these groups, such as the supposed delicacy of birds/women and the virility of horses/men. The metaphors for 'child', on the other hand, are more varied: children are described as young birds (*CHICK*, *NESTLER*, *POULT*), mammals (*KITLING*, *LAMBKIN*, *KID*), amphibians (*TOAD*, *TAD*, *TADPOLE*), and fish (*KIPPER*, *TIDDLER*). This suggests that, in the case of children, it is solely the concept of small size, rather than the perceived

characteristics of these animals, that is perspectivized in these metaphors.

Plant metaphors are most frequent in the 16th century: five coinages (*BUDLING*, *BUD*, *PRIMROSE*, *SLIP*, and *SAPLING*), or 22% of the total, are of this type. In other periods, plant metaphors are sporadic. It is also interesting to speculate about the lack of other types of metaphor, although, of course, negative evidence must be treated with caution. In his article on words for 'child' in other languages, Taylor (1929: 314) concludes that 'from such notions as "small", "foolish" or "dirty", the step to the concept 'child' was not a long one and could be taken in many lands'. However, with a few exceptions (*MOP* meaning both 'fool' and 'baby'; *QUAT* meaning both 'pustule' and 'young person'; and the sense of *SQUIRT* developing from 'diarrhoea' to 'child'), these concepts do not appear to be salient in the lexicalization of 'child' in English. Furthermore, there are very few clothing metonyms, compared to the large number of such metonyms for 'girl/young woman' (*SKIRT*, *PETTICOAT* etc.; see Kleparski 2004), indicating that this is not a salient aspect of children.

In the 19th and 20th centuries, the main development is the number of coinages which distinguish different ages or stages of infancy and childhood, such as *TODDLER*, *PRE-SCHOOLER*, and *PRE-ADOLESCENT*. As noted in section 3 above, 'child' is a rather fuzzy category, and in earlier periods was only 'loosely defined' (Heywood 2001: 17). The development of more specific age-categories in this period reflects the increasingly segmented nature of childhood in the modern age, when age-graded schools and laws necessarily divide children into different groups.

5. Conclusion

The conceptual and chronological arrangement of lexemes in the *HTOED* facilitates a kind of semantic analysis that would previously have been difficult, if not impossible. Using the *HTOED* in conjunction with the *OED*, the lexicalization of childhood has been analyzed from both a historical and a linguistic perspective. In particular, it has been shown that patterns of lexicalization have

Michael Adams (ed.), *"Cunning passages, contrived corridors": Unexpected Essays in the History of Lexicography*. ©2010 Polimetrica International Scientific Publisher Monza/Italy

changed diachronically: in OE the most prominent concept was child-bearing; since the 14th century animal metaphors have been productive; and in the 19th and 20th centuries lexemes have also been based on the activities of children, and on finer distinctions of age-groups. In particular, the LModE coinages indicate an increased attention to children in this period. As Kay (1997: 50) remarks:

> Areas of life which are important to a culture, or to subsections of it, will be highly lexicalized; that is, there will be words or phrases available to describe many aspects of those areas, and to make fine distinctions in meaning.

The evidence of the *HTOED* indicates that 'young person', 'child', and 'baby/infant' have become more highly and finely lexicalized in the LModE period, reflecting the increasing importance of children as a distinct section of society in the modern age.

References

Bäck, H. (1934) *The Synonyms for "child, boy, girl" in Old English: An Etymological-Semasiological Investigation*, Gleerup, Lund.

Bammesberger, A. & J. Grzega (2001) "ModE *girl* and Other Terms for 'young female person' in English Language History" *Onomasiology Online* 2.

Breeze, A. (2002) "Seven Types of Celtic Loanword", in M. Filppula, J. Klemola & H.Pitkänen (eds.) *The Celtic Roots of English*, University of Joensuu Press, Joensuu, pp. 175-181.

Brewer, C. (2000) "*OED* Sources", in L. Mugglestone (ed.) *Lexicography and the* OED: *Pioneers in the Untrodden Forest*, Oxford University Press, Oxford, pp. 40–58.

Crawford, S. (1999) *Childhood in Anglo-Saxon England*, Sutton, Stroud.

Davies, M. (2004-) *BYU-BNC: The British National Corpus*, available online at <http://corpus.byu.edu/bnc> [last accessed 10th October 2009].

Davies, M. (2008-) *The Corpus of Contemporary American English*, available online at <http://www.americancorpus.org> [last accessed 10th October 2009]

Diensberg, B. (1985) "The Lexical Fields *Boy*/*Girl* - *Servant* - *Child* in Middle English" *Neuphilologische Mitteilungen* 86, pp. 328-336.

Grygiel, M. (2006) "On the Cyclicity of Meaning Alterations in English Historical Synonyms of MAN/MALE HUMAN BEING', in R.W.McConchie, O. Timofeeva, H. Tisaari & T. Säily (eds) *Selected Proceedings of the 2005 Symposium on New Approaches in English Historical Lexis (HEL-LEX)*, Cascadilla Proceedings Project, Somerville, MA, pp 60-68.

Heywood, C. (2001) *A History of Childhood: Children and Childhood in the West from Medieval to Modern Times*, Polity Press, Cambridge.

Kay, C.J. (1997) "Historical Semantics and Material Culture", in S.M. Pearce (ed.) *Experiencing Material Culture in the Western World,* Leicester University Press, London, pp. 49-64.

Kay, C., J. Roberts, M. Samuels & I. Wotherspoon (2009) *Historical Thesaurus of the Oxford English Dictionary,* Oxford University Press, Oxford.

Kleparski, G. (1997) *Theory and Practice of Historical Semantics: The Case of Middle English and Early Modern English Synonyms of Girl/Young Woman*, University Press of the Catholic University, Lublin.

Kleparski, G.A. (2004) "CDs, Petticoats, Skirts, Ankas, Tamaras and Sheilas: The metonymical rise of lexical categories related to the conceptual category *FEMALE HUMAN BEING*", in C.J. Kay & J. Smith (eds) *Categorization in the History of English*, John Benjamins, Amsterdam, pp. 71-84.

Liberman, A. (2008) *An Analytic Dictionary of English Etymology: an Introduction*, University of Minnesota Press, Minneapolis, MN /London.

Mintz, S. (2008) "Reflections on Age as a Category of Historical Analysis", *Journal of the History of Childhood and Youth* 1 (1), pp. 91-94.

Taylor, A. (1929) "The Semantics of 'Child'" *Modern Language Notes* 44 (5), pp. 309-314.

Michael Adams (ed.), *"Cunning passages, contrived corridors": Unexpected Essays in the History of Lexicography*, 309-323 ©2010 Polimetrica International Scientific Publisher Monza/Italy

"The various forms of civilization arranged in chronological strata": Manipulating the *HTOED*

Marc Alexander – University of Glasgow
m.alexander@englang.arts.gla.ac.uk

1. Introduction

The complete *Historical Thesaurus of the Oxford English Dictionary* (*HTOED*) is an unparalleled companion to the *Oxford English Dictionary* itself for the study of words. Its database, containing a little under 800,000 sense entries, is a resource which can be used in many different ways beyond the preparation of the printed volumes.[1] This article uses the *HTOED* database to take some steps towards generating data-rich displays of some of the data stored within the thesaurus itself. With access to the database and means of organizing what they find there, researchers can unlock data contained but not accessible within the *OED* (the principal 'parent' of the *HTOED* (Kay *et al* 2009: xiii)) due to its alphabetical arrangement. Charlotte Brewer, identifying the problem of scope, refers to the statement in *The Times* from which this article's title is taken:

> ...even the intensively habitual user [of the OED] could not hope to construct, from an overwhelming multiplicity of individual items, the complete picture, 'the various forms of […] civilization arranged in chronological strata'… (Brewer 2007: 232)

By taking some steps towards visually displaying *HTOED* data, we can glimpse ways in which this thesaurus moves toward giving us a 'complete picture'.

[1] For information about the structure of the original database, see Kay and Chase (1987), Wotherspoon (1992) and Wotherspoon (this volume). The database used here was based on that described in those articles, with a large number of structural modifications by the author.

2. The Data

As outlined elsewhere in this volume (in the chapters by Kay and Wotherspoon), the data stored within the *HTOED* is a fine-grained conceptual hierarchy containing almost all of the recorded words in English, arranged semantically. Each category of words is nested within other, wider categories, so that, for example, the verb category *Live dissolutely* is within *Licentiousness*, itself adjacent to *Guilt* and *Rascalry* and within the wider category *Morality*. This hierarchical structure differs from the organization of many other thesauri, such as that of Peter Mark Roget. While Roget's categories exist in a single linear sequence, *HTOED* categories can relate to others either horizontally (on the same hierarchical level) or vertically (on a higher or lower level, either containing or being contained by another category). In addition, each concept is able to contain a series of subcategories within itself, separate from the main sequence. It is this complex hierarchical structure which helps make the *HTOED* database so useful for visualization: each individual point in the hierarchy can contain both word entries for the concept represented by that point, and also all the conceptual descendants which follow it, each surrounded by siblings of similar meaning.

The size of the *HTOED* also makes it amenable to computational analysis. The current version of the database (as of early 2010) contains 793,747 entries, compared to *OED2*'s 616,500 (Algeo 1990: 137), all within 236,346 categories, each representing a distinct concept. Taking into account each field stored within it, the database itself contains approximately 22.7 million pieces of data.

3. Hierarchy and Visualisation

To best display any data visually, an analyst attempts to increase informational density while simultaneously maximising what de Beaugrande and Dressler call "informativity" (1981: 17ff). This is a careful act because, beyond an ideal 'peak', as information density increases within a fixed space, information transfer rapidly approaches zero (Tufte 2006). *HTOED* data presents a challenge in this area due to its hierarchical nature.

The normal metaphor for visualising hierarchy is a tree-like system, like that often used in organisation charts. *HTOED* data is, however, far too large to be used in such a way – even a spider-like tree or hypertree could not represent the thesaurus, whose largest category alone (for the adverb *IMMEDIATELY*) contains over 250 synonyms. Alternative and emerging tree-like representations were investigated (see Robertson *et al* 1991, 2002 and Furnas and Zacks 1994), but none could appropriately represent the scale of the *HTOED*.

An alternative way of displaying hierarchy two-dimensionally is by representing each category as a nested object on a plane, such as a rectangle. This technique produces a "treemap" (see Shneiderman 2009), wherein each entry in a hierarchy is represented by a rectangle which is large enough to contain smaller rectangles representing its descendants while simultaneously being itself small enough to nest within further rectangles representing its parent categories, rather like a Russian matryoshka doll. In short, a treemap structure takes the organisational chart metaphor of SENIOR IS UP and replaces it with SENIOR IS BIG.

4. *Passion, Love* and *Hate*

Figure 1 shows the *HTOED* data for the transitive verbs for *Affect with passion/strong emotion* (Kay *et al* 2009: 1053) as a treemap. Each rectangle is equal in area, if not identical in dimension, and the algorithm varies these dimensions while keeping area constant in order to ensure it tessellates precisely.[2] The visualisation in Figure 1 shows an extra dimension of data by adding shading to each rectangle, with the shade representing the first cited date of each word, mapped onto a linear scale between 1000CE (black) and 2000CE (white). This necessitates converting the *Old English* value

[2] The algorithm used throughout this paper is Treemap-Squarified, written by Ben Bederson and Martin Wattinberg and available as an open source Java implementation from the University of Maryland Human-Computer Interaction Lab at http://www.cs.umd.edu/hcil/treemap/.

in *HTOED* to an integer value of 1000, and the *Current* value to 2000. These values were chosen to give a reasonable linear scale.

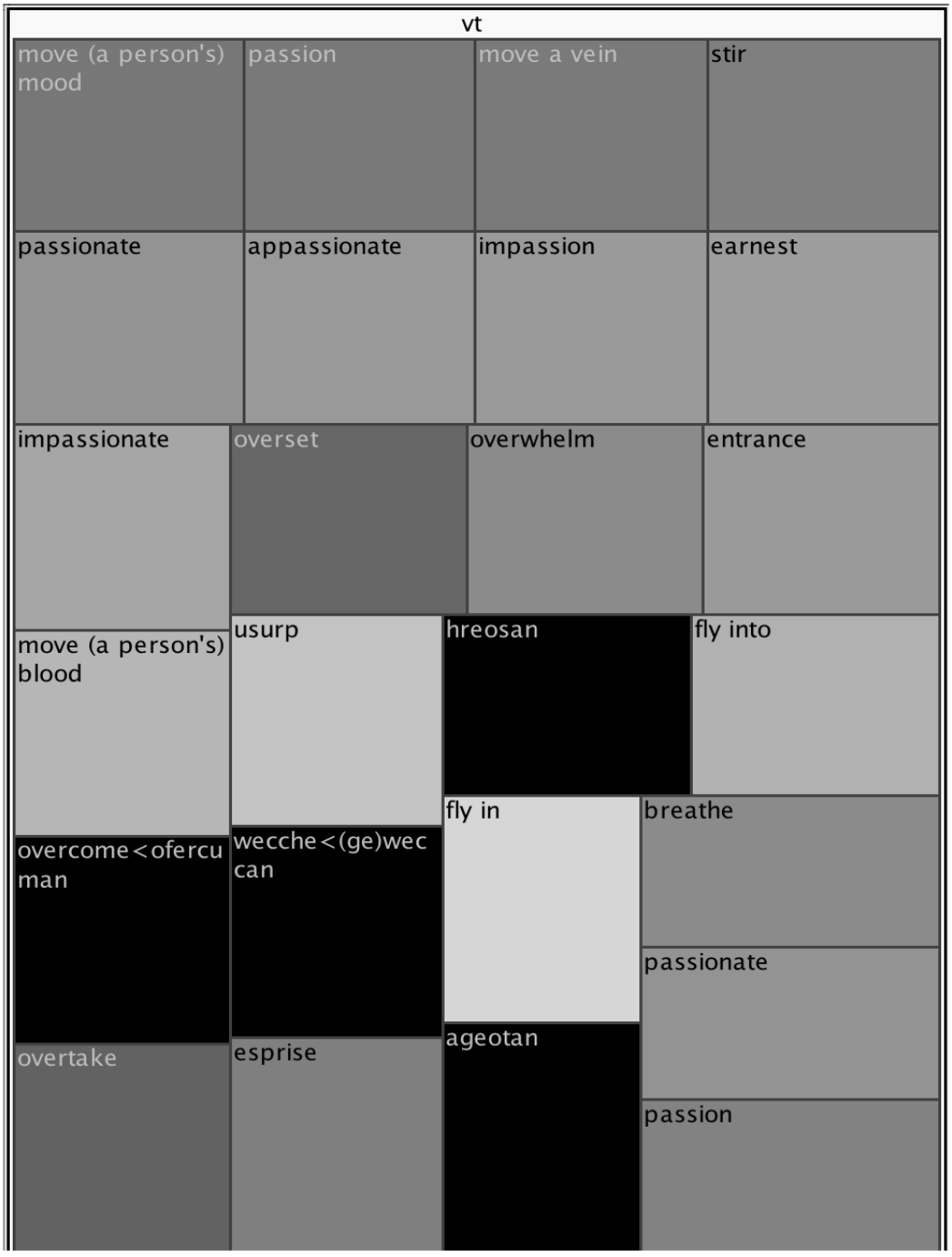


Figure 1. 02.02.15 (vt.): *Affect with passion/strong emotion* by first cited date

This visualization, although large and thus relatively data-poor, displays features of interest. It maintains the principle that word entries should be the most salient part of the image while it attempts to add as much information as feasible, for instance, adding a display of the first cited date and an overlay of the word form, as here. This extra dimension can vary, however: Figure 2 shows all of the *Strong feeling/passion* category with the parts of speech differently coloured (white to black in alphabetical order of adjective, adverb, interjection, noun, verb intransitive, verb reflexive and verb transitive). From the visualization, one quickly perceives that, unlike some categories, *Strong feeling/passion* is dominated by both adjectives and nouns (compare 01.02.08.02 *Beverage*, a category dominated by nouns, or 01.05.08.04.01 *Swift movement*, dominated by verbs).

Figure 2. 02.02.15 (all): *Strong feeling/passion* by part of speech

HTOED data can also be filtered to contextually represent a subset of the data. Figure 3 shows a subset of the *Ardour/fervour* category, displaying only those words which have citation evidence after 1870CE. This date was deliberately chosen to approximate usage in modern times (due to the age of parts of the *OED*). This

allows an analyst to compare the overall category above with the recent category, and to visually inspect areas of interest, such as the small number of words from Old English (eight), and the relatively large number from Middle English.

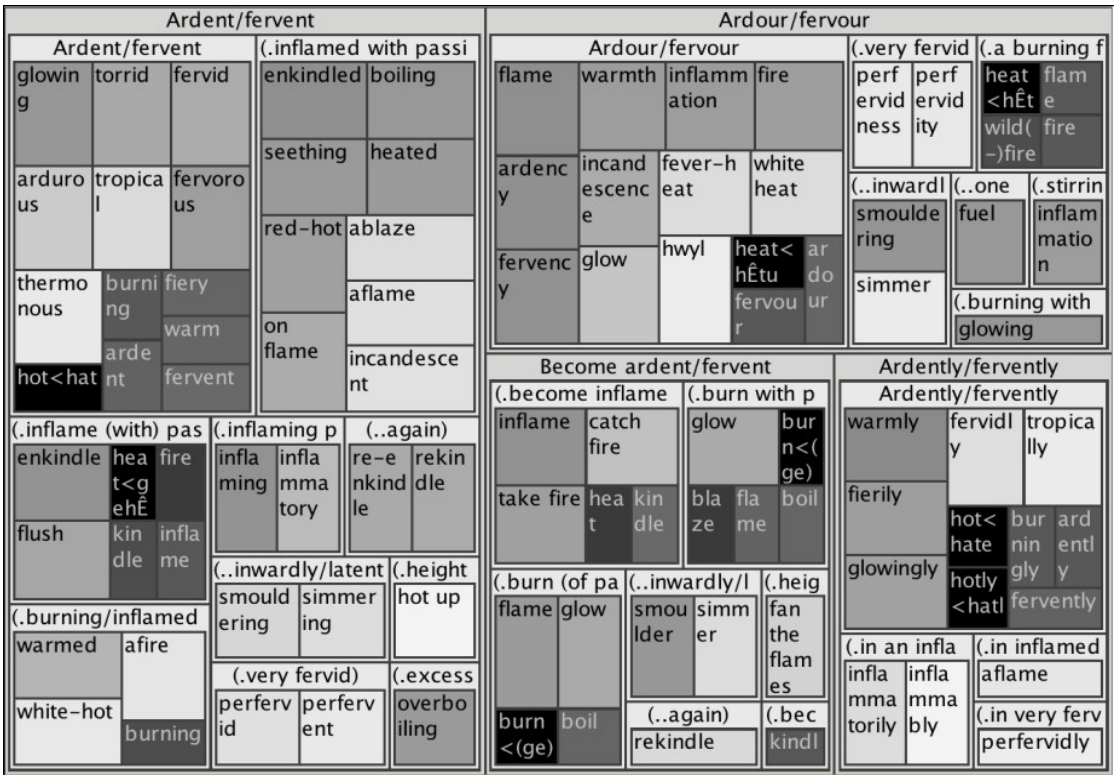


Figure 3. 02.02.15.01 (all): *Ardour/fervour,* current citation evidence only

The other option to add information, alongside colour/shading, is to vary the size of each rectangle. While this is technically possible, it is not ideal for linguistic data. The general principle for visualisation is to ensure that the number of variable dimensions should not exceed the number of dimensions in the data itself; citation dates, parts of speech and other information contained in the *HTOED* are generally linear or categorical, and so unidimensional. As Edward Tufte says:

> There are considerable ambiguities in how people perceive a two-dimensional surface and convert that perception into a one-dimensional number. Changes in physical area on the surface of a graphic do not reliably produce proportional changes in perceived areas. […] These designs cause so many problems that they should be avoided. (Tufte 2001: 71; see also Macdonald-Ross 1977)

Experiments in this area have shown that Tufte's advice is best followed with the *HTOED* visualizations.

Lastly, the final axis which can be varied with respect to these treemaps is scale; that is, the size of each word-rectangle can be reduced, increasing the range of data which each can display. This loses one dimension of data, as small rectangles do not have enough space available for text to be superimposed upon them (such as the word form the rectangle represents), although this can also be an advantage in focusing the visual display on wider, more data-dense patterns. Figure 4 shows the category *Love*, with the first-cited date scale introduced above, and can be compared to Figure 5, *Hate*. It is immediately apparent that *Hate* is a "darker" category than *Love*; in the monochrome scale used here, a category which appears "dark" is one which is relatively early, with most lexical innovation occurring in *Hate* in the Old and Middle English periods (note that Old English is always represented as black, as it is a binary category in the *HTOED* database).

Figure 4 (left). 02.02.22 (all): *Love* by first cited date

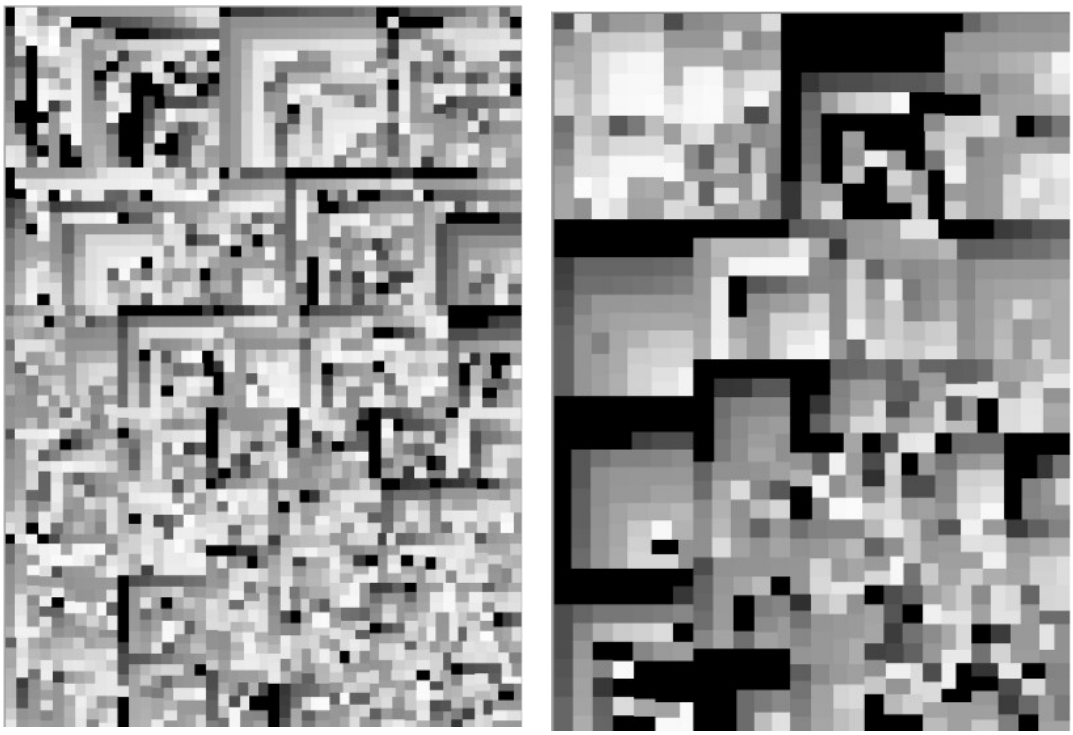


Figure 5 (right). 02.02.23 (all): *Hate* by first cited date

5. Mapping English

The sections above explored the visualization of the *HTOED* at the level of lexical and categorical exploration; what might be termed the "mid-level" of the thesaurus. Moving back further in the hierarchy lets us view the semantic structure of the English language as a whole.

Present-Day English

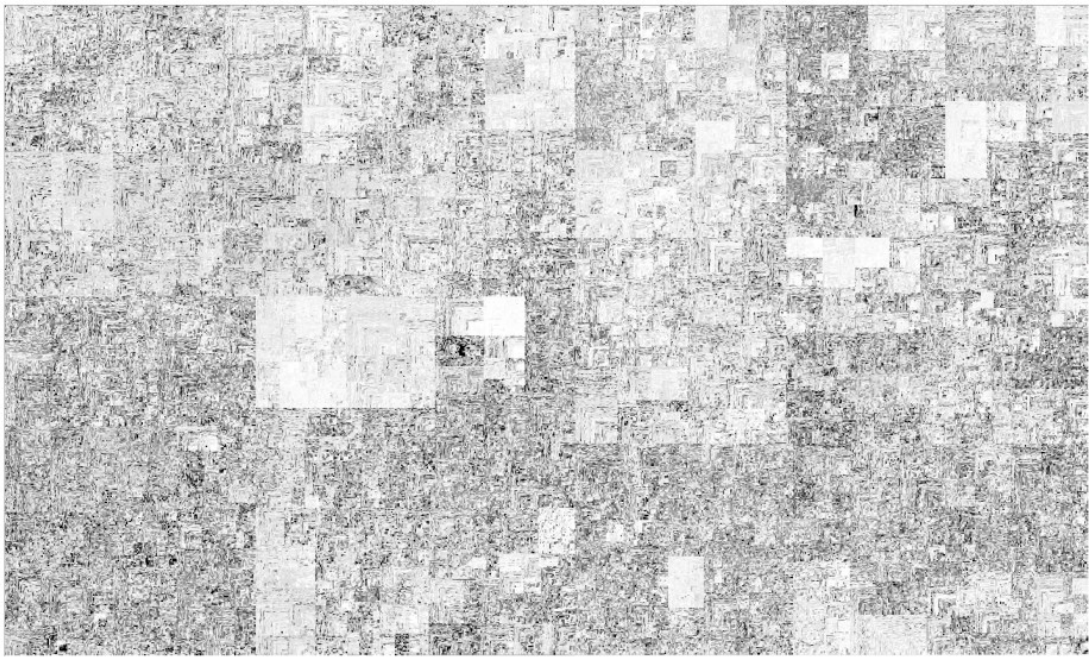


Figure 6. Present-day English as shown by the *HTOED*, shaded by first cited date

Figure 6 is a treemap showing all of present-day English in the *HTOED*, with every word represented by a small dot of ink. Those black dots are present-day words which originated in Old English, and white dots represent those which entered the language much more recently. Again, the map is arranged by semantic field, so words in close semantic proximity are also physically close to one another on the diagram. Such a visualisation is not very useful without a key, which Figure 7 provides.

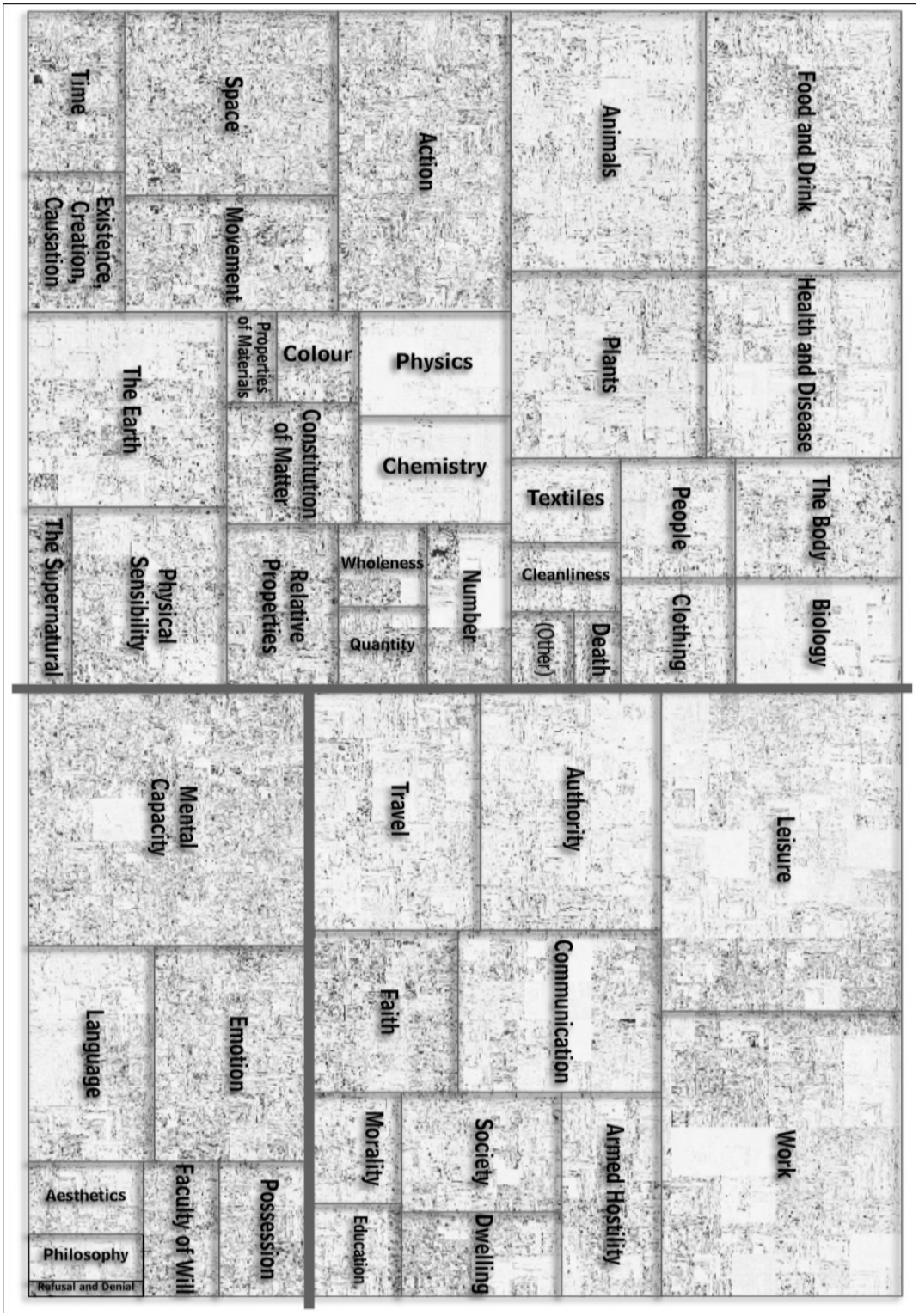


Figure 7. A key for the present-day English treemap

One unintended side-effect of this visualization is that it produces what might be called a "patchwork" effect. Areas such as *Physics* and *Chemistry* are quite light, as are parts of *Number* (which includes *Mathematics*) and *Language*. Although unexpected, this is natural – such light patches are areas of recent lexical innovation, made up of clusters of words first cited in the *OED* in recent years (or, rather, from the late 19th century onwards – recent from the perspective of much of *OED2*, an issue *OED3* will address). Therefore, we can expect a "patchwork" effect in areas affected by rapid social, technological or academic growth, such as *Computing* (inside *Number*, adjacent to *Mathematics*), *Physics, Chemistry, Linguistics, Communication, Travel*, and so on. Conversely, darker and therefore older patches cover existence in *Time and Space, Creation, Causation, Faith, Emotion*, and the parts of *Number* which refer to *Arithmetic* or *Enumeration*.

Diachronic 'Slices'

This effect is pronounced in present-day English, but if other selections of the data are taken, it reduces somewhat. If the present-day data is thought of as a 'slice' of the *HTOED*, then other such slices can be taken between the Old English period and the present day.

Figure 8 shows three such 'slices' of the data, centred on major literary-historical figures of the Late Modern, Early Modern and Middle English periods. Proceeding backwards from the present day, the first treemap is that of Samuel Johnson, and shows those entries first cited before his death in 1784 and last cited after his birth in 1709, a total of 247,933 senses. The second covers those senses we have evidence of being in use during the life of Shakespeare (1564-1616, 207,930 senses) and the third illustrates the same for Geoffrey Chaucer (an approximation of 1340-1400, with 73,432 senses).

Looking at these treemaps, there is a visible reduction between present-day English and that of Johnson in the patchwork effect observed in Figure 6, although it is still observable in places (most notably in the area of *Leisure*). A further reduction is visible in the semantic space of Shakespeare where these shades are much more

evenly distributed, and by the time of Chaucer almost no delineated rectangular patches of innovation can be found.

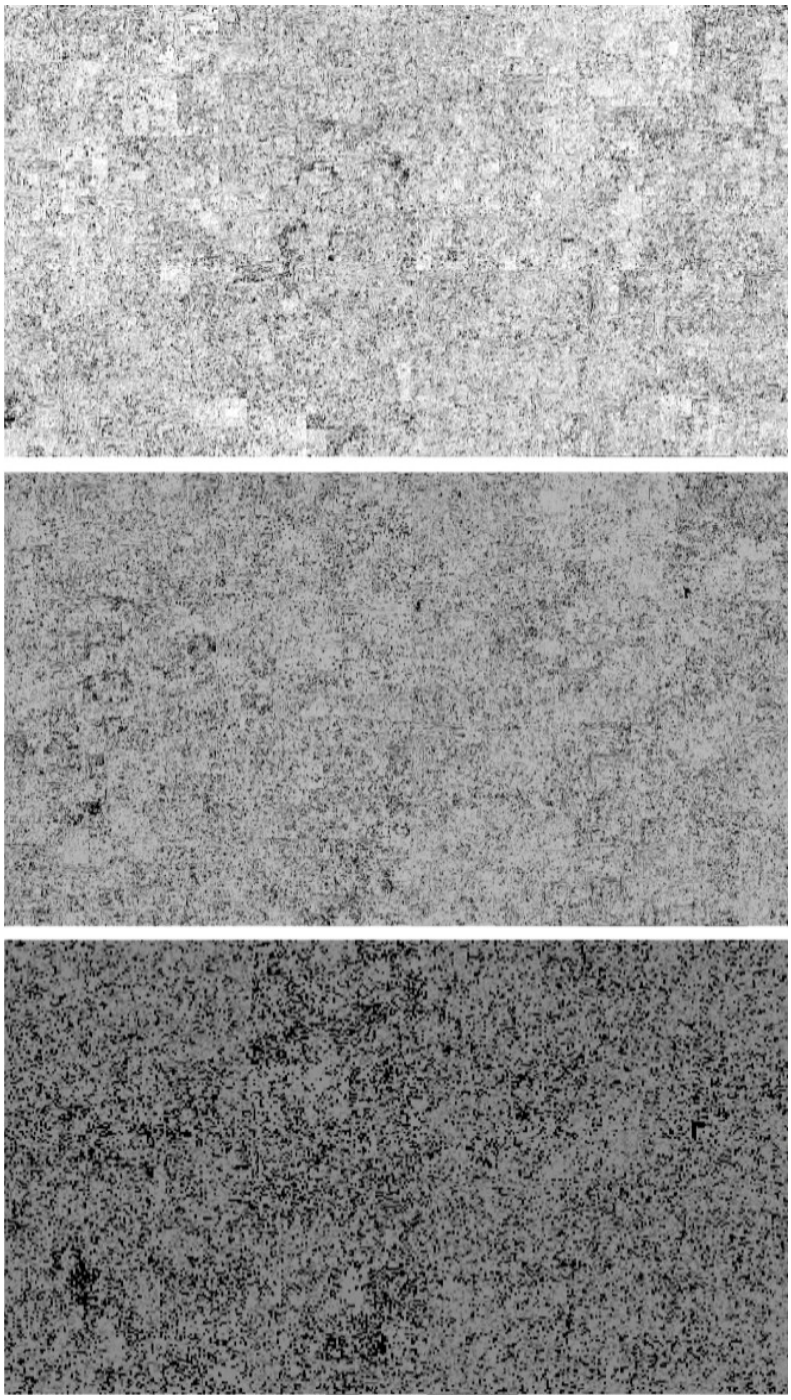


Figure 8. From top to bottom: English during the life of Samuel Johnson, William Shakespeare and Geoffrey Chaucer respectively, shaded by first cited date

One point to note here is that although the images in Figure 8 have been displayed at identical size for clarity's sake, when the size of each word is fixed then the relative size of these images varies greatly (as is apparent from the difference in granularity of the word 'blocks' between Johnson and Chaucer). To disambiguate, Figure 9 shows the relative sizes of the semantic spaces of all four high-density treemaps in this section.

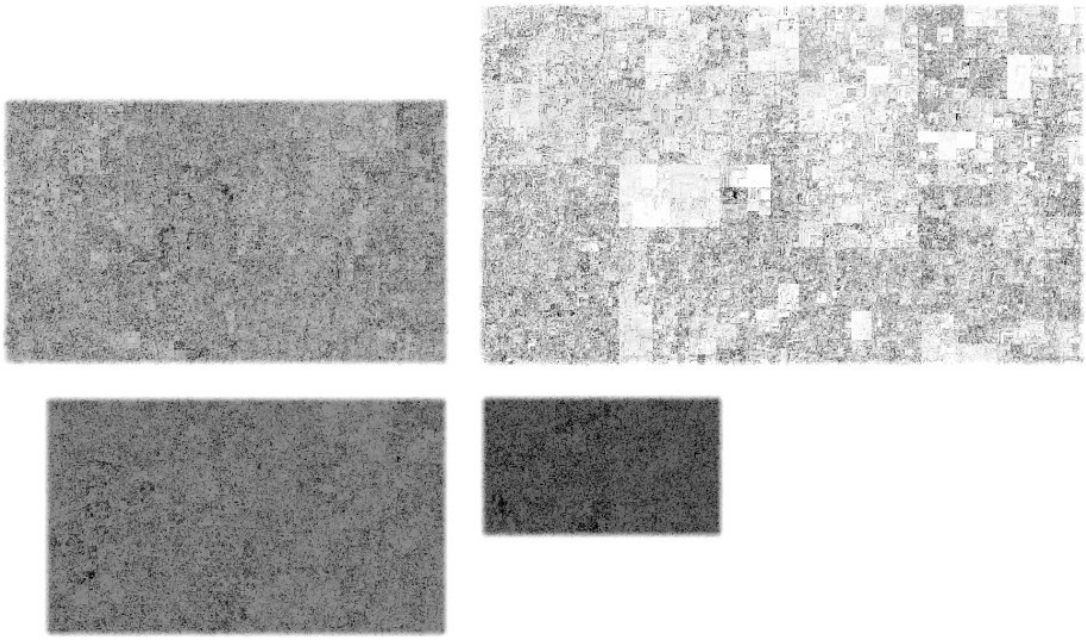


Figure 9. Relative sizes of *HTOED* visualization 'slices'; anti-clockwise from top right, present-day English, English in the age of Johnson, English in the age of Shakespeare and English in the age of Chaucer

Overlays

Finally in this section, it is possible, using two colour scales, to overlay one set of data results on an existing chart. As a demonstration, Figure 10 overlays on the present-day English treemap all those words which have been first cited in the *OED/HTOED* since 1983 (that is, in the lifetime of the current author). Some patterns show up which bear further investigation – a small cluster in *Computing* highlighting the large number of neologisms in that area in the 1980s and 1990s, and a surprisingly large cluster in *Leisure* which shows a jump in music-related neologisms (especially in 03.11.03.01.03.03 (n.) *Pop music*, with

entries such as *GANGSTA RAP* 1990-, *TECHNO-HOUSE* 1991- and *TRIP HOP* 1994-; cf p.1705 of Kay *et al* 2009). Such overlays can show the location of selected senses and link the study of neologisms, word forms or other information with their semantic distribution. Beyond this, they should also be of interest to researchers into the practice of lexicography, giving (as here) information about recent emphases in the collection of neologisms pre-*OED3*, as it is unlikely that the high number of musical terms in the overlay shown in Figure 10 is proportionally correlated with developments in the English language itself, but is rather related to the nature of the *OED Additions* series within the data as a whole.

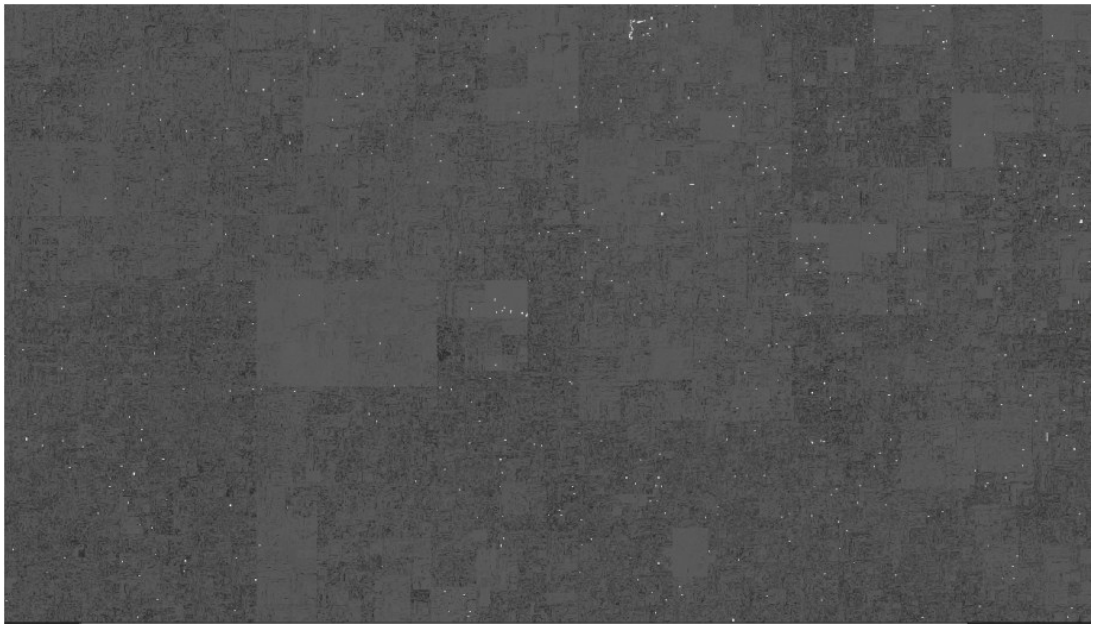


Figure 10. Neologisms since 1983 as recorded in *HTOED*

Similarly, Figure 11 similarly overlays onto present-day English the famously polysemous word-form *set* in *HTOED*, which appears in 313 distinct categories. Its distribution is not as clustered as the neologisms above, with remarkably few clusters – demonstrating the extensive spread of the word form across the semantic space of English.

Figure 11. The senses of the word-form *set*, as recorded in *HTOED*

6. Summary

In the ways outlined above, such visual displays of *HTOED* data can provide useful entry points to a large, complex lexicographical and lexicological dataset. Firstly, in a pedagogical sense, displays could give students and others a new way of looking at lexicological data, and of exploring them as an application of semantic field theory.[3] Secondly, as computer displays and online dictionary interfaces become more polished, new ways of encouraging exploration of lexical data online are needed to replace the lost experience of browsing a printed dictionary, rather than only providing users with a blank search interface. And finally, such visualizations can point analysts towards areas of possible semantic, lexical or cultural interest, whether areas of trauma in the history of the language, areas of rapid growth, or areas of relative stability. All of these could be made easily visible through some enhancement and development of

[3] Adrienne Lehrer (1974: 15) describes semantic fields as the theory that "the words of a language can be classified into sets which are related to conceptual fields and divide up the semantic space or the semantic domain in certain ways." If Figure 7 above shows the semantic space of English, then Figure 6 can be said to show the "word space" of the language, a plane containing lexical items and parallel to the semantic space, but without explicit semantic divisions.

the first steps given here towards displaying and exploring the data stored within the *HTOED* itself.

References

Algeo, J. (1990) "The Emperor's New Clothes: The Second Edition of the Society's Dictionary", *Transactions of the Philological Society*, 88(2), pp. 131-150.

de Beaugrande, R.-A. & W. Dressler (1981). *Introduction to Text Linguistics*, Longman, London.

Brewer, C. (2007) *Treasure-House of the Language: The Living* OED. Yale University Press, New Haven, CT.

Furnas, G.W. & J. Zacks (1994) "Multitrees: Enriching and Reusing Hierarchical Structures", *Human Factors in Computing Systems: Proceedings of the CHI '94 Conference 1994*, pp 330-336.

Kay, C. & T. Chase (1987) "Constructing a Thesaurus Database", *Literary and Linguistic Computing* 2(3), pp. 161–163.

Kay, C., J. Roberts, M. Samuels, & I. Wotherspoon (2009). *Historical Thesaurus of the OED*. Oxford University Press, Oxford.

Lehrer, A. (1974) *Semantic Fields and Lexical Structure*. North-Holland Publishing Co., Amsterdam.

Macdonald-Ross, M. (1977) "How Numbers Are Shown: A Review of Research on the Presentation of Quantitative Data in Texts", *Educational Technology Research and Development*, 25(4), pp. 359-409.

Robertson, G.G., J.D. Mackinlay, & S.K. Card (1991). "Cone Trees: Animated 3D Visualizations of Hierarchical Information". *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Reaching through Technology*, pp.189-194.

Robertson, G.G., K. Cameron, M. Czerwinski, & D. Robbins (2002) "Animated Visualization of Multiple Intersecting Hierarchies". *Journal of Information Visualization*, 1(1), pp.50-65.

Shneiderman, B. (2009) "Treemaps for Space-constrained Visualization of Hierarchies", http://www.cs.umd.edu/hcil/treemap-history/index.shtml, accessed 26 March 2010.

Simpson, J. & E. Weiner (eds.) (1989) *The Oxford English Dictionary*, 2nd. ed., Oxford University Press, Oxford.

Tufte, E. (2001). *The Visual Display of Quantitative Information*, 2nd ed., Graphics Press, Cheshire, CT.

Tufte, E. (2006). *Beautiful Evidence*. Graphics Press, Cheshire, CT.

Wotherspoon, I. (1992) "Historical Thesaurus Database Using Ingres", *Literary and Linguistic Computing* 7(4), pp. 218-225.

Lexicography worldwide:
theoretical, descriptive and applied perspectives

Anche se l'attività lessicografica ha alle proprie spalle una lunga e onorevole tradizione, gli anni recenti hanno visto nuovi e significativi sviluppi nella teoria lessicografica e nella descrizione e produzione dei dizionari, in un periodo in cui il lessico ha riconquistato una certa preminenza nella riflessione linguistica.

La collana "Lexicography worldwide: theoretical, descriptive and applied perspectives" si propone pertanto come forum di discussione sulle nuove prospettive in ogni ambito lessicografico: i nuovi sviluppi nella teoria lessicografica; le analisi critiche dettagliate dei dizionari del passato e del presente; le ricerche sull'interfaccia tra i dizionari, la loro produzione e il loro uso, e gli utenti.

La collana pubblicherà studi monografici, raccolte di saggi o atti di convegni. Ciascun volume della collana è sottoposto a revisione anonima da parte di due membri del Comitato Scientifico e/o del Comitato Consultivo internazionale (attualmente composto da 13 studiosi americani, inglesi, francesi, spagnoli e italiani). Si auspica che la pubblicazione in rete open-access della collana "Lexicography

worldwide" possa incoraggiare il dialogo scientifico internazionale tra i ricercatori, e tra questi e i lessicografi di professione.

~ ~ ~ ~

Though the practice of dictionary-making has a long, time-honoured tradition behind, recent years have witnessed new and significant developments in lexicographical theory and in the description and production of dictionaries, at a time when lexis has regained prominence in linguistic research.

The book series "Lexicography worldwide: theoretical, descriptive and applied perspectives" is therefore meant to be a forum for discussion and debate over new perspectives on any aspect of lexicography: new developments in lexicographic theory; detailed critical analyses of past and present dictionaries; research on the interface between dictionaries, their production and use, and their users.

The book series will publish monographs, collections of essays or conference proceedings. Each book in the series is blind peer-reviewed by two members of the Scientific Committee and/or the international Advisory Board (currently consisting of 13 American, British, French, Spanish and Italian scholars). It is hoped that the online open-access publication of the "Lexicography worldwide" series will encourage international scientific dialogue among researches and between researchers and practising lexicographers.

~ ~ ~ ~

Bien que l'activité lexicographique ait derrière soi une tradition longue et honorable, au cours de ces dernières années tant la lexicographie que la description et production de dictionnaires ont connu un développement considérable, parallèle au renouveau d'intérêt dont le lexique bénéficie au seine de la réflexion linguistique.
Dans une telle optique, la collection « Lexicography worldwide : theoretical, descriptive and applied perspectives » se propose comme un lieu de discussion sur les nouvelles approches dans tous les domaines lexicographiques : nouveaux développements dans la théorie lexicographique ; analyses critiques des dictionnaires d'hier et

d'aujourd'hui ; recherches sur les rapports entre les dictionnaires, leur production et leur usage.

Cette collection accueillira des études monographiques, des recueils d'articles ou des actes de colloques. Chaque volume de la collection sera soumis à la lecture, sous forme anonyme, de deux membres du Comité Scientifique et / ou du Comité de lecture international (celui-ci étant composé, à l'état actuel, de 13 spécialistes américains, anglais, français, espagnols, italiens). On souhaite que la publication de la collection « Lexicography worldwide » sur le réseau Internet en libre accès encouragera les échanges scientifiques internationaux entre les chercheurs, et entre ceux-ci et les lexicographes professionnels.

~ ~ ~ ~

Aunque la actividad lexicográfica cuenta ya con una larga y honrosa tradición, en los últimos años se han producido nuevos y significativos avances, tanto en cuestiones teóricas como en producción de diccionarios, en un momento en el que se puede observar que el estudio del léxico ha alcanzado cierta preponderancia en la reflexión lingüística.

La colección "Lexicography worldwide: theoretical, descriptive and applied perspectives" se propone como fórum de discusión de nuevas perspectivas en ámbito lexicográfico: renovados aspectos teóricos; análisis crítico de diccionarios del pasado y de la actualidad; investigaciones sobre la interfaz entre diccionarios, su produción, utilización y usuarios.

La colección publica estudios monográficos, artículos seleccionados y actas de congresos. Dos miembros, componentes del Comité Científico o bien del Comité del Consulta Internacional (actualmente compuesto por 13 expertos: americanos, ingleses, franceses, españoles e italianos), son los encargados de evaluar de modo anónimo la obra. La colección "Lexicography worldwide" nace con la intención de estimular el diálogo científico internacional entre investigadores y lexicógrafos de profesión, a través de su publicación abierta on-line.

Dear Reader,

many thanks for your interest in our books!

All our activities are aimed at generating, contributing to or disseminating accessible knowledge to anyone in the world who might be interested, without obstacles concerning economic disadvantages.

We're trying to build a **new model of scientific publishing** that embraces economic self-subsistence, openness, and fairness (**Open Access Publications**).

If you think our efforts are worthy, would you please support us **by ordering the printed edition** of this book? You can purchase the book by credit-card at our shopping cart: **www.polimetrica.com**.

For more information you can send an e-mail to our address: **editorial_office@polimetrica.com**.

Thank you for caring enough to get involved.

Sincerely,

Editorial Staff
POLIMETRICA