Digital thesauri as semantic treasure troves: a Linguistic Linked Data approach to "A Thesaurus of Old English"

Stolk, S.S.

Chapter 2

2. Functionalities required of historical language thesauri for research and education

In order to improve the dissemination of historical language thesauri on the Web, not only is their content of importance but also the functionality required by their target audience. This chapter describes functionality required by users in academia, both for research and educational purposes, of editions of these lexicographic resources. In order to gather information on such functionality, a number of sources have been consulted. First, existing editions of historical language thesauri and handbooks on both thesauri and lexicography in general, which were also consulted in Chapter 1, provided valuable input. Second, academic reviews of these thesauri and notable research employing these resources offer further insights. Lastly, additional input has been collected from stakeholders – experts in lexicography, linguistics and philology (amongst other fields) – through dedicated stakeholder meetings, workshops, and feedback based on preliminary results in research and education in a research project titled 'Exploring Early Medieval English Eloquence'.[1] Within the remainder of this chapter, the present set of functionality gathered is detailed in five sections: navigation (2.1), resource views (2.2), extension (2.3), analyses (2.4), and data management (2.5).

2.1. Navigation

The first and foremost functionality required of thesaurus editions is the ability for users to navigate their content. The preface to *HTE* indicates that users have a choice in how to approach that content:

> There are two ways to approach a thesaurus: by familiarizing oneself with its structure and principles of organization, or, more commonly, by using its index to determine in which category or categories a word appears.[2]

Whether the one approach is truly used more commonly than another is less clear cut. Christian Kay, one of the editors of *HTE*, mentions that the alphabetical index is there "for convenience" and may not be needed for "confident and

[1] The research project and its resulting case studies are discussed in Chapter 8. Appendix 2.A lists the functionality required by researchers participating in this project.

[2] *HTE1*, p. ix.

frequent users" of the thesaurus.[3] Similarly, the editors of *TOE* prompt the user to approach their thesaurus "by subject rather than through the alphabetic index".[4] Nevertheless, both means of navigation are offered to users of historical language thesauri and hence warrant further treatment.

2.1.1. Navigation via the thesaurus structure

Users can employ the topical structure of a thesaurus to move from meaning to lexical items conveying that meaning. In order to facilitate this approach, print editions of the historical language thesauri under discussion all include an outline of their topical system.[5] For those historical language thesauri that have three or fewer levels in their topical system, such as *ScT*, this outline typically presents the system in its entirety. For those thesauri with a larger degree of specialization in their topical systems, termed distinctive thesauri by Christian Kay and Marc Alexander, the outline may be limited to the top categories in order to save space.[6] Such a limited outline is offered in the print editions of *TOE* and *HTE*, which contain over ten levels in their topical systems and provide an outline for the first two and three of these levels, respectively.

Through the use of hyperlinks, electronic editions of historical language thesauri can supply users with the means to navigate the topical system in an interactive manner. The majority of the current electronic editions of these thesauri offer an overview similar to those found in the print editions. The overviews of *BTH*, *TOE4*, *HTE2*, and *HTE3* are located on a single webpage, allowing the user to expand or collapse categories to show or hide subordinate ones. Only when selecting a specific category for viewing purposes will this edition show the senses contained within. Such a useful overview has not been offered by all Web-based editions of thesauri. *TOE3*, no longer available, and *HTS* require the user to navigate the structure by taking single steps further down, or up, the hierarchy. A new webpage is loaded in the browser for each position in the topical system that the user passes through. In contrast, all electronic editions indicate the current position in the taxonomy through breadcrumbs, a trail of hyperlinks of the current category and ones superordinate to it.

2.1.2. Navigation via the thesaurus index

The second manner in which users can approach a thesaurus is through an alphabetic index. Reviewers of historical language thesauri have called the availability of such an index "necessary" and even "indispensable" in order to find a word or phrase within the categories of thesauri.[7] The added value of

[3] Kay and Alexander, 'Diachronic and Synchronic Thesauruses', p. 368.
[4] *TOE2*, p. xv.
[5] *TOE2*, pp. v–viii; *LSM*, pp. iv–xvii; *HTE1*, pp. xxix–xxx; *ShT*, pp. xvii–xxv; *ScT*, pp. iv–vii. *DSSPIEL* has separated its outline over the table of contents for the primary categories (p. xix) and the sections that treat each of those categories. These sections begin with a list of the subcategories.
[6] Kay and Alexander, 'Diachronic and Synchronic Thesauruses', p. 370.
[7] Görlach, Review of *TOE1*, p. 399; Momma, Review of *TOE2*, p. 80.

an index seems to have been well understood by the editors of the historical language thesauri of Scots and English, since all their printed editions include an alphabetic index to their contents.

Which parts of the contents are indexed alphabetically can vary between thesauri in print. The majority of the historical language thesauri treated here – *ShT*, *LSM*, *TOE*, and *HTE* – index their lexical senses (through their headforms), but not their categories. The editors of the other two printed historical language thesauri have opted to base their index on English rather than their actual lexicon. As such, the index that complements *DSSPIEL* is one to its categories, or headings, rather than to its categorized lexical items.[8] The index of *ScT* is meant to provide "a detailed signposting in English into the intricacies of Scots vocabulary",[9] listing near English equivalents to the Scottish senses, when available, as well as more generic English concepts taken from its English categories.[10] In other words, this index incorporates elements from both the lexical senses (i.e., their definition) and the topical system in order to assist the reader, who is likely to learn more about Scots through the English language than through the Scots vocabulary directly.

Although they have their use for thesauri in printed from, alphabetic indexes are no longer essential for digital thesauri. Instead, the functionality required to find a word or phrase may be provided by a search system. Indeed, the digital editions of *TOE* and *HTE* sport such search systems and have abandoned an alphabetical index. These search systems offer a vast improvement over an index in three ways. The first is in terms of retrieval speed — going back and forth between index and the thesaurus proper can now indeed take only "a twinkling of an eye".[11] The second improvement is that both categories and lexical senses can be included in searches without any serious drawbacks in this method of access. Where textual editions were criticized for not containing alphabetical indexes to both categories and entries,[12] possibly due to limits on printing space imposed by publishers or to avoid too great a weight of the books for readers to handle them, the digital editions of *TOE* and *HTE* offer this sought-after ease of access through their new search systems. The third improvement offered by digital search systems is that lexical senses can now be located based on the additional information per sense, including the labels they carry. This allows searches to be restricted, for instance, to items of a particular part of speech or to those marked as 'poetical'. Of course, it would be just as interesting to restrict searches to senses that do *not* carry a given label, since these senses are implicitly marked by the editors as belonging outside of the group indicated by the absent label.[13] The possibility to exclude a label from searches would allow one, for example, to locate only non-poetical lexical senses. This functionality of exclusion of labels in searches is not yet provided by the current systems in

[8] *DSSPIEL*, p. 1505.
[9] *ScT*, p. xix.
[10] *ScT*, p. xix.
[11] Busse, 'A Celebration of Words and Ideas', p. 804.
[12] Görlach, Review of *TOE1*, p. 399.
[13] Svensén, *A Handbook of Lexicography*, p. 315.

place for the digital versions of *TOE* and *HTE*.

2.2. Resource views

The second functionality required of thesaurus editions is the ability to view all the information available on a specific category or lexical item that a user chooses to inspect. For lexical items these resource views ideally contain the part of speech, language, and usage features; and for categories, their name and identification string.[14] Moreover, these overviews should indicate relations to other resources where relevant, such as, for categories, a list of the lexical senses that are allocated to the viewed location of the thesaurus taxonomy or, concerning lexical items, an overview of synonyms. These essential resource views are present in historical language thesauri (e.g., as entries in a printed thesaurus), albeit not always as complete as possible.

One piece of information useful for research, noted by both editors and reviewers of thesauri and absent from existing resource views of lexical items, is an overview of the different senses of a single lexeme.[15] Such views of senses related through polysemy facilitate explorations of metaphorical and metonimical uses and offer a measure of ambiguity in utterances in which the lexeme in question occurs. Through these connections of individual lexemes, it is possible to detect which semantic fields in a language have close conceptual ties. For example, terminology surrounding sleep can be used metaphorically for death; temperature for emotions.[16] By mapping out metaphorical connections such as these, it is possible to gain a better understanding of the stylistic impact of metaphors and to grasp which groups of words are more easily used to symbolically represent other meanings.[17] Existing historical language thesauri of Scots and English, rather than including the information on senses related through polysemy in the resource view of a lexical item, rely on the user switching to either the index (in print editions) or the search engine (in digital editions) to obtain this knowledge on the lexical item inspected.

Resource views could also be employed to present details on elements other than categories and lexical items. Labels found in thesauri are valuable elements of information, too, that may warrant their own resource view. Such labels, which benefit from thorough descriptions, are utilized by researchers to investigate the usage features with which those labels are associated.[18] Providing resource views for these elements could assist researchers in finding helpful information on them, including a full name and description instead of only a code or abbreviation. Moreover, providing these resource views for labels can assist in retrieving the elements they mark. One such need is exemplified in Kathryn Allan's work on

[14] See Chapter 1 for a thorough overview of information available in historical language thesauri.
[15] See *HTE1*, p. x; Ilson, 'On the Historical Thesaurus of the Oxford English Dictionary', p. 256.
[16] For both metaphorical ties mentioned, see *Mapping Metaphor with the Historical Thesaurus*.
[17] Ibid.
[18] Brewer, 'Labelling and Metalanguage', p. 493; Norri, 'Regional Labels in Some British and American Dictionaries'. See the discussion in section 1.8 of Chapter 1.

metaphor and metonymy, drawing on information from *HTE*.[19] Lexical senses in *HTE* were marked with labels that represent what Allan calls core concepts. Subsequently these custom labels were used to retrieve all items tagged with a specific core concept, such as ANIMAL or AGE.[20] Although the need for adding custom labels will be explored further in section 2.3, the demand for viewing a list of all items to which a label is applied can be satisfied through a resource view for this kind of element. Despite these reasons, none of the Web-based editions of historical language thesauri offer resource views for elements other than categories and lexical items.

Lastly, resource views would benefit from the means to hide information deemed irrelevant to a user. This aspect has been noted by creators and reviewers of thesauri alike as important.[21] The editors of *HTE*, for instance, state that the very purpose of their thesaurus is "to provide a detailed record of the English vocabulary from the earliest times to the present, with sufficient accompanying information that, for any given period in the past, the user should be able to ascertain the exact state of the vocabulary (i.e., the 'lexical system') which existed at that time".[22] They point out that the *HTE* could thus be used to "act as a thesaurus for any period in the past", allowing one to determine which lexical items will have been available to Shakespeare.[23] However, both digital editions of *HTE* prevent the creation of a subthesaurus based on the recorded diachronic usage features, containing only those lexical senses that were available in Shakespeare's time — or any other subthesaurus based on a set of criteria for that matter. Any sifting is left as manual labour to the user. *TOE*, too, includes valuable tagging information, stating whether its items are found only in poetry or only in glosses, for example. Although *TOE3* allowed viewing subthesauri based on the available tagging information, *TOE4* no longer sports this helpful feature.[24] In short, the existing digital versions of these thesauri lack the ability for their users to view only those items deemed of interest based on usage features and other information available — a filtering ability applicable to both resource views and the mechanisms available to navigate a thesaurus.

2.3. Extension

The third functionality researchers require of thesaurus editions is the ability to extend them, connecting additional information to their content in order to enrich or reuse that existing data. Examples of such extensions are indications of date and dialect, results from corpus searches, and indications whether a word or meaning is found in a particular text, context, or is notable in some other

[19] Allan, *Metaphor and Metonymy.*

[20] Ibid., pp. 21–2.

[21] For an academic review discussing the possibility and desirability of creating subthesauri through such mechanisms of filtering, see Conner, Review of *TOE1*, p. 888.

[22] *HTE1*, p. xiii.

[23] *HTE1*, p. xiv.

[24] For a description of the previously available *TOE* website and its feature to create subthesauri based on tagging information, see Stolk, 'Welcoming the *Thesaurus of Old English Statistics*', pp. 11–14.

qualitative or quantitative way. The functionality to extend thesaurus content offers users the means to have the thesaurus reflect their own interests and to share salient information with others. This section discusses two forms of extension: 1) elaboration on thesaurus content and 2) connecting other bodies of knowledge to a thesaurus.

2.3.1. Elaboration on content

Many scholars desire labelling information in historical language thesauri beyond what has been made available and, for that purpose, the ability to elaborate on thesaurus content. In a review of *ShT*, for instance, Christian Kay expresses that she finds it "frustrating that nothing other than the part of speech of each word is offered to the user".[25] A second example is the existing tagging information per lexical item in *TOE*, which is deemed helpful and efficient,[26] but also thought to be rather limited.[27] Indications of date and dialect, for example, are notably absent. As it stands, all items are treated as belonging to "a single geographically and temporally indistinguishable mass".[28] Researchers certainly possess further insights on these topics: Rolf H. Bremmer Jr asserts that the dialectal origins of many Old English words can be pinpointed globally as Anglian, West-Saxon, or Kentish and that such information would be worthwhile to add to *TOE*.[29] The added knowledge would facilitate research into the impact of regional influences on the Old English vocabulary as it has come down to us and, further along the line, on present-day English. Researchers participating in the workshop series, too, indicated the usefulness of adopting custom labels to add further distinctions in genre, dialect, diachronic usage, and etymology.[30]

The means to elaborate is also desired on elements other than labels for lexical senses. Kathryn Allan, for example, points out that sections of *HTE* available to her lacked lexical items that she thought belonged in these locations and, consequentially, wished to insert there.[31] Similarly, certain lexical items may warrant marking to be stricken from a thesaurus — especially if dealing with words that appear to lack evidence for their existence, so-called ghost words.[32] Another case for elaborating has been made by Nils Århammar, who suggests supplementing *DSSPIEL* with lexical items of various stages of the Frisian language.[33] Additions and refinements to the topical system, too, are valued for research purposes. Christian Kay, in her review on *ShT*, professes that she would welcome a further subdivision in its existing system of rather abstract categories. Such a fine-grained division "would make it easier to identify areas of

[25] Kay, Review of *ShT*, p. 72.
[26] Cavill, 'Names and Things in Anglo-Saxon and Early Norman England', p. 186; Conner, Review of *TOE1*, p. 888.
[27] Bremmer, 'Treasure Digging in the Old English Lexicon', p. 111; Görlach, Review of *TOE1*, p. 399; Dance, Review of *TOE1*, p. 313.
[28] Dance, Review of *TOE1*, p. 313.
[29] Bremmer, 'Treasure Digging in the Old English Lexicon', pp. 111–12.
[30] See Chapter 8 and Appendix 2.A.
[31] Allan, *Metaphor and Metonymy*, p. 20.
[32] Hartmann and James, *Dictionary of Lexicography*, s.v. 'ghost word'.
[33] Århammar, 'A Frisian Supplement to Buck's Dictionary of Indo-European Synonyms?'.

high lexicalisation [i.e., where elaboration of vocabulary is notable], and, with the help of other reference works, to examine areas where Shakespearean innovations are prevalent".[34]

A perhaps more drastic fashion in which to extend thesaurus content is by creating a new topical system to supplant the existing one, reusing and extending only lexical items recorded in a thesaurus. A case in point is Thijs Porck's study of the conceptualisation of old age in early medieval England, which draws on the contents of *TOE*, but replaces the existing categorisation with one more suitable to capture nuances newly established in the semantic field under investigation. Lexical senses from four separate *TOE* categories are grouped together as all belonging to the category of "Human old age".[35] The new categorisation system fashioned for these items contains semantic domains, such as "Positive development", "Wisdom", and "Authority". Similarly, *BTH* indicates that it reuses categories from the topical system of *HTE*, "though these were occasionally modified in order to capture the ways in which conceptions in the Middle Ages differed from those of the present day", and adds words from Anglo French, spoken in England alongside Middle English after the Norman Conquest in 1066.[36] In short, many scholars wish to extend the content of historical language thesauri, to varying degrees, for research purposes.

Unfortunately, any elaboration on thesauri content and subsequent sharing of these additions is currently not facilitated by most of the published forms of the historical language thesauri treated here. Only *HTS*, one of the thesauri available online, sports a comment section in which users can share their thoughts publicly (see Figure 2.1). All other publications of the thesauri analysed do not provide the means for additions to be made by anyone other than their editors. One can scribble additions in the margins of paper copies or refer to bits of information via a textual reference, but these acts do not extend the conceptual thesaurus: The additions either affect only a single copy (in the case of annotations in the margins) or are stored in a different location and format that hampers simultaneous and integrated access to the content of both the original and its additions.

2.3.2. Connecting bodies of knowledge

Expansion can come not only in the form of elaboration on existing thesaurus content, but also by connecting other bodies of knowledge to a thesaurus. The desire for such functionality is expressed both by scholars who utilise or review the thesaurus and the editors themselves. The introduction to the paper editions of *TOE*, for instance, states the following: "Should a precise meaning be wanted, a dictionary is needed".[37] *ScT* contains a similar statement, referring readers to the *Concise Scots Dictionary* for etymology and pronunciation of items from its lexis.[38] The introduction to *HTE*, too, indicates that connectivity between

[34] Kay, Review of *ShT*, p. 73.
[35] Porck, 'Growing Old among the Anglo-Saxons', pp. 68–9.
[36] See *BTH*, section 'About'.
[37] *TOE2*, p. xv.
[38] *ScT*, p. xv.

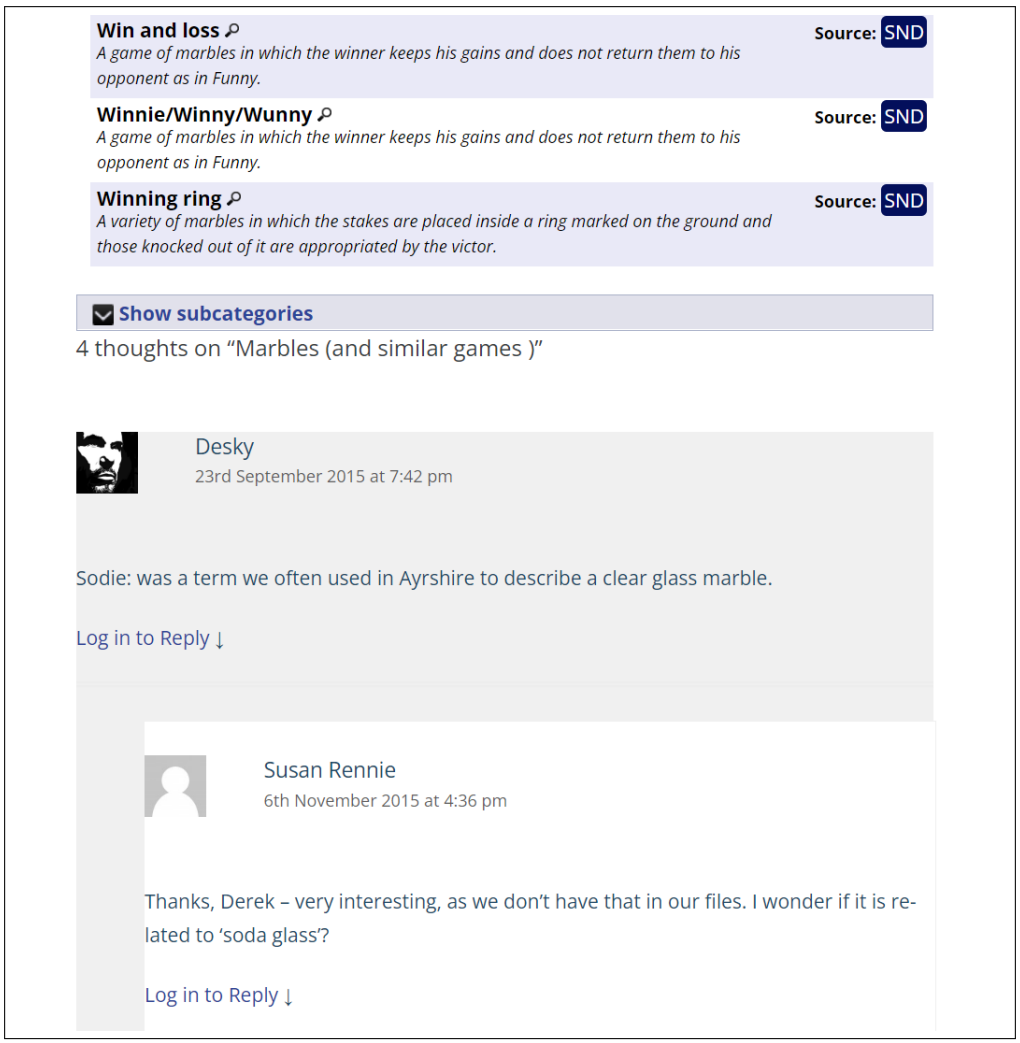

Figure 2.1.: Comments by users on *HTS* category "Marbles (and similar games)".

the thesaurus and a dictionary with supplementing information, the *Oxford English Dictionary*, is desirable.[39] That such relations between bodies are worth exploring is not only suggested by editors but also taken to heart by users. David Crystal, for one, in his explorations of *HTE* content has also turned to related definitions and citations from the *OED*.[40] Likewise, Kathryn Allan has performed analyses of *HTE* items "by looking closely at etymological information supplied in the *OED*".[41] Relations between bodies of knowledge like the aforementioned ones can, as the editor of *ScT* points out, act as "a series of step-by-step doorways into the heart of a national culture".[42]

Relating bodies of knowledge to each other can be achieved in a number of

[39] *HTE1*, p. xiv.
[40] Crystal, *Words in Time and Place*, p. xv.
[41] Allan, *Metaphor and Metonymy*, p. 21.
[42] *ScT*, p. ix.

ways. The extent to which the user is facilitated in retrieving and combining information from multiple bodies varies per solution. The most basic level of relating two bodies is by mentioning to users that such a relation exists, possibly in the introduction. This method is applied in many print editions of historical language thesauri.[43] It is wholly left to the user to access the second body of knowledge and subsequently to locate the desired complementary information. A more user-friendly solution of relating two bodies is by making an explicit mention of where exactly such complementary information can be found. External references per lexical sense do just that.

Although an external reference at a lexical sense directs users to the appropriate location in another body, paper editions still require users to move from body to body in order to accumulate the information sought after. Facilitating users in the fullest manner possible comes from connectivity in a digital environment, such as through hyperlinks. In fact, true connectivity would allow users to access and query knowledge from multiple bodies at the same time, without having to manually move from the one to the other. Unfortunately, as McCracken notes, connecting bodies is "curiously underexplored in lexicography", even though it is "a familiar topic in the context of knowledge bases".[44] By applying conventions found in the context of knowledge bases, it should therefore be possible to remove the somewhat isolated status of historical language thesauri. For users to have complementary information at their fingertips, editors and publishers of thesauri may well want to shift their focus from "aspirations of completeness and comprehensiveness" to "connectivity and interoperability".[45]

2.4. Analyses

The means to perform statistical analyses is another key feature sought after by researchers.[46] Such analyses, utilizing the onomasiological structure of the thesaurus and features of the lexis it contains, facilitate investigations into a range of aspects encoded in the lexicon: cultural elaboration, semantic domains and their cultural connotations, stylistic preferences of authors, use and development of metaphors, and so on.[47] Regrettably, few editions of historical language thesauri offer functionality for even the most rudimentary of analyses.

ShT, Christian Kay observes, sorely lacks a simple count of lexical items under a category.[48] Such counts indicate the degree of lexicalization, also known as cultural elaboration, of semantic concepts.[49] The underlying hypothesis for the

[43] See Chapter 1, section 1.4.1.
[44] McCracken, 'The Exploitation of Dictionary Data and Metadata', p. 513.
[45] Ibid., p. 513.
[46] Evidenced by Appendix 2.A, which indicates such functionality has been required by researchers in the project 'Exploring Early Medieval English Eloquence'.
[47] See, for instance, *ShT*; Crystal, *Words in Time and Place*; *Mapping English Metaphor Through Time*; Porck, 'Growing Old among the Anglo-Saxons', pp. 59–71; Diller, 'Measuring the Growth of Semantic Fields'.
[48] Kay, Review of *ShT*, p. 73.
[49] Wierzbicka, *Understanding Cultures through Their Key Words*, pp. 10-11.

importance of these figures is that domains that are important in a culture are heavily encoded in the language of that community, providing its speakers with a multitude of nuances to discuss the subject. For thesaurus content in a digital environment, this statistic is relatively straightforward to obtain. Even so, many thesauri – both paper and electronic editions – do not include these statistics in their editions, including the paper editions of *TOE* and *HTE*. However, their electronic editions *TOE3*, *TOE4*, and *HTE3* currently indicate the number of lexical senses located at any given category but, regrettably, without presenting the accumulated figure for that category and its subordinate categories, i.e., for the semantic domain.[50] The larger the semantic domains analysed, the more valuable it will be for researchers to have these statistics automatically generated as opposed to calculating them manually. In *TOE*, for instance, 1,348 lexical senses evoke the concept "13.02 War", whereas the nine co-ordinate domains for peace encompass a mere 119 lexical senses.[51]

Statistics more nuanced than a count of all items located at a specific category are absent from all editions of historical language thesauri. This absence hampers researchers in utilizing the semantic information encoded by the topical system for an onomasiological analysis of lexical items with specific characteristics. Salient information on lexis – such as period, region, register, or use by a specific author – can form the basis of investigations into their spread across the semantic hierarchy, be it horizontally over the various semantic fields or vertically between levels of specificity in meaning attributed to the lexical items selected. Thus, the hierarchy of a historical language thesaurus has the potential to act as "summary of the semantic framework" and yield onomasiological profiles for those words and phrases in which researchers are interested.[52] Additionally, researchers participating in the workshop series indicated the desire to contrast analyses between different languages and other features attributed to lexical items.[53]

2.5. Data management

The fifth functionality required for research is the ability to manage the data under investigation, a demand that is largely the consequence of the need to extend thesaurus content.[54] An important component of data management,

[50] In June 2017, *HTE3* was updated to include the means to generate heatmaps and sparklines, visualizations that indicate degrees of lexicalization throughout the recorded history of the English language.

[51] The nine co-ordinate categories for peace are the category "13.01 Peace, state of law and order" and the subcategories "13|01 Peace, absence of dissension", "13|02 Unanimity, concord, agreement", "13|03 n. Good terms, rapport", "13|04 Peace, freedom from hate", "13|05 Peace among dwellings", "13|06 Accordant, not at variance", "13|07 Peacefully, peaceably", and "13|08 To agree, settle".

[52] Kay, 'Food as a Fruitful Source of Metaphor', p. 77.

[53] See Appendix 2.A.

[54] This functionality has been requested and used by researchers in the project 'Exploring Early Medieval English Eloquence' (see Appendix 2.A). For the need to extend thesaurus content, see section 2.3.

necessary to support the extension of data by researchers, concerns personal additions, such as annotations and labelling. Researchers who supplement a historical language thesaurus with such additional knowledge should retain full control over their own data, including the ability to display or hide their annotations, create backups of their additions, and share their data with others.

A second component of data management, applicable to historical language thesauri extended by other data sources, is the means to select which of these sources are deemed relevant for researchers' explorations.[55] Web-based editions of thesauri should allow for sets of information to be combined, based on the user's selection, for viewing and analyses, thus supporting researchers in their onomasiological explorations of lexis and any salient features captured. An advanced form of this functionality ought to support choosing for a certain revision of a data source, too, in cases where multiple are available. Revisions of digital historical language thesauri are certainly not shunned by their editors.

At its initial publication, the first digital edition of *TOE* contained some corrections to its printed counterpart. These included changes of usage information and additions of new words, on the basis of new knowledge, which stemmed "largely from completed sections of the Toronto Dictionary of Old English".[56] The exact alterations were left unspecified, though these will have mostly concerned lexical items starting with the letter F: between the publication of the second impression in print in 2000 and the first publication of the electronic edition in 2005, the Toronto dictionary only published its findings on Old English words starting with that letter.[57] As for any further updates to electronic editions, Kay mentioned that the taxonomy "will, of course, be subject to rolling revision" as the electronic environment allows for such changes.[58] Indeed, the number of items the online *TOE* contains has increased from the 50,706 items Kay mentions in her article for the initial release of the electronic edition to 51,483 items on 26 May 2017.[59]

When digital thesauri are subject to ongoing revisions, it may present difficulties for scholars to work with them.[60] Firstly, it could mean that the contents change whilst the scholar is doing research that involves the contents of the thesaurus. A lexical item could no longer be included under the category it used to be present in, and a category might have moved within the topical system. In effect, such changes during research entails scholars work with material that is ever in progress and for which it becomes difficult to discuss that reference body

[55] This functionality has been requested and used by researchers in the project 'Exploring Early Medieval English Eloquence' (see Appendix 2.A).

[56] Kay, '*A Thesaurus of Old English* Online', pp. 36–40.

[57] The section on the letter F was published in 2004. See the 'Publications' section of *DOE*. Accessed on October 31, 2016.

[58] Kay, '*A Thesaurus of Old English* Online'.

[59] An export of the *TOE* database as it was on 26 May 2017 has kindly been provided to the author by the University of Glasgow.

[60] See also Allan's remark on working with information from *HTE*: "It seemed preferable, and more theoretically justifiable, to work with the data as it existed at a particular stage of *HTE*, whilst acknowledging that this may be incomplete. This is especially the case given the current revision of the *OED*, which will in turn affect *HTE* data and may lead to a number of insertions and changes in later editions" (*Metaphor and Metonymy*, p. 20).

as a whole or even partially. Secondly, such revisions can make it more difficult to ensure verifiability of performed research. One of the important reasons why articles include references to the material they employed is that this practice allows others to review their work, verify whether the source material is accurate and reliable, and that the results are justifiable and could be reproduced if so desired.[61] With a reference body that is unstable, references to its contents may not provide the verifiability desired. Note that for publications in print, no such difficulties are present. Scholars can select an edition available at the moment of writing, employ that particular edition consistently, and have their references include the information on which edition should be consulted for purposes of verification.

In order to ensure that scholars can choose a particular edition or revision of a historical language thesaurus, the publication environment will need to support a form of versioning. That is to say, the thesaurus should be accessible under a system that allows users to view the taxonomy in each of its stages, enabling scholars to view and refer to the taxonomy in the state it was for a certain publication that made use thereof. Some digital reference bodies already provide such means to look into earlier states. Entries of the third edition of the *OED*, for instance, contain links to the corresponding entries from the previous edition.[62] Maintaining separate editions online, similar to how this is the case for publications in print, will also open the possibility to pinpoint which changes have been made from one revision to the next, where they are located in the topical system, and possibly even for what reasons these have been made. Such information could help scholars in determining how up-to-date the taxonomy is, and whether they agree with the adjustments made, rather than having to analyse such matters themselves. Presently, such functionality is missing from all historical language thesauri in an electronic form treated here.

2.6. Conclusion

Drawing on existing historical language thesaurus editions, academic reviews, and research employing these resources, the current chapter has addressed functionality desired by researchers. An overview of the five functionalities that have been discussed is presented in Table 2.1. Inclusion of this functionality varies in the existing editions of the historical language thesauri analysed. Whereas all editions offer navigation (through the topical system and through an index) and resource views, the ability to extend their content is found in only one thesaurus (i.e., *HTS*) and is limited to sharing user comments. The effect of this lack of options to extend these thesauri is that, unsurprisingly, the need and mechanisms for data management are in these cases absent, too. Similarly, analysing thesaurus content through automated means is, across all thesauri, either minimal or vastly underexplored when considering the additional

[61] See, for instance, the common ethics and responsibilities within science regarding the treatment of data and the sharing of research results as discussed in Gauch Jr, *Scientific Method in Brief*, pp. 226–30.

[62] See *OED Online*.

Code	Name	Description
R1	Navigation	Approaching a thesaurus should be possible in two manners: through its overarching taxonomy or through the lexis it organises.
R2	Resource views	Complete overviews of available information are to be presented on any given resource within a thesaurus that a user chooses to inspect.
R3	Extension	Thesaurus content should be extendable, allowing users to connect additional information to existing content.
R4	Analyses	Statistical analyses, utilizing the onomasiological structure of the thesaurus and features of the lexis it contains, should be made possible.
R5	Data management	Users must have full control over their own data and the ability to select which data sources are deemed relevant for their explorations, allowing sets of information to be combined for viewing and analysis.

Table 2.1.: Functionality required by researchers of historical language thesauri.

information that the semantic frameworks formed by these lexicographic works have the potential to yield. In short, the functionality identified here as required for research and education – navigation, resource views, extension, analyses, data management – should open new research avenues when incorporated into the dissemination of historical language thesauri. Sharing this set of functionality with researchers will greatly facilitate exploring the lexis of a historical language through a semantic lens.

References

Allan, K., *Metaphor and Metonymy: A Diachronic Approach*, Publications of the Philological Society 42 (Oxford, 2008).

Århammar, N., 'A Frisian Supplement to Buck's Dictionary of Indo-European Synonyms?', *Us Wurk* 21-22 (1972-1973), 241-5.

Atkins, B. T. S. and M. Rundell, *The Oxford Guide to Practical Lexicography* (Oxford, 2008).

Bremmer Jr, R. H., 'Treasure Digging in the Old English Lexicon', Review of *TOE2*, *NOWELE* 40 (2002), 109–14.

Brewer, C., 'Labelling and Metalanguage', in *The Oxford Handbook of Lexicography*, ed. P. Durkin (Oxford, 2016), pp. 488–500.

BTH = The Bilingual Thesaurus of Everyday Life in Medieval England, eds. (Glasgow, 2019). https://thesaurus.ac.uk/bth/. Accessed: 15 March 2022.

Busse, B., 'A Celebration of Words and Ideas: The Stylistic Potential of the *Historical Thesaurus of the Oxford English Dictionary*', Review of *HTE1* and *HTE2*, *Language and Literature* 21.1 (2012), 84–92.

Cavill, P., 'Names and Things in Anglo-Saxon and Early Norman England', Review of *TOE1* and *Words, Names and History: Selected Writings of Cecily Clark* ed. by Peter Jackson, Nottingham Medieval Studies 41 (1997), 186–91.

Conner, P. W., Review of *TOE1*, *Speculum* 73.3 (1998), 887–9.

Cruse, D. A., *Lexical Semantics* (Cambridge, 1986).

Crystal, D., *Words in Time and Place: Exploring Language through* The Historical Thesaurus of the Oxford English Dictionary (Oxford, 2014).

Dance, R., Review of *TOE1*, *Medium Ævum* 66.2 (1997), 312–13.

Diller, H.-J., 'Measuring the Growth of Semantic Fields: The Case of the English Emotion Lexicon', in *Cognition in Language: Volume in Honour of Professor Elżbieta Tabakowska*, eds. W. Chłopicki et al.(Kraków, 2017), pp. 574–96.

DOE = *Dictionary of Old English: A to I Online*, eds. A. Cameron et al. (Toronto, 2018). https://tapor.library.utoronto.ca/doe/.

DSSPIEL = Buck, C. D., *A Dictionary of Selected Synonyms of the Principal Indo-European Languages: A Contribution to the History of Ideas* (Chicago, 1949).

Gauch Jr, H. G., *Scientific Method in Brief* (Cambridge, 2012).

Görlach, M., Review of *TOE1*, *Anglia* 116.3 (1998), 398–401.

Hartmann, R. R. K. and G. James, *Dictionary of Lexicography* (London, 1998).

HTE = *Historical Thesaurus of English*; see *HTE1*, *HTE2*, *HTE3*.

HTE1 = *Historical Thesaurus of the Oxford English Dictionary*, 2 vols., eds. C. Kay et al. (Oxford, 2009).

HTE2 = *Historical Thesaurus of the Oxford English Dictionary*, eds. C. Kay et al. (2010). http://oed.com/thesaurus.

HTE3 = *The Historical Thesaurus of English*, eds. C. Kay et al. (Glasgow, 2014). http://historicalthesaurus.arts.gla.ac.uk.

HTS = *Historical Thesaurus of Scots*, ed. S. Rennie (Glasgow, 2015). http://scotsthesaurus.org.

Ilson, R. F., 'On the Historical Thesaurus of the Oxford English Dictionary', Review of *HTE1*, *International Journal of Lexicography* 24.2 (2011), 241–60.

Kay, C., Review of *ShT*, *International Journal of Lexicography* 9.1 (1996), 71–6.

Kay, C., '*A Thesaurus of Old English* Online', *Old English Newsletter* 38.3 (2005), 36–40.

Kay, C., 'Food as a Fruitful Source of Metaphor', in *Mapping English Metaphor Through Time*, eds. W. Anderson et al. (Oxford, 2016), pp. 66–78.

Kay, C. and M. Alexander, 'Diachronic and Synchronic Thesauruses', in *The Oxford Handbook of Lexicography*, ed. P. Durkin (Oxford, 2016), pp. 367–80.

Learning with the Online Thesaurus of Old English, eds. C. Hough and C. Kay (Glasgow, 2007). http://oldenglishteaching.arts.gla.ac.uk. Accessed: 31 October 2016.

LSM = Coleman, J., *Love, Sex and Marriage: A Historical Thesaurus*, Costerus New Series 118 (Amsterdam, 1999).

Mapping English Metaphor Through Time, eds. W. Anderson et al. (Oxford, 2016).

McCracken, J., 'The Exploitation of Dictionary Data and Metadata', in *The Oxford Handbook of Lexicography*, ed. P. Durkin (Oxford, 2016), pp. 501–14.

Momma, H., Review of *TOE2*, *Notes and Queries* 50.1 (2003), 79–80.

Norri, J., 'Regional Labels in Some British and American Dictionaries', *International Journal of Lexicography* 9.1 (1996), 1–29.

Mapping Metaphor with the Historical Thesaurus (Glasgow, 2015). http://mappingmetaphor.arts.gla.ac.uk. Accessed: 31 October 2016.

OED = *Oxford English Dictionary*; see *OED1*, *OED2*, *OED3*, *OED Online*.

OED1 = *The Oxford English Dictionary*, eds. J. A. H. Murray et al. (Oxford, 1884–1933).

OED2 = *The Oxford English Dictionary*, eds. J. A. Simpson and E. S. C. Weiner, 2nd edn (Oxford, 1989).

OED3 = *The Oxford English Dictionary*, ed. J. A. Simpson. http://oed.com.

OED Online = *Oxford English Dictionary Online*. http://oed.com. Accessed October 18, 2016.

Porck, M. H., 'Growing Old among the Anglo-Saxons: The Cultural Conceptualisation of Old Age in Early Medieval England' (Ph.D dissertation, Leiden University, 2016). https://hdl.handle.net/1887/39136.

The Scots Thesaurus, eds. I. Macleod et al. (Aberdeen, 1990).

ScT = *The Scots Thesaurus*, eds. I. Macleod et al. (Aberdeen, 1990).

ShT = Spevack, M., *A Shakespeare Thesaurus* (Hildesheim, 1993).

Stolk, S., 'Welcoming the *Thesaurus of Old English Statistics*: *The Thesaurus of Old English* and the Vocabulary of Greetings' (M.Phil. dissertation, Leiden University, 2013). https://hdl.handle.net/1887/31760.

Svensén, B., *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making* (Cambridge, 2009).

TOE = *A Thesaurus of Old English*; see *TOE1*, *TOE2*, *TOE3*, *TOE4*.

TOE1 = *A Thesaurus of Old English*, eds. J. Roberts et al., 2 vols., King's College London Medieval Studies XI (London, 1995).

TOE2 = *A Thesaurus of Old English*, eds. J. Roberts et al., 2 vols., Costerus New Series 131, 2nd edn (Amsterdam, 2000).

TOE3 = *A Thesaurus of Old English*, eds. J. Roberts et al. (Glasgow, 2005). http://libra.englang.arts.gla.ac.uk. Accessed February 1, 2013.

TOE4 = *A Thesaurus of Old English*, eds. J. Roberts et al. (Glasgow, 2015). http://oldenglishthesaurus.arts.gla.ac.uk. Accessed March 15, 2022.

Wierzbicka, A., *Understanding Cultures through Their Key Words: English, Russian, Polish, German, and Japanese*, Oxford Studies in Anthropological Linguistics 8 (New York, 1997).

Appendix 2.A: Functionality elicited in 'Exploring Early Medieval English Eloquence'

This appendix offers an overview of functionality required of historical language thesauri for research. The functionalities listed here have been elicited in the project 'Exploring Early Medieval English Eloquence', which utilizes *A Thesaurus of Old English*, through dedicated stakeholder meetings, workshops, and feedback based on preliminary results in research and education. This project and its case studies are discussed in Chapter 8. Functionality marked as novel in the table below is, at the time of writing, unavailable in any of the thesaurus editions analysed in Chapter 1 and Chapter 2. For each functionality request, a footnote describes whether and when it was implemented in the web application Evoke (described in Chapter 7) and in which of the case studies (discussed in Chapter 8) this functionality was used.

Functionality	Novel
Navigation (R1)	
Navigation via the thesaurus structure: The user must be able to navigate the topical structure of the thesaurus and open a given category in a resource view.[63]	-
Navigation via the thesaurus index: The user must be able to search for specific elements (i.e., lexical items, categories, labels) and open a given one in a resource view.[64]	-
Show subthesaurus: Show only those categories and items of interest, based on a selection of features made by the user. (This requirement holds for resource views as well as navigation.)[65]	+
Advanced search: Restrict searches to items based on a selection of features made by the user.[66]	-
Resource views (R2)	
For any resource, offer basic information: Offer rudimentary information on a resource, such as its name, identifier (i.e., metadata), and position within the topical structure.[67]	-

[63] This functionality has been implemented in Evoke (since v1.0.0) and used in the following case studies described in Chapter 8: Dekker (education); Depuydt and De Does; Fletcher; Khan et al.; Porck; Porck and Stolk (education); Van Baalen; Van de Poel and Stolk.

[64] Implemented in Evoke (since v1.0.0) and used in: Dekker (education); Depuydt and De Does; Khan et al.; Porck; Porck and Stolk (education); Van Baalen; Van de Poel and Stolk.

[65] Not yet implemented in Evoke (v1.4.1). This functionality has been requested by four researchers participating in the EEMEE workshops.

[66] Not yet implemented in Evoke (v1.4.1). This functionality has been requested by one researcher participating in the EEMEE workshops.

[67] Implemented in Evoke (since v1.0.0) and used in: Dekker (education); Depuydt and De Does; Fletcher; Khan et al.; Porck; Porck and Stolk (education); Van Baalen; Van de Poel and Stolk.

Functionality	Novel
For a category, show lexical items allocated to it: Show, for a category, which lexical senses are positioned at this thesaurus category.[68]	-
For a lexical item, show senses related through polysemy: Show, for a lexical sense and for its lexeme, all senses attributed to the lexeme in question.[69]	+
For a lexical sense, show synonyms: Show, for a lexical sense, which synonyms are available in the language. If multiple languages are available, show lexical senses that lexicalize the same concept (i.e., are translations).[70]	-
For a lexical item, show its labels: Show labels in the resource view, both original and ones newly introduced by the user, assigned to lexical items. (This requirement applies to the index as well as to resource views.)[71]	-
For a lexical item, show its language: Show the language in the resource view, besides part of speech, attributed to lexical items. (This requirement applies to the index as well as to resource views.)[72]	-
For a label, show resources marked by it: Show, for a label, which lexical items or other resources were marked with this label.[73]	+
For a category, show associations: Show, for a category, which categories are evoked by other senses of the lexical items found at this thesaurus category (i.e., senses related through polysemy).[74]	+
Extension (R3)	
Annotate and label content: Users should be able to annotate lexical items (and other elements) with custom labels.[75]	+
Link bodies of knowledge: Users should be able to link data from another work to data contained in a historical language thesaurus.[76]	+

[68] Implemented in Evoke (since v1.0.0) and used in: Dekker (education); Depuydt and De Does; Fletcher; Khan et al.; Porck; Porck and Stolk (education); Van Baalen; Van de Poel and Stolk.

[69] Implemented in Evoke (since v1.1.0) and used in: Dekker (education); Depuydt and De Does; Khan et al.; Porck; Porck and Stolk (education); Van Baalen; Van de Poel and Stolk.

[70] Implemented in Evoke (since v1.0.0) and used in: Dekker (education); Porck; Porck and Stolk (education); Van Baalen; Van de Poel and Stolk.

[71] Implemented in Evoke (since v1.2.0) and used in: Fletcher; Khan et al.; Porck; Van Baalen.

[72] Implemented in Evoke (since v1.3.0) and used in: Depuydt and De Does; Van de Poel and Stolk.

[73] Implemented in Evoke (since v1.3.0) and used in: Dekker (education); Fletcher; Khan et al.

[74] Implemented in Evoke (since v1.4.0) and used in: Dekker (education).

[75] Implemented in Evoke (since v1.3.0) and been used in: Dekker (education); Fletcher; Khan et al.; Porck; Van Baalen.

[76] Implemented in Evoke (since v1.0.0) and used in: Depuydt and De Does; Porck; Van de Poel and Stolk.

Functionality	Novel
Match word list: Match an existing list of lexical items, from one knowledge body, against ones found in the thesaurus.[77]	+
Analyses (R4)	
Provide basic analyses: Users should have access to basic onomasiological analyses on the composition of semantic fields (e.g., on the basis of part of speech or number of items found over the various subordinate categories) when viewing a category.[78]	-
Provide advanced analysis: Users should be able to perform advanced onomasiological analyses for the items of their interest, based on a selection of features made by the users themselves.[79]	+
Contrast advanced analyses: Users must be able to contrast onomasiological analyses of one set of features versus another.[80]	+
Top statistics: Provide a top 10 on categories that contain the most lexical items, possibly restricted based on a selection of features made by the user.[81]	+
Data management (R5)	
Manage own data: Users must have full control over their own data: the ability to hide or view their additions, create backups, and restore backups.[82]	+
Select data sources: Users must be able to select which data sources are deemed relevant for their explorations, allowing sets of information to be combined for viewing and analysis.[83]	+

Table 2.A.1.: Initial set of functionality elicited for research.

[77] This functionality is perhaps best facilitated through tooling specifically designed for aligning two bodies of knowledge as opposed to through a thesaurus edition and its functionality. Chapter 7 describes three such tools that have been used to link, or align, different lexicographic works: a custom alignment tool by Porck, Excel spreadsheets by Van de Poel and Stolk, and a Lex'it-based linking tool by Depuydt and De Does.

[78] Implemented in Evoke (since v1.1.0) and used in: Dekker (education); Porck and Stolk (education).

[79] Implemented in Evoke (since v1.3.0) and used in: Porck; Van Baalen; Van de Poel and Stolk.

[80] Implemented in Evoke (since v1.4.0) and used in: Porck; Van Baalen; Van de Poel and Stolk.

[81] Not yet implemented in Evoke (v1.4.1). This functionality has been requested by one researcher participating in the EEMEE workshops.

[82] Implemented in Evoke (since v1.3.0) and used in: Dekker (education); Depuydt and De Does; Fletcher; Khan et al.; Porck; Van Baalen; Van de Poel and Stolk.

[83] Implemented in Evoke (since v1.4.0) and used in: Dekker (education); Depuydt and De Does; Khan et al.; Porck; Van Baalen; Van de Poel and Stolk.