

Evoke: A Web application for thesauri and Linguistic Linked Data

Sander Stolk

Introduction

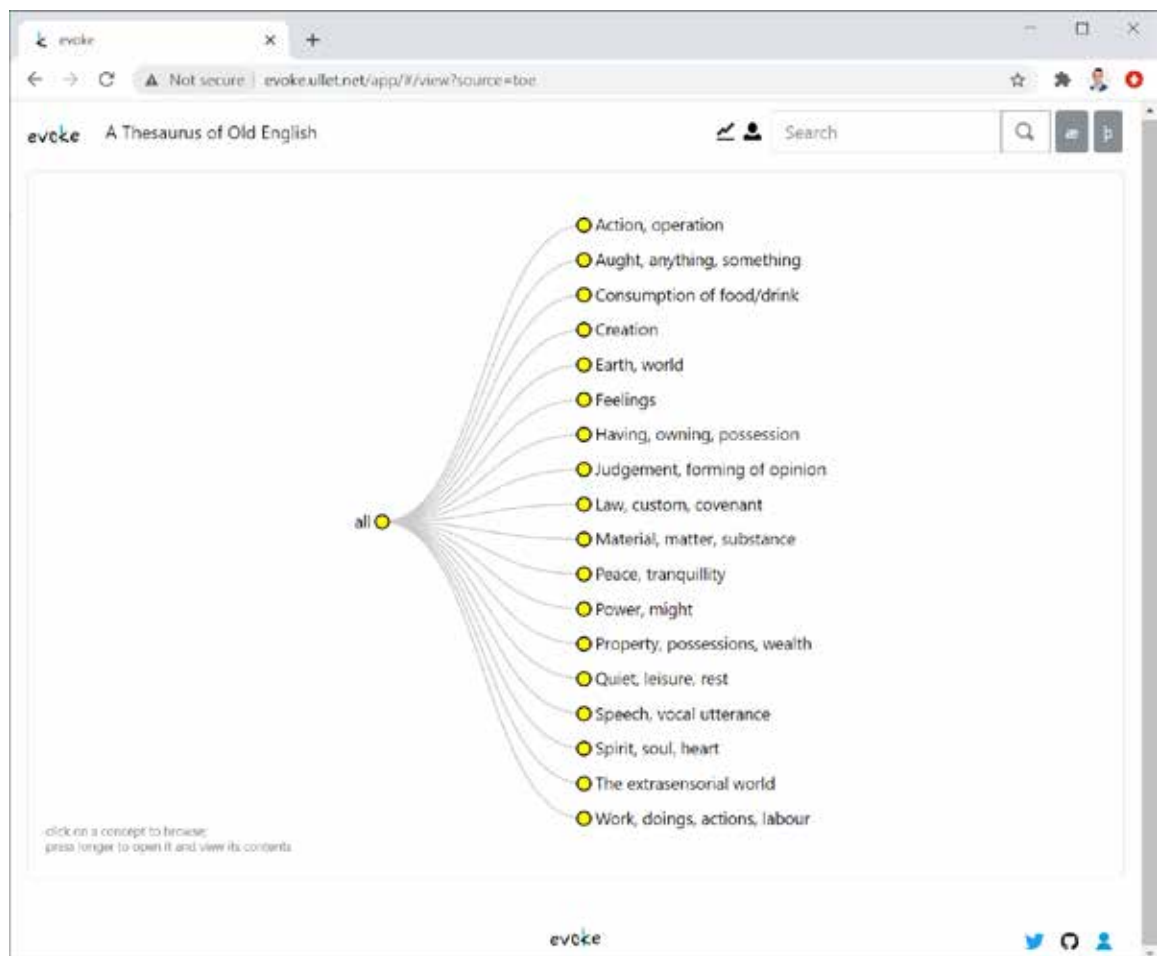
The manner in which information is shared has changed rapidly with the advent of the Web. New forms for sharing lexical data have emerged, too, owing to the transition from ink to Internet. A notable one is Linguistic Linked Data (LLD), which offers several benefits over other formats by taking advantage of decentralized Internet technology and open Web standards. The benefits in using this form – and in applying related Web technologies – are demonstrated in the new Web application Evoke. Evoke allows exploring and interacting with linguistic resources, employing a novel method that enables users to view a thesaurus, extend it – without the risk of infringing on licenses or requiring additional hosting costs with publishers – and perform new and exciting analyses over a combination of datasets. Moreover, the application showcases how a number of mechanisms, built on top of LLD, can be employed to reduce barriers for users to start working with valuable linguistic resources. The development of Evoke has been supported by experts in lexicography, linguistics and philology (amongst other fields), to ensure the software is intuitive and useful for both research and educational purposes. The following sections will discuss (1) the Web application Evoke; (2) *A Thesaurus of Old English*, the first LLD resource made available in Evoke; and (3) current uses in research and education of these two resources.

(1) Evoke

The Web application Evoke allows users to interact with thesauri. These lexicographic works organize words according to meaning – by means of a semantic hierarchy – rather than alphabetically, thus enabling the user to move from meaning to words that express that meaning. Users of Evoke can not only browse and view content captured within these linguistic resources (cf. Figures 1 and 2), but also expand on them and, owing to the organization principle of such works, perform analyses through the semantic lens of their overarching hierarchies. The resulting onomasiological profiles that are generated by Evoke are based on features of interest that the user selects: specific labels, languages, and/or parts of speech. A generated profile contains statistics and charts on the item count of a selection, its degree of ambiguity (indicating polysemy), its degree of synonymy,



Sander Stolk is a PhD researcher in Digital Humanities at Leiden University. He holds an MSc in Computer Science from VU Amsterdam and an MA in English Language and Culture from Leiden University. His research investigates how the dissemination and reuse of thesauri can be improved by utilizing Linked Data technology, having the analysis of information needs of scholars and development of the Web application Evoke at its core, and he is a member of the W3C OntoLex community that works on standards regarding Linguistic Linked Data. He is also Head of Innovation at Semmtech, a company specializing in using Linked Data technology. s.s.stolk@hum.leidenuniv.nl



its distribution over categories in the hierarchy (cf. Figure 3), and a distribution over the depth of the semantic hierarchy (indicating how specific in meaning the selected words are, cf. Figure 4).

Through the functionality of Evoke, users can explore the lexis in a thesaurus, and optionally also in other datasets, with ease and powered by the decentralized mechanisms offered by LLD methodologies (cf. Cimiano et al. 2020). A notable feature of LLD is the use of so-called URIs as data identifiers (that is, using Web addresses to identify words, word senses, labels, and concepts in the semantic hierarchy), which allow one to reference data elements and link additional content. This approach can be likened to how Web pages employ links. Instead of linking Webpages, however, Evoke and its use of LLD enable users to link data. Although use of the Evoke interface is not mandatory for creating such links, users can use it to add information easily by annotating content.

Annotating a specific word, for instance, can be achieved in Evoke by typing a sentence with a label (indicated by a hashtag). Doing so will automatically create a piece of Linked Data annotation that adheres to the Web Annotation standard of W3C (see #riddle47 in Figure

Figure 1. Navigating the semantic hierarchy of *A Thesaurus of Old English* in Evoke

Evoke is available at <http://evoke.ullet.net>. A demonstration of its functionalities can be found at <http://evoke.ullet.net/demo>.

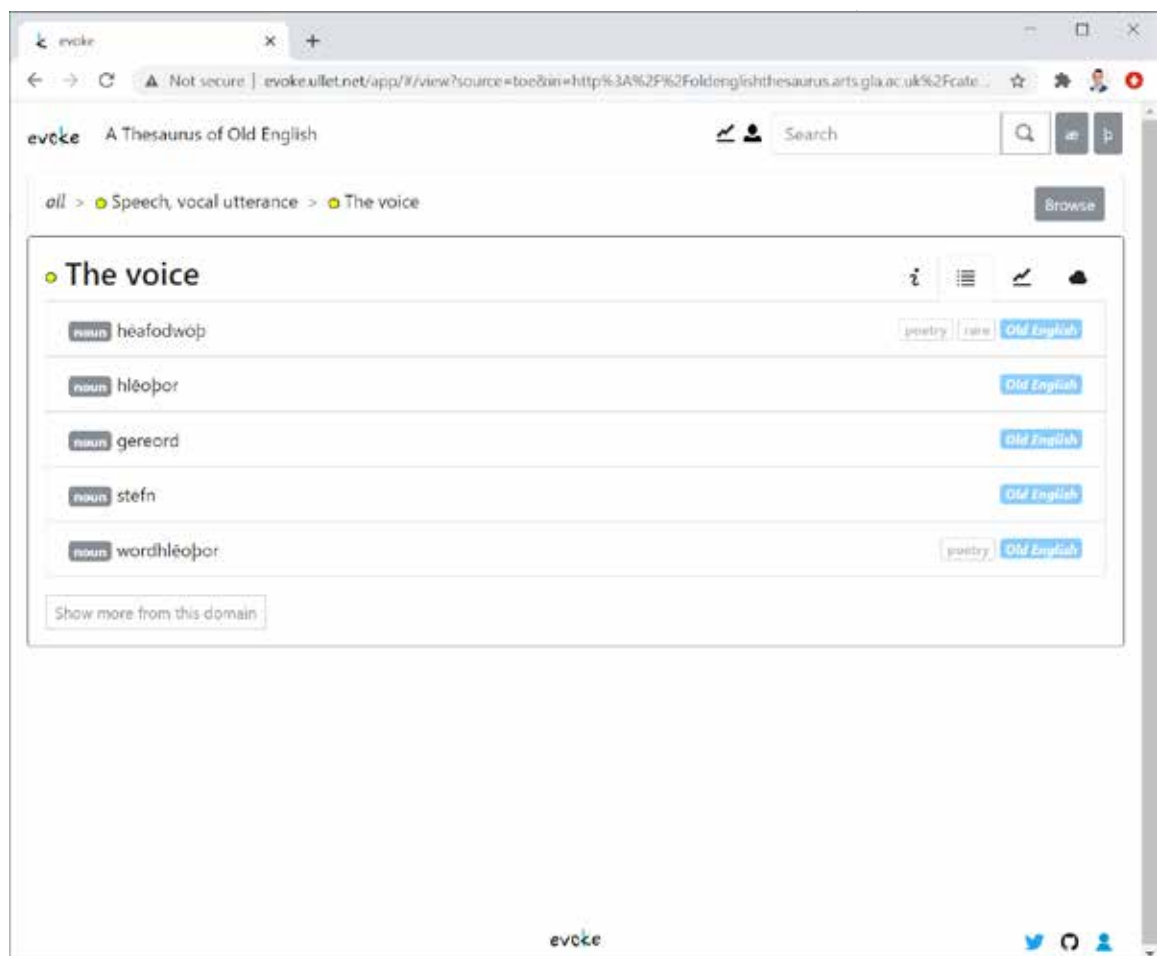
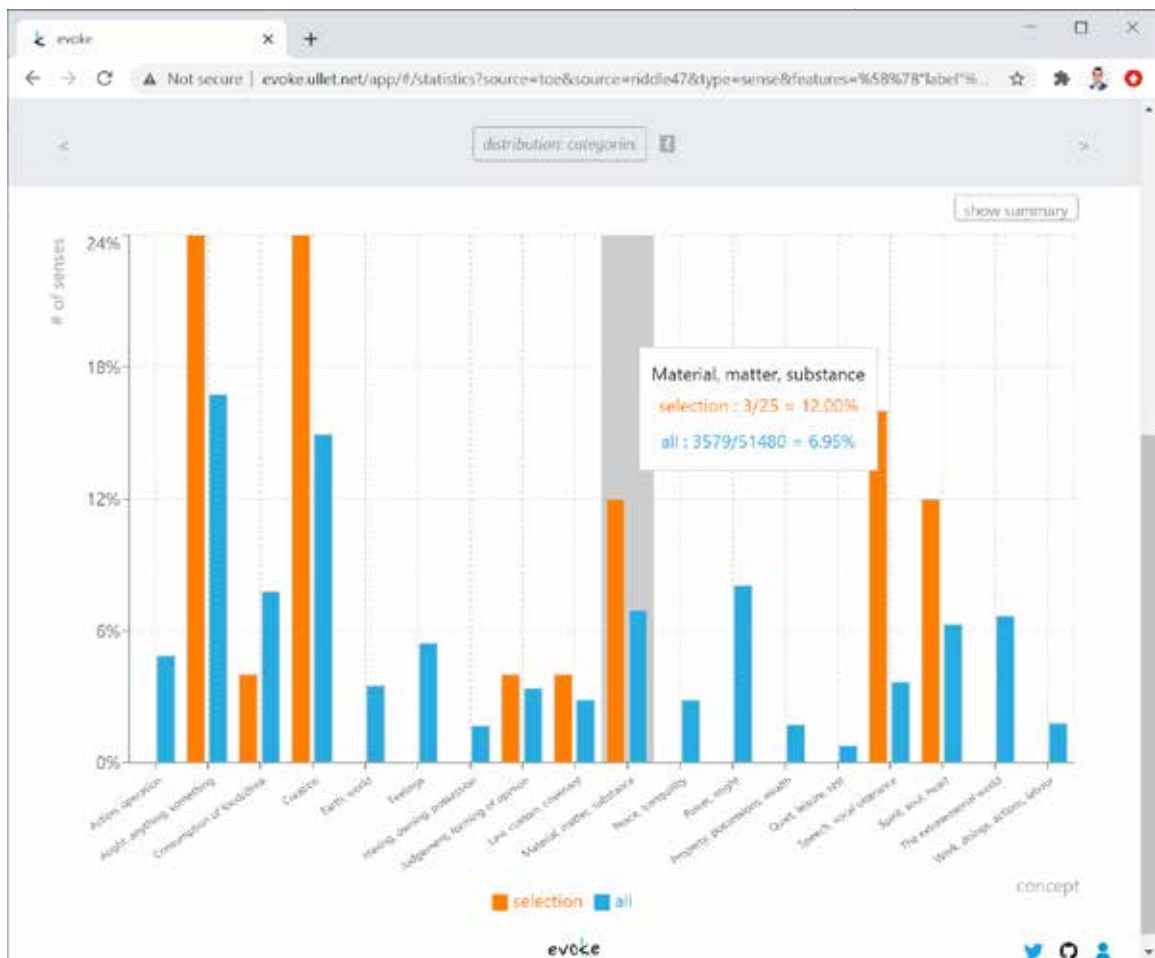


Figure 2. A list of Old English words that denote “the voice”

5; cf. Sanderson et al. 2017). The novelty here is that annotations created in Evoke are not stored in a database but are instead stored locally in the user’s browser. The annotations can be downloaded as a file for backup, and backups can be reactivated in the browser. The publisher of the original content therefore does not need to host these annotations. Moreover, the annotations themselves only contain references to the identifiers, or URIs, of the original content without including that content in the annotation itself. As alluded to before, this methodology can entice users to explore a lexicographic work and interact with it, formulate a plan of research, and only at a later stage take the hurdle in activating an account and getting support for further research – be it in the form of a more open license or in getting access to advanced services from the publisher of the resource.

(2) A Thesaurus of Old English: An LLD resource

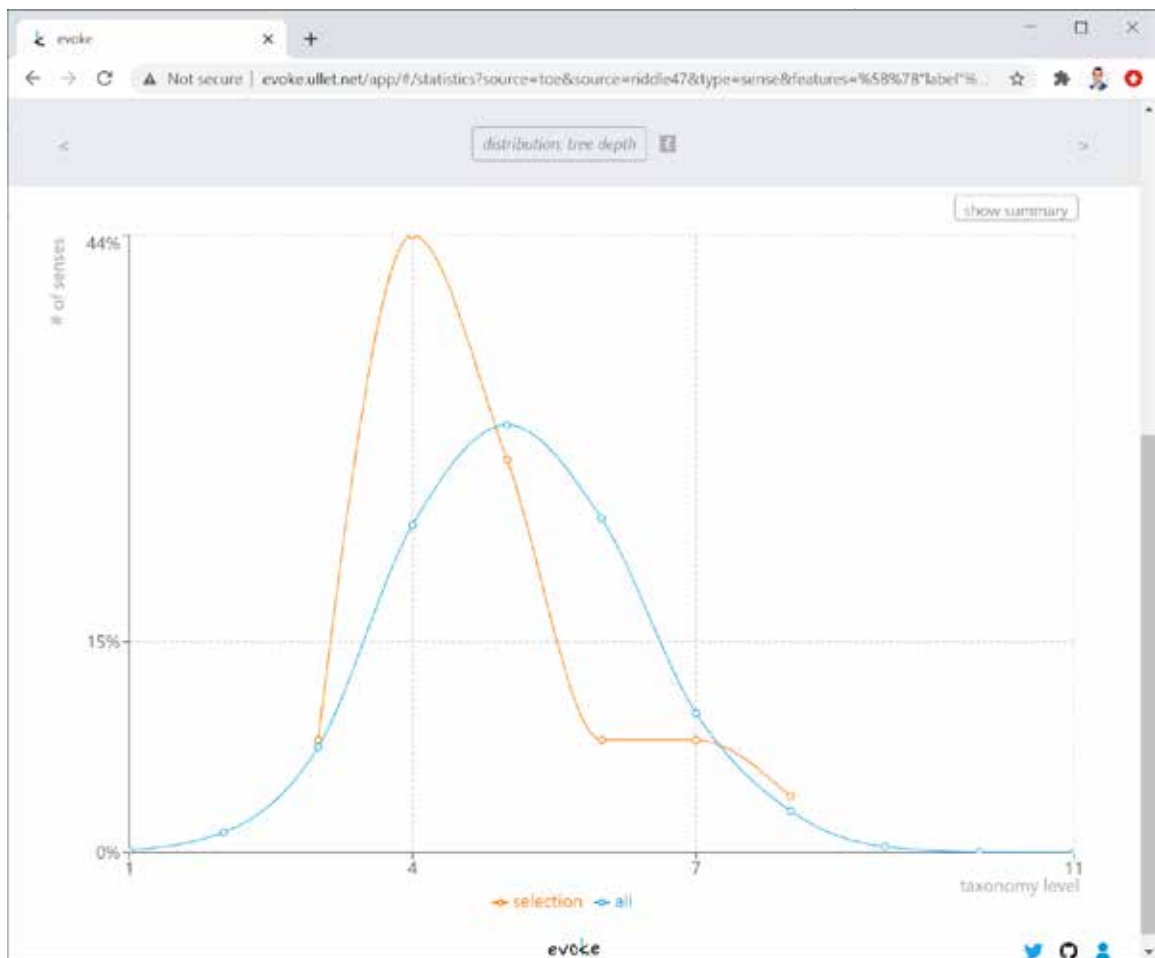
The first lexicographic resource made available in Evoke as LLD is *A Thesaurus of Old English* (Stolk 2019). This thesaurus captures lexis of Old English, the early medieval variant of English spoken between roughly 500 and 1100 AD. Upon its first publication in 1995, this resource has been met with high praise, having been called “the most



important contribution to Old English studies for years”, since the “comprehensive analysis” that it forms allows scholars to “investigate what distinctions Anglo-Saxons felt important enough to make in the lexicon” (Görlach 1998). The value of lexicographic works such as this is evident, then, for the audience which they target. As pointed out, dictionaries and thesauri allow one to explore not just language but also culture. They offer a lens through which users can learn many details on words in a given vocabulary, including (but not limited to) nuances in meaning, part of speech, and restrictions in regional or temporal use.

Many scholars who have used or reviewed *A Thesaurus of Old English* have indicated their wish to include additional information, such as indications of region, etymology, metaphorical use, and so on (e.g. Dance 1997, Bremmer 2002, Anderson et al. 2016). Given the fact that lexicographers cannot cater to every research need, this calls for extendibility of the dataset. That is to say, researchers (and other users) would benefit from the means to add data, form new queries over the combined information, and visualize the connections to acquire new insights. Indeed, the team behind the lexicographic

Figure 3. Chart in Evolve on the distribution of Old English word senses over categories in the semantic hierarchy; a selection of senses found in the text Exeter Riddle 47 (orange) are compared to all known ones (blue).



resource itself could benefit from such engagement with their work, too, being able to accumulate feedback, incorporate more data, and offer services based on their innate knowledge of the structure of the lexicon that is at the heart of the research being done. Such extendibility constitutes the precise functionality of Evoke that was demonstrated to be of great value in both research and educational use, as will be discussed in the next section.

(3) Exploring Early Medieval English Eloquence

To assess the usefulness of Evoke both in research and for education, a research project was formulated with the title Exploring Early Medieval English Eloquence. Based on the case study of a single thesaurus, this project has brought together 17 scholars from universities and lexicographic institutions from across Europe to explore the contents of *A Thesaurus of Old English* using Evoke. The various aspects of research approach the information from the perspective of a specific discipline, namely that of linguistics, literary-criticism, history, lexicography, and philology. In their explorations, the researchers (and in educational settings, their students) set about viewing other material next to that of *A Thesaurus*

Figure 4. Chart in Evoke on the distribution of Old English word senses over the depth of the semantic hierarchy; a selection of senses found in the text Exeter Riddle 47 (orange) are compared to all known ones (blue).

The screenshot shows the Evoke web interface. At the top, there is a search bar with the text 'riddle47'. Below it, a breadcrumb trail reads: 'all > Law, custom, covenant > Lawlessness > An offence > Wrongful taking, theft > A robber, thief'. The main entry is for the word 'þeof', which is a noun. It is described as 'in sense evoking: A robber, thief'. The entry includes a URL: 'http://oldenglishthesaurus.arts.gla.ac.uk/sense/#id=44698'. Under 'Type', it is listed as 'Lexical sense'. Under 'Name', it is 'þeof'. Under 'Synonym', there are four entries: 'hlōpere', 'mōrsceaþa', 'sceācere', and 'þeofsceaþa', each with a 'glosses' button and an 'Old English' tag. An 'Annotation' section at the bottom shows '#riddle47' with a 'riddle47' tag. A modal window on the right shows 'belongs to entry: noun þeof' and 'Add your own #annotation...' with an 'Add' button.

of Old English, some by linking an already existing source and others by using the annotation system in Evoke.

Novel research done within this project include analyses of lexis used in specific texts (e.g. *Beowulf*) or used by specific authors (e.g. Ælfric). Their analyses offer us new insights into the semantic choices made by authors or present in genres. Such onomasiological profiles may act as semantic fingerprints in identifying authorship or in positioning other work. Other research focuses on metaphors associated with anger and their development through the history of the language. Lastly, various researchers work on linking Old Frisian and Old Dutch lexis with the semantic hierarchy of the thesaurus. This will allow us to contrast how many nuances these language communities respectively had, next to those of Old English, in expressing certain concepts such as kinship or greetings. All these researchers have linked up additional information to the original thesaurus content. Results have been presented at the 21st International Conference of English Historical Linguistics (ICEHL21, Leiden University) and will

Figure 5. Information on one of the senses of the Old English word *þeof* [thief] in Evoke, including an annotation from the dataset “riddle47” that marks this specific sense.

be published in a special issue of the journal *Amsterdamer Beiträge zur älteren Germanistik*, entitled *Exploring Early Medieval English Eloquence*, to be published in open access in fall 2021.

As regards the usefulness of thesauri in education, in the past two years students at Leiden University and the University of Groningen have used Evoke and *A Thesaurus of Old English* to explore aspects of Old English language and culture. Dr. Kees Dekker presented his experience on the use of Evoke as part of an introductory course on Old English at the University of Groningen; students at Leiden University participate in a 2-hour workshop that familiarizes them with digital tools and resources for studying Old English (*A Thesaurus of Old English*, Evoke, and the *Dictionary of Old English* and its *Corpus*). Workshop materials, created by Dr. Thijs Porck and Sander Stolk, are available on request. The learning exercises created at both universities will be incorporated in the Evoke website. Courtesy of Prof. Carole Hough (University of Glasgow), these will include units from *Learning with the online Thesaurus of Old English*. The resulting online material is due to be made public in fall this year.

References

- Anderson, W., E. Bramwell and C. Hough (eds.). 2016.** *Mapping English Metaphor through Time*. Oxford: Oxford University Press.
- A Thesaurus of Old English*.** <http://oldenglishthesaurus.arts.gla.ac.uk>.
- Bremmer Jr, R.H. 2002.** Treasure Digging in the Old English Lexicon. *NOWELE* 40, 109–14.
- Cimiano, P., Chiarcos, C, McCrae, J.P. and J. Gracia. 2020.** *Linguistic Linked Data: Representation, Generation and Applications*. Heidelberg: Springer.
<https://www.springer.com/gp/book/9783030302245>.
- Dance, R. 1997.** Review of *A Thesaurus of Old English*. *Medium Ævum* 66:2, 312–13.
- Dictionary of Old English*.** <https://www.doe.utoronto.ca/>.
- Görlach, M. 1998.** Review of *A Thesaurus of Old English*. *Anglia* 116:3, 398-401.
- Hough, C. and C. Kay. (eds.). 2017.** *Learning with the online Thesaurus of Old English*. <https://oldenglishteaching.arts.gla.ac.uk/>



/Instituut voor
de Nederlandse
taal/



Sanderson, R., Ciccarese, P. and B. Young (eds.). 2017. Web Annotation Data Model: W3C Recommendation 23 February 2017. <https://www.w3.org/TR/annotation-model/>.

Stolk, S. 2019. *A Thesaurus of Old English* as Linguistic Linked Data: Using OntoLex, SKOS and lemon-tree to Bring Topical Thesauri to the Semantic Web. In *Proceedings of the eLex 2019 conference, 2019*. 223–247.
https://elex.link/elex2019/wp-content/uploads/2019/09/eLex_2019_13.pdf.

The logo for 'evoke' features the word in a lowercase, sans-serif font. The letter 'v' is stylized with a vertical blue line passing through its center, and the letter 'e' is also stylized with a vertical blue line passing through its center.