

Evoked: exploring and extending lexicographic resources using a linked data approach

Sander Stolk

Centre for the Arts in Society, Leiden University, The Netherlands

E-mail: s.s.stolk@hum.leidenuniv.nl

Abstract. Lexicographic resources such as thesauri contain a wealth of information for research, but their published forms not uncommonly limit the ways in which users can interact with them. The web application Evoked offers users functionality for viewing, navigating, extending, and analysing content of topical thesauri. Its use of linked data mechanisms and a novel architecture (relying on the use of data catalogues, internet browser storage, annotation of URIs) addresses one of the more intractable problems in modern lexicography: allowing users to engage more fully with published lexicographic content without them infringing on licenses or requiring additional hosting. Users of Evoked can engage with lexicographic content by annotating, adding tags, and building custom queries without the need to have full, unlimited access to the entire dataset of the lexicographic resource.

Evoked is one of the first applications that provides a user-friendly interface for working with Linguistic Linked Data resources, opening these resources up to users who may not have advanced knowledge of linked data and RDF technologies. Students as well as established researchers have confirmed the user interface and underlying architecture of Evoked to be of value in answering novel research questions relying on the data of these lexical treasure troves – complemented by their own.

Keywords: Evoked, web application, software, lexicography, thesaurus, Linguistic Linked Data, linked data, data catalogue, browser storage, license

1. Introduction

Lexicographic resources such as dictionaries and thesauri contain a wealth of information for research. As Manfred Görlach points out, such resources allow one to explore not just language but also the culture of a language community.[1] 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.

Reviews and research show that, however well-received a lexicographic work may be, it is common to nevertheless find its published form subject to the following two issues. Firstly, users are unable to query and reuse the information contained within in a way other than its editors (and publishers) had foreseen.[2–5] The physical, rigid structure of paper editions forestalls rearranging or analysing their content, of course, but even those resources that have been made available in an electronic environment are

often limited to a basic set of predefined queries (e.g., browsing, searching). Secondly, researchers desire more information per lexical item than that already present. Examples of such information is whether a word or meaning is found in a particular text, or context, or is notable in some qualitative or quantitative way.[1, 3, 5–7]

It is evident that lexicographers and publishers cannot be expected to add information and functionality to fulfil all requests users may have for their individual contexts and needs. Scholars would benefit, then, from being able to add and view their own data alongside the content from a lexicographic work – the digital equivalent of scribbling in the margins of a paper edition copy. Licensing concerns of publishers, however, often restrict the ways in which users can interact with lexicographic works available online. Full access to the entire underlying database is not always made available. Moreover, there is a cost attached to hosting and moderating user

1 annotations. The two issues mentioned, then, are not
2 easily resolved. Indeed, they are considered to be
3 some of the “more intractable problems in modern
4 lexicography”.¹

5 This article describes the web application Evoke,
6 which adopts a novel approach to address the
7 aforementioned concerns for published lexicographic
8 works.² The application offers functionality to view,
9 navigate, extend, and analyse topical thesauri that
10 have been made available as Linguistic Linked Data.
11 By employing linked data mechanisms, Evoke
12 navigates the interests of both users and publishers of
13 lexicographic resources. This approach aims to reduce
14 barriers for users to start using valuable lexicographic
15 resources and extending them – without infringing on
16 more restrictive licenses or demanding additional data
17 storage from the publisher of the resource.

18 The structure of this article is as follows. Section 2
19 discusses related work. Section 3 describes the web
20 application Evoke: its architecture, use of Linguistic
21 Linked Data, data catalogues, and internet browser
22 storage. Before the conclusion, section 4 details
23 ongoing research that demonstrates the impact and
24 usefulness of Evoke and its design.

27 2. Related work

28
29 Linked data plays “an increasing role in
30 eLexicography”, adopting the Linguistic Linked Data
31 paradigm and relying on the OntoLex-Lemon
32 model.[9] Indeed, several recent initiatives aim at
33 building and maintaining Linguistic Linked Data
34 resources, including the H2020 projects ELEXIS
35 (2018-22), Prêt-à-LLOD (2019-22) and the COST
36 Action Nexus Linguarum (2019-23). Tooling in these
37 initiatives that work with Linguistic Linked Data
38 focus on creation, discovery, transformation, and
39 linking.[9] Cases in point are the tools LingHub[10],
40 which offers discovery of language resources by
41 searching through their metadata; and NAISC³, used
42 for aligning two RDF datasets. Unfortunately, most
43 applications currently available for working with
44 Linguistic Linked Data “come with a considerable
45 entry barrier and they address the advanced user of
46

47
48 ¹Marc Alexander, Director of *The Historical Thesaurus of*
49 *English*[8], personal communication, September 28, 2020.

50 ²Evoke: <http://evoke.ullet.net>. Source code: <https://github.com/ssstolk/evoke>. Demonstration: <http://evoke.ullet.net/demo>.

51 ³<https://github.com/insight-centre/naisc>

1 RDF technologies rather than a typical linguist”. [11]
2 Evoke is one of the first applications, then, that aims
3 to provide a user-friendly interface for such resources
4 and to open them up to a wider audience.

5 Ways to extend resources on the Web by using
6 linked data mechanisms (not specific to Linguistic
7 Linked Data) are pivotal in notable recent work such
8 as the tool hypothes.is⁴, used specifically for
9 annotating webpages, and the ecosystem SOLID⁵,
10 which relies on personal RDF data hubs. Both works
11 use online databases to store information of users,
12 requiring them to login to their account before they
13 can add data. In contrast, Evoke demands no login as
14 user data is stored locally, i.e., in memory of the
15 internet browser, instead. The alternative Evoke offers
16 grants users complete control over their own data and
17 annotations (backup, share, publish), does not demand
18 for that data to be stored online, and requires no
19 account details before interacting with a resource and
20 extending it. This approach both avoids public
21 comments cluttering webpages of annotated resources
22 and encourages users to engage in open science.

23 3. Evoke

24
25
26
27 The web application Evoke allows users to interact
28 with topical thesauri expressed as Linguistic Linked
29 Data. A topical thesaurus is “a work of
30 lexicographical reference which presents lexical facts
31 with semantic domains as its core organizational
32 principle, rather than in alphabetical
33 arrangement”. [12] Users of Evoke, then, can explore
34 these hierarchies of semantic domains (or concepts)
35 and view lexical facts of words positioned in such a
36 hierarchy. Development of the features necessary for
37 exploring such content has been informed by
38 feedback from both researchers and students
39 (gathered through workshops, courses, and
40 questionnaires) since the release of the first prototype
41 of Evoke in 2018. The aim of the application is to
42 allow scholars to explore, extend, and analyse
43 linguistic data captured in a thesaurus – optionally in
44 combination with other datasets – with ease and
45 powered by the decentralized mechanisms that linked
46 data offers.

47 The application allows users to select from a
48 catalogue which datasets they wish to activate and

49
50 ⁴<http://hypothes.is/>

51 ⁵<https://solidproject.org/>

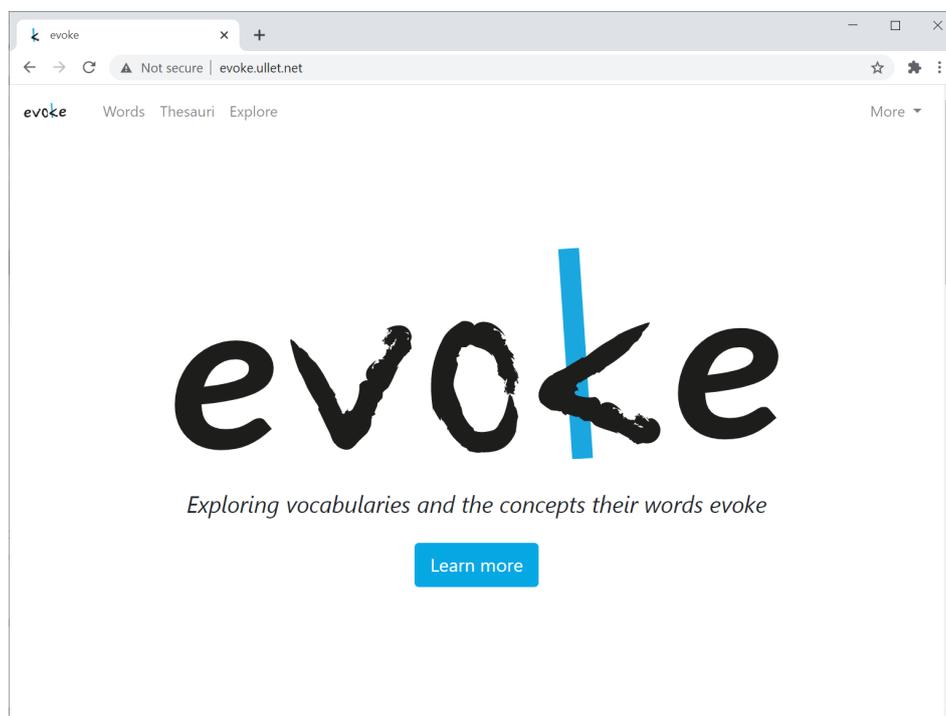


Fig. 1. Homepage of Evoke.

explore (which will be discussed in-depth in section 3.3). Users can view the content of activated datasets in the *browse* section of Evoke, which allows them to follow any data links within or between the datasets. Thesaurus concepts from a semantic hierarchy are shown by means of a tree (see Figure 2). Clicking on a branch allows the user to dive deeper into the hierarchy. By holding down the mouse instead, or clicking on the word 'open' next to a branch, users can open up and view the details of the semantic concept selected (see Figure 3 for 'The Voice'). The default tab lists the various words that express the meaning of the concept in question. Other tabs available provide information on the identification of the concept (e.g., URI and name), statistics on the distribution of the words within this taxonomy branch, and word clouds.

Lexical facts of specific words are available, too, and can be accessed by clicking on a word listed. The default tab then provides the identification of both the word in this specific sense and the word in general, the part of speech, synonyms, tags, and annotations (see Figure 4 for the word *heof* in the sense of "A robber, thief"). The other tabs in this view offer all known meanings of the word as a list or as a word cloud. Users can add their own annotations for any

resource they view, which will be discussed in-depth in section 3.4.

Next to browsing, Evoke can be used to perform more advanced analyses over the data available. The *statistics* section of the application offers the possibility to build custom queries for analysing the use of specific tags, parts of speech, and/or languages (see Figure 5). A selection can be contrasted with all words (default), other parts of speech, other languages (if multiple are present), or another custom selection of features. Analysis results for a query are shown in the form of charts that convey the item count of a selection, its degree of ambiguity (indicating polysemy), its degree of synonymy, its distribution over taxonomy subcategories (see Figure 6), and its distribution over the depth of the taxonomy (indicating how specific in meaning the selected words are; see Figure 7).

Finally, the *user* section of Evoke allows users to make a backup of their annotations, reactivate backups, discard their user data, or publish their data to an online data service.

3.1. Architecture

Evoke employs client-side rendering, which entails that the user's internet browser does not fetch entire webpages from a server but fetches only data necessary to fill out pages that it itself composes. Use of this rendering technique has the advantage of reducing the amount of resources required server-side. The interface of Evoke is rendered using React and Reactstrap (basic HTML) complemented by libraries specifically intended for vector graphics (i.e., D3, D3Pie, Recharts, and Wordcloud).⁶

The application offers a high degree of flexibility in managing which content is to be viewed or analysed, including where individual datasets have been made available. The functionality to realize this hinges on the use of so-called data catalogues, which lists datasets available for querying. Upon start-up, Evoke loads a default data catalogue. Custom data catalogues can, however, be used too. The use of data catalogues in Evoke will be discussed in more detail in section 3.3.

Datasets listed in a catalogue are accessed through Ajax calls to SPARQL endpoints and/or APIs. As Ajax calls to data services are asynchronous, the Evoke web application contains a so-called high-order component that acts as loader. This loader wraps the desired component in one that first awaits data requests. During the loading time, a loading icon is presented. Once the data required has been retrieved, the loader renders the wrapped component with its proper input. This mechanism is used, in combination with the underlying data form of Linguistic Linked Data, to follow links iteratively in fetching further information from all active datasets. Thus, when a user selects a semantic concept of a thesaurus for viewing, a list of words that express that concept is retrieved. For each of those words, their URI is used to retrieve all associated tags available across the active datasets – whether part of the original thesaurus data or a set of annotations created by others.

The application ingests, next to data served remotely via SPARQL or API calls, linked data stored in the internet browser of the user. A dataset in the browser can be stored using either the Turtle serialization or in JSON-LD and is interpreted using

⁶React: <https://reactjs.org/>. Reactstrap: <https://reactstrap.github.io/>. D3: <https://d3js.org/>. D3Pie: <http://d3pie.org/>. Recharts: <https://recharts.org/>. Wordcloud: <https://wordcloud2-js.timdream.org/>.

libraries from the Comunica framework.[13]⁷ The use of this storage method is detailed in section 3.4.

3.2. Linguistic Linked Data

Evoke adopts the Linguistic Linked Data paradigm for the way it accesses, explores, and extends data. This paradigm builds on linked data principles, which advocate the use of Web mechanisms for capturing and sharing data, employing graph-based models (i.e., nodes and relations between them) and identifying data by means of URIs (often Web addresses).[14] The use of URIs allows one to capture and identify data in a language-independent manner, reuse terminology defined elsewhere, and to create links between datasets or nodes within different datasets.

Applying linked data principles to language resources nets a number of benefits.[15] One of these benefits is that its data form enables the merging of datasets in order to obtain a valid combined set of data. Thus, linguistic resources and datasets elaborating on them can be queried in unison. A second benefit is an increased level of interoperability. Using standardized terminology in describing linguistic data increases a shared understanding of that data and facilitates their interpretation by software. Moreover, the use of URIs as identifiers ensures data can be linked without the need for duplication of information from one dataset into the other. The ability to link (or reference) in such a manner is valuable in the setting of Linguistic Linked Data, since it is not uncommon to find lexicographic resources subject to licenses meant for viewing only and stipulating users are not allowed to copy or download a substantial portion of its content (e.g., [8, 16]). By adopting URIs in published lexicographic resources, then, their users can start exploring and extending these resources at an early stage and may be enticed to engage more fully with the content offered without infringing on such licenses.

The Evoke web application, in order to draw on the aforementioned benefits, assumes language resources to adopt Linguistic Linked Data as specified by the W3C OntoLex community,[17] applied specifically to the context of topical thesauri.[18] The use of linked data in Evoke is, however, not limited to solely the *content* of datasets catalogued. The data catalogues

⁷The Comunica library n3 is used for Turtle; jsonld-streaming-parser for JSON-LD; actor-init-sparql-rdfjs for SPARQL queries in browser memory.

1 themselves and offline data stored in the internet
2 browser employ this data form, as will be discussed in
3 the next sections.

4 3.3. Data catalogue

5
6
7 The data catalogues in Evoke adhere to the W3C
8 DCAT vocabulary, an international standard
9 specifically for expressing datasets and the services
10 that provide these sets, including access details.[19]
11 This information is stored in the JSON-LD format and
12 can therefore be read as JSON or, through the context
13 provided, interpreted as RDF.[20] The use of these
14 standards are meant to accommodate a higher level of
15 interoperability with other tooling and services. An
16 example catalogue is shown in Listing 1. Users can
17 add their own datasets or data services to a catalogue,
18 store their catalogue locally as a JSON file, and
19 activate a local catalogue by dragging and dropping it
20 onto the Evoke web application.

21 Access to datasets may or may not need to be
22 limited, depending on the usage license associated
23 with them. In order to ensure that Evoke can work
24 with more restrictive licenses, too, two types of access
25 mechanisms for data services are supported in data
26 catalogues: (1) a SPARQL endpoint and (2) the Evoke
27 API. The former allows services to respond to any
28 query using the standardized querying language for
29 RDF. The latter ensures content can be viewed and
30 browsed in Evoke through a basic set of queries
31 specific for this need, without offering users full
32 access (that is, the possibility to extract or download
33 the full dataset). Distinguishing between these two
34 types of access mechanism in the data catalogue is
35 achieved through different values for the
36 *endpointDescription* attribute of data services.⁸

37 Which datasets listed in the active catalogue will be
38 queried by Evoke depends on the selection made by
39 the user. The Evoke interface allows users to enable
40 (or disable) datasets listed. Only those datasets can be
41 enabled that (1) have a data service associated with
42 them and (2) already have all of their required
43 dependencies enabled. To illustrate, the ‘Riddle 47’
44 dataset contains links to the dataset ‘A Thesaurus of
45 Old English’ and depends on it for analyses. Once a
46 user has enabled this required dataset in the interface,
47 that user can opt to enable the ‘Riddle 47’ dataset, too

48
49
50 ⁸The value of *endpointDescription* denoting a SPARQL endpoint
51 is "http://www.w3.org/ns/sparql-service-description#Service"; the
value denoting the Evoke API is "http://evoke.ullet.net/api".

1 (see the top bars in Figure 4). Datasets served by the
2 same service, though available in different graphs, are
3 queried in unison and allow statistical analyses to be
4 performed.⁹

5 3.4. Browser storage

6
7
8 Any user of the Evoke web application can
9 annotate linguistic content, such as words or semantic
10 concepts, with information relevant to them. Typing a
11 sentence in the annotation component of a page will
12 automatically create a linked data annotation that
13 adheres to the Web Annotation standard of W3C
14 including any extracted tag when a hashtag is used
15 (see “#riddle47” in Figure 4).[21] The novel aspect of
16 this approach is that such an annotation is not stored
17 in an online database but is instead stored locally in
18 the user’s internet browser, employing the
19 *localStorage* attribute of Web storage.[22] Annotations
20 stored in the browser can be downloaded
21 as a file to backup and can be reactivated in the
22 browser – providing users with full control over their
23 created content and allowing them to share it in the
24 manner of their choosing (see Listing 2). Publishers
25 of the original lexicographic resource benefit from
26 this approach, too, as they neither need to moderate,
27 store, or host annotations nor offer users login
28 mechanisms before they can interact with the
29 information. Costs for hosts may thus be substantially
30 reduced for presenting users with this functionality.

31 Annotations contain references to the identifiers, or
32 URIs, of the original lexicographic content without
33 including the raw data of the annotated content in the
34 annotation itself. This approach can entice users to
35 already explore dictionaries and interact with them,
36 formulate a plan of research, and at a later stage take
37 the hurdle in getting support for further research from
38 the publishers – be it in the form of a more open
39 license, getting access to advanced services, or asking
40 assistance from the expertise of lexicographers. Users
41 can be said to have an invested interest in the
42 lexicographic resource at this point. Moreover, their
43 additions are explicit, digital, and can be used in this

44
45
46 ⁹Evoke does not perform statistical analyses over data found at
47 *different* data services, since that would require assembling a full
48 picture of the relevant data either at the user, one of the data services,
49 or a data aggregator (be it query results or parts of datasets), which
50 is an intricate process, time-consuming, heavy on network traffic,
51 and possibly restricted or made impossible by the licenses that are
applicable to the datasets.

form for analyses when queried in unison with the original dataset (facilitated by the characteristics of Linguistic Linked Data).

User data stored in the internet browser can, as with any RDF dataset, be published to a data service and added to a data catalogue for use in Evoke. Indeed, when one publishes through the Evoke user interface, a new data catalogue is created automatically in which the published dataset is listed. This updated catalogue is not made public by the application: as with other user content, the catalogue is stored in the browser. A means to download the updated catalogue is provided to the user immediately after a successful publication. Users can choose to share it with others in a way they see fit or, if they desire to share the newly published dataset publicly, upload it to a public server and/or contact the administrator of the application to request inclusion of the dataset in the default catalogue.

4. Case study: A *Thesaurus of Old English*

To explore the impact of Evoke and the solution suggested, a research project has been organized for a case study of a single thesaurus. This project has brought together 17 scholars from 8 universities and lexicographic institutions from several countries to explore the contents of *A Thesaurus of Old English* using the application.[23] This thesaurus, transformed to Linguistic Linked Data,[24] captures the lexis of the early medieval variant of English that was spoken between roughly 500 and 1100 AD. Each investigation within the research project explores the lexis from the perspective of a certain discipline: linguistics, literary-criticism, history, lexicography, philology. In their explorations, the researchers (and, in some cases, students) set about in viewing, adding to, and analysing content of *A Thesaurus of Old English*. The majority of the researchers employ the annotation system of Evoke to add relevant data. Some, however, utilize linked data mechanisms and link datasets through other software first before viewing and analysing the results in Evoke.

Examples of research done within this project are linking up words (or word senses, rather) from Old Frisian and Old Dutch to the thesaurus taxonomy. Doing so will allow researchers to contrast how many nuances these language communities respectively were known to have, next to Old English, in expressing certain concepts such as kinship or greetings. Other research focuses on adding

information on the occurrence of words in certain medieval texts (e.g., *Beowulf*) or their use by a certain author (e.g., Ælfric), offering insights into the onomasiological profiles associated with authors or genres. These profiles may act as semantic fingerprints in identifying authorship or in positioning other work. The outcomes will be presented at the pre-conference EASE workshop of the 21st edition of the International Conference of English Historical Linguistics (Leiden, June 2021)¹⁰. Its proceedings will be published in a special issue journal of the *Amsterdamer Beiträge zur älteren Germanistik*¹¹ (Fall 2021). Preliminary results show that both students and established researchers – including ones who had no prior knowledge of linked data and RDF technology – are able to utilize Evoke in answering novel research questions.

5. Conclusion

Evoke is one of the first applications that provides a user-friendly interface for working with Linguistic Linked Data resources, opening these resources up to users who may not have advanced knowledge of linked data and RDF technologies. The web application employs linked data mechanisms to offer functionality for viewing, navigating, extending, and analysing content of topical thesauri together with relevant, linked datasets. The novel architecture of Evoke employs elements – i.e., data catalogues, internet browser storage, annotation on URIs – that address some of the more intractable problems in modern lexicography: licensing and hosting concerns that prevent offering functionality to users to interact with the lexicographic resources they value. Evoke demonstrates that, by employing linked data mechanisms, interests of both users and publishers can be taken into account so that users can extend these resources and formulate new queries – without infringing on licenses or demanding additional data storage from the publisher of the resource. The web application Evoke, then, allows for a higher engagement of users with published lexicographic resources in exploring and extending these lexical treasure troves.

Future work on Evoke will aim to incorporate functionality most frequently requested by

¹⁰<https://www.universiteitleiden.nl/icehl21>

¹¹<http://brill.com/abag>

1 researchers. These include the ability for users to hide
 2 content (e.g., filtering based on user-defined tags).
 3 This feature would effectively allow users to create
 4 subthesauri. A second helpful feature is one that
 5 indicates which datasets from a data catalogue
 6 contains statements on a specific URI (be it a word,
 7 word sense, semantic concept, or anything else).
 8 Thus, Evoke should be able to foreground which
 9 datasets may be relevant for a user's context.
 10 Additionally, such a feature will be helpful to locate
 11 works that have engaged with one's own.

12 References

- 13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
- [1] M. Görlach, Jane Roberts and Christian Kay, with Lynne Grundy. *A Thesaurus of Old English*. 2 vols. London: King's College, 1995, xxxv + 1555 pp. £55.00 hb., *Anglia - Zeitschrift für englische Philologie* **116**(3) (1998), 398–401.
- [2] R.F. Ison, On the Historical Thesaurus of the Oxford English Dictionary, *International Journal of Lexicography* **24**(2) (2011), 241–260. doi:10.1093/ijl/lecq032.
- [3] K. Allan, *Metaphor and Metonymy: A Diachronic Approach*, Publications of the Philological Society, Vol. 42, Wiley-Blackwell, Oxford, 2008.
- [4] P.W. Conner, Jane Roberts and Christian Kay, eds., with Lynne Grundy, *A Thesaurus of Old English*, 1: *Introduction and Thesaurus*; 2: *Index*. London: King's College London, 1995. 1: pp. xxxv, 1-716. 2: pp. iv, 717-1555. £55., *Speculum* **73**(3) (1998), 887–889. doi:10.2307/2887551.
- [5] C.J. Kay, Marvin Spevack, *A Shakespeare Thesaurus*. Hildesheim/Zürich/New York: OLMS, 1993. xxv+541., *International Journal of Lexicography* **9**(1) (1996), 71–76.
- [6] R.H. Bremmer Jr, Treasure Digging in the Old English Lexicon, *NOWELE* **40** (2002), 109–114.
- [7] R. Dance, Jane Roberts and Christian Kay, with Lynne Grundy, *A Thesaurus of Old English*, King's College London Medieval Studies 11 (London: King's College London Centre for Late Antique and Medieval Studies, 1995). 2 vols. xxxv + 1555 pp. ISBN 0-9522119-0-4. £55.00., *Medium Ævum* **66**(2) (1997), 312–313.
- [8] C. Kay, M. Alexander, F. Dallachy, J. Roberts, M. Samuels and I. Wotherspoon (eds), *The Historical Thesaurus of English*, 2nd edn, University of Glasgow, 2021. <https://ht.ac.uk/>.
- [9] T. Declerck, J.P. McCrae, M. Hartung, J. Gracia, C. Chiarcos, E. Montiel-Ponsoda, P. Cimiano, A. Revenko, R. Saurí, D. Lee, S. Racioppa, J. Abdul Nasir, M. Orlikowsk, M. Lanau-Coronas, C. Fäth, M. Rico, M.F. Elahi, M. Khaalchik, M. Gonzalez and K. Cooney, Recent Developments for the Linguistic Linked Open Data Infrastructure, in: *Proceedings of the 12th Language Resources and Evaluation Conference*, European Language Resources Association, Marseille, France, 2020, pp. 5660–5667. ISBN 979-10-95546-34-4. <https://www.aclweb.org/anthology/2020.lrec-1.695>.
- [10] J.P. McCrae and P. Cimiano, Linghub: a Linked Data based portal supporting the discovery of language resources., in: *SEMANTiCS (Posters Demos)*, A. Filipowska, R. Verborgh and A. Polleres, eds, CEUR Workshop Proceedings, Vol. 1481, CEUR-WS.org, 2015, pp. 88–91. <http://dblp.uni-trier.de/db/conf/i-semantics/semantics2015p.html#McCraeC15>.
- [11] C. Chiarcos, B. Klimek, C. Fäth, T. Declerck and J.P. McCrae, On the Linguistic Linked Open Data Infrastructure, in: *Proceedings of the 1st International Workshop on Language Technology Platforms*, European Language Resources Association, Marseille, France, 2020, pp. 8–15. ISBN 979-10-95546-64-1. <https://www.aclweb.org/anthology/2020.iwltpl-1.2>.
- [12] C. Kay and M. Alexander, Diachronic and synchronic thesauruses, in: *The Oxford handbook of lexicography*, P. Durkin, ed., Oxford University Press, Oxford, 2016, pp. 367–380. ISBN 978-0-19-969163-0.
- [13] R. Taelman, J. Van Herwegen, M. Vander Sande and R. Verborgh, Comunica: a Modular SPARQL Query Engine for the Web, in: *Proceedings of the 17th International Semantic Web Conference*, 2018. <https://comunica.github.io/Article-ISWC2018-Resource/>.
- [14] P. Cimiano, C. Chiarcos, J.P. McCrae and J. Gracia, *Linguistic Linked Data - Representation, Generation and Applications*, Springer, 2020. ISBN 978-3-030-30224-5. doi:10.1007/978-3-030-30225-2.
- [15] C. Chiarcos, J. McCrae, P. Cimiano and C. Fellbaum, Towards open data for linguistics: Lexical Linked Data, *New Trends of Research in Ontologies and Lexical Resources: Ideas, Projects, Systems* (2013), 7–25. ISBN 978-3-642-31781-1.
- [16] OED Online, Oxford University Press, Oxford, 2020. <https://oed.com>.
- [17] P. Cimiano, J.P. McCrae and P. Buitelaar (eds), *Lexicon Model for Ontologies: Community report*, 10 May 2016, W3C, 2016. <http://www.w3.org/2016/05/ontolex/>.
- [18] S. Stolk, lemon-tree: Representing Topical Thesauri on the Semantic Web, in: *2nd Conference on Language, Data and Knowledge (LDK 2019)*, M. Eskevich, G. de Melo, C. Fäth, J.P. McCrae, P. Buitelaar, C. Chiarcos, B. Klimek and M. Dojchinovski, eds, OpenAccess Series in Informatics (OASiCS), Vol. 70, Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 2019, pp. 16:1–16:13. ISSN 2190-6807. ISBN 978-3-95977-105-4. doi:10.4230/OASiCS.LDK.2019.16. <http://drops.dagstuhl.de/opus/volltexte/2019/10380>.
- [19] R. Albertoni, D. Browning, S. Cox, A. Gonzalez Beltran, A. Perego and P. Winstanley (eds), *Data Catalog Vocabulary (DCAT) - Version 2: W3C Recommendation 04 February 2020*, W3C, 2020. <https://www.w3.org/TR/vocab-dcat-2/>.
- [20] M. Sporny, D. Longley, G. Kellogg, M. Lanthaler, P.-A. Champin and N. Lindström, *JSON-LD 1.1: A JSON-based Serialization for Linked Data*. W3C Recommendation 16 July 2020, W3C, 2020. <https://www.w3.org/TR/json-ld11/>.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51



Fig. 2. Evoke showing the taxonomy of 'A Thesaurus of Old English'.

The screenshot shows the Evoke interface with a search bar and a user profile icon. The breadcrumb path at the top reads 'all > Speech, vocal utterance > The voice'. Below this, the word 'The voice' is highlighted with a yellow dot. To the right of 'The voice' are icons for information, list view, edit, and share. Below the title, there is a list of five words, each in a grey box with a 'noun' label on the left and a blue 'Old English' label on the right. The words and their additional labels are: heafodwōþ (poetry, rare), hlēoþor, gereord, stefn, and wordhlēoþor (poetry). At the bottom left of the list area, there is a button that says 'Show more from this domain'. At the bottom of the interface, the word 'evoke' is centered, and on the right, there are social media icons for Twitter, GitHub, and a user profile.

Fig. 3. Evoke listing words denoting "The voice" from 'A Thesaurus of Old English'.

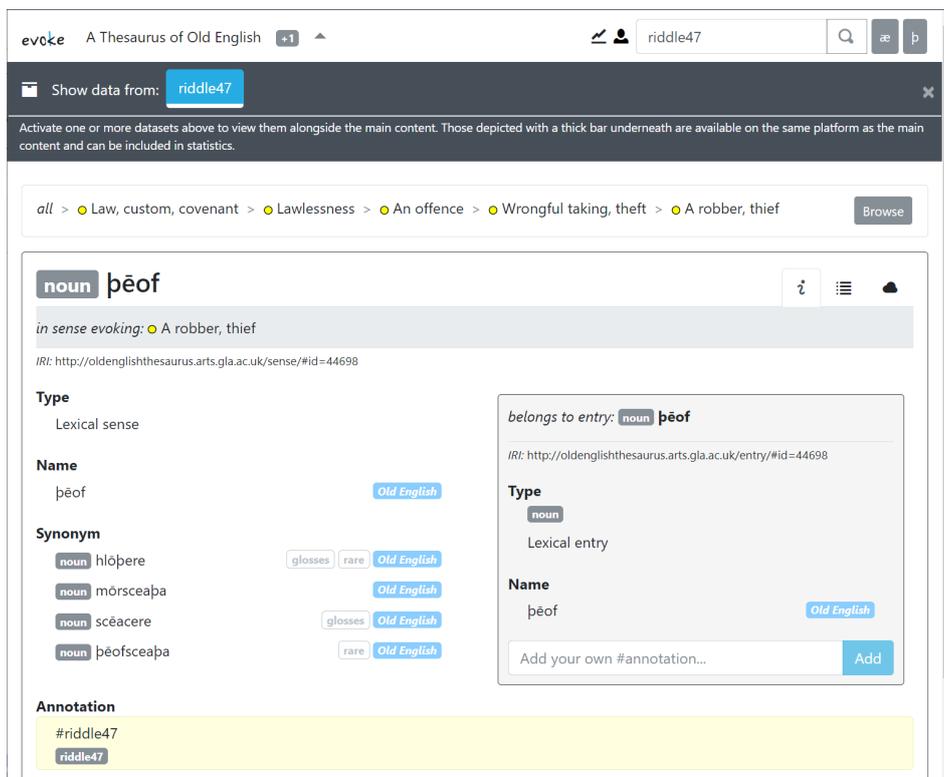


Fig. 4. Evoke showing a sense of the word *þeof* as described in two datasets (i.e., ‘TOE’ and ‘riddle47’).

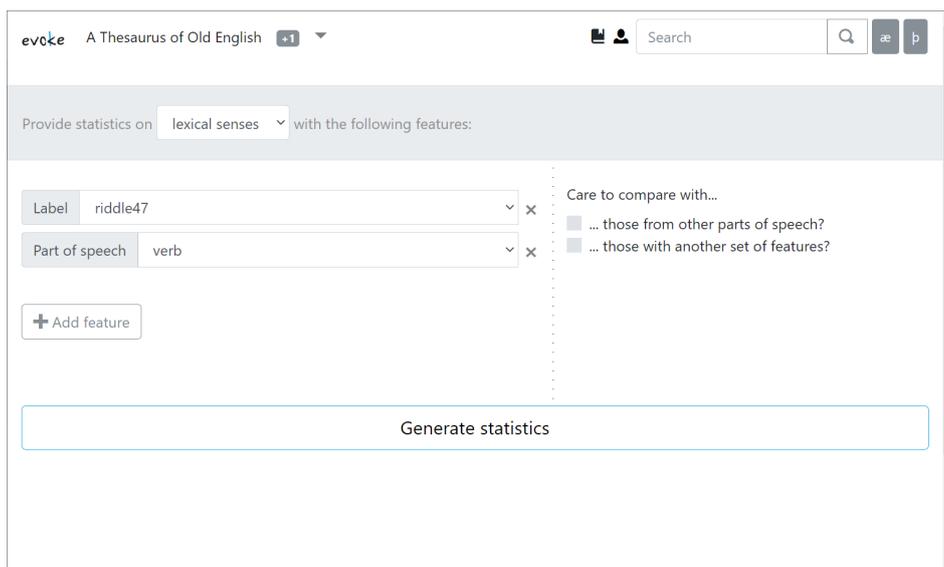


Fig. 5. Evoke showing the form that allows users to select their features of interest in the statistics section.



Fig. 6. Evoke showing the distribution over semantic concepts of word senses labelled 'riddle47' (orange) versus all senses (blue).

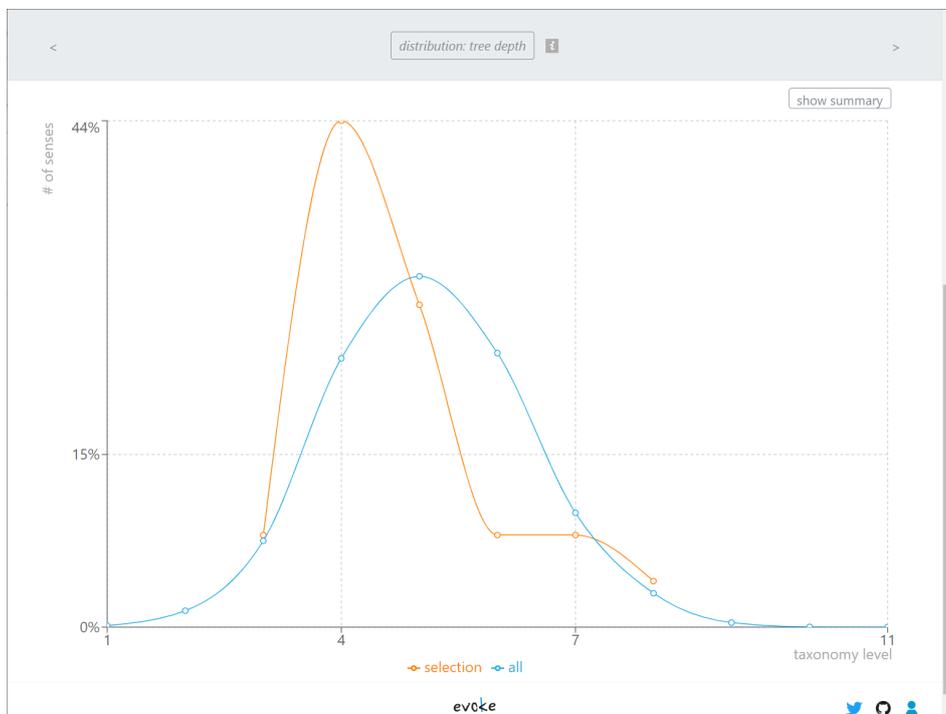


Fig. 7. Evoke showing the distribution over taxonomy depth of word senses labelled 'riddle47' (orange) versus all senses (blue).

```

1  {
2    "@context":
3      "https://raw.githubusercontent.com/ssstolk/DCAT-AP/master/releases/2.0.0/Draft/dcat-ap_2.0.0.jsonld",
4    "$schema": "./catalog.schema.json",
5    "@id": "",
6    "@type": "Catalog",
7    "service": [
8      {
9        "@id": "http://evoke.ullet.net/platform",
10       "@type": "DataService",
11       "title": "Evoke platform",
12       "identifier": "evoke-platform",
13       "endpointURL": "http://142.93.226.251:8081",
14       "endpointDescription": "http://evoke.ullet.net/api",
15       "landingPage": "http://evoke.ullet.net/app/",
16       "mode": "get",
17       "servesDataset": [
18         "http://oldenglishtesaurus.arts.gla.ac.uk/",
19         "https://w3id.org/evoke/set/riddle47"
20       ]
21     }
22   ],
23   "dataset": [
24     {
25       "@id": "http://oldenglishtesaurus.arts.gla.ac.uk/",
26       "@type": "Dataset",
27       "title": "A Thesaurus of Old English",
28       "identifier": "toe",
29       "landingPage": "http://evoke.ullet.net/thesaurus/toe/",
30       "license": "http://evoke.ullet.net/thesaurus/toe/#license",
31       "issued": "2017-05-26",
32       "distribution": {
33         "accessService": "http://evoke.ullet.net/platform",
34         "accessGraph": "http://oldenglishtesaurus.arts.gla.ac.uk/",
35         "mediaType": "application/sparql-results+json"
36       }
37     },
38     {
39       "@id": "https://w3id.org/evoke/set/riddle47",
40       "@type": "Dataset",
41       "title": "riddle47",
42       "identifier": "riddle47",
43       "distribution": {
44         "accessService": "http://evoke.ullet.net/platform",
45         "accessGraph": "https://w3id.org/evoke/set/riddle47",
46         "mediaType": "application/sparql-results+json"
47       },
48       "requires": [
49         "http://oldenglishtesaurus.arts.gla.ac.uk/"
50       ]
51     }
52   ]
53 }

```

Listing 1 Data catalogue serving 'A Thesaurus of Old English' and 'riddle47'

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

[22] I. Hickson (ed.), Web Storage (Second Edition): W3C Recommendation 19 April 2016, W3C, 2016. <https://www.w3.org/TR/webstorage/>.

[23] J. Roberts, C. Kay and L. Grundy, *A Thesaurus of Old English: In Two Volumes*, Rodopi, 2000. ISBN 9789042015630.

[24] S. Stolk, 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, pp. 223–247. https://elex.link/elex2019/wp-content/uploads/2019/09/eLex_2019_13.pdf.

```

1
2
3
4
5
6 {
7   "@context": [
8     "http://www.w3.org/ns/anno.jsonld",
9     {
10      "skos": "http://www.w3.org/2004/02/skos/core#",
11      "Concept": "skos:Concept",
12      "prefLabel": "skos:prefLabel"
13    },
14    {
15      "content": {
16        "@reverse": "rdfs:isDefinedBy",
17        "@container": "@index"
18      }
19    }
20  ],
21  "content": {
22    "Annotation": [
23      {
24        "id": "https://w3id.org/evoke/id/annotation/e6081476-b449-45d1-bec5-31aad6aad367",
25        "type": "Annotation",
26        "created": "2020-06-29T17:41:06.351Z",
27        "motivation": "commenting",
28        "target": "http://oldenglishtesaurus.arts.gla.ac.uk/sense/#id=17981",
29        "bodyValue": "#riddle47",
30        "body": {
31          "id": "https://w3id.org/evoke/id/annotation/e6081476-b449-45d1-bec5-31aad6aad367-body",
32          "type": "SpecificResource",
33          "source": [
34            "https://w3id.org/evoke/id/concept/riddle47"
35          ],
36          "purpose": "tagging"
37        }
38      },
39      {
40        "id": "https://w3id.org/evoke/id/annotation/d17922bc-da28-4bab-82a3-238776c753ab",
41        "type": "Annotation",
42        "created": "2020-06-29T17:42:02.782Z",
43        "motivation": "commenting",
44        "target": "http://oldenglishtesaurus.arts.gla.ac.uk/sense/#id=33789",
45        "bodyValue": "#riddle47",
46        "body": {
47          "id": "https://w3id.org/evoke/id/annotation/d17922bc-da28-4bab-82a3-238776c753ab-body",
48          "type": "SpecificResource",
49          "source": [
50            "https://w3id.org/evoke/id/concept/riddle47"
51          ],
52          "purpose": "tagging"
53        }
54      }
55    ],
56    "Concept": [
57      {
58        "id": "https://w3id.org/evoke/id/concept/riddle47",
59        "type": "Concept",
60        "prefLabel": "riddle47"
61      }
62    ]
63  }
64 }

```

Listing 2 User-created annotations of two word senses, which are tagged to indicate they occur in the Old English text *Exeter Riddle 47*.