11ᵗʰ International Conference on Historical Lexicography and Lexicology – ICHLL 11

June 16-18, 2021
Universidad de La Rioja

BOOK OF ABSTRACTS

UNIVERSIDAD DE LA RIOJA

TABLE OF CONTENTS

PLENARY LECTURES

GENERAL SESSIONS

(Mis)adventures in corpus creation: Lemmatization, PoS tagging, and the creation of a dictionary of forms for Old Spanish

Francisco Gago-Jover

College of the Holy Cross (United States of America)

fgagojov@holycross.edu

For the last four years a small team of hispano medievalist have been working on the development of the *Old Spanish Textual Archive*, a morphologically tagged and lemmatized corpus of more than 25 million words, based on the more than 400 paleographic transcriptions of medieval texts written in Castilian, Asturian, Leonese, Navarro-Aragonese and Aragonese prepared by the collaborators of the Hispanic Seminary of Medieval Studies (HSMS). Once we established the basic architecture of the corpus, we proceeded to evaluate a variety of taggers/lemmatizers—Freeling, LaPOS, Marmot, neoTAG, and TreeTagger—trying to find one that allowed us to preserve the orthographical variation present in medieval texts written in Castilian, Aragonese, Navarrese, and Leonese, without the need to normalize the language prior to processing, as this would have eliminated most variation and with it, the linguistic value of the corpus.

We ended up selecting Freeling as it had already been trained and tested with a small part of the HSMS corpus (Sánchez Marco 2011, 2012), thus avoiding the need to train the other tools. FreeLing's processing pipeline is relatively simple: a set of texts is sent to the analyzer, which processes and enriches the texts with linguistic information using different modules: tokenizer, dictionary, affixer, probabilistic analyzer and unknown-word guesser, and PoS tagger. However, once we processed the whole HSMS corpus with Freeling to evaluate its real-world performance, we detected a fairly large number of problems caused by the dictionary that Freeling employs. Correcting and improving that dictionary became our main task for the last two years. In this presentation I will discuss 1) the origins and structure of the original Freeling dictionary and why unsupervised methods of enrichment can lead to problems, 2) the typology of problems detected and the implications they have on an accurate tagging/lemmatization, 3) the tool developed to edit and improve the dictionary, with its benefits and shortcomings, and 4) other methods used to improve the dictionary, including lemma and PoS extraction from other dictionaries, and the process of manual revision.

Gathering, cultivating and harvesting the data — the evolution of a digital resource for Old Norse prose

Ellert Jóhannsson

University of Copenhaguen (Denmark)

ellert@hum.ku.dk

The main theme of this paper is the data used in historical dictionary work and how technological advances have opened up new avenues for managing, exploring and exploiting lexicographic data.

The focus will be on the development of A Dictionary of Old Norse Prose (ONP), which has gone through several incarnations in its 80 years of existence. ONP was originally conceived of as a supplementary citation collection to its main 19th century predecessor. It then became an independent dictionary project, with its own set of principles and standards, centered on collecting example citations to cover the entirety of the Old Norse prose vocabulary and to register data about the medieval source material. This work then finally bore fruit in the eighties when publication of printed dictionary volumes commenced. After 15 years and several volumes covering about one fourth of the alphabet, the project became all-digital, to be published electronically online from 2010. Since then ONP Online has been improved and enhanced to its current form as a dynamic digital resource with several innovative features. The key to the success of the ONP project is meticulous registration of data and very detailed data management system. ONP sought inspiration from the Middle English Dictionary in its scope and attention to philological details, with great emphasis on the integrity of the source material. For the first forty years of the project a vast amount of data was collected, but management was limited to analogue methods, using paper and pen and several types of slips organized in filing cabinets. It was not until the computer age that it was possible to take full advantage of ONP's archives by digitizing the data and then later linking it together in different ways.

In this paper, I will give an overview of this development, identifying specific milestones in the evolution of the project and their significance. I will show how consistent data management and organization makes it possible through recent technological advancements to bring to light aspects of the data, which have not been discernible until now. This is evident by the latest online platform of ONP, launched in December 2019, which brings together the work of many generations of lexicographers and offers unprecedented look at all the data ONP has collected throughout its existence with extensive linking as well as several enhancements from other digital sources. This allows the user to interact with ONP and its source material in previously impossible ways, using digital text editions, scanned manuscript pages, other lexicographic resources and even Google translate.

'Impersonal' and 'Reflexive' constructions: Verb features peculiar to Old and Middle English

Michiko Ogura

Chiba University (Japan)

ogura.dainagon@jcom.home.ne.jp

When Old English appeared in a written form for the first time, it had already lost inflections like optative, hortative, perfective, passive, etc. Making up for these morphological forms, it started, again before it was written and preserved, using periphrastic expressions with modal auxiliaries, *habban*, *beon/wesan*, *utan*, *ongan*, etc. Without having middle voice, it used 'impersonal' and 'reflexive' constructions (the single quotes mean that they included quasi-impersonals and quasi-reflexives in the real sense of the words). In this paper I focus on some such verbs as *lician*, *lystan*, *sceamian*, *þyncan* and *wer(g)ian* with their native and/or loan synonyms like *(dis)plesen*, *joien*, *remembren*, *repenten*, *semen*, etc. and their constructions used in Old and Middle English so as to maintain that their peculiar features reflect compensatory devices of the lost function before the appearance of Old English. The phonetic-morphological-syntactic merger, which occurred from late Old to early Middle English, is also discussed, in order to see the results of the rivalry of some synonyms into Modern English.

Lemmatization in corpus-to-dictionary systems: The case study for Old Church Slavonic

Ilia Afanasev (Petrozavodsk State University/Saint-Petersburg State University, Russia)

Olga Lyashevskaya (Higher School of Economics/Vinogradovite Institute of the Russian Language, Russian Academy of Sciences, Russia)

szrnamerg@gmail.com, olesar@gmail.com.

Attempts at compilation of electronic dictionaries with corpora have significantly grown in last two decades (Rundell and Kilgarriff 2011) (Paquot 2012). The natural continuation of this approach is to build a system that creates both corpus and dictionary simultaneously, or, at the very least, immediately forms dictionary from corpus after the creation of the latter one.

The main aim is to develop a system that may perform corpus-to-dictionary transformation from scratch as effectively and fast, as it is possible. This implies both the necessary implementation of universal lemmatizer development, and collecting the lemmatization results, if provided during the creation of corpus. So, attention specifically is paid to this problem.

The proposal is to build the corpus of one Old Church Slavonic text, and compare the results of specialized lemmatizer for this language (Afanasev 2021), lemmatizer, designed to deal with multiple languages (Kanerva *et al*. 2020), and naïve lemmatizer, operating from scratch. The comparison of efficiency of these three tools is to be measured with string similarity distances, Levenshtein distance, Damerau-Levenshtein distance, and Jaro-Winkler distance.

The most interesting results this system may show on the text with very specific linguistic features, like Kiev Folia (Kamphuis 2020). Each of lemmatizers demonstrates approximately similar results, however, the time they take differs, which is going to be demonstrated.

As a result of this research, system is enhanced with three possible lemmatization options. Recommendations of their efficiency, based on the comparison results, are provided.The module of the system that is able to form dictionary from corpus, is prepared.

References

Afanasev, I. 2021. 'A Hybrid Lemmatiser for Old Church Slavonic'. In *HSE WP BRP Series: Linguistics / LNG*. Moscow: HSE Working Papers, vol. 106. 1–19.

Kamphuis, J. 2020. *Verbal Aspect in Old Church Slavonic*. Brill: Leiden. 329.

Kanerva, J., Ginter, F., and T. Salakoski. 2020. 'Universal Lemmatizer: A sequence-to-sequence model for lemmatizing Universal Dependencies treebanks'. In *Natural Language Engineering*. Cambridge: Cambridge University Press. 1–30.

Paquot, M. 2012. 'The LEAD dictionary-cum-writing aid: An integrated dictionary and corpus tool'. In Sylviane, G., and M. Paquot (eds.), *Electronic Lexicography*. Oxford: Oxford University Press. 163–185.

Rundell, M. and A. Kilgarriff. 2011. 'Automating the creation of dictionaries: where will it all end?'. In F. Meunier, S. De Cock, G. Gilquin, and M. Paquot (eds.), *A Taste for Corpora: In honour of Sylviane Granger. Studies in Corpus Linguistics*. Amsterdam: John Benjamins Publishing Company. 45: 257–282.

Vocabularium Bruxellense. Towards electronic edition of a Medieval Latin vocabulary

Renaud Alexandre (Comite Du Cange – IRHT, France)

Iwona Krawczyk (Institute of Polish Language, Poland)

Krzysztof Nowak (Institute of Polish Language, Poland)

renaud.alexandre@irht.cnrs.fr,

iwona.krawczyk@ijp.pan.pl, krzysztof.nowak@ijp.pan.pl

Vocabularium Bruxellense is a name attributed by scholars to a Medieval Latin vocabulary preserved in a manuscript conserved at the Royal Library of Belgium (Ms. Bruxelles, B.R., II 1049.). This lexicographic compilation dated to the mid-12th century is believed to be the work of an anonymous Cistercian monk (Weijers 1996).

The text, until now unpublished, was transcribed by H. F. Reijnders, a Dutch medievalist who provided it also with a basic apparatus. Although critical edition of the text will still require a considerable effort, the team of philologists have already started work to make this important text available to the scholarly community in the near future.

The present paper reports on the first results of the work which consisted of both data modeling and its manual and semi-automatic XML annotation, but we focus on challenges that this product of medieval erudition posed to TEI modeling.

Vocabularium Bruxellense is similar to other works of the medieval lexicography in that it does not lend itself to easy TEI representation, although one can take inspiration from recent major electronic editions, such as *Liber glossarum* (Grondeux & Cinato 2016). Grouping of morphologically related words under single heading, fuzzy boundaries between definition and quotation, to name only few, make the encoding schema inevitably complex.

The application of the schema is, also, all but trivial. Considerable interpretative effort on the part of the annotator is required, for instance, to distinguish between different types of illustrative material or to discriminate proper definitions or equivalents from introductory formulas such as 'dicitur etiam'.

The TEI XML edition of the transcript, as soon as it is proofread and its sources identified, will be freely available to the public through a TEI Publisher application and will become a basis for subsequent critical edition.

References

Grondeux, A. and F. Cinato. 2016. '*Liber Glossarum Digital*' *Latinitatis Medii Aevi*. 54. 233-238.
[http://liberglossarum.huma-num.fr.]

Weijers, O. 1996. 'Notice Sur Le «Vocabularium Bruxellense»'. *ALMA. Archivum*

The Velum Project. Building a corpus of Medieval Latin for lexicographic purposes

Renaud Alexandre (Comite Du Cange – IRHT, France)

Krzysztof Nowak (Institute of Polish Language, Poland)

Bruno Bon (Institute of Polish Language, Poland)

renaud.alexandre@irht.cnrs.fr, krzysztof.nowak@ijp.pan.pl, bruno.bon@irht.cnrs.fr

It has been exactly 100 years since the project of the dictionary of Medieval Latin "*Novum Glossarium Mediae Latinitatis*" (= NGML) was launched in 1920 that seeks to describe vocabulary of texts produced across the Europe between 800 and 1200 AD (Bon 2015). The dictionary is based on manually excerpted quotations and systematic queries in digital text collections. Both sources are, however, considerably biased: while the former highlight rare words to detriment of frequent ones, the latter is skewed toward literary texts compared to documents.

The present paper discusses outline and preliminary results of the project "Velum. Visualisation, exploration et liaison de ressources innovantes pour le latin médiéval" which seeks to build a large representative and balanced corpus of Medieval Latin. Planned to contain 100M words, the corpus will also include three subcorpora corresponding to periods before and after 800-1200 AD, namely Antiquity, Patristic, and Scholastic writing. The texts will be annotated both on text (genre, place and time of composition etc.) and token level (part of speech, lemmatization etc.).

In the project's first year, works to be included in the corpus were selected, based on a thorough analysis of the Index scriptorum, an open database of ca. 10000 references of the NGML sources [https://www.glossaria.eu/scriptores]. Whenever possible, they were retrieved in the TXT or XML format from existing repositories. In more than half the cases, however, scans were only available. Although the processing was automatized to some extent, it also faced a number of obstacles since, for example, image retrieval workflow would greatly vary depending on the source of the file. The images thus acquired were next optimised, OCR-ed, and eventually batch proofread. Retrieved text will be now structured in order to clearly separate Latin text and paratext (editorial content, page numbers, footnotes etc.).

References

Bon, B. 2015. 'Histoire et perspectives du *Novum Glossarium Mediae Latinitatis*. Un « Nouveau Du Cange'. *Archivum Latinitatis Medii Aevi*. 73: 297–308.

The tradition of the *Vocabularia infirmitatum* in the Middle Low German language area

Chiara Benati

University of Genova (Italy)

chibena@tin.it

Pharmaco-botanical and medical specialized bilingual glossaries are quite well-represented in the Middle Low German language area. The largest part of these glossaries focuses on medicinal plants drugs and can be ascribed to either the tradition of the so-called *Artemisia-Dictionary*, an alphabeticalLatin-German glossary of officinal plants going back to the thirteenth century (Schnell 2017:17), or to *Synonima apotecariorum*, an alphabetical pharmaco-botanical dictionary synoptically providing Latin, Greek, Arabic, Hebrew and German names for over 500 vegetal and mineral medical substances (Schnell 1995: col. 557-559).

On the other hand, only three wordlists can be described as specifically or mainly focusing on the names of the diseases. These are Greifswald, UB, Ms. 683, fols 118v-120r (*Vocabularius infirmitatum*), the pathological section of the topical glossary in Paderborn, Erzbischöfliche Bibliothek, Theodoriana Ba. 16, fol. 204r, and Copenhagen, KB, GKS 1663 4to, fols 179r-204r whoseheadwords mainly belong to the semantic field of pathology (Benati 2016: 55). All these three glossaries are *codices unici*, which suggests they never enjoyed particular popularity or wide circulation.

In this paper, these three pathological glossaries, which –apart from Copenhagen GKS 1663 4to– have never been studied and are still inedited, will be taken into consideration within their manuscriptcontext and contrasted in order to ascertain if they can be considered completely independent from one another or if, on the other hand, they are somehow genetically related or rely on analogous sources, as some corresponding batches could indicate.

References

Benati, C. 2016. 'A medical dictionary for personal usage: the Latin-Low German glossary on fol. 179r-204r of Copenhagen, Royal Library, GKS 1663 4to'. In M.V. Domínguez-Rodríguez, A. Rodríguez Álvarez and G. Rodríguez Herrera (eds.), *Words across History: Advances in Historical Lexicography and Lexicology*. Las Palmas de Gran Canaria: Servicio de Publicaciones y Difusión Científica. 54-66.

Schnell, B. 1995. 'Synonima apotecariorum'. In W. Stammler and von K. Langosch (eds.), *Die deutsche Literatur des Mittelalters. Verfasserlexikon*. Berlin: De Gruyter. 557-559.

Schnell, B. 2017. 'Die deutsche Medizinliteratur des Mittelalters. Vorüberlegungen zu einer historischen Darstellung'. In J. Riecke (ed.), *Sprachgeschichte und Medizingeschichte. Texte–Termini–Interpretationen*. Berlin: de Gruyter. 3-20.

The ghostword OCS *sęštь*/*sęštъ* 'wise' or:
The long shadow of Franz Xaver (Ritter von) Miklosich (Miklošič)

Harald Bichlmeier

Sächsische Akademie der Wissenschaften zu Leipzig (Germany)

harald.bichlmeier@uni-jena.de

In newer Indo-European literature we find several times the adjective OCS *sęštъ* (*LIV²* 533,klug, weise'; Meiser 2003: 121; *SEJL* 552f. 'mądry'; *SEJL²* (2018) 1264 'wrażliwy, mądry'; *SEJL²* (2019) 1334 'wrażliwy, mądry') bzw. *sęštъ* (*EDLIL* 554; *EDSIL* 450 [own lemma]; *EDPG* 437 'sensible, wise'). It is usually reconstructed as PIE **sent-i̯o-* and said to belong together with Lat. *sentīre* 'to think, to feel, to mean' etc. On the other hand, there are several books etc., where we might expect that word, but where we do not find it (*NIL* [PIE. **sent-* nottreated]; *EDAIL* 636f.; Kluge/Seebold 2011: 850; *EDBIL* 399; *ALEW* 916f.). In (etymological) dictionaries of OCS the word is also missing (Sadnik/Aitzetmüller 1955 [1989]; Kurz/Hauptová1966–1997 [2006]; *ESJS*; Cejtlin et al. 1994).

After a closer look at the word, it becomes clear that it is a clssical ghostword: It came into being by a misreading respectively an emendation. The presumed lexeme appears as a *hapaxlegomenon* in the Codex Suprasliensis. Instead of the real form OCS *sǫštęję* of the codexMiklosich gives it in his first edition of the respective text of 1845 and in the first full edition of the Codex Suprasliensis of 1851 as *sęštęję* 'prudens', because we find at the same place of the Greek original the adjective *φρονίμου*. But in fact the adjective *φρονίμου* of the Greek version has simply not been translated. Later one this ghostword has made a rather impressivecareer: After it had been given in Miklosich's full edition of the Codex Suprasliensis of 1851, he gives it in about half a dozen grammars and dictionaries. From there it was taken over into Indo-Europeanist literature at the the end of the 19th sentury and in the 20th century goes on to be quoted among others in Walde-Pokorny (1927-34) und Pokorny (IEW) and from there goes on into the etymological dictionaries of the 21st century.

The word, however, does not appear in slavicists' literature, as the new edition of the Codex Suprasliensis by Sever'janov 1904 has the correct reading at that place. Thus, the word simply did not exist anymore in slavistic studies.

Indo-Europeanists simply did not get that news. The reason appears to be quite simple: *Slavicanon leguntur*.

References

Cejtlin, R.M., Večerka, R., and E. Blagova (Eds.). 1994. *Staroslavjanskij slovar' (po rukopisjam X-XI vekov)*. Moskva: Russkij jazyk.

EDAIL. Martirosyan, H.K. 2010. *Etymological Dictionary of the Armenian Inherited Lexicon*. (Leiden Indo-European Etymological Dictionary Series 8). Leiden-Boston: Brill.

EDBIL. Derksen, R. 2015. *Etymological Dictionary of the Inherited Baltic Lexicon*. (Leiden Indo- European Etymological Dictionary Series 13). Leiden-Boston: Brill.

EDLIL. de Vaan, M. 2008. *Etymological Dictionary of Latin and the other Italic Languages*. (Leiden Indo-European Etymological Dictionary Series 7). Leiden-Boston: Brill.

EDPG. Kroonen, G. 2013. *Etymological Dictionary of Proto-Germanic*. (Leiden Indo-European Etymological Dictionary Series 11). Leiden-Boston: Brill.

EDSIL. Derksen, R. 2008. *Etymological Dictionary of the Slavic Inherited Lexicon*. (Leiden Indo-European Etymological Dictionary Series 4). Leiden-Boston: Brill.

ESJS. Havlová, E. *et al. Etymologický slovník jazyka staroslověnského*. Fasz. 1–14. Praha: Academia 1989–2008, Fasz. 15–18, Brno: Tribun EU. 2010-2016.

Fick, A. 1870. *Vergleichendes Wörterbuch der indogermanischen Sprachen. Ein sprachgeschichtlicher Versuch. Erste Abtheilung*. 2. umgearbeitete Auflage des 'Wörterbuchs der indogerm. Grundsprache' Göttingen 1868. Göttingen: Vandenhoeck & Ruprecht's Verlag.

Fick, A. 1874. *Vergleichendes Wörterbuch der indogermanischen Sprachen sprachgeschichtlichangeordnet. Dritter Band enthaltend den Wortschatz der Germanischen Spracheinheit mit einemBegleitwort von Dr. A. Bezzenberger*. 3. umgearbeitete Auflage. Göttingen: Vandenhoeck & Ruprecht's Verlag.

IEW. Pokorny, J. 1959. *Indogermanisches etymologisches Wörterbuch*. I. Band. München, Bern:Francke.

Kluge, F. 2011. *Etymologisches Wörterbuch der deutschen Sprache*. Bearbeitet von Elmar Seebold. 25., durchges. u. erw. Aufl. Berlin, Boston: de Gruyter.

Kurz/Hauptová 1966–97 [2006]. Kurz, J. and Z. Hauptová (Eds.), *Slovník jazyka staroslověnského / Lexicon linguae palaeoslovenicae*. Praha: Academia. 1966-97. [Reprint: *Slovar' staroslavjanskogo jazyka*, Sankt Peterburg: Izdatel'stvo Sanktpeterburgskogo Universiteta 2006]

LatEW. Walde, A. and J.B. Hoffmann. 1930-1954. *Lateinisches Etymologisches Wörterbuch* (2 Bde.). Heidelberg: Winter.

LIV². Rix, H. *et al*. 2001. *Lexikon der indogermanischen Verben. Die Wurzeln und ihre Primärstammbildungen*. Zweite, verb. u. erw. Aufl. Wiesbaden: Reichert.

Miklosich, F. 1850. *Lexicon linguae slovenicae veteris dialecti*. Vindobonae apud Guilelmum Braumüller.

Miklosich, F.von. 1862-65. *Lexicon palaeoslovenico-graeco-latinum. Emendatum, auctum*. Wien:Wilhelm Braumüller. [2. Neudruck der Ausgabe 1862–65. Aalen: Scientia Verlag 1977]

Miklosich, F. 1886. *Etymologisches Wörterbuch der slavischen Sprachen*. Wien: Wilhelm Braumüller.

NIL. Wodtko, D.S., Irslinger, B. and C. Schneider. 2008. *Nomina im Indogermanischen Lexikon*. Heidelberg: Winter.

Sadnik, L. and R. Aitzetmüller. 1955. (1989). *Handwörterbuch zu den altkirchenslawischen Texten*. Heidelberg: Winter.

SEJL. Smoczyński, W. 2007. *Słownik etymologiczny języka litewskiego*. Wilno: WydawnictwoUniwersytetu Wileńskiego.

SEJL² 2018. Smoczyński, W. 2018. *Słownik etymologiczny języka litewskiego*. Wydanie drugie, poprawione i znacznie rozszerzone. o.O. Version vom 20.1.2018 (www.rromanes.org/pub/alii/Smoczyński W. Słownik etymologiczny języka litewskiego.pdf; abgerufen im März 2018).

SEJL² 2019. Smoczyński, Wojciech. 2019. *Słownik etymologiczny języka litewskiego*. Wydanie drugie, poprawione i znacznie rozszerzone. Współpraca redakcyjna: Michaił Osłon. o.O. Version vom 9.3.2019 (www.rromanes.org/pub/alii/Smoczyński W. Słownik etymologiczny języka litewskiego.pdf; abgerufen im März 2019).

Walde, A. 1910. *Lateinisches Etymologisches Wörterbuch*. Heidelberg: Winter.

Walde, A. and J. Pokorny. 1927.1932. *Vergleichendes Wörterbuch der indogermanischen Sprachen*. 3 Bde. Berlin – Leipzig: Walter de Gruyter & Co.

Plagiarist's fatigue: The case of Ann Fisher's plagiarism of John Entick's *New Spelling Dictionary*

Alexander Bocast

Berkeley Bridge Press (United States of America)

alexander.bocast@gmail.com

This study is a follow-on to a paper by Rodríguez-Álvarez and Rodríguez-Gil (2006), which asked whether Ann Fisher (1773) plagiarized John Entick's *Dictionary* (1765). While their dictionaries exhibit only mild *alphabet fatigue*, we found three additional effects of *alphabet fatigue* in Ann Fisher's *Dictionary*.

We drew our sample from every letter in Fisher's dictionary, avoiding where possible overlap with the Rodríguez sample. We compared the Rodríguez sample to corresponding entries in Entick (1767) and then added those re-examined entries to our sample. Our combined sample contains 1991 words.

We investigated three questions: 1) How many entries in each letter did Fisher take from Entick?; 2) Of those entries, how many did she copy exactly?; 3) Of the words in those entries she took from Entick, how many words did Fisher retain?

We expected to find that Fisher plagiarized Entick more extensively than the Rodríguez paper (2006) suggested. However, we were surprised that these proportions all seem to increase from A to Z in Fisher's dictionary: the proportion of entries from Entick, the proportion of exact copies, and the proportion of words retained. Simple linear regressions confirm these impressions.

Fisher attempted to obscure her plagiarism of Entick by introducing entries from other dictionaries and by reworking Entick's entries. However, while she worked her way through the alphabet, our data suggest that Fisher progressively drew less and less from other sources; she directly copied more of Entick's entries; and she reworded fewer of his other entries.

Thus, a plagiarist may also suffer from *alphabet fatigue*. As Fisher worked her way through the alphabet, she diverged less and less from Entick; at the end, at Z, they are nearly identical. In other words, as she grew more and more fatigued, Fisher put less and less effort into obfuscating her plagiarism.

References

Entick, J. 1765. *The New Spelling Dictionary*. London: Edward & Charles Dilly.

Entick, J. 1767. *The New Spelling Dictionary*. London: Edward & Charles Dilly.

Fisher, A. 1773. *An Accurate New Spelling Dictionary, and Expositor of the English Language*. 2nd ed. London: Ann Fisher.

Rodríguez-Álvarez, A. and M.E. Rodríguez-Gil. 2006. 'John Entick's and Ann Fisher's Dictionaries: An Eighteenth-Century Case of (Cons)Piracy?' *International Journal of Lexicography,* 19. 287-319.

The specificities of encoding encyclopedias: Towards a new standard?

Alice Brenon (École Normale Supérieure de Lyon, France)
Denis Vigier (Université Lyon 2, France)
alice.brenon@ens-lyon.fr, denis.vigier@ens-lyon.fr

The genre of dictionaries and encyclopedias is a well-known topic in digital humanities (Roe *et al.* 2016; Williams 2017; Vigier *et al.* 2019…). Over the past years, several projects among which *ARTFL, BASNUM, COLLEX- LGE* and *NENUFAR* have focused on them, especially in diachrony. A major standard, XML-TEI, has proposed a module containing specific elements and attributes to encode dictionaries. This module has however shown some limitations when used to encode «La Grande Encyclopédie» (LGE), a late- nineteenth century encyclopedia written by a team of specialists including Marcellin Berthelot (Jacquet-Pfau (2015)).

This paper describes those limitations, proposes an alternative encoding and discusses the need for a modification of *XML-TEI* to describe encyclopedias. We will present manually sampled articles from *LGE* of varied length, polysemic or not, with a flat or nested structure and show the encoding issues each raised. The size of our corpus makes encoding manually impossible in practice, so we tried to map their structure to an encoding that is regular enough to be used as output for an automated tool. The elements provided in the TEI module for dictionaries raised some difficulties which we will discuss. We will finally comment on three possible encoding scenarios and explain the reasons we used to choose between them.

Our study identified specific structures that are hard if not impossible to encode with XML-TEI and proposed an encoding scheme that is XML-TEI compliant but outside of the dictionaries TEI module. This questions the relevance of this module for encyclopedias and the need for either an upgrade or the design of a new one. Confronting the various possibilities with other encyclopedic texts, in particular ones closer to dictionaries like the «Grand Dictionnaire Universel» by Larousse would certainly prove useful.

References

ARTFL. [http://artfl-project.uchicago.edu/]

BASNUM. [https://basnage.hypotheses.org/]

COLLEX- LGE [https://www.collexpersee.eu/projet/disco-lge/]

Jacquet-Pfau, C. 2015. 'Élaboration et destinée d'une encyclopédie à la fin du XIXe siècle: les trente-et-un volumes de La Grande Encyclopédie.Inventaire raisonné des sciences, des lettres et des arts par une Société de savants et de gens de lettres. *(1885-1902)'. ÉLA, Études de linguistique appliquée*, 177. 85-100.

NENUFAR. [https://cahier.hypotheses.org/nenufar]

Roe, G., Gladstone, C. and R. Morrissey. 2016. 'Discourses and Disciplines in the Enlightenment: Topic Modeling the French Encyclopédie'. *Frontiers in Digital Humanities,* 2.

Vigier, D., Moncla, L., Joliveau, T., Mcdonough K. and A. Brenon. 2019. 'GeoDISCO: Encyclopedic Geographical Discourse in France from the En- lightenment to Wikipedia'. *GIR'19, 13th International Workshop on Geographic Information Retrieval.* Lyon, France.

Williams, G. 2017. 'Mapping a dictionary: Using Atlas ti and XML to analyse a late XVIIth century dictionary'. *RiCOGNIZIONI. Rivista di Lingue e Letterature straniere e Culture moderne*, 4 (7). 161-180.

"*I was put to bed of a baby girl*": The language of giving birth in English from the 16th century onwards

Nuria Calvo Cortés

Universidad Complutense de Madrid (Spain)

ncalvo@filol.ucm.es

The experience of childbirth has changed through history. Although many women continued giving birth at home, lying-in hospitals became popularplaces for children to be born in the 18th century in England (Cody 2004; Wilson 2018) and in the 19th century in America (Martell 2000; Wertz andWertz 1989). This resulted in a higher intervention in the process of child delivery of not only female midwives, but also male midwives and physicians.

The present study aims at analysing English 'giving birth' expressions fromthe 16th century. These were first extracted from the *Historical Thesaurus of English* and checked in the *Oxford English Dictionary Online*. They were next searched in five British and American corpora and collections of books online (*Early English Books Online, the Corpus of Late Modern English Texts Extended Version, the British National Corpus, the Corpus of Historical American English* and *the Corpus of Contemporary American English*). Finally, the expressions were examined in a relatively small corpus of petitions signed by 18th and 19th century women who had just had a baby.

The results show that when studying lexis diachronically the availability of digital materials today contributes to a more accurate description of the evolution of the terms analysed. They also indicate that small specialised corpora can be more useful than large corpora when studying very specificterms. Finally, the conclusion suggests that corpus-based studies from different periods of the language can enrich online dictionaries, which are more easily updated nowadays and can incorporate new meanings and references without going through a lengthy re-editing process.

References

Cody, L.F. 2004. 'Living and dying in Georgian London's lying-inhospitals'. *Bulletin of the History of Medicine*, 78 (2). 309-348.

Martell, L.K. 2000. 'The hospital and the postpartum experience: Ahistorical analysis'. *Journal of Obstetric, Gynecologic, & Neonatal Nursing*, 29 (1). 65-72.

The Historical Thesaurus of English. [https://ht.ac.uk]

The Oxford English Dictionary Online. [www.oed.com].

Wertz, R.W. and D.C. Wertz. 1989. *Lying-in: A history of childbirth inAmerica*. New Haven: Yale University Press.

Wilson, A. 2018. *The making of man-midwifery: childbirth in England, 1660-1770*. London: Routledge.

Semiotic approaches to the semantic web: Toward a *Thesaurus linguae Latinae* 2.0

Massimo Cè

The Bavarian Academy of Sciences and Humanities (Germany)

ce@thesaurus.badw.de

This paper reimagines the world's most comprehensive dictionary of ancient Latin for the Semantic Web. Despite the origins of the *Thesaurus linguae Latinae* (TLL) in the late 19th century, the lexicon's analytical construction makes it a suitable candidate for machine-readable encoding. Through the first-ever interpretation of the lexicon's hierarchical "organization" (*Gliederung*) in semiotic terms, I demonstrate how the dictionary can be integrated with the OntoLex-Lemon Model, a key mechanism to represent lexical relationships in the Semantic Web. Building on recent attempts to apply linguistic theory to Latin lexicography, my approach contributes to the dynamic conversation about the digital transformation of the TLL.

The typical structure of a TLL article is analytical in three principal ways. First, the "head" (*Kopf*) of the lexicographical article systematically assembles general information about the lemma. The key semiotic distinction here is between *langue* and *parole*, i.e., the basic form of a word (e.g. *regina* "queen") and its actual graphic and phonetic realizations, including nonstandard variants (e.g. *recina*, ρηγινα). Second, the body of the article begins with the word's primary definition. Here the signifier (e.g. *regina*) is paired with one or more signifieds, including definitions and synonyms (e.g. *ea quae regnat* "a female being who rules," βασίλισσα "queen"). Third, the article's individual subgroupings link specific occurrences of the word to concrete usages (e.g. *uxor imperatoris* "emperor's wife"), again associating instances of *parole* with categories of *langue*. Consequently, the three components of a TLL article can be mapped onto the elements of the OntoLex-Lemon model: the headword corresponds to the model's "Lexical Entry"; the lemma's primary definition is reflected in the "Lexical Concept"; and the semantic subgroupings are assigned to discrete "Lexical Senses." The result is a radically new conception of the TLL, based on philological, semiotic, and digital approaches.

The cultural word "shawl" and its derivatives in the languages of the Near East and Europe

Alexandra Cherkunova, Elena Molchanova,

Institute of Linguistics, Russian Academy of Sciences (Russia)

cherkunova@iling-ran.ru, e_molchanova@mail.ru ,

We would like to start our report by mentioning the famous Russian artist of the "Silver Age" Konstantin Korovin (1861-1939).

In the picture (pic.1), our attention is drawn by the shawls worn by the girls, in Spanish *chal*. Having been borrowed from Persian, this word can be traced in many European languages; however, it had initially originated in India. The etymology of Persian *šāl* can be found in [2] the word is adopted from Hindi; detailed references are given to Sanskrit, Pali, Iranian, and other European languages. In Persian, the word is presented as an independent lexeme, as well as in its derivatives and compounds.

One of the derivatives (Persian *šāl-me* "Indian turban") testifies to one more, in addition to the Indian, source of influence: the suffix -me from the Türkic * -ma. Dehkhoda dictionary contains an article with the headword *čālme* and the label "Turkism". One of the meanings: "a kind of turban worn by the Hindus". How did the Indian word get into European languages? One can assume the mutual influence of various languages over the course of the Great Silk Road for many centuries.

The Russian historian E.E. Kuzmina wrote:

> "The belt of Eurasian steppes stretches 7,500 km from the Danube to the Great Wall of China. It used to be the zone through which goods, innovative technologies, new religious ideas and images of art, and, finally, individual ethnic groups that determined the process of ethnogenesis of various peoples, including Indo-European, were spreading over the millennia. According to A. M. Petrov (1995: 46), the Great Silk Road is "not a road at all [...] It is a huge historical and cultural space, capable of moving in time, through which in antiquity and the Middle Ages land international communication went from the extreme limits of Asia to the countries of the West".

One might think that the Spanish lexeme *chal* has also experienced numerous relocations, including the Great Silk Road.

References

Kuzmina E.E. 2010. *The prehistory of the Silk Road*. 240. (in Russian)

Dehkhoda – Dehkhoda A. Loghatnāme (Encyclopedic Dictionary) (ed.) M. Moʻin and J. Shahidi, in 14 volumes. [https://dehkhoda.ut.ac.ir/fa/dictionary] (in Persian)

Hassandoust, M. 1393-2014. *Farhang-e riše-šenāxti-ye zabān-e fārsi*. Tehrān: Farhangestān-e zabān va adab-e fārsi, (in Persian). Vol 1-5.

Picture 1

Picture 2

Postcolonial perspectives on populations in the early American Spanish lexicography: A sociolinguistic approach to population terms in Ricardo Palma's *Papeletas lexicográficas* (1903)

Piero Renato Costa León

Università degli Studi di Verona (Italy)

pierorenato.costaleon@univr.it

The purpose of this paper is to detect and classify the main social features the Peruvian intellectual Ricardo Palma (1833-1919) attributed to the usage of demographic and geographic terms such as *quechuista*, *incásico*, *amazónico*, *bolivianizado*, *americanizable*, *europeizador*, *polonizador*, *gringo*, *londonense*, etc. in his *Papeletas lexicográficas* (1903). This dictionary was elaborated as part of the first generation of American Spanish dictionaries of *provincialismos*, whose aim, by the end of the 19th century and the beginning of the 20th century, was to register both lexical forms and meanings common in the Americas but poorly or not included (by that moment) in the traditional and influential *Real Academia Española*'s (RAE) dictionaries.

Considering thus this historic value of Palma's lexicographic work, this paper aims to provide an insight into the way in which the use of the aforementioned set of terms was reported in such an early, non-normative Spanish language dictionary, elaborated from anon-entirely hegemonic centre of linguistic and political power. More concretely, the interest in these relational entries deriving from geographical points of reference and/or ethnic groups leads us to analyse here the semantic, sociolinguistic and even some ideological contents Ricardo Palma offers in his definitions and, consequently, to propose a number of trends in the sociolinguistic distribution of the former, according toPalma's social and linguistic views.

Therefore, and thanks to a corpus built upon those demography-related articles in the dictionary, this paper aims to classify the information in the entries according to criteria such as Palma's remarks on geographic predominance and origin, different American isoglosses, endogenous/exogenous and current/archaic usage, pejorative values and metaphorical innovations. Consequently, with the classification of the data lexicographically available, it is intended hereby to emphasise the fruitful contributions the dictionaries of *provincialismos* can offer to the reconstruction of (part of) the history of Spanish lexicon.

References

Palma, R. 1903. *Papeletas lexicográficas*. Lima: Imprenta La Industria.

Derivational morphology and dictionaries of English: An historical perspective

Janet DeCesaris

Universitat Pompeu Fabra (Spain)

janet.decesaris@upf.edu

This paper explores the relationship between word formation and dictionary representation in general purpose monolingual dictionaries of English from an historical perspective, concentrating on dictionaries published between the mid-19th century and today.

Lexicographers differ in terms of the knowledge of morphology they assume on the partof their target audience, and consequently have differing policies toward derived words. For example, two American dictionaries from the mid-20th century, the *American College Dictionary* (1947) and *Webster's New World Dictionary* (1953), expressed quite different views on this, with the latter proclaiming in its front matter: "This dictionary contains no run-on forms" (1953: viii) while the former states "Words which are simple derivatives of the main entry, and which present no meaning problem, are run on after the etymology" (1947: xl). Views on derived words also affect the order of words defined because the nesting of derived words under the base headword (adopted by the *Chambers Dictionary*, for instance) has consequences for presenting words in (non)alphabetical order. The lemmatization of words that are used in more than one category as the result of conversion (e.g., limit[noun], limit[verb]; deliverable[adjective], deliverable[noun]) is yet another issue on which dictionaries differ: older works, such as the Webster-Mahn edition of Webster's dictionary and *The Century Dictionary*, prefer to present separate headwords, whereas some current dictionaries of English prefer to present all categories under a single headword.

General purpose dictionaries of English have long afforded affixes, combining forms, andother bound morphemes headword status, but the type of information given for these elements differs greatly from dictionary to dictionary. Some dictionaries classify all word-forming elements as affixes; others discriminate more. Some treat etymologically different sources of a single affixal form as different senses of a single affix, whereas others provide a more detailed −and often etymologically based− analysis.

In sum, we aim to present the varying approaches taken by dictionaries to derived words and word-forming elements, to determine which have left their mark on the lexicographic tradition of English.

From *dama* to *rapariga*, from *cavalheiro* to *rapaz*: Lexicographic (im)partiality

Laura Do Carmo (Fundação Casa de Ruy Barbosa, Brazil)

AlinaVillalva (University of Lisboa, Portugal)

Esperança Cardeira (University of Lisboa, Portugal)

laurafcarmo@gmail.com, alinavillalva@campus.ul.pt, ecardeira@hotmail.com

The lexicographic record of words related to man and woman in Portuguese dictionaries, over the course of two centuries, reveals that the portrait of men is rather positive, whereas women are generally nastily depicted. Meanings, concepts, phrases, examples of use, definitions, they all show stereotypes that remained as a lexicographic memory and a sociological profile. This biased portrait of males and females is endorsed by the dictionary in a range of words that were initially intended to differentiate individuals in terms of age, marital status and social condition. These names include, among others: *cavalheiro, dono, dom, dama, dona, donzela, donzel, senhor, senhora, moça, moço, rapaz, rapariga, mancebo*. Their English equivalents are not easy to establish, but they will somehow map into words like, for instance, 'gentleman', 'dame', 'maiden', 'lord', 'lady', 'young person', 'girl', 'boy'. The transposition is not as straightforward as what might be expected, since these words, especially those that concern female individuals, carry meanings that refer to the private domain of affections and sexual roles.

Our research is about looking back in time through the mirror glass of dictionaries. We will analyse the relevant entries in a handful of critically selected Portuguese dictionaries, since the early 18th century (Rafael Bluteau 1712-1728, *Vocabulário português e latino*) to the early 20th century (Cândido de Figueiredo 1913, *Novo dicionário da língua portuguesa*, 2nd edition), and, in between, we will also consider two editions of Antônio de Morais Silva's *Dicionário da língua portuguesa* (1789 and 1890), as well as Caldas Aulete's (1881) *Dicionário contemporâneo da língua portuguesa*. Continuity and changes, additions or omissions, all the contrasts will reveal the (un)witting partiality in the lexicographic making.

References

Aulete, C. 1881. *Diccionario contemporaneo da lingua portugueza*. Direcção confiada a Antonio Lopes dos Santos Valente. Lisboa: Imprensa Nacional. 2 vols.

Bluteau, R. 1712-1728. *Vocabulario portugues e latino*. Coimbra: Collegio das Artes da Companhia de Jesu; Lisboa Occidental: Off. de Pascoal da Sylva. 10 vols.

Figueiredo, C.d. 1913. *Novo diccionario da lingua portuguesa*. Nova ed. corr. e copiosamente ampl. Lisboa: A.M. Teixeira. 2 vols.

Silva, AdM. 1789. *Diccionario da lingua portugueza*. Lisboa: Off. S.T. Ferreira. 2 vols.

Silva, AdM. 1890-1891. *Diccionario da lingua portugueza*. 8. ed. Lisboa: Empreza Litteraria Fluminense. 2 vols.

Conceptually and alphabetically organized lexicon: Correlation of meaning and form in Old English

Nataliia Drozhashchikh

Tyumen State University (Russia)

ndro2012@gmail.com

The paper is dedicated to the topic of organizing conceptual information in Old English lexicon with relation to its alphabetical structure. The topic is connected with the issue of form-meaning systematicity (Monaghan *et al.* 2014; Blasi *et al.* 2016; Pimentel *et al.* 2019) and develops the framework of the global organization of the English lexicon (Ogura and Wang 2012). The hypothesis of the study is that semantic categories of words starting with the certain consonantal grapheme (sound) form a small world network within a definite dictionary article (e.g. the words with /s/ onset) and outside it. The aim of the study is to compute the concepts of alphabetically organized words and show the interconnectedness/independence of their meanings. The data for the study was the dataset of /w/ monomorphemic words extracted from *Old English Corpora* (Pintzuk and Plug 2000; Taylor *et al.* 2003). All /w/ words from the corpora were collected and lemmatized; semantic categories of most frequent words were identified and aligned with *A Thesaurus of Old English* (TOE). To show significantly relevant dependencies between the meanings of alphabetically organized words, we employed such statistical methods as chi-square test, correlation coefficient, and some others. The study was carried out using Python libraries. The results of the computation were listed in the tables and plotted in Gephi. /W/ words with higher frequency refer to such categories as 'existence', 'life and death', 'mental faculties', 'social interaction' out of 18 general categories in TOE. The study can be continued examining words with other consonantal onsets to show the scope of conceptual capacity and constraints on form-meaning connections. The results of the study are relevant for future research on semantic changes in historical lexicology since it traces dependencies in vast lexical data and reveals conceptual transformations in different historical periods.

References

Blasi, D.E., Wichmann, S., Hammarström, H., Stadler, P.F. and M.H. Christiansen. 2016. 'Sound-meaning Association Biases Evidenced across Thousands of Languages. *Proceedings of the National Academy of Sciences*, 113 (39). 10818-10823.

Monaghan, P., Shillcock, R.C., Christiansen, M.H., and S. Kirby. 2014. 'How Arbitrary is Language?' *Philosophical Transactions of the Royal Society B-Biological Sciences*, 369 (1651).

Ogura, M. and W.S-Y Wang. 2012. 'The global Organization of the English Lexicon and its Evolution'. In H. Sauer and G. Waxenberger (eds.), *English Historical Linguistics.* vol. II: Words, Texts and Genres. Amsterdam/Philadelphia: John Benjamins. 65-83.

Pimentel, T., McCarthy, A.D., Blasi, D., Roark, B. and R. Cotterell, 2019. 'Meaning to Form: Measuring Systematicity as Information'. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence. 1751-1764.

Pintzuk, S. and L. Plug. 2000. *The York-Helsinki Parsed Corpus of Old English Poetry*.

Taylor, A., Warner, A., Pintzuk, S. and F. Beths. 2003. *The York-Toronto-Helsinki Parsed Corpus of Old English Prose.*

TOE. *A Thesaurus of Old English.* Glasgow. University of Glasgow.

Pidgin dictionaries: The history of Tok Pisin lexicography as colonial history

Stefan Engelberg

The Leibniz Institute for the German Language (Germany)

engelberg@ids-mannheim.de

Tok Pisin is a pidgin-creole language spoken in eastern New Guinea, the Bismarck Archipelago and the northern Solomon Islands. It has its roots in English-based pidgin varieties in the southwestern Pacific and formed in the late 19th century under German colonial rule. It is nowadays one of the official languages of Papua New Guinea.

As with many pidgin languages, the bad reputation Tok Pisin had among Europeans stood in sharp contrast to its indispensability for colonial economy as well as for the nation-building process in post-colonial times. This explains the fairly extensive lexicographic history of Tok Pisin with several dozens – partly unpublished – dictionaries and numerous smaller vocabularies.

The talk will present the history of Tok Pisin lexicography as a part of colonial history. The particular needs and purposes of different interest groups in colonial times shaped the dictionaries they produced: (i) European scientists (e. g. anthropologists) who relied on Tok Pisin as a lingua franca in an area where several hundred indigenous languages are spoken, (ii) German catholic missionaries who – confronted with this language diversity – turned to Tok Pisin for missionary work, and, finally, (iii) native speakers of English who compiled dictionaries when it became clear that Tok Pisin would not simply disappear under Australian rule (after WWI), in particular when in WW II efficient communication between military and civilians became necessary.

After missionary lexicography produced an excellent printed dictionary in 1971 and Papua New Guinea became independent in 1975, no larger lexicographic efforts were made for some decades. It was only in the current century that Tok Pisin lexicography has been showing a revival, partly in form of independent internet lexicography, partly as printed dictionaries from publishing houses (Oxford University Press).

Imitative words in Old Norse: A study of the *Altnordisches etymologisches Wörterbuch* by Jan de Vries.

Maria Flaksman

St. Petersburg State Electrotechnical University/University of Jena (Russia)

mariaflax@gmail.com

Imitative (onomatopoeic, mimetic) words are present in the vocabularies of modern languages from various families (Hinton et al. 1994, Voeltz et al. 2001) and may even be a language universal (Voronin 2006: 179). Therefore, there is no reason to suggest that they were absent at the earlier stages of the languages' development. Old Norse is one of the best recorded languages of the IndoEuropean family as it survives in numerous texts – Edda, sagas, skaldic poems. Therefore, one may expect to find traces of old imitative words preserved in these texts, which might give an insight into the sphere of sound-denotation of the ancient world.

Indeed, we find, for example: [Egil's saga] Egill segir, at ekki var ort. 'Hefir hér setit svala ein við glugginn ok klakat í alla nótt…'. — *Egil said that nothing was done. 'Here,' said he, 'has sate a swallow by the window and twittered all night ...'*; [Snorri's Edda] …þá hljóp ór skógi nǫkkurum merr ok at hestinum ok hrein við. — *… a mare lept from a wood towards the horse and whinnied.*

The research is conducted on the *material* of the *Altnordisches etymologisches Wörterbuch* by Jan de Vries, textual examples drawn from literary sources (*Egil's saga* and Snorri's *Edda*). The research *methods* are historical-comparative analysis, etymological analysis and phonosemantic analysis (Voronin 2006: 91). The ***questions*** to be discussed are: 1) how many ON imitative words are registered in the dictionary; 2) to what extent does the ON imitative vocabulary differ from the Modern Icelandic one; 3) were ON imitated words affected by the later regular sound changes; 4) were there any noticeable semantic shifts exhibited by the ON imitative words; 5) were there any borrowed imitative words in Old Norse and 6) what are the general tendencies of expressivity loss in Old Norse onomatopoeic lexicon.

References

Hinton, L., Nichols J. and J.J. Ohala (eds.). 1994. *Sound Symbolism*. Cambridge: Cambridge University Press.

Icelandic-English Dictionary. Orðabók.is. [http://www.ordabok.is/index.asp]

Voeltz, E.F.K. and Ch. Kilian-Hatz (eds.). 2001. *Ideophones. Typological Studies in Language 44.* Amsterdam-Philadelphia: John Benjamins.

Voronin, S.V. 2006. *Osnovy Phonosemantiki*. (The Fundamentals of Phonosemantics). Moscow: Lenand.

Vries, J. de. 1962. *Altnordisches etymologisches Wörterbuch*. Leiden: Brill.

What new words were recorded in English medical dictionaries in the 1650s? or, discovering new lexicographical knowledge from existing sources using digital tools

Jonathan Furner

University of California (United States of America)

jfurner@ucla.edu

Thanks to hundreds of person-years of collaborative scholarship, finding the earliest known uses of English words occurring in historical dictionaries is a simple matter of looking up those words in the OED and noting the first citations. Determining which of the headwords in those dictionaries were new at the time of each dictionary's publication is, on the face of it, also relatively straightforward, since OED.com offers the opportunity to search on authors, titles, and dates of citations. This method suffers, however, because some of the OED's citation data is outdated and/or incorrect: as a consequence, some older words are included in the results, and some new words are missing.

The method described in the proposed paper involves the following steps: 1) use OED.com to search for entries in which headwords and explanations are supported by first citations to named dictionaries; 2) use Worldcat.org to identify digital copies of dictionaries, e.g., at Archive.org, ECCO, EEBO, Google Books, HathiTrust; 3) use those digital copies and LEME to verify citations and to search for headwords in earlier and later dictionaries; and 4) use EEBO-TCP to search for occurrences of headwords in earlier and later works through 1700.

The method has been tested for Culpeper's *Physical Dictionary* (1655) and Tomlinson's *Physical Dictionary* (1657; McConchie 2019), and full results will be presented in the proposed paper. This work has three outcomes: it verifies existing candidates for new words introduced by dictionaries; it identifies new candidates for earliest written word-uses previously thought to have been associated with dictionaries, and makes proposals of updates for entries in the OED; and it identifies new candidates for source texts used by dictionary authors, and thereby enhances our knowledge of the circumstances of production of those dictionaries.

References

McConchie, R. 2019. 'The Physical Dictionaries of 1655-1678,'. In *Discovery In Haste: English Medical Dictionaries and Lexicographers, 1547 to 1796*. Berlin: De Gruyter. 68-77.

The adjective *gesælig* in Old English prose: Towards a characterization of a lexical field of holiness in Old English

Ondřej Fúsik, Alena Novotná

Charles University (Czech Republic)

Ondrej.Fusik@ff.cuni.cz, novotna4alena@gmail.com

(s1)   *þu eart swiðe gesælig*
    thou art very happy/blessed
    (coboeth,Bo:34.94.16.1814)

The adjective *gesælig* is one of the representatives of the lexical field of Holiness in Old Englishin the perspective of Trier's (1931) theory and its later modifications (Schwyrter 1996), alongside with other adjectives and participles (*gebleotsod, gehalgod, halig* as defined in *TOE*online: 'Holiness: Characterized by holiness, blessed."). The present study aims to analyze the meaning of the word *gesælig* in the context of prose texts as they are presented in *York Helsinki Toronto Parsed Corpus of Old English* (by using the query [word="(ge)?sæli?g?(a|u|o|e)?(n|m|s)?t?" & tag="ADJ.*"]) through the identification of its conceptual fields, i.e. who or what can be *gesælig* or who/what can be granted this quality. A further focus of the study is to determine whether the meaning of *gesælig* shifts according tothe syntactical use of the adjective, i.e. whether it is used predicatively or attributively as was suggested in Fúsik (2018), and to identify other adjectives from the same lexical field which may potentially be preferred in the same syntactical role. The analysis also includes the comparative and superlative forms of the adjectives. Lastly, the presentation shall also include a comparative study of the meanings of *gesælig* and *halig* as similar research for the latter adjective has previously been conducted (Fúsik 2018). This section will furthermore attempt to answer the question of why or how the adjective *gesælig,* meaning "blessed", developed into the modern day meaning of the word *silly* as well as to trace any possible indications that the word is strictly used in Christian collocations.

Figure 1.

References

Fúsik, O. 2018. 'Old English prose adjectives meaning "holy": towards a characterization of a lexical field'. PhD dissertation. Univerzita Karlova, Filozofická fakulta, Ústav anglického jazyka a didaktiky. Vedoucí práce Čermák, Jan. Prague.

Schwyter. 1996. J. R. 'Old English Legal Language: the Lexical Field of Theft'. *North-Western European Language Evolution,* 15. Odense: Odense University Press.

Taylor, A., Warner, A., Pintzuk, S., and F. Beths. 2003. *The York-Toronto-Helsinki Parsed Corpus of Old English Prose* (YCOE). Department of Linguistics, University of York.

TOE. *A Thesaurus of OldEnglish*. Glasgow: University of Glasgow.

Trier, J. 1931. *Der deutsche wortschatz im sinnbezirk des verstandes: Von den anfängen bis zumbeginn des 13. Jahrhunderts*: Volume 31 of Germanische Bibliothek: Untersuchungen und Texte; Volume 1 of Der deutsche Wortschatz im Sinnbezirk des Verstandes: Die Geschichte eines Sprachlichen feldes, Jost Trier Volume 31 of Germanische Bibliothek. The University of Michigan: C. Winter.

Getting to terms with the sea: maritime terminology in the 1701 *Dictionnaire Universel*

Ioana Galleron (Université Sorbone Nouvelle, France)
Clarissa Stincone (Université Sorbone Nouvelle, France)
Geoffrey Williams (Université Grenoble Alpes, France)
ioana.galleron@sorbonne-nouvelle.fr, clarissastincone@live.it, geoffrey.williams@orange.fr

The *Dictionnaire universel* (Furetière 1690) constitutes an important stage in French historical lexicography by the attention it gives to linguistic domains, and particularly the terminology of arts and crafts. Subsequent editions (1701, 1702 and 1708), heavily revised byHenri Basnage de Beauval (Furetière 1701), amplify this movement and prove to be importantlandmarks towards the publication of modern encyclopedias.

In this paper, we will present a case study allowing to demonstrate this evolution. We use a digitized version of the DU 1701, built within the French ANR project BASNUM. Mapping the entries to the sources and to later works, including the *Encyclopédie'* Diderot and Alembert (Diderot et Le Rond d'Alembert 1751) and specialised dictionaries as the *Dictionnaire de Marine* by (Aubin 1702) and the *Dictionnaire de la Marine Française* of (Romme 1792) are part of the endeavour.

So far, we found that Basnage's DU adds 249 maritime terms to the 1040 found in the 1690 edition of Antoine Furetière. Additionally, there are 63 '*terme de mer*' as well as maritime related terms to be found in other areas as carpentry, mechanics, fortification and war. 'Marine' terms in the 1701 *Dictionnaire universel* are mostly either nouns (803) or verbs (877);by comparison, adjectives are only a handful (37).

An indepth study by (Fennis 1988) using the 1702 edition showed how Basnage both used the1690 and copied verbatim from the *Dictionnaire des termes propres de marine* by (Desroches 1687). Whilst Fennis (1988) looked at dictionary sources, our approach takes a linguistic and terminological angle. Thus, we are interested in how maritime entries can be managed within a terminological framework, and onomasiologically structured through a bottom-up ontologyso as to allow users to extract lexical units thematically.

References

Aubin, N. 1702. *Dictionnaire de Marine: contenant les termes de la navigation et de l'architecture navale.* Amsterdam: Pierre Brunel.

BasNum. [https://anr.fr/Projet-ANR-18-CE38-0003] [https://basnage.hypotheses.org]

Desroches, N. 1687. *Dictionnaire des termes propres de la marine*. Paris: Amable Auroy.

Diderot, D. and J. Le Rond d'Alembert. 1751. *Encyclopédie ou Dictionnaire raisonné dessciences, des arts et des métiers*. 17 vols. Paris.

Fennis, J. 1988. 'Les sources du vocabulaire maritime dans le Furetière de 1701'. *Travaux de linguistique et de philologie*, XXVI (1). 75-94.

Furetière, A. 1690. *Dictionnaire Universel, contenant généralement tous les mots françois tant vieux que modernes et les termes des sciences et des arts*. La Haye and Rotterdam.

Furetière, A. 1701. Dictionnaire universel, contenant généralement tous les mots françois tantvieux que modernes, & les termes des sciences et des arts. *Tome 1; par feu messire Antoine Furetière;2e édition revue, corrigée et augmentée par M. Basnage de Bauval*. La Haye et Rotterdam: Arnoud et Reinier Leers.

Romme, C. 1792. *Dictionnaire de la marine française*. La Rochelle: P. L. Chauvin.

Lability in Old English verbs: Chronological and textual distribution

Luisa García García, Esaúl Ruiz Narbona

University of Sevilla (Spain)

luisa@us.es, esaulrn@gmail.com

This paper researches the chronological and textual distribution of labile verbs in Old English. Present-Day English displays an unusual spread of labile coding, both numerically and in the types of verbs that allow it. Thus, for instance, some unergatives such as *walk*, *jump*, or *march* admit labile coding in English (*Andrew walked the dog*), which is typologically exceptional (Haspelmath 2016: 52). Incipient favouring of labilization as valency coding strategy can be detected already in Old English, which not only has comparatively more labile verbs than any other Germanic language (Hermodsson 1952), but which also, unlike these, adopts labile coding in a considerable share of formerly causative verb pairs (García García 2020). Old English labile verbs coming from Germanic causative oppositions are precisely the sample on which we base our study. The choice ensures that the verbs in question have undergone labilization in Old English, and not before, and allows us to map the expansion of this linguistic innovation in the earliest stages of the English language. With this, we hope to contribute to a better understanding of the origins and development of labilization in English. All in all, we address questions such as the increase of lability within Old English, possible textual restrictions and what can be inferred from these, semantic restrictions for labilization in Old English and whether they are similar to those in Present-day English as well as whether the tendency towards labilization can be considered native or not.

References:

García García, L. 2020. 'The Basic Valency Orientation of Old English and the Causative *ja*-formation: A Synchronic and Diachronic Approach'. *English Language and Linguistics* 24: 153-177

Haspelmath, M. 2016. 'Universals of Causative and Anticausative Verb Formation and the Spontaneity Scale'. *Lingua Posnaniensis* 58: 33-63.

Hermodsson, L. 1952. 'Reflexive und intransitive Verba im älteren Westgermanischen'. Unpubl. PhD dissertation, University of Uppsala.

Old Romanian Lexicography in the Digital Era: The eRomLex Project

Ana-Maria Gînsac, Alex Moruz, Mădălina Ungureanu

Alexandru Ioan Cuza University (Romania)

anamaria.gansac@gmail.com, alex.moruz@gmail.com, madandronic@gmail.com

Apart from a few Romanian glosses written on a Slavonic manuscript, the first Romanian lexicographical works date back from the seventeenth century, in the form of several bilingual lexicons of which one is Latin-Romanian, another Greek-Romanian and six Slavonic-Romanian. The last ones are based on the most important Slavonic dictionary of the time, Pamvo Berynda's Slavonic-Ruthenian Lexicon (Kyiv 1627). The six Slavonic-Romanian lexicons, all manuscripts, are not individual writings, but they are modified copies either of one of them, correlated with the Slavonic model, or of a prototype that has not been preserved. The existence of the common model (Lex.Ber.), demonstrated by the (partial) coincidences of lexical inventory and word definitions, led us to the idea of an electronic comparative edition of these lexicons, allowing for the simultaneous display of parallel entries from all the lexicons, as well as for complex interrogation, extraction and exhaustive comparison of the data, in a manner which a standard, printed edition, could never allow and providing thus the basis for quantitative and qualitative processing of the material. It is the edition proposed in the eRomLex project (www.scriptadacoromanica.ro).

In this paper, we will outline the main difficulties that have occurred so far in the process of digitizing the corpus within the eRomLex project. We will also take a closer look at the specific interventions of the Romanian scribes. We will analyze and compare parallel excerpts from all the lexicons, observing the differences between them and the model, both in terms of the inventory of the Slavonic entries and of structure of the Romanian explanations. This work was supported by a grant of the Romanian Ministry of Education and Research, CNCS-UEFISCDI, project number PN-III-P1-TE-2019-0517, within PNCD III.

References

Nimciuk, V. 1961. 'Рамво Веринда і його « лексіконъ славенорѡсскїй и именъ тлъкованїє »'. In Nimčuk, V. V. (ed.). *Pamvo Berynda, Leksikon slavenorosskij i imen tl'kovanïe*, Kiev.

The overall structure of early polyglot dictionaries

Anna Grzeszak

University of Warsaw (Poland)

anna.grzeszak@uw.edu.pl

Multilingual dictionaries, both ancient and contemporary, may contain several sections differing in the choice of the source language (e.g. the five-language version of Calepino's *Dictionarium* printed in 1592 in Venice, Decimator's *Thesaurus linguarum*, Wiktionary, Glosbe, Xobdo). Bilingual dictionaries share the same property and are therefore classified either as unidirectional (e.g. a Spanish-French dictionary) or bidirectional (e.g. a Spanish-French, French-Spanish dictionary). The structure of multilingual dictionaries may be much more complex; however, the existing dictionary typologies, designed predominantly to describe mono- and bilingual wordbooks, do not take into consideration such a high level of structure complexity. In an attempt to fill this gap, the present study aims to explore types of 16th- and 17th-century polyglot dictionaries distinguished on the basis of their overall structure.

The study material consisted of 51 general-purpose dictionaries printed between 1501 and 1700 which paired five different languages or more. The polyglots were analysed according to the variability of their source languages and the variability of languages paired in subsequent dictionary sections. The results indicate that four types may be distinguished: 1) polyglots with a constant set of paired languages and only one source language, 2) polyglots with a changing set of paired languages and only one source language, 3) polyglots with a changing set of paired languages and more than one source language, where not every tongue has this function, 4) polyglots with a changing set of paired languages where every tongue is a source language in at least one dictionary section.

Implementing a corpus-based approach to lexicographical work: Lemmatisation of the Old English comparative adverbs

Yosra Hamdoun Bghiyel

University of La Rioja (Spain)

yohamdou@unirioja.es

The aim of this paper is to present the lematisation process of the Old English non-verbal categories. The main Old English dictionaries of reference, including Bosworth and Toller's (1973) *An Anglo-Saxon Dictionary,* Clark Hall's (1996) *A Concise Anglo-Saxon Dictionary* and Sweet's (1976) *The Students Dictionary of Anglo-Saxon*, fail to compile a full inventory of lemmatised inflectional forms in a systematic way. The singularities of Old English, a language that presents considerable spelling inconsistency partly due to the existence of several dialects and the absence of a written standard, make this pending task even more necessary.

This research contributes with a new lemmatisation methodology implemented in the OE comparative adverbs. The starting point of this study is the automatic extraction of the forms morphologically tagged with the ADVR label (comparative adverbs). The inflectional forms inventory was automatically extracted from *The York-Toronto-Helsinki Parsed Corpus of Old English Prose* and *The York-Helsinki Parsed Corpus of Old English Poetry* which include annotated corpora. The lemmas are taken from the database of Old English *Nerthus*. Secondly, a lemma provided by the lexical database of Old English *Nerthus* is manually assigned to each of inflectional form. Thirdly, the results are contrasted with two lexicographical sources of reference: Seelig (1930) and the *Dictionary of Old English* in order to verify the lemma assignment and disambiguate doubtful cases. The conclusions insist on the applicability of the lemmatisation method to all non-verbal categories of Old English. This piece of research constitutes an advance in the description of the Old English language, and offers a fully lemmatised inventory of Old English comparative adverbs.

References

Bosworth, J. and T. N. Toller. 1973 (1898). *An Anglo-Saxon Dictionary*. Oxford: Oxford University Press.

Clark Hall, J. R. 1996 (1896). *A Concise Anglo-Saxon Dictionary*. Toronto: University of Toronto Press.

Healey, A. diPaolo (ed.). 2018. *The Dictionary of Old English in Electronic Form A-I*. Toronto: Dictionary of Old English Project, Centre for Medieval Studies, University of Toronto.

Martín Arista, J. (ed.) L. García Fernández, M. Lacalle Palacios, A. E. Ojanguren López, and E. Ruiz Narbona. 2016. *NerthusV3*. Online Lexical Database of Old English. *Nerthus Project*. Universidad de La Rioja. [www.nerthusproject.com].

Pintzuk, S. and L. Plug (ed.). 2001. *The York-Helsinki Parsed Corpus of Old English Poetry*. Department of Language and Linguistic Science, University of York.

Seelig, F. 1930. *Die Komparation der Adjektiva und Adverbien im Altenglischen.* Heidelberg: Winters.

Sweet, H. 1976 (1896). *The student's Dictionary of Anglo-Saxon*. Cambridge: Cambridge University Press.

Taylor, A., A. Warner, S. Pintzuk and F. Beths (eds.). 2003. *The York-Toronto-Helsinki Parsed Corpus of Old English Prose*. Department of Language and Linguistic Science, University of York.

Icelandic as depicted in the Basque-Icelandic glossaries of the 17th century: The first occurrences of Icelandic in a lexicographical context

Anna Hannesdottir (University of Gothenburg, Sweden)

Joakim Liljergen (University of Gothenburg Library, Sweden)

anna.hannesdottir@svenska.gu.se, anna.hannesdottir@svenska.gu.se

Four manuscripts containing Basque-Icelandic glossaries from the 17th century are known (Etxepare & Miglio 2015, Deen 1937). The number of lexical entries is altogether 900, mostly single words but some phrases too. While the glossaries have aroused due interest from Basque linguists (e.g. Knörr 2006), little attention has been paid to the Icelandic language as reflected by the equivalents (however, see Guðmundsson 1979, Erlingsson 1995). These are the focus of this study.

Early Icelandic lexicography was first and foremost historical. A large Icelandic-Latin dictionary was published 1650 with the aim of facilitating the Danes' comprehension of the remarkable Icelandic medieval literature (Worm 1650, Faulkes & Ingólfsson 2010). The glossaries, on the other hand, focus on verbal communication between Basque whalers and Icelandic farmers about everyday matters. A point of interest from a lexicographic perspective is how the different lexical functions of the Icelandic items in Worm's dictionary vs. the glossaries manifest in the vocabulary. In the former, the Icelandic entries are carefully chosen in accordance with the aim of the dictionary while, in the glossaries, the Icelandic equivalents are simply evoked by the selection of the Basque entries.

In our presentation, we compare the Icelandic units occurring in the glossaries with the vocabulary presented in Worm's dictionary. For 2/3 of these items, the first occurrence in a lexicographic context is in these glossaries. We will relate the Icelandic vocabulary in the glossaries vs. in the dictionary to the different aims of the respective work and also to the function as source vs. target language. Also, we will exemplify some lexical patterns in the Icelandic vocabulary of the early 17th century found in the first lexicographic occurrences revealed in our survey.

References

Deen, N. G. H. 1937. *Glossaria duo Vasco-Islandica*. H. J. Paris: Amsterdam.

Erlingsson, G. 1995. Við strandir norður staðnæmdist sú strákaþjóð: þrjú basknesk-íslensk orðasöfn frá 17. öld. (Unpublished BA thesis.)

Etxepare, R. and V. G. Miglio. 2015. 'The newly-discovered fourth Basque-Icelandic glossary'. In X. Irujo and V. G. Miglio (eds.). *Basque Whaling in Iceland in the XVII century: Legal organization, cultural exchange and Conflicts of the Basque fisheries in the North Atlantic*. Basque Law Series 2. Santa Barbara: University of California.

Faulkes, A. & G. Ingólfsson, (eds.). 2010. *Specimen lexici runici and Glossarium priscae linguae Danicae*. (Orðfræðirit fyrri alda 5.) Reykjavík: Stofnun Árna Magnússonar í íslenskum fræðum.

Guðmundsson, H. 1979. 'Um þrjú basknesk-íslenzk orðasöfn frá 17. öld'. *Íslenskt mál og almenn málfræði*. vol. 1. 75–87.

Knörr, H. 2006. 'Basque fishermen in Iceland: Bilingual vocabularies in the 17th and 18th centuries'. *Euskera*. vol. 51. 491–499.

Worm, O. 1650. *Specimen lexici runici* […]. Martzan: Hafniæ.

Challenges in the process of retro-digitization of Croatian grammar books before Illyrism

Marijana Horvat, Ana Mihaljević, Martina Kramarić

Institute of Croatian Language and Linguistics (Croatia)

mhorvat@ihjj.hr, ana.mihaljevic@stin.hr, mkramar@ihjj.hr

In the history of the Croatian language, there is a long tradition of writing grammar books. So far none of the projects of retro-digitization have included Croatian grammar books from the pre-standard period of the Croatian language (before the establishment of the common standard language and orthography). Within the current project *Retro-digitization and Interpretation of Croatian Grammar Books before Illyrism–RETROGRAM*, funded by Croatian Science Foundation, the most important grammar books from this period will be digitized.

In this talk, we will briefly present aims and sources for this project. In more detail, we will discuss methods, challenges that we faced so far, and first results. The main goal of the project is to create an open-access web portal which will include: facsimiles of selected grammar books, their transcription or translation and an index of historical grammar and linguistic terminology, and will be equipped by thematic searching possibilities on the morphological level.

The digitization process includes transcriptions or translations of the selected grammar books, and the annotation on the text level, on the terminological level, and on the morphological level using TEI tags. Sources include eight printed grammars written in Croatian, Latin, or Italian language by Croatian authors in the period from the beginning of the 16th to the beginning of the 19th century.

We will describe the obstacles and challenges we faced so far in the digitization process, starting with the creation of the TEI Header and defining the common structure of all grammars in order to compile instructions for annotators. The main challenge of the projectis the fact that the TEI module for grammars has not been established yet, so our work through establishing this standard may have a significant impact on future similar undertakings.

References

RETROGRAM. *Retro-digitization and interpretation of Croatian grammar books before Illyrism*. Zagreb: Institute of Croatian Language and Linguistics. [http://retrogram.jezik.hr/]

The looking glass: Historical lexicography and identity in contemporary Mesoamerica

Kathryn M. Hudson
University of Buffalo (United States of America)
khudson@buffalo.edu

This paper explores the ways in which historical lexicographic materials contribute to modern conceptualizations and reconceptualizations of identity in contemporary Mesoamerica by providing access to knowledge lost during the colonization process; by offering a sense of legitimacy through connections with a recognized and documented history; and by facilitating claims of cultural and linguistic competence. In this view, which is rooted in a critical lexicographic perspective, the original cultural source and precise referent of a lexeme is less significant than its occurrence in a historical lexicographic record, which symbolizes historical validity and indigeneity for many modern populations. Such materials thus function as sources of cultural, historical, and linguistic information – most commonly in relation to faunal, floral, and topographic referents – that can be used to develop and/or assert an indigenous identity, particularly when non-native alternatives are available; the documents themselves are also viewed as proof of identity and as a tangible connection to a past that has been largely erased. These functions of are particularly significant in El Salvador, Honduras, and Nicaragua, where the effects of colonial and postcolonial histories on indigenous identities have been especially dramatic. Here the use of historical lexicographic records contributes to the formation of two distinct but closely related categories of identity: a generalized pan-indigenous identity and one of several more specific local or regional identities connected to historically documented groups. The author's ethnographic and linguistic work with communities in this area includes documentation of these processes, how they reflect and relate to ongoing formulations of identity and assertion, and the sociocultural contexts in which they occur.

Historical and genre labels in the dictionary of the 16th-century Slovenian literary language

Alenka Jelovšek

Fran Ramovš Institute of Slovenian Language (Slovenia)

jelovsek.alenka@gmail.com

The Dictionary of the 16th-Century Slovenian Literary Language, the first book of which will be published in 2021, is a synchronic historical dictionary, covering a relatively short period of the Slovenian Reformation between 1550 and 1603. It is based on a complete extract of material from 53 works of Slovenian Protestants (approx. 3 million words). The explanatory dictionary tries to present in detail the semantic structure of lexemes in the given period.

As the material for the dictionary is limited in terms of text types (mainly biblical translations and theological texts, describing (and interpreting) either the ancient or the 16th-century reality), the usage of some specific labels that place lexemes in either a historical or genre context is necessary to mark specific senses, sub-senses and semantic nuances, thus drawing attention of the user to the historically or ideologically specific use of lexemes (eg. *in Old Testament*, *Protestant*, *Catholic*, *Protestant about Catholic*), which the dictionary cannot fully describe, despite relatively extensive semantic definitions, as it would require a broader encyclopedic explanation. Other labels are mostly related to the genre of biblical texts, e.g. *in parables*, *in Revelation*; as lexemes in these texts often have an additional metaphorical meaning which is not explicit within the texts and often cannot be unambiguously interpreted, the labels indicate such symbolic meaning without describing it fully. When the symbolic meaning becomes an independent sense in the dictionary, the labels serve to mark the semantic distinction between the original biblical text and its interpretation in the exegetical commentaries (eg. *beast* in Apocalypsis 'being in the form of an unusual horned animal, who turns people away from God; beast' vs. in commentaries 'servant of Satan who turns people away from faith in Christ; antichrist').

Dictionary as historical evidence: A case study of English-Latvian dictionaries published in DP camps after WWII

Laura Karpinska

University of Latvia (Latvia)

laura.karpinska@lu.lv

Dictionary can serve as historical evidence of the experiences of its compilers and target group. The study focuses on the analysis of English-Latvian dictionaries published in the extraordinary conditions of Displaced Persons (DP) Camps from 1945 to 1948.

Scared of the return of the Soviet army a large number of people left Latvia at the end of WWII. Most of them found shelter in DP camps in Germany. There were several factors which contributed to active cultural life in Latvian DP camps: the basic needs of the refugees were secured, there was plenty of free time since it was difficult to find work, many of the refugees were well educated and eager to participate in various cultural activities. Despite the shortage of paper and printing equipment, a considerable number of books, including six English-Latvian dictionaries, were published in DP camps. Since most of Latvian DP camps in Germany were located in the British and American zones, most Latvian refugees later emigrated to various English-speaking countries.

The aim of the study is to analyse the publishing details and the characteristic features of the mega-, macro- and microstructural level of the English-Latvian dictionaries published in DP camps to detect how these dictionaries reflect the conditions and needs of Latvian refugees living in these camps after WWII.

The study reveals that the dictionaries are small and of poor printing and binding quality, may provide lists comprising encyclopaedic information in their back matter, various inconsistencies and limitations can be observed on both the macro- and microstructural level of the dictionaries. The number of English-Latvian dictionaries published during this period reveals that there was a considerable demand for these dictionaries in DP camps. The dictionaries also reflect the hardships that Latvian refugees encountered and their aspirations to learn the English language, acquiring both linguistic and encyclopaedic information.

On lemmas and dilemmas again: Problems in historical dialectal lexicography

Georgia Katsouda, Ioanna Manolessou

Academy of Athens (Greece)

katsouda@academyofathens.gr, manolessou@academyofathens.gr

The proposed paper discusses a well-known issue in historical lexicography: thespecific form that should be adopted for the headword, so that it may successfully subsume all variant forms of a lemma, in a way that allows easy retrieval by the user (Reichmann 2012). The problem is exacerbated in historical dialectal lexicography, where the variant forms are even more distanced from the standard formof the word/lemma (Rys and Keymeulen 2009).

The discussion is based on two major Greek lexicographic projects, using specific examples: a) the *Historical Dictionary of Modern Greek* of the Academy of Athens (ILNE), which includes the standard language and its dialects (Manolessou and Katsoudaforthcoming) and b) the historical dictionary of the Cappadocian dialects, a project of the University of Patras, which covers the sub-dialects of Cappadocian, formerly spoken in Asia Minor and nowadays spoken by 3rd generation immigrants in Central and Northern Greece.

In the case of Greek, the distance between variant forms considered (on etymological reasons) to be part of the 'same' word/lemma can be considerable, given: a) the long attested history of Greek; b) the phenomenon of diglossia, which results in the 'artificial'retention in the Standard of forms not having undergone language change and c) the existence of dialects isolated from the main body of Greek through historical circumstances for more than 1000 years (e.g. Cappadocia, Pontus and South Italy).

Criteria for the selection of a specific headword form include (see the *Manual of the Historical Dictionary*, ILNE-MR): a) its presence/absence in the standard language; b)its relative geographical spread; c) the number of phonetic/morphological changes it hasundergone and d) its degree of recognizability by the user. Given that the new HistoricalCappadocian Dictionary is compiled as a parallel digital/online and print dictionary, thefindability problem is less acute.

References

Historical Dictionary of the Cappadocian Dialects. University of Patras.

ILNE. *Ἱστορικὸν Λεξικὸν τῆς ΝέαςἙλληνικῆς, τῆς τε κοινῶς ὁμιλουμένης καὶ τῶν ἰδιωμάτων, τ. 6 (Α- διάλεκτος)* (*Historical Dictionary of Modern Greek, both of the Standard Language and the Dialects*). Athens: Academy of Athens.

ILNE-MR. 2011. *Κανονισμός Συντάξεως του Ιστορικού Λεξικού της Νέας Ελληνικής (Manual of Regulations of the Historical Dictionary of Modern Greek).* Athens: Academy of Athens.

Manolessou, I. and G. Katsouda. Forthcoming. 'The making of the Historical Dictionary of Modern Greek: Problems and solutions in the domain of historical and dialectal lexicography'. In H. van der Velde, A. Bocast and F. Dolezal (eds.), *Proceedings of ICHLL 10-Leewarden 2019*. Berkeley Bridge Press.

Reichmann, O. 2012. *Historische Lexikographie. Ideen, Verwirklichungen, Reflexionen an Beispielen des Deutschen, Niederländischen und Englischen.* Berlin/Boston: De Gruyter.

Rys, K. and J. van Keymeulen. 2009. 'Intersystemic Correspondence Rules and Headwords in Dutch Dialect Lexicography'. *International Journal of Lexicography,* 22. 129-150.

Knowledge-oriented extended author's dictionary

Margit Kiss

Research Centre for the Humanities (Hungary)

kiss.margit@btk.mta.hu

In our project we created a unified system to store the author's oeuvre, an extended author's dictionary, the critical annotations and data from external knowledge sources all linked together usingthe RDF-technique. We processed the oeuvre of Kelemen Mikes (a Hungarian author of the 18th century) turning it into a vast, computer-based knowledge source for humanist researchers.

For improving the efficiency of textual analysis of historical texts, we transformed the corpus into XML TEI form, and we created an extended author's dictionary attached to the corpus (Kiss and Mészáros 2018a). We extended headwords by adding additional lexicographical knowledge, linking relevant critical notes and referencing external knowledge entities using the Linked Open Data techniques (Kiss and Mészáros 2018b).
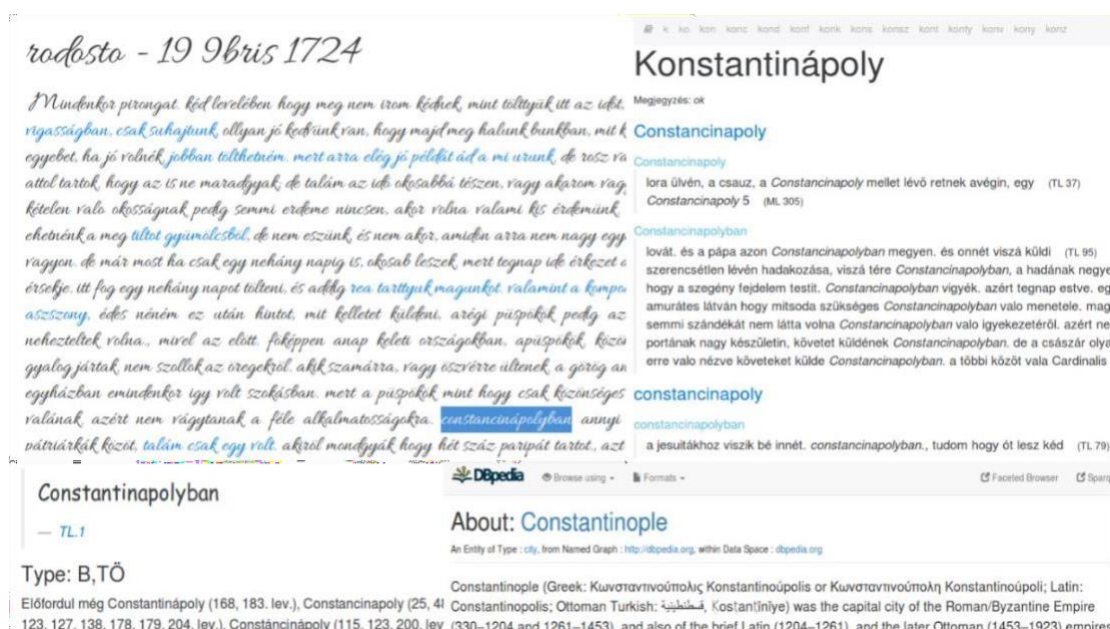
Figure 1. Linking the corpus, dictionary entries and DBpedia data together

After creating this vast data source (corpus – extended dictionary – external knowledge sources) we extracted knowledge pieces and transformed them into RDF triplets. This way we created a system that stores the corpus, the author's dictionary and an RDF knowledge base in a unified way.

In order to acquire more knowledge from the critical annotations we also developed a new method toauthor them. This is based on the idea of using controlled natural languages to write text notes that are processable by NLP tools and can be automatically converted into a computer-based knowledge representation. This method greatly simplifies extracting knowledge from the research notes (Kiss and Mészáros 2018c).
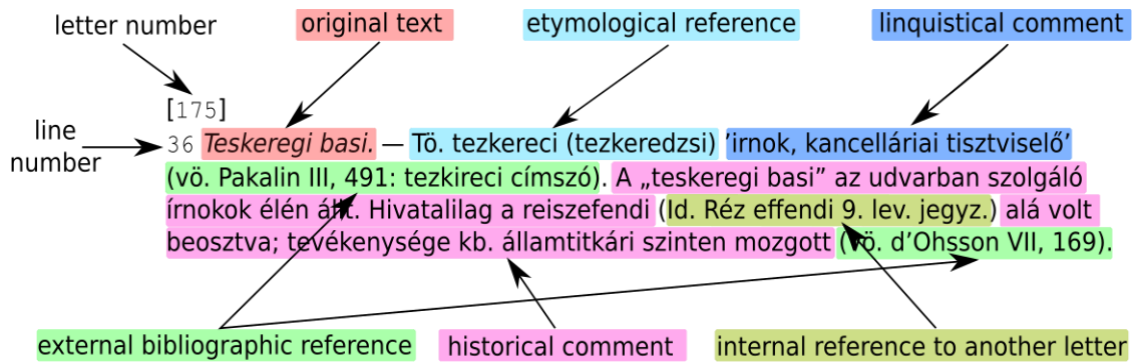
Figure 2. The structure of a critical annotation

The extended knowledge-oriented system gives possibility to access much more semantic content than before, and it also enabled the development of new techniques for data accessand document analysis (Bartalesi *et al.* 2018).

References

Bartalesi V. *et al*. 2018. 'A web application for exploring primary sources: The Dante Sourcescase study'. *Digital Scholarship in the Humanities,* 33 (4) 2. 705-723.

Kiss, M. and T. Mészáros. 2018a. 'Rethinking the role of digital author's dictionaries in humanities research'. In J. Čibej, V. Gorjanc, I. Kosem and S. Krek (eds.), *Lexicography in Global Context*, *Proceedings of the XVIII EURALEX International Congress*, Ljubljana University Press. 871-879.

Kiss, M. and T. Mészáros. 2018b. '*The DHmine Dictionary work-flow: creating a knowledge-based author's dictionary'.* In J. Čibej, V. Gorjanc, I. Kosem and S. Krek (eds.), *Lexicography in Global Context*, *Proceedings of the XVIII EURALEX International Congress*, Ljubljana University Press. 77-86.

Kiss, M. and T. Mészáros. 2018c. 'Knowledge Acquisition from Critical Annotations'. *Information* 9,179. 1-10.

Mikes Kelemen. 1966-1988. *Összes művei.* H. Lajos (Ed.), Akadémiai Kiadó, Budapest.

Creating a historical dictionary with an automated corpus annotation system

Margit Kiss

Research Centre for the Humanities (Hungary)

kiss.margit@btk.mta.hu

Creating an author's dictionary is a highly labour-intensive process (Arnold 2006). Each word from an automatically generated concordance list need to be analyzed and processed separately: researchers define word-forms, variations, dictionary words etc. This task is even harder for agglutinating languages and for historical texts (Kiss and Mészáros 2018a). We are processing a large oeuvre (1,5 million words) of an 18th-century Hungarian author (Kiss and Mészáros 2018b). Firstly, we created the concordance list with all word occurrences; then, researchers started to process this huge list to create dictionary entries (Prószéky and Kis 1999). In order to facilitate and speed up this process we developed new methods and computer-based tools.

Firstly, we created a new workflow for creating dictionary entries based on an annotated version of the corpus. Instead of processing the concordance list researchers read and annotate the corpus with dictionary information. This way each sentence is only read once that already speeds up the process by the factor of their average length in words. We developed a corpus annotation scheme that can describe all dictionary information and a software tool that can automatically create the dictionary from the annotated corpus.

Secondly, a software tool has been created to automatically annotate the corpus based on already authored dictionary entries. We also use machine learning techniques to annotate word occurrences not present in the dictionary. Researchers finalize these annotations. This greatly simplifies the extension of dictionary entries with new occurrences and also helps in creating new headwords.

Our method has many advantages: 1) the textual interpretation of each word is much easier than from a concordance list; 2) annotation tags are inserted automatically; 3) many words are annotated automatically; 4) many time-consuming and redundant work phases are avoided. We estimate that this method allows us to finish the remaining five-year manual work of the Mikes project within a year.

References

Arnold, B. 2006. 'Night Watches on the Computer: Creating an Author's Dictionary with Computational Means'. *Literary and Linguistic Computing,* 21. 5-14.

Kiss, M. and T. Mészáros. 2018a. '*The DHmine Dictionary work-flow: creating a knowledge-based author's dictionary'.* In J. Čibej, V. Gorjanc, I. Kosem and S. Krek (eds.), *Lexicography in Global Context*, *Proceedings of the XVIII EURALEX International Congress*, Ljubljana University Press. 77-86.

Kiss, M. and T. Mészáros. 2018b. 'Rethinking the role of digital author's dictionaries in humanities research'. In J. Čibej, V. Gorjanc, I. Kosem and S. Krek (eds.), *Lexicography in Global Context*, *Proceedings of the XVIII EURALEX International Congress*, Ljubljana University Press. 871-879.

Prószéky, G. and B. Kis. 1999. 'A Unification-Based Approach to Morpho-Syntactic Parsing of Aggluti-native and Other (highly) Inflectional Languages'. *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics, ACL '99.* College Park, Maryland: Association for Computational Linguistics. 261-68

Latin plant names in Sergiy Shelonin's Lexicon (3rd edition)

Kira Kovalenko

Institute for Linguistic Studies, Russian Academy of Sciences (Russia)

kira.kovalenko@gmail.com

Sergiy Shelonin is a famous Russian scholar, who lived in the Solovki monastery in the 17th century and created the biggest lexicon of his time. Its third edition comprises more than 16,000 word entries, which explain foreign and Russian obsolete words and phrases. One of the sources taken by Sergiy for his lexicon wasthe Russian translation of '*Liber de arte distillandi de simplicibus'* ('Kleines Destillierbuch') written by Hieronymus Brunschwig, which formed the first part of '*Medicinarius'* (1505). In the wordlist, Sergiy included all the plant names (mostly German) that he found in this medical treatise with equivalents presumably in Latin, and sometimes in Greek, Arabic and Russian. Latin names often contain a lot of mistakeswhich appeared during transliteration into Cyrillic and rewriting from one manuscript to another, nevertheless they served for popularization of Latin terminology which penetrated to the Moscovy by means of foreign doctors who wrote out their prescriptions in Latin, as well as translated herbals and phrasal books(Cuba 1485; Ridley 1996). The amount and variety of sources, containing Latin terminology created good basis for new phytonyms to take root in the Russianbotanical terminology and in the vernacular language. In the paper, Latin plant namesfrom the Sergiy Shelonin's Lexicon (3rd edition) are going to be analysed and their usage in the Russian language of later periods to be traced. The research was made on the materials represented at the Russian phytonyms database *PhytoLex*.

References

Cuba, J. von. 1485. *Gart der Gesundheit*. Augsburg.

Medicinarius. 1505. *Das buch der Gesundheit. Liber de arte distillandi Simplicia et Composita.* Strassburg.

PhytoLex. *Russian Phytonymy of the 11th-17th centuries.* [https://phytolex.iling.spb.ru].

Ridley, M. 1996. *A Dictionarie of the Vulgar Russe Tongue.* Köln: Böhlau Verlag.

Shelonin, Sergiy. 1650s. *An Alphavit giving explanations of foreign words discovered in Russian holy books, having the beginning from A.* Manuscript of the Russian National Library, Solovki collection 18/18. In folio.

Learning in Old English: Information and knowledge acquisition

Miguel Lacalle Palacios

University of La Rioja (Spain)

miguel.lacalle@unirioja.es

This paper aims at providing an inventory of Old English verbs belonging to the class of learning on the grounds of their common semantic components and shared grammatical behavior. Verbs of learning depict states of affairs comprising a Cognizer who learns some Content, as in *John learned how to play the guitar*; or an Effector that causes a Cognizer to get to know some Content, as in *John learned how to play the guitar thanks to the master's teachings*.

The framework of verb classes and alternations (Levin 1993) and Role and Reference Grammar (Van Valin and LaPolla 1997; Van Valin 2005) provide the theoretical basis for this study, which takes the line that for verbs to belong to a certain class they must convey similar meaning components (described by means of *Aktionsart* types and thematic roles) and show similar grammatical behavior (understood as the morphosyntactic realization of arguments). This research direction, therefore, contributes to a synchronic organization of the Old English lexicon that takes the syntactic range of verbs into account.

Textual and lexicographical sources as *The Dictionary of Old English Web Corpus* and the *Dictionary of Old English* respectively and the lexical database *Nerthus* have been used to select the data for the analysis of the linking between semantics and syntax with these verbs.

The main conclusion is that, considering the constructions and alternations in which they participate, the Old English verbs *(ge)frignan*, *(ge)leornian*, *ofācsian* and *onfindan* are the best candidates for membership of the verbal class of learning.

References

Healey, A. diPaolo (ed.) with J. Price Wilkin and X. Xiang. 2004. *The Dictionary of Old English Web Corpus*. Toronto: Dictionary of Old English Project, Centre for Medieval Studies, University of Toronto.

Healey, A. diPaolo (ed.) 2018. *The Dictionary of Old English in Electronic Form A-I*. Toronto: Dictionary of Old English Project, Centre for Medieval Studies, University of Toronto.

Levin, B. 1993. *English Verb Classes and Alternations*. Chicago: University of Chicago Press.

Martín Arista, J. (ed.), L. García Fernández, M. Lacalle Palacios, A. E. Ojanguren López, and E. Ruiz Narbona. 2016. *NerthusV3. Online Lexical Database of Old English*. *Nerthus Project*. Universidad de La Rioja [www.nerthusproject.com].

Van Valin, R. 2005. *Exploring the Syntax-Semantics Interface*. Cambridge: Cambridge University Press.

Van Valin, R. and R. LaPolla. 1997. *Syntax: Structure, meaning and function*. Cambridge: Cambridge University Press.

Depriving in Old English: A RRG analysis

Miguel Lacalle Palacios

University of La Rioja

miguel.lacalle@unirioja.es

The aim of this paper is to identify the constructions and alternations in which Old English verbs of depriving partake on the basis of the system of lexical decomposition of Role and Reference Grammar as well as the verbal classes distinguished by Van Valin and LaPolla (1997), Van Valin (2005), Pavey (2010) and Cortés Rodríguez (2014) who proposes a framework of verbal classes and alternations, according to which class membership implies verbs participating in the same alternations of argument realization; while verbs with similar grammatical behavior tend to express similar meanings.

Previous research has focused on the morphological case that these verbs license (McLaughlin 1983; Visser 1984; Mitchell 1985; Molencki 1991) and indicated that they tend to select a direct case and an oblique one, for instance an accusative object of person and a genitive of thing as in *forhwan hine se cyng ealles benæmde* [ChronE (Irvine) 162100 (1104.15)] 'for which the king deprived him of all the possessions'.

The verbs analyzed from *A thesaurus of Old English* and *Historical thesaurus of the Oxford English dictionary* include those that primarily mean 'to deprive, to take away' The textual data were retrieved from *The Dictionary of Old English Web Corpus*, while the syntactic analysis relies on the *York-Toronto-Helsinki Parsed Corpus of Old English Prose.*

Regarding methodology, on the morphosyntactic side, grammatical case, prepositional government, and voice have been considered. Semantically speaking, the number and class of semantic participants, the *Aktionsart* type and the type of possession (alienable vs. inalienable) have been taken into account.

Conclusions show a correlation between a given alternation and different constructions when both the morphosyntax and semantics are considered. Moreover, the perspective on the state of affairs (former possessor vs. new possessor) determines the morphosyntactic aspects of the expression.

References

A. Cameron, A. Crandell Amos, A. diPaolo Healey et al. (ed.). 2007. *The Dictionary of Old English: A to G* online. Toronto: Dictionary of Old English Project.

A. diPaolo Healey with J. Price Wilkin and X. Xiang. 2009. *Dictionary of Old English Web Corpus.* Toronto: Dictionary of Old English Project.

Cortés-Rodríguez, F. 2014. 'Aspectual features in Role and Reference Grammar: A layered proposal'. *Revista Española de Lingüística Aplicada* 27: 23–53.

Kay, C., Roberts, J., Samuels, M. &, I. Wotherspoon. 2009. *Historical thesaurus of the Oxford English dictionary*. Oxford: Oxford University Press.

McLaughlin, J. 1983. *Old English syntax: a handbook*. Tübingen: Max Niemeyer Verlag.

Mitchell, B. 1997 (1985). *Old English syntax volume I: concord, the parts of speech, and the sentence*. Oxford: Oxford University Press.

Molencki, R. 1991. *Complementation in Old English*. Katowice: Uniwersytet Slaski.

Pavey, E. 2010. *The Structure of Language: An Introduction to Grammatical Analysis*. Cambridge: Cambridge University Press.

Roberts, J. and Kay, C., with L. Grundy. 2000 (1995). *A thesaurus of Old English in two volumes*. Amsterdam - Atlanta: Rodopi.

Mining terms in English: A linguistic and socio-cultural analysis

Julia Landmann (University of Heidelberg, Germany)

Andreas Landmann (Independent researcher, Germany)

Julia_Landmann@web.de, DrALandmann@mineral-fascination.biz

Mining is a wide and complex field. The search and exploitation of minerals is unceasing in many countries worldwide. Before its collapse towards the close of the twentieth century, coalmining in the United Kingdom, for instance, which dates back to Roman antiquity, occurred in various regions of the country, such as Northumberland and Durham, North and South Wales, Yorkshire, Kent, the East and West Midlands. An essential aim of the present paper is to give an overview of the plethora of mining terms in English which have been coined due to progress and advance in this field over time, ranging from terms to do with boring, drillingand extraction, to ventilation, pipelines and hauling. The *Historical Thesaurus* of the *Oxford English Dictionary Online* and specific sources such as *Elsevier's Dictionary of Mining and Mineralogy* constitute valuable tools to identify these types of lexical items. In all, this study relies on the analysis of a comprehensive lexicographical sample of about 4000 mining and mineralogy terms documented in English. A specific focus of this paper will be on the impact of German on the English language in the field of mining. Since the beginning of mining, Germany has been one of the most important European mining countries. German mining flourished from the late Middle Ages to well into the early modern era on the basis of salt and ore extraction. At the turn of the twentieth century, the production of coal and steel enabled Germany to become a world economic power. The present paper will show that German has provided English with a variety of words from this domain. A comparison will be made between the semantics and useof the German-derived lexical items in English and their equivalents in the original donor language. In addition, the multi-layered socio-cultural and historical background of the German-English contact situation will be taken into account, which led to the borrowing of a number of mining terms into English down the ages.

References

Dorian, A. F. 1997. *Elsevier's Dictionary of Mining and Mineralogy*. Elsevier: Elsevier Science Ltd.

Murray, J., Bradley, H., Craigie, W. and C. T. Onions (eds.). 1884-1933. *The Oxford English Dictionary*; Burchfield, R. (ed.). 1972-1986. *Supplement*; Simpson, J. and E. Weiner (eds.). 1989. 2nd edition.; Simpson J., Weiner, E. and M. Proffit (eds.). 1993-1997. *Additions Series*; Simpson, J. (ed.). (In progress). Third edition, OED Online (including the *Historical Thesaurus of the OED*). [www.oed.com]

The origins of English locative inversion: Fronted locatives in Old English

Sergio López Martínez
University of Oviedo (Spain)
lopezmsergio@uniovi.es

Locative Inversion occurs cross-linguistically, including in Present Day English. In clauses exhibiting Locative Inversion, the subject appears post verbally, and a locative element, a prepositional phrase in most cases, is proposed, as in *Next to it sood a silver urn* [...] (Biber *et al.* 1999: 912). Constructions similar to this can be traced back to the Old English period. Therefore, one of the aims of this paper will be to provide an overview of the distribution of fronted locatives in Old English, analysing the *YCOE*. Interestingly, not only can we find occurrences of this syntactic word order, reminiscent of PDE Locative Inversion, in OE main clauses, but also in OE embedded clauses, where V2 word order with preverbal non-subject elements would not be expected, as in (1) below:

(1)   Forðæm eac wæs *ðæt ðe* [***beforan ðæm temple***]^{PP} *stod* [***æren ceac***
        Because also was that   before   the temple   stood brass cauldron
        onuppan twelf ærenum oxum]^{SUBJ}
        upon   twelve brass   oxen
        'Because it also was that a brass cauldron upon twelve brass oxen stood
        before the temple'   (CP:16.105.1.687)

This paper will also study the differences between the syntactic structure of clauses with fronted locatives in Old English and those with locative inversion in PDE. Based on research on other languages such as Spanish, Ojea (2019: 1) believes that Locative Inversion is not a root phenomenon, since it can be "found in many contexts where standard topicalization would be ruled out". It is true that examples with embedded XSV word order (i.e. with the verb in final position and the subject in second position), which could be ascribed to topicalization, are also found in the corpus, as in (2):

(2)   swa hit Romane   selfe   sædon þæt [***under hiera anwalde***]^{PP}
        so   it   the Romans in this way said   that under their authority
        [***nan bismerlecre dæd***]^{SUBJ} *ne gewurde*^{SUBJUNCTIVE}
        no   shameful   deed   no happen^{SUBJUNCTIVE}
        'so the Romans said in this way that no shameful deed would happen
        under their authority'   (Or, 5:3.116.27.2447)

Finally, this paper will attempt to clarify the motivation behind each of these two types of word order (XVS and XSV). As regards embedded XVS clauses, there is strong evidence to suggest they are the result of the influence of discourse factors, partially given the discourse status of the locative as a highly deictic element, linking to the previous discourse. On the other hand, concerning embedded XSV clauses, we can assume that unaccusative verbs, lacking an external argument, allow the topicalization of these PPs, which surface in the leftmost position of the clause.

References

Biber, D., Stig Johansson, G., Leech, S. C. and E. Finegan. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education Limited.

Ojea, A. 2019. 'EPP-satisfaction on discourse grounds: the case of Locative Inversion'. In *Syntax*.

Taylor, A., Warner, A., Pintzuk A. and F. Beths. 2003. *The York Toronto Helsinki Parsed Corpus of Old English Prose*. Electronic texts and manuals available from the Oxford Text Archive.

Agile lexicography: Rapid prototyping of historical dictionaries with Shiny, with examples from Sanskrit and Tibetan

Ligeia Lugli

King's College London (United Kingdom)

ligeia.lugli@kcl.ac.uk

This paper advocates the benefits of adopting R Shiny and agile software development to build online historical dictionaries. Shiny is a package for the statistical programming language R that makes it very easy to create online applications (Chang et al. 2018). Agile is an approach to software development that favours a fast iterative cycle from minimal product prototyping, to user feedback and further prototype development (Williams and Cockburn 2003).

Drawing on two ongoing projects in Sanskrit and Tibetan lexicography, this paper argues that the combination of an agile workflow with the technical simplicity of Shiny offers three main advantages for historical digital dictionary development. First, the ability to incrementally build and scale up dictionary prototypes allows lexicographers to adjust and update their digital infrastructure during the entire lifecycle of the project, with minimal software development costs. Second, it offers the opportunity to release frequent ad interim versions of larger projects, thus helping lexicographers to gather feedback and showcase their work to funders regularly (Lugli 2019). Third, Shiny allows lexicographers to experiment with innovative dictionary features better to address the challenges of historical lexicography, such as data uncertainty and interpretability. This paper illustrates how the uncertain periodisation of Sanskrit sources (Lugli 2018) has been addressed in a Shiny-based Sanskrit dictionary allowing users to customise the chronological metadata associated with the dictionary sources. It also shows how a Shiny-based diachronic Tibetan dictionary builds on the Gdex framework (Kilgariff et al. 2008) to allow users to define their own criteria for what constitutes a 'good' example in each of the period featured in the dictionary.

The paper concludes with an overview of the limitations of Shiny and some potential drawbacks of its application to lexicography.

References

Chang, W., Cheng J., Allaire, J.J., Xie, Y. & J. McPherson. 2018. *Shiny: Web Application Framework for R*. [https://CRAN.R-project.org/package=shiny]

Kilgarriff, A., Husak, M., McAdam, K., Rundell, M. and P. Rychly. 2008. 'GDEX: Automatically Finding Good Dictionary Examples in a Corpus'. In E. Bernal and J. DeCesaris (eds.) *Proceedings of the Thirteenth EURALEX International Congress.* Barcelona: Institut Universitari de Linguistica Aplicada, Universitat Pompeu Fabra. 425–432.

Lugli, L. 2018. 'Drifting in Timeless Polysemy: Problems of Chronology in Sanskrit Lexicography'. *Dictionaries: Journal of the Dictionary Society of North America*, vol. 39(1). 105–129.

Lugli, L. 2019. 'Smart lexicography for under-resourced languages: lessons learned from Sanskrit and Tibetan'. In I. Kosem, T. Zingano Kuhn, M. Correia, J. P. Ferreria, M. Jansen, I. Pereira. J. Kallas, M. Jakubíček, S. Krek, and C. Tiberius (eds). *Smart Lexicography: eLex 2019*. 198–212.

Williams, L. and A. Cockburn. 2003. 'Agile software development: it's about feedback and change'. *Computer*. vol. 36(6). 39-43.

Example, usage variant and linking between the dictionary and corpus data

Olga Lyashevskaya

Higher School of Economics (Russia)

olesar@gmail.com

We report on the digitization effort to create a TEI-compatible electronic version of the Dictionary of Russian Language of the 11-17th cc. (DRL11-17 1975), with the more strategicobjective to build an integrated online platform that would link the dictionary to the Old Russian and the Middle Russian Corpus (RNC; Sichinava 2014; Lyashevskaya 2018).

For each word attestation in the dictionary examples, and for each linked corpus exampleprovided as supplementary data (Fuertes-Olivera 2016), we distinguish among several types of usage variants: 1) the earliest attestation; 2) representations in later documents of the period that may reflect orthographic, grammatical, and lexical perception of the word and its context by scribes; 3) representations that reflect perception of the word usage by publishers in the 19th-21st cc. (including normalization efforts; such representations are often attested in the corpus); 4) representations that reflect the policy of the dictionary editors (including those in revisions made by different editors); 5) representations employed by the users of the online platform (e.g. searches using modern shape of the word, representations with wildcards for forgotten words, etc.).

To handle such variation, we use a set of attributes on the TEI tags <cit>, <note> and <gloss> in the electronic dictionary. Both forms and labels are then transferred to the database, in which uses from the corpus and representations from the users' model for search also mapped and classified. In addition, all variants are mapped at the level of token, lemma, hyperlemma, and associated lemma. The resulting resource allows one to reconstruct the network of usage variants associated with the target word and can be used in the task of linking to other historical digital resources.

References

DRL11-17. 1975. *The Dictionary of Russian Language of the 11th-17th centuries*. vol. 1-31.Moscow.

Fuertes-Olivera, P.A. 2016. 'A Cambrian Explosion in Lexicography: Some Reflections for Designing and Constructing Specialised Online Dictionaries'. *International Journal of Lexicography*, 29 (2). 226-247.

Lyashevskaya O. 2018. 'Towards a TEI model of historical dictionaries of Russian'. *El'Manuscript*.

RNC. *Russian National Corpus*. [http://www.ruscorpora.ru/new/en/index.html.]

Sichinava D. V. 2014. 'Historical corpora of the Russian National Corpus as a tool for diachronic grammatical studies' [Istoricheskie korpusa Natsional'nogo korpusa russkogojazyka kak instrument diakhronicheskikh issledovanij grammatiki]. In V.A. Baranov, V. Zheljazkova and A.M. Lavretiev (eds.), *Textual heritage and information technologies*.

Lexicographic analysis of the *Russian-Tajik Dictionary* (1933-34)

Abdusalom Mamadnazarov

Tajik National University (Tajikistan)

abdusalom.mamadnazarov@gmail.com

The purpose of this paper is to study the lexicographic peculiarities of the Russian-Tajik Dictionary (published in 1933-34) (further: RTD-1933-34), which is little known not only to a wide range of readers, but also to many linguists and lexicographers. The dictionary under study has now become a bibliographic rarity, which can only be found in private collections. Even the National Library does not have a copy of this publication.

A review of the literature indicates that almost nothing is known about this dictionary other than general information, that is, the dictionary has not been the object of a research until today.

From the lexicographic analysis of the RTD-1933 it can be stated that despite its shortcomings, the dictionary is an important lexicographic, historical, and written monument of the Tajik language with its own characteristics, both in terms of its macrostructure and in terms of the microstructure of its dictionary entries. The dictionary was compiled in Samarkand and published in Leningrad in the *Comintern Publishing House.*

In macro-structural terms, it is revealed that the dictionary consists of: 1) Introduction, entitled "From the compilers" [p. 8-10] in Tajik and Russian languages; 2) a list of references used (19 items) in Russian and Tajik languages [p. 7, 10]; 3) word list of the first volume in alphabetical order of the Russian language from the letter А to О [p. 11-325]; 4) a word list of the second volume from the letter П to Я [p. 331-626]; and 5) the Appendix [p. 627-633] consisting of the alphabetical glossary of foreign words, expressions and aphorisms used in the original language in the Russian press, posted at the end of the dictionary.

As a result of lexicographic analysis, it can be stated that the dictionary entry consists of a Russian word in bold, with a capital letter without a paragraph and a Tajik translation, an equivalent or equivalents without any grammatical and stylistic labels and illustrative examples. A dictionary entry may consist of one lexical unit, a word combination, or of two or three synonymous words separated by a comma or semicolon. Numbers of meaning and interpretation separates polysemy words.

From a number of technical tools for formatting a dictionary entry the use of a bold font for the entry units and a regular font for translations and interpretation is observed. To save space, a bold dash sign (—) is used instead of reflexive verb form of the entry unit.In the dictionary, in one entry, the Imperfect and Perfect forms of the verb are fixed as the entry units. The usual dash sign is also used in the dictionary after transcription of borrowed foreign words, before their interpretation. The definition of the entry units in the dictionary is carried out by using translations, finding Tajik equivalents, by describing the entry word, using synonyms or a number of synonyms, and by cross reference to other dictionary entries.

RTD-1933-34 contains a large glossary of foreign words (about 400 – words, expressions, and aphorisms in the form of sentences from Greek, Latin, English, French, German, and other languages used in the Russian language in the media before the beginning of the XX century, without changes and translation into Russian. The structure of the dictionary entry in the appendix consists of words, expressions or aphorisms as an entry unit, labels, transcription in the Tajik language, translation or interpretation into the Tajik language. Further detailed lexicographic and lexical research of the peculiarities of this dictionary is of great importance for the history of Tajik lexicography.

The lemmas and inflections of Old English nouns: The letter L

Laura Martín de las Pueblas Yanguas
University of La Rioja (Spain)
laura.martin-pueblas@alum.unirioja.es

The aim of this paper is to provide a list of Old English nominal lemmas for textual forms beginning with the letter L, as well as a classification by type of nominal inflection. This research is relevant because the *Dictionary of Old English*, whose headword entries offer, among other things, the lemma and the inflection type of the noun in question, has published the letters A to I. This precludes studies of Old English that require a lemmatised corpus because the inflectional forms in the *Dictionary of Old English Corpus* cannot be related to their respective lemmas and, consequently, extensive analyses cannot be carried out. The method is based on the comparison with the York Corpus of Old English (both prose and poetry), which provides the morphological tags that allow us to select only nominal inflections from the letter under analysis. Then, the inflectional forms are compared with a list of nominal lemmas extracted from *An open access annotated parallel corpus Old English-English.* To continue, the nominal forms are classified by declension type, which is assigned on the basis of Campbell (1987) and Hogg and Fulk (2011). Conclusions are expected in the areas of declension consistency, morphological class and spelling variation.

References

Campbell, A. 1987. *Old English Grammar*. Oxford: Oxford University Press.

Healey, A. diPaolo (ed.). 2018. *The Dictionary of Old English in Electronic Form AI.* Toronto: Dictionary of Old English Project, Centre for Medieval Studies, University of Toronto.

Hogg, R. M. and R. D. Fulk. 2011. *A Grammar of Old English. Volume 2: Morphology.* Oxford: Blackwell.

Healey, A. diPaolo (ed.) with J. Price Wilkin and X. Xiang. 2004. *The Dictionary of Old English Web Corpus.* Toronto: Dictionary of Old English Project, Centre for Medieval Studies, University of Toronto.

Martín Arista, J., Domínguez Barragán, S., García Fernández, L., Ruíz Narbona, E. Torre Alonso, E. and Vea Escarza, R. (comp.). 2021. *ParCorOEv2. An open access annotated parallel corpus Old English-English*. *Nerthus Project*. Universidad de La Rioja. [www.nerthusproject.com].

Pintzuk, S. and L. Plug. 2001. *The York-Helsinki Parsed Corpus of Old English Poetry*. Department of Language and Linguistic Science, University of York.

Taylor, A., Warner, A., Pintzuk, S., and F. Beths. 2003. *The York-Toronto-Helsinki Parsed Corpus of Old English Prose.* [http://www.helsinki.fi/varieng/CoRD/corpora/YCOE/].

Prime identification in historical languages: The Old English exponent of the semantic prime LIVE

Raquel Mateo Mendaza

University of La Rioja (Spain)

raquel.mateo@unirioja.es

The Natural Semantic Metalanguage theory (Wierzbicka 1996; Goddard 2008 and 2015) relies on the assumption that complex concepts can be explained in terms of simpler and irreducible ones. These simpler concepts –semantic primes– are universal, as they can be found in every language displaying the same meaning and with similar syntactical complementation patterns. To prove their universality, semantic primes have been identified in several living languages. However, there is also an open line of research on prime identification in historical languages developed by Martín Arista and Martín de la Rosa (2006) and de la Cruz Cabanillas (2007). Recent research on this field has dealt with the identification of the Old English primes included in the category *Actions, events, movement, contact* (Author 2013; 2016a; 2016b; forthcoming) and have established a proper methodology for this kind of studies. Against this background, this paper aims at developing this line of research by identifying the Old English exponent for the semantic prime LIVE, included in the category *Life and Death*. Taking the methodology proposed in previous research, a search on the *Historical Thesaurus of the Oxford English Dictionary* (Kay *et al*. 2009) and *Nerthus* database (www.nerthusproject.com) for those verbs conveying the meaning 'to live' provides the list of candidates for prime exponent. These candidates are then examined in terms of their morphology, textual frequency, semantics and syntax. With this purpose, the *DOE* (Healey *et al.* 2018), the *DOE* corpus (Healey *et al*. 2009) and *Nerthus* database have been accessed. The results obtained from these criteria concur on the verb *(ge)libban* as the accurate candidate for the Old English exponent for the semantic prime LIVE. The identification of examples of *(ge)libban* within the alternative syntactic configurations –*valency options*– associated to LIVE in terms of the NSM theory confirm the selection of this verb as prime exponent.

References

Author. 2013. 'The Old English exponent for the semantic prime TOUCH. Descriptive and methodological questions'. *Australian Journal of Linguistics,* 33 (4). 449-466.

Author. 2016a. 'The search for Old English semantic primes: the case of HAPPEN'. *Nordic Journal of English Studies*, 15 (1). 71-99.

Author. 2016b. 'The Old English exponent for the semantic prime MOVE'. *Australian Journal of Linguistics*, 36 (4). 542-559.

Author. Forthcoming. 'Semantic primes in historical languages. The identification of the Old English exponent for do'.

de la Cruz Cabanillas, I. 2007. 'Semantic Primes in Old English: a Preliminary Study of Descriptors'. *SELIM* 14. 37-58.

Goddard, C. (ed.). 2008. *Cross-Linguistics Semantics*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Goddard, C. 2015. 'Verb classes and valency alternations (NSM approach), with special reference to English physical activity verbs'. In A. Malchukov and B. Comrie (eds.),*Valency Classes in the World's Languages.* Berlin: Mouton de Gruyter.

Healey diPaolo, A. (ed.) Price Wilkin, J. and X. Xiang. 2009. *The Dictionary ofOld English Web Corpus*. Toronto. Dictionary of Old English Project.

Healey diPaolo, A., Cameron, A., Crandell Amos, A. *et al.* (eds.). 2018. *Dictionary of Old English: A to G online*. Toronto. Dictionary of Old English Project.

Kay, C., Roberts, J., Samuels, M. and I. Wotherspoon. 2009. *The Historical Thesaurus of the Oxford English Dictionary*. University of Glasgow: Oxford University Press.

Martín Arista, J. and M.V. Martín de la Rosa. 2006. 'Old English Semantic Primes: Substantives, Determiners and Quantifiers'. *ATLANTIS,* 28 (2). 9-28.

Martín Arista, J. (ed.), García Fernández, L., Lacalle Palacios, M., Ojanguren López, A.E. and E. Ruiz Narbona. 2016. *Nerthus V3*. Online Lexical Database of Old English. *Nerthus Project*. Universidad de La Rioja [www.nerthusproject.com.]

Wierzbicka, A. 1996. *Semantics. Primes and Universals*. Oxford: Oxford University Press.

Histories of words. Turning the page with *Wortgeschichte digital*

Nathalie Mederake

Academy of Sciences Göttingen (Germany)

nmedera@gwdg.de

Every word has its own story, evolving, sometimes significantly, over time. The project *Wortgeschichte digital* aims to describe the historical changes of meaning in a selected range of German vocabulary from about 1600 to the present day (Harm, in print). As a purely digital project, the entries are not processed alphabetically, but according to thematic contexts. In the first phase of work, important words in the key topic, politics and society will be covered using novel approaches. Instead of the lexicographical presentation formats that have been common up to now and which are essentially limited to a sequence of meanings and references, readable word stories will be introduced to describe the semantic developments in a narrative, thus making them more comprehensible. At the same time, digital linking and search possibilities will be used in an optimal way with the aim to develop and operate a digital information system that provides both a comprehensive and reliable description of German vocabulary and its continual changes. A series of articles from the thematic cluster of politics and society will be used to illustrate the structure of *Wortgeschichte digital*. We can observe that several words move towards the centre of this cluster asthey are influenced by political and intellectual undercurrents on the threshold of modernity. This is especially true for words that were initially describing physical surroundings, for instance atmosphere, climate, environment and milieu. A closer look to entry structure, presentation method as well as digital tools and content will illustrate how *Wortgeschichte digital* is retracing the evolutionof a vocabulary against the background of historical developments.

References

Harm, V. (in print). 'Beyond the "Grimm": German Historical Lexicography after Deutsches Wörterbuch'. In F. Dolezal and H. Van de Velde (eds.), *Broadening Perspectives in the History of Dictionaries and Word Studies* (selected papers from ICHLL 10). Newcastle upon Tyne: Cambridge Scholars.

Wortgeschichte digital. [https://www.zdl.org/wb/wgd/]

Electronic lexicographic description of Croatian Church Slavonic: Challenges and perspectives

Ana Mihaljević, Ana Šimić, Jozo Vela

Old Church Slavonic Institute (Croatia)

amihaljevi@gmail.com, ana.simic@stin.hr, jvela@stin.hr

Although the *Dictionary of the Croatian Redaction of Church Slavonic* (RCJHR 2000) is being compiled in the "golden age of lexicography" (Bailey 1996) it does not reflect this "goldenness" in the way it is compiled: as a simple Word document.

However, it is not easy to find an appropriate electronic tool to improve and modernize the compilation process due to the complexity of the overall context surrounding it: 1. the language itself – Croatian Church Slavonic language (CCS), the first Croatian literary language usedfrom the 11th until 16th century written in the Glagolitic script; 2. the parallel corpus exists onlyas a paper-card file (it contains CCS translations and the Greek and Latin source-texts); 3. five idioms-four scripts dictionary entry and 4. its rich microstructure (Vukoja 2014; Šimić 2016). Nevertheless, being well aware of all the advantages of modern e-lexicography (e.g. Atkins and Rundel 2008; Granger and Paquot 2012), the authors and editors consider the possibilities that thedigital era offers: the digitalization of the corpus on which the dictionary is based,retrodigitalization of the already published dictionary volumes, and using a professionaldictionary writing system (DWS), e.g. TLex, DPS, Lexonomy, Lexique Pro. Working in a DWS –which would support the preparation of the dictionary text– would enable linking entries andentry elements within the dictionary and possibly also with external sources. It would also enable different ways of searching through the dictionary. However, due to the complex nature of the entry, it is important to find a software that would meet the specific needs of this dictionary. Using descriptive and comparative methods, the authors will present the state-of-the-art of RCJHR as well as a model for its digitalization with an aim to update the compilationof RCJHR by engaging digital tools.

References

Atkins, B. T. and M. Rundell. 2008. *The Oxford Guide to Practical Lexicography.* Oxford: Oxford University Press.

Bailey, R.W. 1996. 'Old Dictionaries, New Knowledge'. *Digital Studies/Le Champ Numérique*, 4.

Granger, S. and M. Paquot. 2012. *Electronic Lexicography*. Oxford: Oxford University Press.

RCJHR. 2000. *Rječnik crkvenoslavenskoga jezika hrvatske redakcije. I. svezak: a¹ – vrêdь*. Zagreb: Staroslavenski institut; 2016. *Rječnik crkvenoslavenskoga jezika hrvatske redakcije. II. svezak: vrêdьnь – zapovêdnica*. Zagreb: Old Church Slavonic Institute; 2016. *Rječnik crkvenoslavenskoga jezika hrvatske redakcije. 21. svš.: zapovêdь – zemльnь*. Zagreb: Old Church Slavonic Institute; 2018. *Rječnik crkvenoslavenskoga jezika hrvatske redakcije, 22. svš.: zemльnь – i*). Zagreb: Old Church Slavonic Institute.

Šimić, A. 2017. 'Mikrostruktura i istraživački izazovi obrade Rječnika crkvenoslavenskogajezika hrvatske redakcije'. *Filologija,* 69. 99-128.

Vukoja, V. 2014. 'The corpus of the Croatian Church Slavonic texts and the current state ofaffairs concerning the Dictionary of the Croatian redaction of Church Slavonic compiling.' In A. Abel, C. Vettori, and N. Ralli, (eds.), *Proceedings of the XVI EURALEXInternational Congress: The User in Focus.* 1221-1235.

Embedded voice in seventeenth- and eighteenth-century dictionaries in England

Linda Mitchell

San Jose State University (United States of America)

linda.mitchell@sjsu.edu

In dictionary entries, voice reflects the beliefs and personalities of lexicographers who attempted to distinguish their work from that of their competitors. These efforts sometimes resulted in colorful and entertaining dictionary entries. Examining how seventeenth- and eighteenth-century lexicographers established a voice in their dictionaries moves the study of dictionaries from the mechanics of how dictionaries were constructed to an expanded view of what else was going on in the minds of these lexicographers. Looking at the use of rhetorical devices such as tropes and figurative language helps explain how lexicographers achieved some degree of individuality in their work. It also provides a larger picture of the social, cultural, religious, and political attitudes of the period.

Tropes were convenient for exaggeration and comparisons, especially in entries on politics and religion in which the lexicographer was arguing passionately. In *A Complete Christian Dictionary* (1678) Thomas Wilson uses an analogy as he refers to the Pope in Rome as "A monstrous person, neither man nor woman, God nor Devil; but partaking in a devilish and brutish nature." Another rhetorical device, the enthymeme, was often used to associate the authority of a well-known figure with the accuracy of the entry. In the *Complete English Dictionary* (1772) Frederick Barlow supports his statement with a quotation from Milton's *Paradise Lost*: "Ácceptable is accented on the first syllable by most moderns, who have likewise the authority of Milton on their side. Sometimes it implies merit, and that the thing to be given is worthy to be received with approbation. 'So fit, so acceptable, so divine.' Paradise Lost, Book. iii." The enthymeme from this dictionary entry might be constructed in this way: *Milton's quotations are authoritative. [Missing premise: This entry is a Milton quotation.] Therefore, this entry is authoritative*.

Figurative language, especially imagery, was also used to establish voice and persona. In the *Ladies Dictionary* (1694) John Dunton sets a light, playful tone, while at the same time establishing a more serious persona. In this example, he warns young women not to become overweight: "Bodies of an unwieldy Bulk are to many unpleasing...no one can think it a very pleasing Sight to see a Soul struggling under a mountainous [sic] Load of Flesh." Throughout the dictionary, he relies heavily on figurative language to maintain his persona of being the wise teacher to young ladies.

Thus, most people assume that dictionaries have always been impersonal without a hint of an author's style, beliefs, or attitudes. To the contrary, some lexicographers in seventeenth- and eighteenth-century dictionaries in England, knowingly or not, crafted a voice or persona. Often, the voice was to project authority. Sometimes, it was a creative voice to achieve some degree of individuality. Lexicographers were clever and effective in their use of rhetorical devices to achieve the desired voice in their dictionary entries.

References

Barlow, F. 1772. *A Complete English Dictionary*. London.

Dunton, J. 1694. *Ladies Dictionary*. London.

Wilson, T. 1678. *A Complete Christian Dictionary*. London

Building up a corpus to analyze the use of military terminology during World War Two

Rodrigo Muñoz Cabrera

Nankai University (China)

rodrigo.munozcabrera@unir.net

Military terminology suffered a dramatic evolution during WWII, as soldiers created new strategies, weapons and tactics that required new conceptuological items to denote them. At the same time, and due to several reasons, a variety of synonyms -not only in English but also in other languages- were used instead of the normalized ones. Due to our interest in studying the evolution of military terminology, we carried our such research through Corpus Linguistics and its analysis tools, as they allowed us to analyze and discern the terms used by soldiers at that particular time.

We aimed at observing the evolution of military terminology using journals edited by the US Army during WWII to perform a synchronic analysis, based on the terms normalized in the glossaries published by the US War Department in 1943 and 1944. We compiled a corpus that consisted of 98 texts, and 2,160,147 words. With the aid of *AntConc toolkit*, we aimed whether to confirm or not that the terminological units used in these periodicals were the same ones that those included in the US Army glossaries.

Thanks to the results obtained, we claim that some terminological units evolved -both, from a conceptuological and a lexical point of view- and new ones surfaced; some in English or in the language they were conceived. Besides, the use of some of these terms became so widespread that, nowadays, they are no longer considered terminological items, becoming lexical ones and included in post-war English dictionaries.

The vocabulary of Easter Island (ca 1770) and the indigenous voices collected in the Spanish and English scientific expeditions of the 18th century

Luis Pablo Nuñez
University of Granada (Spain)
luispablo@ugr.es

This communication aims to show an ongoing project: the creation of a multilingual database of voices from indigenous languages collected in the Spanish, English and French scientific expeditions of the 18th-19th centuries. Based on other previous work carried out by ourselves and by others (Alvar Ezquerra 2013; Fernández Rodríguez 2013; Hernández Hernández 2006, 2018; Pablo Núñez 2010, 2015, 2019, 2020a and 2020b), we will describe the origins of the project, its objectives and, as a sample of the philological difficulties found in the original materials, we will analyze the vocabulary of Easter Island (Spanish-Rapanui, ca. 1770), compared with the lexical samples collected by English navigators for this same Malay-Polynesian language.

References

Alvar Ezquerra, M. 2013. 'El desconocido Vocabulario del idioma de los habitantes de Nutka'. In M. L. Calero Vaquera (ed.), *Métodos y resultados actuales en historiografía de la Lingüística. Congreso de la Sociedad Española de Historiografía Lingüística.* Münster: Nodus Publikationen, vol. 1. 15-24.

Fernández Rodríguez, R. 2013. 'The Nootka and Sandwich vocabularies in Relación de la entrada de San Lorenzo de Nutka (1789)'. *Sprachtypologie und Universalienforschung - STUF Language Typology and Universals*. 314-327.

Hernández Hernández, E. 2006. 'Un proyecto de investigación para la historia de la lexicografía hispano-amerindia'. *Boletín de la Sociedad Española de Historiografía Lingüística*, 5. 29-40.

Hernández Hernández, E. 2018. *Lexicografía hispano-amerindia, 1550-1800: Catálogo descriptivo de los vocabularios del español y las lenguas indígenas americanas.* Madrid: Iberoamericana-Vervuert.

Pablo Núñez, L. 2010. *El arte de las palabras, Diccionarios e imprenta en el Siglo de Oro*. 2 vols. Badajoz: Editora Regional de Extremadura.

Pablo Núñez, L. 2015. 'Las exploraciones oceánicas españolas del siglo XVIII y los envíos de plantas y animales desde América'. In M. DeBeni (ed.), *De los descubrimientos a las taxonomías, La botánica y la zoología en la lengua española del Renacimiento a la Ilustración*. Mantova: Universitas Studiorum. 229-260.

Pablo Núñez, L. 2019. 'La importancia de las imágenes en la divulgación de las grandes expediciones científicas'. In M. De Beni (ed.), *Imagen y discurso técnico-científico en español, miradas interdisciplinarias*. Mantova: Universitas Studiorum. 155-178.

Pablo Núñez, L. 2020a. 'Los exploradores europeos del siglo XVIII y el comienzo del estudio de las lenguas indígenas de Norteamérica y Polinesia'. In A. Gallego Cuiñas, A. López and A. Pociña (eds.), *El libro: Reflexiones interdisciplinares sobre la lectura, la edición y la biblioteca*. Granada: EUG. 293-306.

Pablo Núñez, L. 2020b. 'La recopilación de informaciones lingüísticas en las expediciones científicas del siglo XVIII y el método filológico'. *Revista Diálogos* (RevDia, ISSN 2319-0825, Universidade Federal do Mato Grosso, Brasil). 8/2 (Monográfico: O fazer filológico em sctricto e lato sensu). 75-97.

The semantics and syntax of Old English *Prevent* and *Forbid* verbs

Ana Elvira Ojanguren López
University of La Rioja (Spain)
ana-elvira.ojanguren@unirioja.es

This paper deals with the Old English verbal classes *Prevent* and *Forbid* and aims at classifying a set of Old English verbs into these two classes on the grounds of their semantic and syntactic behaviour. The analysis is based on the framework of verb classes and alternations and the theory of Role and Reference Grammar. Levin's (1993) framework of verbal classes and alternations stresses the correlation between the semantics and the syntactic behaviour of verbs, in such a way that the meaning components and the morpho-syntactic realization of the arguments of a group of verbs have to be similar for these verbs to give rise to a verbal class. The analysis of the morphosyntactic similarities and differences is carried out on the basis of Role and Reference Grammar (Van Valin and LaPolla 1997; Van Valin 2005) and comprises *Aktionsart*, macrorole assignment, privileged syntactic argument assignment, nexus and juncture. The inventory of verbs under analysis comprises *āgǣlan, belēan, bewerian, forbēodan, forfōn, forhabban, for(e)sacan, forwiernan, gǣlan* and *tōcweðan.* The main conclusion is that *Prevent* verbs are characterised by the construction of coordination, while *Forbid* verbs are found in subordination constructions and admit passivisation.

References

Levin, B. 1993. *English Verb Classes and Alternations*. Chicago: University of Chicago Press.

Van Valin, R. 2005. *Exploring the Syntax-Semantics Interface*. Cambridge: Cambridge University Press.

Van Valin, R. and R. LaPolla, 1997. *Syntax: Structure, meaning and function*. Cambridge: Cambridge University Press.

The diachronic dimension of the semantic shift 'to take up' → 'to abolish' in the light of typological data

Mariia Orlova

Institute of Linguistics, Russian Academy of Sciences (Russia)

mriaorlova@gmail.com

The presentation is a result of research conducted within the *Database of Semantic Shifts* project, aiming to collect semantic shifts in languages of the world. The electronic database, currently containing 3987 shifts, has been developed since2002 in the Institute of Linguistics in Moscow by a research group under the guidance of Anna A. Zalizniak. This presentation investigates the semantic evolution of verbs meaning 'to raise, lift; to take up' in 11 languages (Romance, Germanic, Greek), attempting to determine its cognitive grounds.

According to the proposed scheme, the development of senses is based on the notion of removing an object from a given space, so that with each subsequent shift the distance between the object and its original location becomes greater. The shifts form a chain of 5 meanings: (1) 'to take up' → (2) 'to take off (e.g., a lid)' → (3) 'to take away' → (4a) 'to destroy', (4b) 'to abolish (e.g., a law)'. The said distance comes into focus at step (2), increases at step (3) and becomes ultimate at step (4) inits concrete (metonymical) and abstract (metaphorical) realizations (4a and 4b), asthe object ceases to exist.

The presentation highlights the divergences between conceptualization models in different languages: some links in the chain are generally omitted and the corresponding situations are denoted by other lexemes (e.g. in Germanic languages, where German *aufheben*, Dutch *opheffen*, Swedish *häva*, Danish *hæve* and Norwegian *heve* only express the meanings (1) and (4b)). An ideal realization of the chain has been discovered only in Ancient Greek, where the lexeme αιρω can denote all the five meanings. The presentation compares the various forms that the schemetakes and suggests some preliminary considerations on its cognitive nature.

References

DatSemshift 2.0. Zalizniak A.A., Bibaeva M., Bobrik M., Gruntov I., Mikhailova T., Orlova M., Russo M., Smirnitskaya A. and M. Voronov. *The Catalogue of semantic shifts in languages of the world*. [http://datsemshift.ru/]

Linking historical sources to modern databases

Tiina Paet, Lydia Raadik

Institute of the Estonian Language (Estonia)

tiinapaet@gmail.com, lydia.raadik@eki.ee,

The most consistently published dictionary in Estonia and still the most popular one is *Õigekeelsussõnaraamat / Dictionary of Standard Estonian*. The past century has seen eleven revised editions of the dictionary. The first was published in 1918, the most recent one appeared in 2018. The spelling dictionary of 1918 laid the foundations of a cultivated and codified national language as well as a tradition of collective corpus planning in Estonia.

The aim of the paper is to introduce a web application from 2018, called *The Spelling Dictionary of Estonian 1918* – a Commented Online Dictionary, which provides comfortable online access to this lexicographic monument of 1918. This application creates a bridge between the linguistic standards of 1918 and 2018. The innovative online dictionary enables a comparison of the entry words across the two dictionaries, thus revealing how much of the linguistic material has been preserved and what a century has done to the spelling of words as well as to their inflection and semantics. This makes the dictionary interesting research material not only for linguists, but also for anyone interested in changes in standard Estonian over a hundred years.

The online dictionary contains 18,012 entries. Every entry word that has been changed, but is still in use has been supplemented with its modern shape, a link to the word as presented in the dictionary of 2018, a comment, and a link to relevant text examples from 1900–1920. This online dictionary is a pilot project aiming to test the potentials and applications of historical lexicography in search of the best way to connect the written monuments of Estonian with modern databases, thus putting studies of standard Estonian into a historical perspective. The experience acquired in the process of this study enables a future supplementation of the online application with data from forthcoming editions of the Dictionary of Standard Estonian.

References

Eesti keele õigekirjutuse-sõnaraamat. 1918 [The Spelling Dictionary of Estonian 1918]. Tallinn: Rahvaülikool.

The Spelling Dictionary of Estonian 1918–A Commented Online Dictionary [http://www.eki.ee/dict/qs1918/].

Õigekeelsussõnaraamat ÕS 2018 [The Dictionary of Standard Estonian DSE 2018]. [http://www.eki.ee/dict/qs/].

Processing optimality in Old English complex sentences containing relatives

Rodrigo Pérez Lorido

University of Oviedo (Spain)

lorido@uniovi.es

This paper focuses on the processing strategies adopted by Old English speakers when embedding or extraposing relative clauses with object antecedents, as illustrated in (1) and (2) respectively:

(1)   Embedded relative clause
     gyf ðu   þas wyrte [ðe man ricinum nemneð] on þinre æhte   hafast
     if   you   this plant   that man ricunum call   in you possession have
     'if you have in your possession this plant that people call 'ricinus'…'

                                                            (coherbar,Lch_I_[Herb]:176.1.2541)

(2)   Extraposed relative clause
     gyf hwa   þas wyrte mid him hafað [þe we peristereon nemdon]…
     if   whoever   this plant   with him has   that we peristereon called
     'if whoever has with him this plant that we called 'peristereon'…'

                                                            (coherbar,Lch_I_[Herb]:67.1.1164)

The analysis assumes the dictates in Hawkins (1994, 2007) *Early Immediate Constituents* (ERC) approach, according to which the structural characteristics of the superordinate and relative clauses in those constructions interact to produce different constituent recognition domains (CRDs) that ultimately determine the choice between extraposition or embedding in terms of processing optimality. Given that the human language processor prefers linear orders that minimise constituent recognition domains, the analysis of the interaction of these parameters in both embedded and extraposed object relative clauses in Old English will provide insightful information about the processing strategies employed by OE speakers in parsing complex sentences, and also about the relevance of processing issues in the change OV →VO in English.

     The study is based on the analysis of an extensive corpus of Old English texts taken from the *York-Toronto-Helsinki Corpus of Old English Prose* and on the comparison of the OE findings with other west Germanic languages with SOV order, such as Modern German.

References

Hawkins, J. 1994. A *Performance Theory of Order and Constituency*. Cambridge: Cambridge University Press.

Hawkins, J. 2007. *Efficiency and Complexity in Grammars*. Cambridge: Cambridge University Press.

From dictionaries to corpus: lemmatization and tagging of Old Czech verbs

Jirí Pergler

Ústav pro jazyk český Institute of the Czech Language (Czech Republic)

pergler@ujc.cas.cz

The aim of the paper is to present our work on lemmatization and tagging of verb forms in *Old Czech Text Bank*, a 6 million corpus of Old Czech being built in the Czech Language Institute of Czech Academy of Sciences.

We will start with a brief overview of the Czech language and the periodization of its development, with special emphasis on the verbal morphology. Czech verbs have rich paradigms of forms, which express morphological categories including person, number, mood, tense or aspect; in Old Czech the repertoire of forms was even larger than it is now.

In the rest of the paper we will concentrate on our work, which follows similar research on common nouns, (Synková 2017; Jínová, Lehečka, and Oliva 2014). In order to perform automatic morphological analysis, we need to generate a list of all possible forms of all Old Czech verbs in all phonological variants, each form being supplemented with the information about lemma and morphological categories. To do so, we make use of modern dictionaries of Old Czech and existing historical grammars, as well as of the internal version of the corpus through which we check and complement the information provided by the older descriptions. The resulting description consists of four parts: (1) the description of conjugation patterns in terms of the distinct endings of respective verb forms, (2) the description of morphophonological alternations, i. e., the formal changes of the verb bases, (3) the list of verb lemmas, where each lemma is conjoined with the conjugation pattern and the type of alternation, and (4) the description of phonological changes in order to generate different phonological variants. In the end, we will show what the final lemmatization and tagging look like in the corpus.

References

Elektronický slovník staré češtiny [Electronic dictionary of Old Czech]. 2006. Prague: Czech Language Institute. [http://vokabular.ujc.cas.cz].

Jínová, P., Lehečka, B., and K. Oliva. 2014. 'Describing Old Czech declension patterns for automatic text analysis'. *Mundo Eslavo: revista de cultura y estudios eslavos*, 13: 7–17.

Staročeská textová banka [Old Czech Text Bank]. Prague: Czech Language Institute. [https://korpus.vokabular.ujc.cas.cz/]

Staročeský slovník [Old Czech Dictionary]. 1968–2008. Prague: Academia.

Synková, P. 2017. *Popis staročeské apelativní deklinace (se zřetelem k automatické morfologické analýze textů Staročeské textové banky) [Description of Old Czech Common Nouns Declension (with regard to Automatic Morphological Analysis of Texts in Old Czech Text Bank)]*. Prague: Faculty of Arts, Charles University.

Erazm Rykaczewski's *Słownik polsko-angielski* … (1851) and its sources:
A comparative analysis

Mirosława Podhajecka

University of Opole (Poland)

mira.podhajecka@wp.pl

Erazm Rykaczewski's *A Complete English-Polish dictionary*…(1849)/*Słownik polsko-angielski*…(1851) was the first biscopal English-Polish/Polish-English dictionary. As well asTL equivalents for SL headwords, it offered a rich selection of illustrative examples. Compilinga bilingual dictionary requires fluency in both one's native language and the foreign language,but Rykaczewski's knowledge of English was limited. This explains why, in the making of theEnglish-Polish volume, he drew on other dictionaries, rendering German and French lexicographical material into Polish.

In the light of the above, how he compiled the Polish-English volume remains unresolved.Based on a handful of examples, Podhajecka (2016: 99) suggests that he might have used the same dictionaries. This paper tests the above hypothesis empirically. The methodology was twofold. Firstly, a sample of English citations was checked against Google Books to select potential sources. Secondly, the bilingual dictionaries which included most of the citations, and in the closest structures (e.g. *he is looking for news by the first post*), were selected for a more extensive analysis. Fleming and Tibbins' *Royal Dictionary English and French…* (1844-1845) transpired to be Rykaczewski's major source. The qualitative and quantitative data resulting from this research will be discussed in detail.

The research indicates that many English examples of usage in Rykaczewski's volume may be found in bilingual dictionaries of English and French (Boyer 1771; Chambaud 1815; Wilson 1833), German (Ebers 1799; Küttner and Nicholson 1813; Flügel 1843; Hilpert 1845), Italian (Baretti 1813), and Portuguese (Vieyra 1773). All this leads one to claim that Rykaczewski's dictionary was deeply rooted in European lexicography. Since searches in digitized resources are time-consuming and prone to error, only a multilingual database of eighteenth- and nineteenth-century bilingual dictionaries would allow for fully-fledged comparative analyses. For the time being, however, it remains a dream for the future.

References

Baretti, G. 1813. *Dizionario delle lingue italiana ed inglese: Italiano ed inglese*. London: F.C.E.J. Rivington.

Boyer, A. 1771. *The Royal Dictionary, Abridged in Two Parts: French and English, and English and French...* London: Printed for Mrs. Bathurst *et al*.

Chambaud, L. 1815. *Dictionnaire francois-anglois & anglois-francois...* Londres: Cadell and Davies .

Ebers, J. 1799. *The New and Complete Dictionary of the German and English Languages…* Leipzig: Breitkopf and Härtel.

Fleming, C. and J. Tibbins. 1844-1845. *Royal Dictionary English and French…* Paris: Didot.

Flügel, J.G. 1843. *Flügel's Complete Dictionary of the German and English Language*s... Adapted by C.A. Feiling and A. Heimann. 2nd edition. London: Whittaker and Co.

Hilpert, J. L. 1845. *A Dictionary of the English and German, and the German and English Language.* Carlsruhe: Printed for and by Th. Braun.

Küttner, C.G. and W. Nicholson. 1813. The *New and Complete Dictionary of the German Language for Englishmen…* Lepzig: Printed for E.B. Schwickert.

Podhajecka, M. 2016. *A History of Polish-English/English-Polish bilingual Lexicography (1788-1947)*. Opole: Wydawnictwo Uniwersytetu Opolskiego.

Rykaczewski, E. 1849-1851. *A Complete English-Polish Dictionary…/Słownik polsko-angielski…* Berlin: B. Beck.

Vieyra, A. 1773. *Dictionary of the Portuguese and English Languages in Two Parts*. London: Printed for F. Wingrave *et al*.

Wilson, J. 1833. *A French and English Dictionary: Containing Full Explanations…* London: Joseph Ogle Robinson.

A first approach to Emil Hahn's *German-Spanish Dictionary of Medical Terms* and its relevance for the TELEME project

Carmen Quijada Diez (University of Oviedo, Spain)

Marta Gómez Martínez (University of Cantabria, Spain)

quijadacarmen@uniovi.es, marta.gomezm@unican.es

Emil Hahn's *German-Spanish Dictionary of Medical Terms* (1914) deserves our attention because it was the first bilingual German-Spanish medical terminology dictionary published in Spain. There are at least two medical encyclopaedias translated directly into Spanish from German, such as that of Hugo von Ziemssen (*Handbuch der speciellen Pathologie und Therapie*), as well as the exceptional case of Albert von Eulenburg's encyclopaedic dictionary (*Real-Enzyclopädie der gesamten Heilkunde*), both rendered into Spanish in the last years of the nineteenth century. Hahn's case is particularly interesting both for the date of its publication (1914) as for the presumable interest that probably led to its creation: the emergence of other languages, such as English and German, as opposed to the hegemonic position of French in the medical field. In thispaper we aim at shedding light on this work, its author, the context in which the work was conceived, its recipients and its basic content.

By presenting this work we also intend to present its usefulness within the TELEME project (*Spanish Medical Lexicographic Thesaurus*), an ongoing project that will be explained during this presentation: it consists of a searchable electronic corpus of nineteenth-century Spanish medical dictionaries that contains not only Spanish-written medical dictionaries and encyclopaedias, but also covers the contribution to medical knowledge made by translated works of this kind and bilingual dictionaries, such as Hahn's work.

References

Hahn, E. 1914. *Diccionario alemán-español de Términos de Medicina*. Madrid: Librería Gutenberg de José Ruiz/Ruiz Hermanos Sucesores

The sources of *Andreas* 1334-41

Tom Revell

University of Oxford (United Kingdom)

tom.revell@balliol.ox.ac.uk

The search for sources has been a perpetual interest in Old English scholarship. Identifying texts that influenced later writers can illuminate compositional technique and provide greater insight into the minds of early medieval English poets. However, most of this research has looked to texts in other languages for comparanda - Latin, Greek, Old High German, Norse - while verbal patterns and parallels between Old English poems have been seen (for the majority of the last century) as a product of an ultimately oral-formulaic tradition.

This paper identifies and analyses the influence on the *Andreas*-poet of earlier poetry in Old English; not of *Beowulf* or the signed poems of Cynewulf as has been recently put forth elsewhere, but of *Guthlac A*. Using a number of digital tools available through the University of Oxford's *CLASP* project, this paper close-reads *Andreas* 1334-41 against the backdrop of the entire surviving corpus of Old English poetry, and suggests that a tight cluster of very rare lexical items in these lines in *Andreas* correspond to *Guthlac A*, in a such manner, number and specificity as to eschew arguments of oral coincidence. This paper goes on to suggest other verse sections that might have influenced the *Andreas*-poet here, and to analyse what such interrelationships can provide to our understanding and interpretation of Old English poetry.

"My principal guide...for the dialectical words of this county": The contribution of *Horae Subsecivae* to James O. Halliwell's *A Dictionary of Archaic and Provincial Words* (1847)

Javier Ruano-García

University of Salamanca (Spain)

fjrg@usal.es

This paper examines the contribution of *Horae Subsecivae* (now Bodl. MS Eng. lang. d. 66) to James O. Halliwell's *A Dictionary of Archaic and Provincial Words* (1847). *Horae Subsecivae* is an obscure manuscript glossary that was possibly compiled by Robert Wight of Wotton-under-Edge in *c*.1777-78 and is now preserved amongst Joseph Wright's papers at the Bodleian Library. Even though it has gained little scholarly attention, *Horae Subsecivae* has a substantial dialect element, with a large number of words cited from Cornwall, Devon, Dorset, Gloucestershire, Somerset, Wiltshire and Worcestershire (Ruano-García 2021). The manuscript, which was conceived as an addition to Junius's *Etymologicum Anglicanum* (first published by Edward Lye in 1743), went unnoticed by other eighteenth-century collections, but it drew the attention of later lexicographers like James O. Halliwell, who listed it amongst his Devonshire sources.

This paper analyses the impact of *Horae Subsecivae* on Halliwell's coverage of Devonshire vocabulary. I combine archival material with quantitative and qualitative methods of analysis in order to determine the proportion of words that are cited from the manuscript, and to assess the treatment they are given. The first (1847), second (1852) and third editions (1855) of the dictionary are scrutinised so as to identify possible changes behind Halliwell's reliance on this unprinted compilation. The findings suggest that Halliwell cited the glossary second-hand from a transcript of the 1830s issued by Sir Frederick Madden (now BL Add MS 78754), which was reproduced carefully in a small number of entries despite being advertised as Halliwell's "principal guide" for Devonshire vocabulary. Inspection of the dictionary shows that he also drew on the manuscript for words used in other western counties (e.g. *leavance*, Gloucestershire), and that he treated it as the only source for some terms (e.g. *dudman*) and senses (e.g. *ford*). This presentation thus seeks to add to previous work on Halliwell's under researched dictionary (e.g. Hultin 1995, Penhallurick 2009), while it shows the usefulness of *Horae Subsecivae* for research on historical lexical dialect variation more generally.

References

Grose, F. 1790. *A Provincial Glossary with a Collection of Local Proverbs and Popular Superstitions. Interleaved and annotated by Sir Frederic Madden*. British Library, Add MS 78754.

Halliwell, J.O. 1847. *A Dictionary of Archaic and Provincial Words*. 2 Vols. London: John Russell Smith.

Halliwell, J.O. 1852. *A Dictionary of Archaic and Provincial Words*. 2 Vols. 2nd edition. London: John Russell Smith.

Halliwell, J.O. 1855. *A Dictionary of Archaic and Provincial Words*. 2 Vols. 3rd edition. London: Thomas and William Boone.

Hultin, N.C. 1995. ''To Shine with Borrowed Splendour': J.O. Halliwell-Phillipps, Thomas Wright, and Victorian lexicography'. *Dictionaries: Journal of the Dictionary Society of North America* 16 .109-150.

Junius, F. 1743. *Etymologicum Anglicanum.* Edward Lye.

Penhallurick, R. 2009. 'Dialect Dictionaries'. In A.P. Cowie (ed.), *The Oxford History of*

English Lexicography. vol. 2. 290-313. Oxford: Oxford University Press.

Ruano-García, J. 2021. '*Horae Subsecivae*, that remarkable glossary of West Country words'. *International Journal of Lexicography*.

Wight, R. *c*.1777-78. *Horae Subsecivae sive additamenta quaedam vocum plerumque topicarum ad Francisci Junii Etymologicum Anglicanum editum ab Eduardo Lye*. Bodleian Library, MS Eng lang d. 66.

The inflection of Latin proper names in *The Old English Martyrology*

Esaúl Ruiz Narbona

University of Sevilla (Spain)

esaulrn@gmail.com

Latin loanwords are a major concern in the study of Old English vocabulary (Kastovsky 1992: 301-17; 2006: 220-6). While general works concentrate on the type of words that entered the Old English language and when, others (Baker 1998) examine more formal aspects, for instance, the type of inflection these loans present in Old English. This paper focuses on the morphological inflection of a specific type of Latin loanwords in Old English, i.e personal names (Marcus or Iulius). These words present a lot of variation in their inflection, ranging from purely Latin morphology to Old English one or a mixture of the two (Campbell 1959: 219). By examining the inflection and linguistic context of the circa 1000 tokens of Latin proper names available in *The Old English Martyrology* (Herzfeld 1900, Rauer 2013), this paper aims at shedding light on fundamental issues concerning the morphology of Latin proper names in Old English: Is the number of Latin-inflected names higher than Old-English inflected ones? How do they adapt to the Old English morphological system? Is the choice of inflection name-dependent or context dependent? Are some morphological cases more readily adapted to Old English than others? Preliminary results show that feminine names normally adopt Old English inflections, while variation is higher in masculine names. Additionally, names tend to adopt a set of Old English inflections based on their gender and not on their original Latin declension. Certain morphological cases such as the dative are more prone to show Old English inflections than the accusative. Furthermore, contextual factors such as adjacency to other names or the adjective *sanctus* seem to influence the type of inflection displayed by names.

References

Baker, P.S. 1998. 'The Inflection of Latin Nouns in Old English Texts'. In P. S. Baker and N. Howe (eds.), *Words and Works: Studies in Medieval English Language and Literature in Honour of Fred C. Robinson*. Toronto: University of Toronto Press. 187-206.

Campbell, A. 1959. *Old English Grammar.* Cambridge: Cambridge University Press.

Herzfeld, G. (ed.). 1900. *An Old English Martyrology.* London: EETS. 116.

Kastovsky, D. 1992. 'Semantics and Vocabulary'. In R. M. Hogg (ed.), *The Cambridge History of the English Language. Volume I: The Beginnings to 1066*. 290-480.

Kastovsky, D. 2006. 'Vocabulary'. In R. M. Hogg and D. Denison (eds.), *A History of the English Language.* 199-270.

Rauer, C. (ed.). 2013. *The Old English Martyrology: Edition, Translation and Commentary.* Cambridge (USA): D.S. Brewer.

A comprehensive lexicographic study of the history, distribution and frequency of relexified diminutives in modern Italian (with comparisons to Spanish)

John Ryan (University of Northern Colorado, United States of America)

Victor Parra-Guinaldo (American University of Sharjah, United Arab Emirates)

john.ryan@unco.edu, egeria381@gmail.com

Following a previous analysis for Spanish (Author 2016), this study for Italian is the second phase of a larger project that examines relexification of diminutives across the Romance languages. As such, this paper comprises a quantitative lexicographic analysis of the entirety of diminutives, relexified in the history of Italian, by utilizing three separate dictionaries: 1) Merz's online *Dizionario inverso dell'italiano moderno* (2004) for isolation of potential diminutive forms; 2) the state-of-the-art digital online *Grande dizionario Garzanti della lingua italiana* (2019) to assess the range of meanings for each entry and to eliminate from the count words that do not originate in diminutives (e.g., *clandestino* 'clandestine' or *benedetto* 'blessed'); and 3) Pharies's *Diccionario etimológico de los sufijos españoles y otros elementos finales* (2002), supplemented by Rohlfs' *Grammatica storica della lingua italiana e dei suoi dialetti. Sintassis e formazione delle parole* (1969) to assess the frequency and relative timing of entry of historically based diminutive forms, both *ad hoc* and relexified, into the Italian lexicon. When compared to results for Spanish by Author (2016), lexicographic data of this study suggests that Italian followed a very different trajectory of diminutive relexification from Latin than that found for Spanish. In other words, Italian appears to have developed a preference for alternate *ad hoc* diminutive suffixes such as -ino/a and -etto/a (based on non-diminutive Latin forms) at a much earlier period than did Spanish, allowing for both greater absorption and the time necessary for these to relexify. Lexicographic data for Spanish, on the other hand, suggests that this language instead continued to favor reflexes of the original Latin diminutive suffixes. The authors propose the reasons for this divergence is the relatively early colonization of the Iberian Peninsula and continued preference for traditional Latin diminutive endings over innovative endings that were being adopted Empire-wide, beyond Castile, including other regions of Hispania.

References

Author. 2016. 'Classification and History of Relexified Diminutives in Modern Spanish: A Lexicographic Approach'. In M.V. Rodríguez Domínguez *et al.*, *Words across History. Advances in Historical Lexicography and Lexicology*. Las Palmas de Gran Canarias: Servicio de Publicaciones y Difusión Científica de la ULPGC. 364-380.

De Agostini Scuola S.p.A. Garzanti Linguistica. 2018. *Il grande dizionario Garzanti della lingua italiana*. Milano.

Merz, G. 2004. *Dizionario inverso dell'italiano moderno*.

Pharies, D. 2002. *Diccionario etimológico de sufijos españoles y otroselementos finales*. Madrid: Editorial Gredos.

Rohlfs, G. 1969. *Grammatica storica della lingua italiana e dei suoi dialetti.Sintassis e formazione delle parole*. Piccola Biblioteca Einaudi.

Is there a place for orthographic dictionaries in the 21st century?

Ana Salgado, Rute Costa, Toma Tasovac

University of Nova de Lisboa (Portugal)

anacastrosalgado@gmail.com, costamrv@gmail.com, ttasovac@humanistika.org

Defining orthographic dictionaries as works that present orthographic information (Hartmann and James 1998) does not facilitate a clear understanding of what these works really are, since any language dictionary does the same. Our first concern was to clarify what an orthographic dictionary is, and in asking this question, we explored the main characteristics of orthographic dictionaries. Language dictionaries can be categorised into general language dictionaries and specialised dictionaries. An orthographic dictionary belongs to the latter subcategory since its object is a specialised element of the linguistic description, that is the orthographic form of the lemma, whereas normativity is the central functional feature (Buchmann 2015). Our second concern was related to Portuguese orthographic dictionaries as they still have not been the subject of in-depth study that considers the long and rich tradition of orthographic vocabularies (Hartmann and James 1998) in Portugal. This contribution also aimed to offer an overview of the history of Portuguese orthographic vocabularies (Verdelho 2013) where they play(ed) an essential role as a normative instrument. On the other hand, we sought to question the usefulness of this type of works in a digital environment in which they are overtaken by general dictionaries themselves and spellcheckers.

This paper has three main sections: 1) First, we address the theoretical framework, focusing on orthographic dictionaries. The notions of orthography and normativity are discussed, and we question if there is a place for orthographic dictionaries in this century. 2) Then, we engage in the macrostructural and microstructural analysis of the *VOLP-1940 – Vocabulário Ortográfico da Língua Portuguesa* [Portuguese Language Vocabulary] (4: 5), published by the Lisbon Academy of Sciences, and explore the main characteristics of this work. 3) We discuss the pertinence of this type of dictionary and its role and place in a digital environment, aiming to preserve the original and historical work, pointing out the added value that these works bring compared to contemporary spellcheckers.

References

Academia das Ciências de Lisboa. 1940. *Vocabulário Ortográfico da Língua Portuguesa*. Lisboa: Imprensa Nacional.

Buchmann, F. 2015. *Spelling Dictionaries. The Oxford Handbook of Lexicography*. DOI: 10.1093/oxfordhb/9780199691630.013.45.

R. R. K. Hartmann and G. James. 1998. *Dictionary of Lexicography*, London and New York: Routledge. 130.

Salgado, A., and R. Costa. 2020. 'O projeto 'Edição Digital dos Vocabulários da Academia das Ciências': o VOLP-1940'. *Revista da Associação Portuguesa de Linguística*, 7. 275–294.

Verdelho, T. 2012. 'A História e a Importância dos Vocabulários Ortográficos da Língua Portuguesa'. *Vocabulário Ortográfico Atualizado da Língua Portuguesa*. Lisboa: Imprensa Nacional Casa da Moeda.

Lexicon leponticum – An online dictionary of Cisalpine Celtic

Corinna Salomon

University of Vienna (Austria)

corinna.salomon@univie.ac.at

Lexicon Leponticum (LexLep) was created 2009–10 in the course of a research project led by Prof. David Stifter at the University of Vienna. The project's aim was to create an online dictionary of the Celtic languages – Lepontic and Cisalpine Gaulish – which are encoded in the inscriptions of the *Cisalpine Celtic corpus of Iron Age Northern Italy*. In an effort to provide complete sources for all lexemes, the database in its original form ended up as more of an epigraphic edition than a lexicon, and remained unfinished after the end of the project. Starting in March, LexLep will be completed and updated the international research project CCeLL – *Cisalpine Celtic Languages and Literacy*, funded by the Austrian Academy of Sciences (2020–22). CCeLL will tackle both the epigraphic and the linguistic aspects of LexLep, in that a major focus lies on the completion of the lexicon of forms and the establishment of reliable criteria by which to distinguish between the two linguistic strata.

The software which underlies LexLep is MediaWiki with a semantic extension – all data is annotated and amenable to semantic search. The rigid structures imposed by semantic properties which are encoded on the structural level (in opposition to more flexible annotation methods such as TEI/EpiDoc) have offered some challenges to the effective mark-up and display of lexicographical information. The lexicon in its original (and present) state has issues concerning the way it deals with variation, uncertainty and optional features, as well as the encoding of non-standardisable data.

The talk will discuss the lexicon's problems and possible solutions, including the treatment of variant readings, abbreviations and incomplete forms, methods for linking related forms, and options for encoding levels of uncertainty in ascription and relations between graphematic, phonological and morphological elements.

Thesaurus linguæ Latinæ 1890-2021

Josine Schrickx

The Bavarian Academy of Sciences and Humanities (Germany)

j.schrickx@thesaurus.badw.de

The Thesaurus linguae Latinae is probably the most comprehensive dictionary in the world and the oldest still in progress. It was founded in the 19th century with the aim of covering all extant Latin literature up to around 600 AD. Not only literary works were to be included, but everything that has been handed down: graffiti, inscriptions, manuals for doctors, soldiers' letters, theological treatises, etc. For this purpose, an extensive archive of paper slips was established, which is still in use today. The dictionary is being compiled by an international team based at the Bavarian Academy of Sciences and Humanities. It consists of a core group of more or less permanent staff and changing junior researchers, who thus get the chance to gain lexicographical experience.

In this talk, I will briefly present the the history of the TLL and the project's working method. Then I will explain how the project has tackled the digital challenges of the last three decades. Of course, we now use digital text databases in addition to slips of paper. Furthermore, the dictionary is published in an online database by the publishing house De Gruyter in addition to the print publication in book form. In addition, we offer open-access PDFs of the volumes published so far. Our next challenge is to open ourselves even further to the digital world.

References

De Gruyter Database. [https://tll.degruyter.com/]

Thesaurus linguæ Latinæ. Open access.
        [https://publikationen.badw.de/en/thesaurus/lemmata]

TLL. *Thesaurus linguæ Latinæ*. 1890-2021. [http://thesaurus.badw.de/en/project.html.]

Australian food slang: A study of restaurant menus

Dana Serditova

University of Freiburg (Germany)

danaserditova@gmail.com

This paper deals with Australian food slang and its representation in Australian restaurant menus, corpora, food blogs and fora, publicist texts. What is understood by *food slang* in the context of this research are words formed by reduction and affixation – addition of suffixes such as "-ie", "-y", "-o" (Plag 2018) to form words like *brekkie*, *benny*, *avo*. In order to provide a comprehensive account of food slang words found in Australian menus, this paper tackles such issues as what slang is: its definition, purpose, characteristics (Partridge 1933; Adams 2009; Coleman 2012); what kind of slang words can be found in Australian restaurant menus; regional variation patterns in the use of food slang; and discrepancies in dictionary definitions compared to corpora examples and the data gathered during the menu research.

The research consists of three main stages: the menu research (Australian cuisine restaurants); the classification of the observed data (including the sociolinguistic maps compiled by the author) and the analysis of dictionary entries (the *OED, Macquarie Dictionary,* etc.) and corpora examples (*Corpus of Global Web-Based English, the iWeb Corpus,* etc.); as well as a social survey conducted among Australians of different age, gender, and from various parts of the continent, with the goal of supporting or refuting the observations made during the previous stages. This methodology allows us to see whether there are any inaccuracies or gaps in dictionary entries and definitions as opposed to real-life use based on the menus and corpora research.

As a result, a compilation of Australian food slang with definitions, use, and regional variation was created, which can be used for further completion of dictionary entries. Some examples include *barra, poachie, schnitty, qukes, lollies*.

References

Adams, M. 2009. *Slang*: *the people's poetry*. Oxford: Oxford University Press.

Coleman, J. 2012. *The Life of Slang*. Oxford: Oxford University Press.

Partridge, E. 1933. *Slang To-Day and Yesterday*. London: Routledge & Kegan Paul Ltd.

Plag, I. 2018. *Word-Formation in English.* Second edition. Cambridge: Cambridge University Press.

Latin lexical-semantic thesaurus of the end of the XIIIth century from Soldaia

Alexander Shaposhnikov

Vinogradov Institute of the Russian Language, Russian Academy of Sciences (Russia)

possidima@gmail.com

The report is dedicated to the remarkable lexico-semantic thesaurus of the 1290s compiled in the office of the Venetian consul in the Republic of St. Stefan (the city of Soldaia Taurica) and included in the incunabula known as Codex Cumanicus (1320s, the city of Kafa, colony of Ghazaria), pages 70 to 110, between grammar and the "book of Catholic missionaries".

The following lexical-semantic fields (groups) are distinguished in Thesaurus: 1) Nomina rerum quae pertinent Deo et ad serviendum ei [70.5]; 2) Haec sunt elementa [71.1]; 3) Haec sunt complexiones [71.6]; 4) Haec sunt tempora [71.11]; 5) Sentimenta quinque [72]; 6) Qualitates temporis [73.1]; 7) Res contrariae [73.25]; 8) Qualitates rerum [75.19]; 9) Nomina rerum que communiter inueniunt [78.1]; 10) Nomina artium et quae pertinent eis [80.1]; 11) Haec continent de speciario et speciaria [80.19]; 12) Faber et ferarius [84.1 + 79.18-21]; 13) Peliparius [84.22]; 14) Sartorius [85.18]; 15) Calegarius [86.1]; 16) Magister asciae [86.1]; 17) Barberius [86.26]; 18) Scriba [87.3]; 19) Nobilitas hominum et mulierum [90.5]; 20) Mercimonia quae pertinent ad mercatore [91.1]; 21) Nomina lapiduum precioxorum [93]; 22) Haec sunt membra hominum et mulierum [93]; 23) De parentatum (parentela) homins [97.1]; 24) Complementa hominum [97.23]; 25) Defecta hominum [98]; 26) Res quae pertinent ad bellum [100.1]; 27) Res oblitae necessariae pro domo [100.11]; 28) Res quae pertinent ad vestes [101.1]; 29) Res quae pertinent ad domum [101.11 + 79.15-16]; 30) Res quae pertinent ad axmisium hominis [101]; 31) Res quae pertinent equo [102]; 32) Res quae pertinent ad lectum [103]; 33) Res quae pertinent ad mensam [104]; 34) Res quae pertinent ad cochina (coquina) [104 +79.17]; 35) Victualia qua nascunt Nomina arborum et fructuum [105 + 88.28-30; 89.1-3]; 36) Nomina herbaria [106.21]; 37) Nomina bestiarum [107]; 38) Nomina vermium [108]; 39) Nomina volucrum [109]; 40) Victualia quae nascunt [110.1].

In addition, several more lexico-semantic fields are included in the above sections without highlighting the headings: 18a) alia ministeria [87.14-31, 89.4-10]; 18b) murator [88.1-6]; 18c) impinctor [88.7-10]; 18d) ligator [88.11-17]; 18e) fornarius [88.18-27]; 18f) cantator [89.11-22]; 18g) meretrix [89.23-25]; 18h) magister scholarum [89.26-29, 90.1-4].

The repertoire of the lexico-scemantic groups of Thesaurus and their lexical content can be supplemented from the combined index of Latin glosses of the entire Codex and the terminology of the Massaria of Cafa of the XIVth century.

In its full form, this Thesaurus adequately reflects the linguistic picture of the world of Venice, Genoa and its colony of Ghazaria at the end of the XIIIth century – early XIVth century.

Principles of shooting terminology formation in Evenki dialects

Karina Sheifer

Institute of Linguistics, Russian Academy of Sciences (Russia)

Institute for System Programming, Russian Academy of Sciences (Russia)

School of Linguistics, Higher School of Economics (Russia)

kar.sheifer@gmail.com

Evenki, an endangered Tungusic language spread in Siberia and Far East, is a language of indigenous hunters and nomads. The semantical field of shooting is highly detailed in Evenki. The variety of its lexicon is also enhanced with the strong dialectal division. Evenki is split into fifteen dialects, which in turn are subdivided into fifty local varieties in total. Besides the richness of the vocabulary, the lexical competence of the native speakers is rapidly sinking into oblivion. The reasons are the abandonment of traditional way of life and transfer to Russian as the everyday language.

The dictionaries compiled throughout the XX century attest shooting vocabulary at very superficial level. The latest edition of the Russian-Evenki dictionary (Vasilevich 2005) contains the lexicon referred to flintlock firearms. These weapons were in wide use in the early-to-mid 19th century. Nevertheless, in Siberia they remained in use in the twenties and thirties of the twentieth century, the epoch of the active codification of the Evenki language. Since the Evenki shooting vocabulary has been no longer studied until recently, the contemporary dictionary indicates the Evenki term for "flintlock mechanism" as a translation for "breechblock".

The lexicon regarded in the present study was gathered throughout the interviewing and elicitation of the native speakers of different Evenki dialects. It was also extracted from the narratives that centralize the hunting stories. The systematized vocabulary includes the verbal and nominal lexicon denoting actions performed by arms (e.g. to recoil) and by shooter (e.g. to pull the trigger). It concerns the nature of hitting shot target and bullet behavior (e.g. to hit below the target). The lexicon also includes weapon component and complement names.

References

Boldyrev B. V. 2000. *Evenki-Russian dictionary.* Part 1, 2. Novosibirsk: Siberian Branch of the Russian Academy of Sciences.

Klyachko, E., Kazakevich O. *Evenki text corpus.* Russian Academy of Sciences Project [www.gisly.com].

Shirokogorov, S. M. 1935. *Psychomental Complex of the Tungus.* London: Kegan, Paul, Trench, Trubner.

Vasilevich G. M. 1958. *Evenki-Russian dictionary.* Moscow: State publishing house of foreign and national dictionaries.

Vasilevich G. M. 2005. *Russian-Evenki dictionary.* Part 1, 2. Saint-Petersburg: Branch of publishing house "Prosveshchenie".

The catalogue of semantic shifts and the case of kinship terms in Old Tamil

Anna Smirnitskaya

Institute of Oriental Studies, Russian Academy of Sciences (Russia)

nyushas@gmail.com

The concept of "semantic shift" is a notion on the crossroads of the fast-developing directions in nowadays linguistics (Haspelmath 2003; Juvonen and Koptjevskaja-Tamm 2016). Historical lexicology, etymological studies, theory of grammaticalization and semantic typology are using similar notions as an instrument for analysis (Vanhove 2008; Zalizniak *et al.* 2012). One of the projects in this theoretical field is *The Catalogue of Semantic Shifts in languages of the world* being developed by the group of researchers under the guidance of Anna A. Zalizniak in the Institute of Linguistics (Moscow). A "semantic shift" in this methodology is understood as a "cognitive proximity of two meanings A and B", connected either synchronically or diachronically (Zalizniak 2018). The realizations of semantic shift can be: synchronic polysemy, diachronic semantic evolution, cognates, morphological derivation and borrowings [ibid]. By now, more than 4000 semantic shifts from more than 300 languages of the world are described in *The Catalogue of Semantic Shifts* (*DatSemshift 2.0.*).

In this work we consider semantic shifts in kinship terminology in Old Tamil. Contrary to popular view, kinship terms change over the centuries. Kinship terms in OT expressed possessive grammatical meanings not preserved in Modern Tamil: OT *entai* 'my father', *nuntai* 'your father' > MT *tantai* 'father', OT *yāy* 'my mother' > *tāy* MT 'mother' (Lehmann 1997). We examine semantic shifts from kinship meanings to other domains: *mother > main, mother > goddess, mother > interjection of surprise, companion > sister,* together with the semantic shifts within the domain of kinship: *mother's brother > father-in-law, grandfather > grandson* etc. We take into consideration diverse examples from other Dravidian kinship systems from Proto-Dravidian to Kannada and Brahui (*STARLING*), as well as from genetically unrelated languages from the *TheCatalogue of semantic shifts.*

References

DatSemshift 2.0. Zalizniak A.A., Bibaeva M., Bobrik M., Gruntov I., Mikhailova T., Orlova M., Russo M., Smirnitskaya A. and M. Voronov. *The Catalogue of semantic shifts in languages of the world*. [http://datsemshift.ru/]

Haspelmath, M. 2003. 'The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison'. In M. Tomasello (ed.), *The New Psychology of Language*. vol. 2. 211-243.

Juvonen P. and M. Koptjevskaja-Tamm (eds.). 2016. *The lexical typology of sematic shifts*. Berlin: De Gruyter Mouton.

Lehmann T. 1997. *Old Tamil, from: The Dravidian languages.* Routledge: S.B.Steever.

STARLING. *Etymological databases*. Starostin S.A. [http://starling.rinet.ru/]

Vanhove M. (ed.). 2008. *From polysemy to semantic change. Towards a typology of lexical semantic associations*. (Studies in language companion series, 106). Amsterdam.

Zalizniak, A.A. 2018. 'The Catalogue of Semantic Shifts: 20 Years Later'. *Russian Journal of Linguistics,* 22 (4). 770-787.

A diachronic study of the late emergence of deverbal -some adjectives in OED3

Chris A Smith

Université de Caen Normandie (France)

chris.smith@unicaen.fr

This exploratory lexicographic study of 262 -some adjectival derivatives aims to study the lateemergence of deverbal some adjectives in OED3.

In a first step, using the OED3 data to generate 262 -some adjectives, a timeline of output is proposed to determine the historical output of the formation, with a view to identifying potential productivity. A morphological analysis shows that -some output is divided into denominal adjectives (Nsome: 116), deverbal adjectives (Vsome: 68), deadjectival adjectives (Adjsome: 32). Further categories include N/V some 26, and OTHER (20, including some unknowns and numerals).

A semantic analysis is then carried out, using the OED definitions to identify key senses of the adjectival -some forms, producing 8 major semantic behaviours: 1) RELATED TO N/V (tastesome, wintersome, toysome); 2) TYPICAL OF (jokesome, marmsome, mothersome, friendsome); 3) TENDENCY TO (muddlesome 1887, moansome, picksome, growsome 1579); 4) FULL OF N (quirksome, perilsome); 5) CAUSING V or N (picklesome 1885, chillsome); 6) ADJ ISH (lithesome, roughsome, weirdsome); 7) GROUP OF N (twosome, threesome, foursome); 8) INTENDED FOR V (relishsome, cuddlesome, wieldsome).

These 8 semantic patterns display a high morphosemantic correlation between semantic type and base word type. They also display some overlap, showing semantic extension through metonymical processes (mainly CAUSE/EFFECT). A discussion of the 8 types is provided with a view to identifying primary and secondary senses.

In a second step, a correlation between morphosemantic types and timeline output is carried out to detect historical patterns of output. The patterns of of Vsome and Nsome output are shown to exhibit major increases at two different time periods. Notably, Nsome consistently outperforms Vsome until 1750, at which point Vsome briefly but considerably outperforms Nsome until 1900.

Finally, this paper aims to focus on the sudden late development of Vsome adjectives and their specificity compared to Ving adjectives and Vable adjectives such as boresome vs. boring, likesome vs. likeable.

Corpus PalaeoHibernicum – **A deeply annotated lexicographic database of Early Irish**

David Stifter (Maynooth University, Ireland)
Bernhard Bauer (Universität Graz, Austria)
Elliott Lash (Georg-August-Universität Göttingen, Germany)
Fangzhe Qiu (University College Dublin, Ireland)
Nora White (Maynooth University, Ireland)
David.Stifter@mu.ie, Bernhard.w.bauer@gmail.com, elliottjamesfrick.lash@uni-goettingen.de, fangzhe.qiu@ucd.ie, nora.white@mu.ie

The ERC Consolidator Grant-funded project *Chronologicon Hibernicum* (*ChronHib*) at Maynooth University in Ireland (2015–2021) had the aim of advancing the knowledge of language variation and change in Old Irish (broadly defined 600–950 A.D.) and refining the absolute chronology of Early Irish linguistic changes, be they phonological, morphological or syntactic. One of the major outcomes of the project is the deeply annotated lexicographic database cum corpus *Corpus PalaeoHibernicum* (*CorPH*). It currently includes 78 texts, mostly from the Old Irish, but also from the early Middle Irish period. The entire lexicon of these texts has not only been annotated for traditional categories such as morphology or PoS, but the *ChronHib* team also developed a system of variational tags, i.e. encoding language change in the annotation. The powerful search interface of *CorPH* allows very easy and very detailed diachronic studies of Old Irish. Combined with newly-developed Bayesian statistical methods in order to handle the often not-so-big data, CorPH permits better insights into the chronological development of Irish on a micro-level. In this presentation, we want to introduce CorPH and we want to demonstrate how the principles underlying its structure lead to new insights into the description of a medieval language.

Digitalisation of the Romanian lexicographic research

Elena Tamba

Alexandru Ioan Cuza University Iași, (Romania)

isabelle.tamba@gmail.com

Lexicography around the world has undergone an extensive process of change, modernization of the means of writing, consulting, etc., through approaches that involve the interconnection of different fields of research. Romanian lexicography is also marked by this change.

In the last years, in Romania, measures have been taken for creating electronic instruments and resources that are necessary for supporting the Romanian language and culture on a transnational level, in the general context of the computerization of the fundamental academic research. The Romanian professionals in linguistics and in computational linguistics have initiated research projects by which they want to valorise the non-digitized resources by acquiring them in electronic formats and to create new resources and instruments for the automatic processing of the language. Electronic dictionaries and texts corpora, structured as databases facilitate knowing, preserving and maintaining the cultural identity on a linguistic level and they allow the inclusion of a national language in the field of interest of the digitalized research of natural languages, at a global level.

This paper aims at outlining the main projects and results in the field of computational Romanian lexicography, for both contemporary and premodern dictionaries.

Bosworth-Toller's Anglo-Saxon Dictionary Online

Ondrej Tichy

Charles University (Czech Republic)

ondrej.tichy@ff.cuni.cz

The *Anglo-Saxon Dictionary* has been the leading lexicographical resource in the study of Old English since its publication. The online version at *bosworthtoller.com* has attempted to bring all the paper dictionary has to offer in a more accessible digital form but also introduce new features made possible by its transformation. This paper presents the history and progress of the ongoing digitization project as well as the brand new beta version of the dictionary. We introduce it expecting to generate feedback that would benefit the new version that is to replace the current site in summer 2020, but we also hope to showcase research that stems from and is part of the dictionary. The introductory part focuses on highlighting the technical, but especially the lexicographical difficulties we have met during more than a decade of development: like the structural inconsistency of the dictionary, the level of fidelity in its digital representation, the disambiguation of some of its data and the reliability of resources that testify to a different age of information processing. The following part shows how the dictionary has been used as a resource for further research and how this can benefit the future of both the dictionary and other projects (e.g. the tools for automatic lemmatization and search). In the last part, the new version is presented in contrast to the current one and in comparison to similar projects, such as the DOE, MED or DSL. In addition to lexicographical features, the technical and design aspects are discussed.

References

Bosworth, J. *An Anglo-Saxon Dictionary Online*. Eds. T. Northcote Toller *et al.* Comp. S. Christ and O. Tichý. Faculty of Arts, Charles University, Prague. [http://www.bosworthtoller.com/].

Mapping the Murray Papers in the Bodleian Library

Stephen Turton, Charlotte Brewer

Hertford College, Oxford (United Kingdom)

stephen@turton.co.za, charlotte.brewer@hertford.ox.ac.uk

During his editorship of the *Oxford English Dictionary*, J. A. H. Murray posted so many letters in pursuit of his inquiries that the British Post Office installed a pillar box outside his Oxford house. Research conducted via letter-writing played a vital role in the construction of this revolutionary new dictionary, published in instalments by Oxford University Press between 1884 and 1928. Many of the resulting documents now occupy the best part of 9 linear metres of shelf space in the Bodleian Library. They include correspondence with famous individuals (such as the poet Tennyson or the British prime minister William Gladstone) along with hosts of lesser known individuals - volunteer readers, advisers, and ordinary members of the public. Apart from vividly testifying to the *OED*'s status as a largescale collaborative enterprise, these letters are a unique and invaluable resource for researchers of late 19th- and early 20-century English language and literature, the social and intellectual networks of the OED's compilers, and the late modern history of lexicography, linguistics, and other disciplines.

Though individual letters from the Bodleian's collection have been quoted in a number of articles and monographs on the history of the *OED* and its compilers (see references), the Murray papers have never been systematically edited or published. The authors of this proposal have recently embarked on a pilot digital edition of a sample of the correspondence, and our paper will review some of the key issues involved in editing and digitizing the documents it contains. In particular, we'll discuss decisions on web design and content, on the thorny issue of copyright, on criteria for choosing which letters to publish – for example, relating to *OED*'s history and methodology, to issues of linguistic correctness, and to literary censorship – along with the project's envisioned outputs.

References

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

Gilliver, P. 2016. *The Making of the Oxford English Dictionary*. Oxford: Oxford University Press.

Mugglestone, L. 2000. *Lexicography and the OED: Pioneers in the Untrodden Forest*. Oxford: Oxford University Press.

Mugglestone, L. 2005. *Lost for Words: The Hidden History of the Oxford English Dictionary*. New Haven and London: Yale University Press.

Murray, K.M.E. 1977. *Caught in the Web of Words: James Murray and the Oxford English Dictionary*. New Haven and London: Yale University Press.

Ogilvie, S. 2013. *Words of the World: A Global History of the Oxford English Dictionary*. Cambridge: Cambridge University Press.

'And (from) him who takes away your cloak, do not withhold your tunic either': Old English ditransitives for Retaining

Juan Gabriel Vázquez-González

University of Huelva (Spain)

juan.gabriel@dfing.uhu.es

One of the major verbal categories associated with the ditransitive construction in Old English is Retaining (Barðdal, Kristoffersen and Sveen 2011: 53-104; Vázquez-González and Barðdal 2019: 586-587). This conceptual domain is made up of two verb subclasses, namely verbs of hindrance (*He sette him weorca mæstras*, 'He [the Pharaoh] put slave bosses in charge of them.' *Exodus* 1:11) and constraining (*Ðæt he him Norþ-Wealas gehyrsumode*, 'that he might make the North Welsh obedient to him.' Chr. 853). Even though the weight of this domain is drastically reduced in Middle English times (Zehentner 2016) and onwards, the category still operates nowadays in verbs of refusal (*refuse*, *deny*, etc.), which are acknowledged as sense C (*Agent causes recipient not to receive patient)* in Goldberg's (1995: 38) polysemous analysis for the ditransitive.

In semantic terms, this category profiles notions such as obedience, slavery, obligations and/or duties in Old English. Some of the verb-specific constructions involved may imply no transfer, others refuse transfer directly (see above) and the rest somehow elaborate on transfer malefactively. Accordingly, they are usually typologized as belonging to the Malefactive Source Construction (Malchukov, Haspelmath and Comrie 2007). However, it is possible that notions like, for instance, *obeying* or *serving* may also be conceptualized benefactively and the constructions involved may then ascribe to the External Possession Construction instead (cf. also Allen 2019: 254-256). In this paper, we will undertake a systematic analysis of all the quotations available in the *Dictionary of Old Engish Corpus* (Healey 2000) for the 26 Retaining terms we have proposed (Vázquez-González and Barðdal 2019: 586-587) in order to verify this possibility. Our results will prove that even though malefactives are the norm, there is also room in the cited corpus for using some of these verb-specific constructions benefactively.

References

Allen, C. 2019. *Dative External Possessors in Early English*. Oxford Studies in Diachronic and Historical Linguistics. Oxford: OUP.

Barðdal, J., Kristoffersen K.E. and A. Sveen. 2011. 'West Scandinavian ditransitives as a family of constructions: With a special attention to the Norwegian V-REFL-NP construction'. *Linguistics,* 49 (1). 53-104.

Goldberg, Adele E. 1995. *Constructions: A construction grammar approach toargument structure*. Chicago: University of Chicago Press.

Healey, A. diPaolo (ed.). 2000. *The dictionary of Old English corpus in electronic form*. Toronto: Pontifical Institute of Medieval Studies, University ofToronto.

Malchukov, A., Haspelmath, M. and B. Comrie. 2007. *Ditransitive constructions: A typological overview*. First draft.

Vázquez-González, J.G and J. Barðdal. 2019. 'Reconstructing the ditransitive construction for Proto-Germanic: Gothic, Old English and OldNorse- Icelandic'. *Folia Linguistica Historica,* 40 (2). 555-620.

Zehentner, E. 2016. *On Competition and Cooperation in Middle English Ditransitives*. Vienna: University of Vienna. Ph.D. Dissertation.

Lexical functions and meaning definition: QUANTITY, SIZE and RANK

Raquel Vea Escarza

University of La Rioja (Spain)

raquel.vea@unirioja.es

This study explores the derived nominal and adjectival lexicon of Old English with a view to completing and systematizing the meaning definitions of affixes in dictionaries. Considering the limited, and somewhat inconsistent, information on affixes available in the Old English dictionaries of reference, this work takes a lexicological and semantic approach to this part of the derived lexicon and proposes the application of functions and macrofunctions in order to gain accuracy on this matter. The starting point are Old English derived nouns and adjectives -as stored in the lexical database *Nerthus*- that perform a lexical function comprised under the semantic categories QUANTITY, SIZE and RANK. A total of twenty-five affixes, fifteen of which are prefixes (*and-, arce-, eall-, for-, fore, forð-, fram-, frēa-, full-, ofer-, sin-, sub-, twi-, ðri-, under-*) and ten are suffixes (*-en*, *feald*, *-ful*, *-icge, -iht, -incel, -le, -ð, -ung, -welle*), have been identified with this semantic value, forming 223 predicates. For each affix participating in these semantic categories, a definition has been provided on the basis of a linguistic analysis that brings together all the related meanings associated with the affix in question and that also includes information regarding the source and target categories involved in the derivation process and the full list of predicates that conform to the rule. The main contribution of this study is, therefore, the arrangement of meaning definitions in an organized and improved system that permits to enhance the lexicological and semantic foundations of dictionaries.

The loss of directional deixis in the history of English

Letizia Vezzosi (University of Firenze, Italy)

Ekkehard Koenig (Freie Universität Berlin, Germany)

letizia.vezzosi@unifi.it, koenig@zedat.fu-berlin.de

Just like most of the Romance languages, Modern English has lost most of its expressions denoting directional deixis in the course of a general reduction of its deictic expressions and especially of its system of demonstratives. Old English, by contrast, still manifests a clear distinction between 'location' and 'direction' in its systems of demonstratives and interrogatives, like many other Germanic languages and within the category 'direction' OE alsodifferentiated between the categories '(movement to a) goal' (*hider* 'hither' - *þider* 'thither') and '(movement away from a) 'source' (*heonon* 'from here' - *þanon* 'from there'), opposing aproximal to a distal exponent in each case (Fig. 1). The relevant cognate forms of these demonstratives are still found in EME (*hither* vs. *thither* and *hence* vs. *thence*), just like the semantically analogous interrogative pronouns (*whither* vs. *whence*), but are marginal, if not totally obsolete, in Modern English.

	person	entity	time	location	goal	source	manner degree	quality
proximal	ic/þū	þes	nū	hēr	hider	heonan	swā, þus	swilc
distal	hē, hēo, hit	se	þā	þær geond	þider	þanon		
interrog.	whāt	whæt	whonne	whær	whider	whanon	hū	hwilc

Figure 1. The system of demonstratives in Old English

In this paper we will take a closer look at the system and use of spatial demonstratives in Old English as a starting point for historical processes of simplification and loss similar to the ones found in the development of inflectional categories.

After a brief characterization of demonstratives and their essential properties, our paper will examine three stages in the historical development of directional demonstratives: A) The stage of OE with a four-term distinction differentiated in terms of the opposition 'proximal-distal' in the deictic dimension and the opposition 'source' vs. 'goal' in the content dimension (Fig. 1); B) The preservation of this for-term system in EME ('source': *hence, thence, whence* – 'goal': *hither thither, whither*); C) The loss of the relevant meaning distinctions in ModE, where *hither* and *thither* arestill employed in a directional sense, but hence and thence are exclusively found in temporal or 'reasoning' interpretations; D) The partial renewal of this system of distinctions with the help of a basic demonstrative and a grammaticalized noun in American dialects: *this-a-way* vs. *that-a-way*.

Finally, it will be shown that the reduction and simplification in the system of demonstrativesis only one aspect of the nearly complete loss of the distinction between 'location' and 'direction' in the structure of Modern English.

Rhyming position as a factor in the persistence of Early Middle English lexis

Johanna Vogelsanger

University of Zurich (Switzerland)

johanna.vogelsanger@es.uzh.ch

The appearance of French loan-words in early ME and the simultaneous disappearance of native words is a well-known phenomenon (see Dekeyser and Pauwels 1990 for an overview), and although it has been re-examined from various perspectives in recent years (Timofeeva and Ingham 2018), one factor that has rarely been considered in this development is rhyming position. Gardner (2014), Käsmann (1961), and Timofeeva (forthcoming) mention its importance in individual cases, but so far there has been no systematic study on a larger scale.

The aim of this paper is to study the impact of rhyme on the survival of native and the adoption of borrowed lexis. My data consists of a list of lexemes ending in French and English rhyming suffixes (*-ity, -ery, -y/-ie,* and *-ship, -ness, -hood,* respectively) taken from the *Linguistic Atlas of Early Middle English* (*LAEME*) and previous work by Gardner (2014). For every item on this list, I collected native synonyms from the *Thesaurus of Old English* and the *Historical Thesaurus of English* and compared their frequencies and rhyme-to-frequency ratios in *LAEME* with those of their rhyming synonyms. The LAEME data (spanning ca 1150-1325) was divided into five subperiods to allow for more fine-grained analysis.

Preliminary results indicate that French loans and words with French rhyming suffixes occur more often in rhyming position than those ending in a native suffix. Similarly, French loans occur more often in rhyming position than their native-origin synonyms. Among native lexemes, those ending in *-ness* tend to have a higher rhyme-to-non-rhyme ratio than others.

References

Dekeyser, X. and L. Pauwels. 1990. 'The Demise of the Old English Heritage and Lexical Innovation in Middle English: Two Intertwined Developments'. *Leuvense Bijdragen,* 79. 1-23.

Gardner, A. 2014. *Derivation in Middle English. Regional and Text Type Variation.* (Mémoires de la Société Néophilologique de Helsinki XCII). Helsinki: Société Néophilologique.

Käsmann, H. 1961. *Studien zum kirchlichen Wortschatz des Mittelenglischen, 1100-1350. Ein Beitrag zum Problem der Sprachmischung* (Buchreihe der Anglia, Zeitschrift für Englische Philologie, 9. Band). Tübingen: Max Niemeyer.

LAEME. *A Linguistic Atlas of Early Middle English*.

Timofeeva, O. and R. Ingham. 2018. 'Special Issue on Mechanisms of French Contact Influence in Middle English: Diffusion and Maintenance'. *English Language and Linguistics,* 22 (2). 197-205.

Timofeeva, O. Forthcoming. '*Nithe and onde, þat is here broþer*: The lexicalisation of a Middle English binominal'.

Towards a classification of lexical obsolescence and loss in Early Modern English

Veronika Volná

Charles University (Czech Republic)

veronika.volna@ff.cuni.cz

In general, literature dealing with the development of the English lexicon is primarily concernedwith the new additions to the language's vocabulary – borrowings, coinages and word formationstrategies, while the topic of lexical obsolescence and loss remains under-researched (Tichý 2018: 81). In the Early Modern period, the rate at which new lexical items appeared in the English language was unprecedented, especially in the years 1590-1620, which is documentedin the *OED*'s *Timelines* feature (2014). At the same time, the obsolescence or complete disappearance of forms was taking place, affecting a significant portion of the English word-stock, including those short-lived coinages or borrowings which had entered the language only several decades or centuries earlier.

Using authentic examples from the *EEBO* corpus (*Early English Books Online*), this paper provides an insight into the role of lexical obsolescence and loss in the development of Early Modern English. Based on frequency data, a list of candidates for obsolescence has beengenerated, including the following: 1) *sacerdote* (lat.), 'priest'; 2) *travalier* (fr.), 'one who travails or labours'; 3) *breastlap* (en.), 'an obsolete synonym of breast-plate'.

In these examples, it is easily conceivable that the reason for the word's disappearance was competition with another already existing, more widespread and perhaps more native-sounding synonym. Although near synonymy and polysemy are identified as the basic catalysts forobsolescence, other factors include disappearance of the word's original referent, weakening of emphasis through overuse, and political correctness (Görlach 1991: 140), i.e. both internal and external to the language.

The most interesting candidates will be selected and the conditions signaling the decline of these forms analysed and discussed. In conclusion, a classification of lexical obsolescence will be proposed, with regard to both language-internal and relevant language-external processes.

References

EEBO. *Early English Books Online.* 2014. Prague: Institute of the Czech National Corpus, FF UK. [https://www.korpus.cz]

Görlach, M. 1991. *Introduction to Early Modern English*. Cambridge: Cambridge University Press.

OED *Online*. 2014. Oxford University Press. [http://www.oed.com/]

Tichý, O. 2018. 'Lexical obsolescence 1700-2000'. In J. Kopaczyk and J. Tyrkkö (eds.), *Patterns in text: Corpus-driven methods and applications.* Studies in Corpus Linguistics, John Benjamins.

Women in the *Oxford English Dictionary:* Citation evidence and the question of representativeness

David-Antoine Williams

St Jerome's University in the University of Waterloo (Canada)

david.williams@uwaterloo.ca

Lindsay Rose Russell has recently described a 'scholarly consensus... that mainstream lexicography, past and present, is shot with sexism and androcentrism' (2018: 15). Scholars particularly interested in the *Oxford English Dictionary* (OED) have been interested, among other things, in the dictionary's treatment of female-authored sources of quotation evidence. Because author gender is not annotated in the digital text of the dictionary, holistic analyses of gender in the OED have not yet been undertaken. Instead, scholarship has tended to focus on one or more highly-cited authors (e.g., Brewer 2012 and 2015, Baigent *et al.* 2005), making comparisons between those authors and analogous male counterparts (determined either in qualitative or quantitative terms). However, like any study attending to a superlative population, such an approach misses the large majority of female authors, and female-authored writing, recorded in the OED. In OED2 (1989), for example, there are about 4,000 women authors cited in a total of 85,000 quotations, with the vast majority of these authors only cited only a handful of times—indeed, 30% are cited only once, and 75% ten times or fewer. The top-twenty most- cited women authors, with between 502 (Adeline Whitney) and 3,149 (George Eliot) citations apiece, contribute about 27% of all female-authored citations in OED2. The picture in the revised portions of OED3 is more fragmented: by the start of 2020, about 87,000 female- authored quotations had been added, with the top twenty authors accounting for about eight percent of all of these quotations. I come to these figures thanks to a data annotation project at St Jerome's University in the University of Waterloo (Canada), which since 2011 has been annotating every OED2 quotation with metadata for, among other things, author gender (Williams 2017 and 2018). In this paper I will explore this dataset, as well as a recent iteration of the OED3 revised text, to sketch out ways in which writings by women have been incorporated into the dictionary from the first edition until today. After briefly discussing what I mean by 'gender' and 'authorship', I begin with an (also brief) technical account of the annotation project and how it came to identify women-authored quotations in OED1 (1928), the two Supplements (1933 and 1972–86), and OED3. The main part of the talk will present and discuss the data comparatively across all editions and Supplements, analyzed mainly with respect to salient features of the quotations— principally period, genre, and subject matter—but also with respect to features of the headwords and sense sections they illustrate. In the final part of the paper I will tentatively propose one wayto arrive at an independent benchmark against which one might begin to measure how 'representative' OED has been of women's writing in general over the decades. Building on Forster 2015, Underwood *et al.* 2018, and Williams 2018, I will propose some techniques to exploit open text repositories and library catalogs, such as the *HathiTrust Library* and the *Library of Congress*, as a source for comparable benchmark data.

References

Baigent, E., Charlotte B. and V. Larminie. 2005. 'Gender in the archive: Women in the Dictionary of National Biography and the Oxford English Dictionary'. *Archives* (Journal of the British Records Association), 30. 13–35.

Brewer, C. 2012. 'Happy Copiousness? *OED's* Recording of Female Authors of the

Eighteenth Century'. *Review of English Studies,* 63. 258. 86-117.

Brewer, C. 2015. 'That Reliance on the Ordinary: Jane Austen and the Oxford English Dictionary. *Review of English Studies,* 66. 276. 744-65.

Forster, C. 2015. 'A Walk Through the Metadata: Gender in the HathiTrust Dataset'. [https://cforster.com/2015/09/gender-in-hathitrust-dataset/].

OED. *Oxford English Dictionary*. 1928. 1989 (2nd edition); 1933 and 1972-86 (Supplements); 3rd edition. Oxford University Press.

Russell, L.R. 2018. *Women and Dictionary Making: Gender, Genre, and English Language Lexicography*. Cambridge UP.

Underwood, T., Bamman D. and S. Lee. 2018. 'The Transformation of Gender inEnglish Language Fiction.' *Cultural Analytics.*

Williams, D-A. 2017. 'Getting More Out of the Oxford English Dictionary (by Putting More In)'. *Dictionaries: Journal of the Dictionary Society of North America.*

Williams, D-A. 2018. 'Advanced Collaborative Support Award Report'.

Historical dictionaries and the semantic web: The possibilities of linking words and texts

Tarrin Wills

University of Copenhagen (Denmark)

tarrin@hum.ku.dk

Developing historical dictionaries for the era of the semantic web presents a number of problems, in particular, many dictionaries use distinct methodologies and have their own data models as a result. There are existing standards for digital dictionary entries, in particular TEI and Ontolex. These have limitations for historical lexicography. The key problem is that historical dictionaries include citations that reference authoritative texts (editions and/or manuscripts). The existing standards do not handle or standardise these references in a way that allows the kind of semantic, two-way referencing that builds the Semantic Web. The *Dictionary of Old Norse Prose (ONP)* was designed before the emergence of digital lexicographic standards but is built on long-standing historical lexicographic practice. The data model it uses includes two-way linking between citations, articles and sources, not only the Old Norse sources but also in many cases foreign parallels for translated works. The rich relationships between words and texts can therefore be made available in new ways. For example, without the addition of any further data, *ONP* can produce parallel glosses of all its texts by exploiting this two-way linking. It also uses linked data techniques to dynamically connect its own datawith external projects, linking from external source digital texts to produce concordances and linking other dictionaries to supplement it where it is currently incomplete. This paper will demonstrate some of these capabilities as well as discuss the potential of extending the existing standards to encompass these possibilities.

References

ONP. *Dictionary of Old Norse Prose*. Department of Nordic Studies and Linguistics, University of Copenhagen. [https://onp.ku.dk/onp/onp.php].

Syllabification of Buchanan's digitized orthoepic dictionaries

Franck Zumstein (University of Paris, France)
Jean-Louis Duchet (University of Poitiers, France)
franck.zumstein@u-paris.fr, jean-louis.duchet@univ-poitiers.fr

James Buchanan published two dictionaries in 1757 (1757a and 1757b) and a third one in 1766.They are very different from one another in terms of macro- and micro-structure. Though different in their structures, the three dictionaries have two common features: 1) the pronunciation of the headword is systematically indicated in each entry and is common to all three dictionaries: 2) syllable division, however, is not indicated.

Buchanan could have marked it thanks to the rules of syllabification in the introductory pages of his dictionaries (Emsley 1933: 1163). Yet, marking syllable divisions was not yet a usual practice and was introduced by Kenrick in 1773 only.

Our presentation aims at developing an algorithm the objective of which is to introduce syllable divisions in each entry of Buchanan's dictionaries according to his own rules. Syllable division is crucial to interpret vowel qualities and stress patterns in the pronunciation clues provided by orthoepic dictionaries. And this information, still considered an issue in 21st century pronouncingdictionaries (Wells 2008, 3.5, xxvii; Roach 2011, 3.6, xiv) was used by the next generation of orthoepists whose dictionaries we intend to compare with Buchanan's (Kenrick 1773; Sheridan 1780; Walker 1791).

An obvious example of the consistency which we have to add to Buchanan's own rules concerns thesyllabification of words ending in consonant+<le>: a strict application of his basic principle of syllable boundary insertion between two consonants would produce syllabifications such as *beet-le,brid-le, embez-le* [sic].

But Buchanan's own observation that *(l)* is a "half vowel" (1757a: 15) the sound of which "mightseem to be expressed by *(il),* yet [is] always to be marked with *(le)*" shows that he observes a syllabic pronunciation of <l> in spite of the spelling with which he must comply.

The algorithm we have in mind must apply an ordered hierarchy of rules to make Buchanan's worka consistent representation of English pronunciation in the 1760s.

References

Buchanan, J. 1757a. *Linguæ Britannicæ Vera Pronunciatio, or a New English Dictionary.* London: A. Millar in the Strand.

Buchanan, J. 1757b. *A New Pocket-Book for Young Gentlemen and Ladies, or a Spelling Dictionary of the English Language*. London: R. Baldwin at the Rose in Paternoster Row.

Buchanan, J. 1776. *An Essay Towards Establishing a Standard for an Elegant and Uniform Pronunciation of the English Language, throughout the British dominions, as practised by the most learned and polite speakers.* London: E.N.C. Dilly.

Emsley, B. 1933. 'James Buchanan and the Eighteenth Century Regulation of English Usage'. *PMLA*, 48 (4). 1104-1166.

Jones, D. 2011. *English Pronouncing Dictionary,* ed. by Peter Roach *et al.* 18th ed. Cambridge University Press.

Kenrick, W. 1773. *A New Dictionary of The English Language: Containing, not only The Explanation of Words,With Their Orthography, Etymology, and Idiomatical Use In Writing; but likewise their orthoepia or pronunciation in speech, according to the present practice of polished speakers in the metropolis*. London.

Sheridan, T. 1780 *A General Dictionary of the English Language, one main object of which, is, to establish a plain and permanent standard of pronunciation, to which is prefixed a rhetorical grammar.* London: J. Dodsley, C. Dilly & J. Wilkie. (Digital edition by Véronique Pouillon, Université de Paris, 2017).

Walker, J. 1791. *A critical pronouncing dictionary, and expositor of the English language,* 6th stereotype edition, 1809 (Digital edition by Nicolas Trapateau, Université de Poitiers, 2015).

Wells, J.C. 2008. *Longman Dictionary of Pronunciation.* Longman, 3rd ed.