Jan Kochanowski University Press

This is a contribution from *Token*: *A Journal of English Linguistics*
Volume 13/2021.
Edited by John G. Newman, Marina Dossena and Sylwester Łodej.
© 2021 Jan Kochanowski University Press.

A database of early modern first citations from the OED: Religious and geographical terminology[1]

Angela Andreani* and Daniel Russo**

**Università degli Studi di Milano*
***Università degli Studi dell'Insubria*

ABSTRACT

In the wake of the Reformation, intellectuals from all parts of the religious spectrum read, studied and translated Christian sources, not only the Scriptures but also ancient and modern patristic sources, sermons, commentaries, chronicles. The users of these texts – translators, theologians, controversialists – were highly experimental and lexically innovative, as demonstrated by the appearance of many of them amongst the first 1000 sources of the *OED*. In our paper we propose a corpus-based study of their lexical competence to assess their impact on the development of the English vocabulary 1500-1650. This is a pilot study intending to test the use of "sources" in the *OED* for corpus-building, and to combine digital databases and corpus-query systems (*OED*, *EEBO*, SketchEngine) for the diachronic study of lexis. Our study points out a prevalence of church-related vocabulary as a specialised terminology, but it also focuses on other secondary domains such as demonyms and geographical terms.

Keywords: Reformation, corpus-based lexicography, church-related vocabulary, geography.

1. Introduction

As is well known, the early modern period is a key moment in the development of the English lexicon, during which the vocabulary displays the fastest

[1] Both authors are responsible for the overall planning and research for this paper. In particular, Angela Andreani is responsible for sections 2.1 and 3.1 while Daniel Russo for section 2.2 and 3.2. Sections 1 and 4 were written jointly by the two authors.

DOI: 10.25951/4820

growth with a peak observed between 1570 and 1630 (Görlach 1991: 136-137; Barber 1997: 219; Nevalainen 2000: 336; and Durkin 2014: 305-306). Word-formation and increased borrowing both contributed to this growth. The use of written English for most purposes, the expansion of literacy and the spread of printing, combined with the increased mobility to and from England, pushed the creative potential of the language and nurtured a continuous influx of new concepts and foreign words. Since Latin had remained the main language for theology, scholarship and the church for centuries, English had not yet fully developed the vocabulary nor the style of religious debate in a highly dynamic religious context; however, during the early modern period this changed, as the vernacular came to be used "in an increasing range of functions, especially as a language of learning and of religious discourse" (Durkin 2014: 306). As has been remarked, "Nothing reveals the deficiencies of a language more surely than translating into it" (Kay – Allan 2015: 14), which suggests that translators of the Tudor and Stuart era were at the forefront of processes of lexical enrichment. Not only was an immense body of classical Greek, Latin and Hebrew sources turned into English during this period, but the entire vocabulary of the church and religion was discussed, re-codified, and significantly enriched, also through the contact with other vernaculars: "previously 'dogmatic' words like *heresy*, *enormity* and *abuse* became relative and plural in meaning, as their use became dispersed among the disputants" (Hughes 1988: 113). Intending to map the influence of religion onto the history of the English lexicon, in this paper we try out a combination of resources and methods that will enable us to explore the intersection between lexicography, translation and religious writing.

2. Materials and methods

2.1 Materials and sources

Our materials were retrieved starting from the *OED's* top 1000 sources and identifying a group of translators, theologians and controversialists active between 1500 and 1650. The data that can be accessed using the *OED's* sources have already proven valuable for linguistic research; Giles Goodland (2013) has investigated the use of neologisms in early modern literature by focusing on a selection of canonical authors retrieved using the "sources" function of the *OED*, while Julie Coleman (2013) has shown what

2021 Jan Kochanowski University Press. All rights reserved.

can be gained from a close analysis of the *OED's* sources combined with an awareness of the limits of this function.

In fact, working with the *OED's* sources opens up a number of methodological questions. One of the main concerns for scholars is the representativeness of the quotations used in the *OED* (Schäfer 1980; Brewer 2010 and 2013; Considine 2009; and Coleman 2013); however, this limit of the *OED's* sources does not prejudice our research, since our starting point is the study of the contribution and legacy of a selected category of writers. Another limit, pointed out by Charlotte Brewer, is that the data searched through the *OED* are not stable since "every quarter, the identical search will produce a different set of results, as the lexicographers upload a new batch of revised entries to the dictionary and remove the corresponding unrevised ones" (2013: 115). This means that there may be discrepancies between our data and the information published on the *OED Online* when the entries involving our source authors are revised. In fact, we may have spotted a couple of such instances working with our data (see §2.2).

Our selection of authors was based on background knowledge and on information we could verify using the *ODNB*. From the list of the *OED's* sources we selected authors whose written output and profession indicates lifelong interaction with Biblical and patristic sources, in the original or in translation. Included are works that cannot be classified as translations proper, and people to whom the professional label of translators cannot be applied. In fact, what constitutes translation, citation or paraphrase in this period is fuzzy, but our assumption is that operating across languages and cultures was standard intellectual practice for our authors. Our sampling includes a combination of established and less canonical figures. In chronological order, our source authors are:

- John Bale (1495-1563), reformed clergyman, bishop of Ossory in Ireland, active evangelical polemicist and author of a commentary of the Book of Revelation;
- John Foxe (1517?-1587), the renowned author of the *Acts and Monuments*, he had a deep knowledge of early Christian historians on which he based sections of his work, was the author of controversial tracts and collaborated with Continental reformers;
- Thomas Cooper (c. 1517-1594), bishop of Winchester, theologian, editor of Thomas Elyot's dictionary and himself an important English-Latin lexicographer;

2021 Jan Kochanowski University Press. All rights reserved.

- John Jewel (1522-1571), bishop of Salisbury, the chief apologist of the Church of England who confuted the authenticity of the Roman Church based on the Patristic sources of the early centuries of Christianity;
- John Daus (c. 1516-1602), chaplain and later schoolmaster and preacher, translator of the works and sermons of prominent European Protestants, and allegedly of Eusebius' *Ecclesiastical Histories*;
- Arthur Golding (1535/6-1606), translator of a series of major works by Calvin, of the Lutheran commentaries on the New Testament from Latin and of numerous works by Continental writers such as Beza, Bullinger, Augustin Marlorat and Philippe Duplessis-Mornay;
- William Fulke (1538-1589), one of the most important controversialists of the Elizabethan age, chaplain and college head who published an extensive confutation of the Rheims translation of the Vulgata in English and engaged in controversy over the translation of the Bible;
- Richard Hooker (1554-1600), clergyman, deputy professor of Hebrew at Oxford, and the most prominent theologian of the period, author of *Of the Laws of Ecclesiastical Polity*;
- James Bell (d. 1606?), translator from Latin of religious writings by John Foxe, Martin Luther, and Walter Haddon;
- Thomas Newton (d. 1607), clergyman and translator, the most eclectic in our sources, he translated and published on a wide range of subjects, mainly secular such as translations of Cicero and Seneca, was also author of *An Herbal for the Bible*;
- Thomas Tymme (d. 1620), a clergyman who published translations of theological works and devotional writings from Latin, French, alongside his own devotional writings;
- William Sclater (c. 1575-1627), clergyman, author of several sermons and of a treatise on justification, best known for his Expositions of the Thessalonians;
- Thomas Taylor (1576-1632), clergyman and a very prolific writer, author of several sermons, religious treatises, and a commentary of Paul's epistle to Titus.

2.2 Method

The study is based on the concepts, frameworks and methods of corpus-based terminology (Cabré 1998; Gamper – Stock 1998), corpus-based analysis of language variation and use (Biber 2009), specialised discourse (Gotti 2005) and specialised translation (Gotti – Šarčević 2006). The methodological

2021 Jan Kochanowski University Press. All rights reserved.

framework of this research project rests upon the extraction and analysis of terminology from a lexical source, i.e. the *OED Online*. Ahmad – Rogers (2001: 584) define automatic term extraction as "the processing of texts using computer programs in order to identify strings that are potential terms"; the most valuable result of term extraction is thus the lexical material that can be used to create terminology databases through a process of examination, testing and validation before items are inserted into lexical resources such as dictionaries. If structured collections of texts are an extremely important source of data in the study of terminology for indexing purposes, one may question the validity of extracting terms that were in turn extracted and processed by the compilers of a dictionary. Rather than showing the contribution of one specific author, the aim of this project is to show the lexical impact of a profile of scholars in the early modern period, especially in unexpected lexical areas; thus, even though this study might be affected by the same possible biases in selection criteria of the *OED's* compilers (Coleman 2013), we believe that working with big data (Weikum *et al.* 2012) and fuzzy sets (Ma 2011) can compensate for this issue.

Extensive research has been conducted on the methodology for lexical extraction (see Pantel – Lin 2001; Jang *et al.* 2021), which according to Mei *et al.* (2016) can be assigned to three macro approaches: rule-based methods, statistical methods and hybrid methods. In the rule-based methods, words are extracted from a lexical resource (a text, a corpus of texts or a dictionary) based on predetermined criteria, which can be linguistic in nature (e.g. morphological categories), but also textual (author, topic, date, etc.). This method is particularly suitable for extracting new or unindexed words (Isozaki 2001; Stanković *et al.* 2016). Statistical methods are based upon statistical linguistic features and usually pair up with machine learning algorithms to extract words in vast corpora; this method is particularly effective when studying collocations (see Pecina 2010) or linguistic patterns and semantic shifts (see Boukhaled *et al.* 2019). Hybrid methods are the combination of rule-based methods and statistical methods and are mainly employed in text mining in language-specific domains (see Hadni *et al.* 2014). This paper is based on a rule-based approach, the rule being *OED* entries listed as first citations assigned to a pre-established list of authors (see §2.1). For most rule-based methods, the definition of rules may be a difficult task resulting in poor systemic flexibility, but as the selection rules employed for this paper are domain-specific and extralinguistic, this issue does not arise.

Paraphrasing Wright – Budin (2001: 726), text corpora are a valuable source of evidence when studying the variation of occurrence and use of

2021 Jan Kochanowski University Press. All rights reserved.

a language for specific/special purposes (LSP) and its terminology in three main aspects: across specialistic domains, between levels of communication, and diachronic change in relation to competing morphological forms, spellings, and terms. Although our lexical extraction pursues the same purposes, we prefer referring to our collection of lemmas as a *database* and to the proper collection of texts of the *OED's* compilers and to the reference collection of early modern English books as a *corpus*. For decades, the notion of corpus has been understood in linguistics as an electronic corpus, which is stored, processed and analysed automatically or semi-automatically by specialised software systems (Baker 2006: 25-26). While commenting on the relationship between lexicography and translation theory and practice in Tognini-Bonelli (1996), Hanks focuses on the distinctions between corpus-based and corpus-driven lexical research: the aim of corpus-based is to force the lexical evidence of a corpus to fit into preconceived theories through the use of "judiciously selected examples", whereas corpus-driven studies attempt to approach data "with an open mind and to formulate hypotheses and indeed, if necessary, a whole theoretical position on the basis of the evidence found" (2012: 417). On the other hand, Xiao (2008) maintains that this sharp distinction found in the literature between the corpus-based approach and the corpus-driven approach is largely overstated. We support this less polarising view at least for the purposes of this project: as will be described in detail below, the extraction phase of this pilot study is completely corpus-driven; however, the analytical phase must be corpus-based as the texts and the lemmas are examined diachronically and belong to a period wherein terminological approaches, writing practices and semantic prosodies varied considerably; a more manual analysis is thus fundamental to establish synchronic and diachronic connections that would otherwise be overlooked.

The main methodological aspect involved in this study is the extraction of lemmas from the *Oxford English Dictionary Online*. The assessment of any term-extraction method must comport with an evaluation of the corpus that is being analysed, not simply as in traditional terminology management in relation to the authority of the authors of texts, but also with regard to the structure and processing of the corpus (Ahmad – Rogers 2001: 585). The automatic extraction of lexical items is one of the most significant problems in Natural Language Processing (NLP): normally in corpus-based studies the aim of word extraction is to isolate sets of terms and expressions with a certain meaning in a collection of text strings (this only partially applies to the purposes of this paper, as will be discussed below). Applications of

2021 Jan Kochanowski University Press. All rights reserved.

computer-aided term extraction include information retrieval, lexicography, parsing, computer-assisted and machine translation, and lexical databases. In an effort to pursue this last application, this paper sets out to build and examine a lexical database of first occurrences of lemmas extracted from the sources listed in §2.1 from the *OED Online*. Instead of an evaluation system relying mostly on human assessments of the quality of extracted terms, we intend to combine automatic extraction and human analysis. The lemmas listed as first occurrences of the authors described in §2.1 in the *OED Online* (section Sources>[author's name]>first entry)[2] are extracted through a Python script (Hammond 2020) and entered into a spreadsheet database, which stored the following data: author, lemma, definition, work-title, date. The script was executed twice in order to verify whether the data were consistent over time – on 4 November 2019 and on 5 February 2021 – and it proved that the great majority of first citations were not amended during the period concerned, only a very small number of first citations had been re-assigned to another (mainly coeval) source during the months from the first extraction to the second, e.g. *Christianlike* was assigned to Newton (1574) in 2019 and to Taverner (1540) in 2021 and *Bohemian* (sense b), formerly Fulke (1579), is now Golding (1562). The second phase involves the classification of the dictionary entries in relation to their semantic field. This process has produced a list of 1,919 lemmas that are indexed with the following information: author, definition, title of first occurrence, date of first occurrence. Although a certain number of entries show some level of classification in the definition section (e.g. "anatomy"), most lemmas do not; therefore, this phase required manual processing, which was time-consuming and, in a few cases, implied terms being classified in more than one category. We have identified several prevailing semantic domains in the extracted database; therefore, in this pilot study we will focus on one expected domain in the corpus, i.e. religious terminology, and on one unexpected domain, i.e. geographical terms. In the third phase, we relied upon another digitised corpus – the Early English Books Online (*EEBO*); through this corpus, which can be efficiently browsed on the online corpus linguistics platform Sketch Engine (Kilgarriff *et al.* 2004 and 2014), we manually evaluated the lexical and semantic aspects (especially in the form of concordances) of the terms in our database in order to obtain a comparative evaluation of term usage in competing forms, spelling variation and, possibly, dating with the extended corpus of early modern English sources.

[2] https://www.oed.com/sources.

2021 Jan Kochanowski University Press. All rights reserved.

3. Results and discussion

This section is divided into two subsections, one for each lexical macrodomain discussed in this paper. The main aspects tackled in our approach are occurrences, spelling, morphology, competing expressions, semantics and etymology. Several intersections are discerned in these seemingly unrelated fields. The years indicated in brackets in the citations below are those reported in the *OED* and extracted into our database; *EEBO* references are indicated with their TCPIDs (Text Creation Partnership ID), which unambiguously identify the source.

3.1 Religion and church-related vocabulary

Church related vocabulary amounts to 177 lemmas ranging from words connected with the writing and the study of the Bible, to words to indicate behaviours against the church, members of the clergy, God, or the sacraments. To make sense of this diversity, the lemmas have been organised into semantic fields and categories, drawing from the classes and senses used in the *OED Historical Thesaurus* (*HTOED*). Three macro-categories have been identified: "faith", the "supernatural", and "other". "Supernatural" defines words and attributes for God and deities (i.e. *petty goddess*, *theandric* and *unitrine*[3]); "other" includes a variety of words from various semantic fields, which have developed (or preserved) senses connected with religion and the church (e.g. *church story*, *disvesture*, *ministership*); and "faith", the first category with 144 lemmas, includes the fields in Table 1 below, arranged from the most to the least numerous:

Table 1. Church-related vocabulary > "Faith": fields and nr. of lemmas

Field	nr. of lemmas	Field	nr. of lemmas
sects	41	canon law	2
church government	33	creed	2
sacrament, communion	9	architecture	1
paganism	8	prayer	1
liturgy, ritual	8	error	1

[3] For the purposes of this paper, words extracted from our database are marked in italics, whereas glosses, definitions, translations, mentioned words and phrases, etc. appear between double quotation marks.

2021 Jan Kochanowski University Press. All rights reserved.

sacrament, ordination	5	Catholicity	1
scripture	5	atheism	1
consecration	4	heresy	1
benefices	4	offence	1
apostasy	3	orthodoxy	1
sacrifice	3	religion	1
canonization	3	sectarianism	1
sacrilege	2	spirituality	1

The data photograph a very rich and composite situation. In what follows, the discussion will be limited to selected examples illustrative of lexical enrichment in particular fields, with attention to morphological experimentation and semantic shifts.

The number of lemmas that indicate religious sects is staggering. In our database we have a variety of new entries and derived forms. A number of lemmas are based on aspects of discipline or of doctrine, such as *flagellant* and *anabaptistry*, and several are derived from the names of their founders, such as *Arianism*, but also *Christianlike*, *Calvinist* and *Mahometical*. In two cases, provenance defines particular sects: *Saxonian* and *Bohemian* (see §3.2 below).

Sometimes we can clearly detect the influence of patristic sources. The noun *Donatian* (Sclater 1627) is a Latin loan whose entrance into English was mediated by the work of Jerome and Augustine. This variant had limited use in the early modern period (12 hits in *EEBO*) and is now obsolete, while "Donatist", much more frequent in the *EEBO* corpus (497 hits), is in current use. *Donatian* may be a zero derivation, since the adjective is attested four decades earlier in the *EEBO* corpus: "as S. Augustine sommetime saide to the Donatian Heretiques" (A04468). *Marcosian* (Fulke 1580) is derived from Greek and like the names of several other sects it entered into English through the popular early Christian work on heresiology written by the bishop of Lyon Irenaeus (*c*. 130-202), *Adversus Haereses*: "Transubstantiation of the wine into blood in Marcus and the Marcosians *Irenaeus lib. 1 cap. 9.*" (A01325, original italic). Another channel was the compendium by the bishop of Salamis Epiphanius (d. 403) known as the *Panarion*, or *Adversus Haereses*: "Likewise the Marcosians when they baptized, vsed to speake certaine Hebrue wordes, […] *Epiph. lib. 1. Tom. 3. haer. 34.*" (A01335, original italic).

The close contacts established by English reformers with Continental communities favoured influences across the vernacular languages.

2021 Jan Kochanowski University Press. All rights reserved.

A number of terms of classical origin may have been modelled on coeval forms in French, German or Italian; this was the case for the term *Confessionist* (Fulke, 1570, but the earliest occurrence in *EEBO* is 1565 in A04474), from French *confessioniste*, used as a synonym for "Lutheran", although much less frequently (23 vs 6,635 hits). Another example may be the term *Calvinist* (Fulke 1579), for which we find the equivalent *calviniste* in French.

Examples of coinages from internal derivation processes are *Anabaptistry* (Foxe 1570), from "Anabaptist", *Lutheranism* (Daus 1560) from "Lutheran", *Puritant* (Fulke 1580) from "Puritan", and possibly *Calvinist* (Fulke 1579), which may have been modelled on the slightly earlier "Calvinism" or derived directly from the name of John Calvin (see *EEBO* A20661 for earlier occurrences dating to 1564, e.g. "how happeneth it that the Caluinistes and the Lutheranes agre not").

Finally, the data illustrate to what extent the separation from the Church of Rome triggered the lexical inventiveness of polemicists. "Popery" became a derogatory catchword for Roman Catholicism, which was framed as a false and idolatrous religion in sermons, pamphlets and treatises. The first lemma in our database is *papistry* (Bale 1543) derived from "papist" (OED s.v. "papist, A. n. 1"), and with its 901 hits in the *EEBO* corpus the most frequent keyword for anti-Catholic slander coming from our sources:

(1) not onely defending the vngodly worship, papistry, and false religion. (*EEBO* A04696)

(2) euen so they that are droonke with the hereticall doctrine of Papistry. (*EEBO* A01327)

(3) The religion of papistry being a Catholick Apostasie from God. (*EEBO* A20740)

The oldest and most frequent alternative by far is "popery" (the spellings popery + poperie retrieve 26,073 hits, with the earliest attestation dating to 1528), not present in our database. A borrowing from Latin, *papism* (Bale 1550) is another relatively frequent term of polemical slander (328 hits), while the adjectives *popan* (Bell 1580) and *papane* (Bell 1581) are variants for the much more frequent and established "popish" and "papal" (respectively 23,410 and 6,041 hits in *EEBO*); in particular, *popan* might be a nonce form by Bell, but a search of *papane* (overall 15 hits) retrieves earlier attestations of

2021 Jan Kochanowski University Press. All rights reserved.

the term in the sense of "Pope's dominion" (A17662) and as a synonym for "papal" (A01130).

Morphology reveals processes of selection and acceptance in the history of several words: in our database *Lollery* (Bale 1547, 4 hits including the spelling "lollerie") occurs as a variant for "lollardry" (4 hits) and "lollardy" (39 hits including the rarer spellings "lollardye" and "lollardie"). In the *EEBO* corpus "Calvinian" appears alongside the variant from our database *Calvinist* (Fulke 1579) although with less freqency (520 vs 1767 hits respectively). With 56 hits *Wycliffian* (Foxe 1570) supersedes the alternative variants "Wycliffist" (18 hits) and "Wycliffite" (21 hits); it may be noted that all appear with that they spelling "Wick-" in *EEBO*.

In order to place the terminology denoting religious sects in the broader context, we have turned to the *HTOED*, which reveals that these terms entered the English vocabulary from different channels and into stages. A consistent portion entered through the translation of the medieval collections of saints' lives, while the 16th century additions may be explained in part as the effect of the recovery of patristic sources and their translation into English, and in part with the need to make sense of an increasingly fragmentary religious situation through lexicalisation. Another semantic field that emerges from our database is in fact that of "sectarianism" (e.g. *interimist*, Daus 1560). If we expand the search for related terminology in the *HTOED* we see that cognate words (e.g. "sectary", "sectator", "sectuary", "sectist") and semantically related forms such as "separatist", "conventicler", and variants, are all additions dating between the 1550s and 1600, signalling the particular development of this area of the English lexicon during the period under review.

Our database highlights another area of special significance in the lexical repertoire of religious authors and translators: Eucharistic terminology. From Latin we have the verb *transcorporated* (Foxe 1570), seemingly a nonce usage proposed as an alternative to the older and more common "transubstantiated" (511 hits in *EEBO*, earliest attestation 1549). The verb *inaccidentated* (Fulke 1579) in our database appears to be another nonce usage. Neologisms of this kind seem to be a distinctive feature of controversial literature; compare the term "iniesuated" (*EEBO* A02617) and further interesting coinages by William Fulke present in our database:

(4) but he [i.e. Christ] is not to be worshipped in bread & wine, or in *the* accidents of bread & wine, because he is neither impanated, nor inuinated, nor inaccidentated, that is, not ioyned to any of them in a personall vnion. (Fulke 1579).

2021 Jan Kochanowski University Press. All rights reserved.

Inaccidentated was derived by affixation from the word "accident", perhaps after the model of the loan *impanated*, which occurs alongside *invinated*, introduced by Fulke in 1579, and likely derived from an earlier form "invinate" already attested in 1550 (*EEBO* A19571).

The occurrence of the pair *consubstantiation/consubstantiate* in our corpus (Hooker 1597) reflects the development of the Eucharistic debate. These words were in fact specialist controversial terminology, in that they helped define and identify different theological positions, as the citation from Hooker's *Of the laws of ecclesiasticall politie* makes clear:

(5) They […] are driuen either to Consubstantiate and incorporate Christ with elements sacramental, or to Transubstantiate & change their substance into his. (Hooker 1597)

and further:

(6) So that they all three do plead Gods Omnipotency: Sacramentaries, to that Alteration, which the rest confess he accomplisheth; the Patrons of Transubstantiation, over and besides that, to the change of one substance into another; the Followers of Consubstantiation, to the kneading of both Substances, as it were, into one lump. (*EEBO* A44334)

These examples enable us to appreciate how morphological experimentation could convey key religious meanings: neologisms with prefix *in-* and denominal suffix *-ate* could be used humorously and/or to convey polemical and derogatory overtones (example 4); the prefixes *con-* and *trans-* could encode specific doctrinal positions regarding the understanding of the body of Christ in the communion (example 6). As a process of word formation, therefore, derivation is not only particularly productive but also of special significance in the lexis of religion, as beliefs, groupings and outlooks became lexicalised.

In the iconoclastic setting of Tudor and Stuart England, words referring to images acquired negative connotations too. An entire vocabulary derived from originally neutral terms such as "image" and "idol" became bywords for paganism, heresy and a false Christianity. *Idolatrous*, whose first evidence is provisionally found in Bale in the *OED* (1550, see s.v. "idolatrous, adj."), is in fact an older presence in the English vocabulary: "a supersticious and idolatrous kynde of worshippyng" (1542, *EEBO* A06710). This variant is the one that has become established in English, and with 9,606 hits in *EEBO* it

2021 Jan Kochanowski University Press. All rights reserved.

proves to be already well-attested in our period. Both apparently introduced by Bale, the adjectives *idolous* (Bale 1546), and *mammetrous* (Bale 1546) are considerably less frequent with 5 and 1 hits respectively. *Mammetrous* is derived from the 14th century loan "mammetry", which indicated idolatry and non-Christian practices. The term is in fact a loan from Anglo-Norman *maumeterie*, a reduced form of *mahumetterie*, ultimately derived from the name of the prophet Muhammad (*OED* s.v. "mammetry, n."). By the time they entered English, *mammetrous* had evidently lost all connections with Islam, so that in our database we find the new entries *Mahometical* (Daus 1561), borrowed from French and Latin, and *Mussulman* (Foxe 1570) borrowed from Persian, Arabic or Turkish (*OED* s.v. "Mussulman, n. and adj."). The new loans are not associated with idolatry, as may be expected, but are nonetheless connotated as blasphemous practices by our sources: the phrase "Mahometicall corruption" appears in a translation of Bullinger's sermons (Daus 1561), and *Musulman* as a term for a "Turkishe priest" (Foxe 1570, on *Turkish* see 3.2 below). With 154 hits in the alternative spellings *Mus*(*s*) *ulman*(*s*), this term superseded *Ma*(*c*)*hometical*(*l*) with its 51 hits. One final coinage in this field, the compound *image-worshipping* (Bale 1544), appears to have been used very limitedly (10 hits) in comparison with the older and well-established "idolatry", a borrowing via French (over 40,000 hits).

Words that have been grouped under the field "church government" display processes of pejoration, especially those related to the field of monasticism, such as *abbey-like* (Foxe 1570) and *cloistered* (Bell 1581), which show that monastic lodgings were framed as places of corruption: "Shewing, The Canterburian Cathedrall to bee in an abbey-like, Corrupt, and rotten condition" (*EEBO* A35353); "these Cloistered Friers, who now grown to the height of their sinnes" (*EEBO* A12738). The word *greasling* (Golding 1583), a derogatory term for Catholic priests derived from "greasing", used contemptuously to refer to the practice of "anointing" in religious ceremonies of the Roman Catholic church:

(7) their popish greasing which they vse only when a man is desperatly sicke. (*EEBO* A01325)

Another interesting lexeme used in our database to denote priests of the Roman Catholic Church is the compound formed within English *mass-monger* (Bale 1551), denoting a "dealer" or a "trafficker" in masses:

(8) For our Massemongers haue Masses in store for all kynde of thinges good or badde. (*EEBO* A06652)

2021 Jan Kochanowski University Press. All rights reserved.

Compared with its current meaning, the term *seminarist* (Fulke 1583) in our data has markedly negative connotations, clearly due to its association with Roman Catholicism:

(9) than all the popish Seminaries, and Seminarists, shall be able to hinder it, iangle of grosse & false translations. (Fulke 1583)

This term and the more frequent compound "seminary priests" are often paired with "Jesuit", when not used as synonyms:

(10) These Seminarists Jesuits, and other Priests. (*EEBO* A20820)

In the late 16th century, they represented in fact the quintessential seminary priests, trained on the Continent, especially at the English college of Douai, which, since 1574, had been the fulcrum of the reorganisation of militant English Catholicism:

(11) the flocking of so many Iesuits and Seminaristes, as so many trompets and bellowes of sedition into England. (Fulke 1583)

In its current sense of "member of a seminar" (OED s.v. "seminarist, n."), the term has undergone secularisation. Other examples of secularisations concern the words *customariness* (Cooper 1608), originally denoting "perfunctory worship", but whose extended use is already attested in the 17th century, and *renouncer* (Bale 1547), denoting especially renouncers of God, the Truth, or the Church and often paired with "abiurer" and "apostate" in the *EEBO* corpus. A common adjective in Present Day English, *ritual* is another term from our database (Foxe 1570) that may be said to have undergone secularisation, as it originally referred to the performance of rites, often intended as empty ceremonies:

(12) Of these solemnities & feastes we reade that they belonged & were inioined to the Iewes vnder the law, were meerly ceremonial & ritual, […] neither are to be reteined in the church or ministerie of CHRIST. (*EEBO* A05025)

One final example worthy of attention is *superintendent*, another of several terms attributed to Bale. According to the *OED*, this is a loan from post-classical Latin after the ancient Greek *ἐπίσκοπος*, found in Jerome to indicate

2021 Jan Kochanowski University Press. All rights reserved.

a "superintendens bishop". In Continental churches and in the reformed Church of Scotland it denotes a chief presiding minister (still in use), and an official appointed to ordain ministers and to oversee a territory (obsolete), but the sense "superintendens bishop" is specific to the English context, and it was used by both reformers and Catholics, with opposite connotations, to indicate the bishops of the Church of England. The term in English was apparently modelled after German *Superintendent*, and its adoption shows both the influences across vernacular languages and the use of early Christian texts as sources for a vocabulary to describe the reformed Church government. A series of examples retrieved from *EEBO* illustrate the gradual establishment of the term "superintendent" in our period, through definition, synonymy and explanation:

(13) Episcopus is as moche to saye as a superintendent or an ouersear, whose offyce was in the prymatyue Churche purelye to instructe the multitude in the wayes of God. (Bale 1544)

(14) And the word (superintendent) being a very latin word made English by vse / should in tyme haue taught the peple by the very etymologie & and proper signification. (*EEBO* A10777)

The final example is particularly cogent coming from an author clearly of Catholic leanings attacking the Reformation as (also) terminological subversion:

(15) They had throwen doune altars, ouerthrowen Churches, denyed all outward Priesthod, changed Bishops into superintendents, Priests into ministers, altars into tables, the chaste clergy into the vnlauful mariage of votaries […] (*EEBO* A11445)

3.2 Demonyms and geography

The total number of lemmas concerning geographical entities is fifty; amongst these lexical items we have further identified the following subcategories: toponyms, demonyms, geographical entities, geo-political institutions, expressions deriving from geographical references.

Oddly enough, proper place names are not significantly represented in the database, with a total of three entries, all of them being related to

2021 Jan Kochanowski University Press. All rights reserved.

classical Greco-Roman heritage. In the preface of the English translation of Levinus Lemnius's *De habitu et constitutione corporis* (1561) by Newton (1576) we can read the placename *Camaryne* (today's Camarina in Sicily, Italy) in the obscure idiom "wade into the very Gulphe & Camaryne of mannes apparaunt wilfulnesse". In order to decipher this expression, it is necessary to read Strabo's account of the marsh of Camarina (in Jones 1978 [1917]: 59-82): before the Carthaginians destroyed Camarina in 405-401 BCE, its inhabitants were plagued by malaria caused by a nearby marsh; once they dried it, the disease stopped spreading; however, there was no longer anything stopping Hannibal's army from razing the city. In this sense, *Camaryne* becomes the metonym of "marsh", and the term is thus defined by the *OED* as "a fetid marsh or swamp". The next place name that can be identified in the database is *Sarum*, which first occurs in Foxe (1570) in various collocates: "dioces(se) of Sarum", "Bishop of Sarum", "Chancellor of Sarum", etc. both in English and in Latin. *Sarum* is a latinised form of "Sar" a medieval abbreviation of Salisbury (Mills 2003); both in Foxe and in other contemporary works found in *EEBO* (e.g. A07139, A16292, A05547), *Sarum* does not only refer to a geographical entity but more specifically to the so-called Use of Sarum (or Sarum Use), i.e. the Latin liturgical rite developed at Salisbury Cathedral from the late 11th century until the English Reformation (Cheung Salisbury 2009). Even more prominently, the third toponym extracted in the database is highly symbolic, *Sodom*, which appears in Bale (1550); in this text, the biblical reference, which is also spelt as "Sodome" and "Sodoma", is mainly used to portray Rome as the place of Papal corruption (16):

(16) why so tyrannouslye bynde ye them, to that fylthye Sodome, withoute redempcyon? (Bale 1550)

In *EEBO* the spelling "Sodom" outnumbers (7,092 hits) both "Sodome" (3,048 hits) and "Sodoma" (211 hits), which confirms today's spelling.

What stands out when browsing the list of geographical lemmas is certainly the peculiar prevalence of demonyms and adjectival phrases related to geographical entities, often with competing variations. Considering the religious background of the time and the main influences in the Protestant reform, there is a high occurrence of first citations in lemmas related to the German-speaking areas of Europe. First and foremost, the word *Dutchland* in Bale (1547), which in *EEBO* seems to be a calque from German *Deutschland* and a less common alternative (99 hits) to "Germany" (31,024 hits), "Germanie" (5,216 hits) and the French inspired "Almaine" (1,048 hits)

2021 Jan Kochanowski University Press. All rights reserved.

and "Alemaine" (20 hits). The fact that these place names were perceived as synonyms is clear in "Germany, is a country called of some Dutchland, of some Almaine" (*EEBO* A05237). However, the confusion between this old and the current meaning of "Dutch" is also apparent in the corpus, and the authors of our database also provide first occurrences of two adjectival phrases that are used to distinguish between Germans and the Dutch: *High Dutch*, i.e. High German or *Hochdeutch*, which is found in Daus (1560), and *Low Dutch*, i.e. Low German or *Niederdeutch,* which first appears in Newton (1576). The distinction between these varieties of Germanic languages is widely understood in the scholars of the time, as can be observed in (17):

(17) Although I bee well acquainted with the high and low Dutch tongue, yet I must confesse that in this ancient Frison language I vnderstand nothing. (*EEBO* A68345)

Other German-based competing demonyms in the database are *Saxonian* in Hooker (1599) and *Saxonish* in Bale (1549); the former carries a geographical yet religious connotation in Hooker (18); for this reason, the *OED* assigns the definition "a Protestant of Saxony" to this entry. Moreover, these competing variations have low scores in *EEBO*: *Saxonian* 11 hits (which *OED* lists as an obsolete form of *Saxon*), *Saxonish* 10 hits (marked as archaic in the *OED*), whereas *Saxon* has 22,896 hits in adjectival phrases.

(18) the French Protestants took Arms against their King, […], the Belgick, the Helvetian, the Bohemian, the Saxonian, the Swevian, the English, as consenting for Obedience to their Soveraigns. (*EEBO* A27046)

Similarly, in his English translation of Sleidane's *Commentaries* (1560), Daus uses both *Suevical* and *Swevical* to identify Swabian Protestants, but this appears to be his own coinage, as there is no other evidence of these adjectives in works other than the *Commentaries*. Furthermore, in the same translation Daus refers to Slavic peoples as *Slavonish*, which has only two other occurrences in *EEBO*, mainly in relation to the Slavic peoples settled in the Balkans. Our database also contains another term denoting a Slavic people – *Bohemian* – in Golding (1562), as can be seen in (19); once again, this term designates a geography-based religious entity, as these "Bohemians" are Bohemian Protestants, or Hussites. *EEBO* lists 1,119 occurrences of *Bohemian* alongside the competing variation "Bohemish" (5 hits).

2021 Jan Kochanowski University Press. All rights reserved.

(19) Thus we are hable to allege Luther, Melancthon, Bucer, and that learned Bohemian, for the indifferencie of the Communion to be ministred either vnder one kinde or bothe. (Golding 1562)

The second large area covered by the first occurrences in the geographical section of our database includes the lands and countries bordering the Mediterranean Sea. A small number of first citations concern places in Italy, with obvious references to the conflict with Roman Catholic clergy, e.g. in Foxe (1563) *Etruscan* is found in the phrase "Etruscan tyrant", which needs to be contextualised. Foxe portrays Bishop Bonner as the perpetrator of the most vicious cruelties and injustices against English Protestants under the Catholic government of Mary I of England; his victims included Thomas Tomkins, whose hand was burned following the bishop's orders. Tomkins' faith was tested by Bonner; likewise, Scaevola's valour was tested by the "Etruscan tyrant" Porsenna: Foxe employs this comparison as a means to invest Tomkins with a heroism comparable to that of a legendary champion. The adjective *Italish* stands out in the database as a first citation in Bale (1544); however, in *EEBO* this appears in two collocates that can be traced back solely to Bale (1548) – "Italish warre" and "Italish préest" (A68202) – thus we can conclude this form is likely to be his own coinage as an alternative to "Italian", which occurs extensively in the same pages (e.g. "Italian prouerb"); however, in this text "Italian" prevails mainly as a noun, e.g. "in the yéere of Christ 1368: which yéere the Italians count 1367" and "the ambassador of France was also present with another stranger an Italian". This distinction is not confirmed in *EEBO*, where "Italian" occurs as both a noun and adjective as in contemporary English. Two competing adjective forms are present for Adriatic Sea: *Adriatical*(*l*) in Cooper (1549) and *Adrian* in Newton (1575); as might be expected, in both cases these adjectives collocate only with the noun "sea". *EEBO* shows that these adjectives were indeed competing variants: *Adriatic* has 38 occurrences, *Adriatical*(*l*) 20 occurrences, and *Adrian* (sea) 15 occurrences; in modern usage, *OED* marks "Adriatical" as obsolete, "Adrian (sea)" as poetic and rare. A similar consideration can be made in relation to the word *Turcian* that appears in Foxe (1570): whilst *Turcian* seems to be a nonce word in *EEBO*, two other forms are in competition – "Turkish" (4,629 hits) and "Turkic" (3,165 hits), which in contemporary English ended up conveying different meanings ("relating to Turkey" and "related to the Turkic language family" respectively). The *OED* also lists a very peculiar usage of "to turkish" as a verb meaning "to transform, especially for the worse; to pervert; to turn into something different" from (20):

2021 Jan Kochanowski University Press. All rights reserved.

(20) sayeth how the turkyshed seede is sowen abroade in England, and in Germany, signifying the doctrine that is contrary to the byshop of Rome. (Daus 1560)

Finally, there are a few other first citations belonging to the Mediterranean area: *Mozarabical* in Newton (1575); *Costantinopolitan* in Fulke (1577); *Ephesine* in Fulke (1555); and *Hierosolymitan* in Bale (1538), the last three being originally geographical terms modelled after Romance adjectives and used in these writings as religious references to Christian denominations and ecumenical councils. In *EEBO*, there is only one occurrence of *Mozarabical* by Newton (A19712) alongside 12 occurrences of the competing form "Mozarabic(k)", which mostly collocate with "liturgy", "use" and "office" to identify a liturgical rite of the Latin Church once used generally in the Iberian Peninsula; there are only two occurrences of *Constantinopolitan*(*e*) as purely geographical references; there are 520 occurrences of *Ephesine* and 355 occurrences of the competing form "Ephesian" (which would later become the primary adjective referring to Ephesus); there are 31 occurrences of *Hierosolymitan* and one occurrence of "Jerusalemite" (which is today's most common adjective relating to the city of Jerusalem). It needs to be noted that these words are still used today although they are in some cases marked as dated, but their semantic value have shifted from mere geographical to mostly historical and religious.

A special mention needs to be made for the first occurrences of terms related to Graeco-Roman geography in Golding's translations of Caesar's *Commentaries* (1563), Ovid's *Metamorphoses* (1565), and Pomponius Mela's *Geography* (1583). These expressions include *Parnassian*, from the Greek mountain Parnassus; *Pylian*, the inhabitants of the ancient Greek town of Pylos; *Pythian*, the demonym of Delphi (whose ancient name was Pytho), whose root allegedly derives from the word "python" in (21); "Salentine", the demonym of the ancient tribe of Messapians, also known as Sallentini in ancient Rome (22).

(21) Python […] Which of the serpent that he slue of Pythians bare the name (Golding 1583)

(22) Spartanes buylt, and Cybaris, and Neaeth salentine, And Thurine bay, and Emese, and éeke the pastures fyne Of Calabrye (Golding 1583)

In this translation of the *Geography* we can also find mentions of the *Seres* (390 occurrences in *EEBO*) along with the correlate adjective *Seric*, which

2021 Jan Kochanowski University Press. All rights reserved.

are respectively a loanword and a calque of the Greek and Latin *Seres*/*sericus* (ultimately from the word "silk" in various Eastern Asian languages, wherefrom the current English word "seric" derives) to identify the Chinese, as in (23):

(23) We vnderstand that the first men in Asia Eastward, are the Indians, Seres, and Scithians. The Seres inhabite almost the middle part of the East, the Indians and Scithians, the two vttermost partes: both peoples extending farre and wide, and not onelie toward the East Occean. (Golding 1583)

In addition to this, our database includes the first mention of the demonym *Asian*(*e*) as a noun (1,225 occurrences in *EEBO*): it is found in Bale (1548) in (24):

(24) These were of all nacions of the earth, of al peoples of the world, and of all languages vnder heauen, Gréekes, Latines, Hebrues, Caldeans, Parthyans, Medes, Elamites, Capadocians, Asianes, Phrigian, Egiptianes, Arabianes, Syrians, Africanes and Indians. (Bale 1548)

Our database is also populated with a relatively small number of first citations of foreign local institutions, most of which are borrowings or calques from contemporary non-classical languages. The most remarkable case is the triplet *Sorbonne*, *Sorbonist* and *Sorbonical*(*l*) – the first and the second occur in Daus (1560), the third in Bale (1543) – which highlights how deeply the Sorbonne became involved as a reputable institution with the intellectual struggle between Catholics and Protestants in the 16th and 17th centuries (Conway 2009). Other expressions in this section include *Archduchy*, more specifically the "Archduchy of Austrich" (Foxe 1563) which was possibly a calque of French from Latin (in *EEBO* this term occurs exclusively in the collocation "Archduchy of Austrich", "Archduchy of Austria" and "Archduchy of Insbruck"); *burgrave*, a calque from German *Burggraf* (a military governor of a German town of castle in the Middle Ages), which first occurs in Bale (1551) and is found in *EEBO* in 64 concordances in the collocates [burgrave] + [of] + [German city]; *calfam*, probably a corrupted version of "caliph" is found in Bale (1550); *vaivode*, a borrowing from Slavic *воєвода*/*vojvoda* ('army leader' or 'duke'), appears in Daus (1560) and *EEBO*'s concordances show a prevalence for the collocation [vaivode] + [of] + [Valachia/Transilvania] (with one curious exception "Vaivode of Athens"); *vergobret*, a magistrate in ancient Gaul, which appears in Golding (1563); *piazza*, a borrowing from Italian, in Foxe (1583).

2021 Jan Kochanowski University Press. All rights reserved.

Finally, on a more trivial note, our database comprises the first occurrences of the words *bugger* in Daus (1560) and *buggerage* in Bale (1548). Although these words have nowadays lost any spatial reference, these terms originally have a geographical connotation, more specifically Bulgaria, wherefrom the Bogomil heretics were thought to have originated and spread around the 11th century; abominable rituals were imputed to Bogomils and this association is still rooted in today's use of the word. Considering the time frame, the very nature of the works in the corpus, and the profile of the authors under consideration, we support the idea that Daus and Bale cannot have been completely oblivious to this connection.

4. Conclusion

The research hypothesis of this study is that translators, theologians and controversialists active between 1500 and 1650 were leaders in processes of lexical enrichment. This is supported by the data stored in our database of first citations, as shown in the list of first occurrences discussed in this paper. Not only were these scholars innovative in their own field of expertise, but they influenced terminology in a variety of domains, as the examples in the realm of geography showed. They were even confident enough in their abilities to control the morphological aspects of lexis that produced a variety of (co-existing) possibilities, as can be seen in the analysis of the vocabulary of the Eucharist and of controversial neologisms as well as in the adaptation of the loans for religious sects and in the analysis of demonyms that emphasised how suffixes denoting entities belonging to countries, nations and regions from different linguistic sources – i.e. *-ish* (Germanic), *-(i)an* (Norman French), *-ic(al)* (Latin) – used to be employed interchangeably, to some extent at least. The set of terms considered in this study shows a clear prevalence of derivational strategies (60%) especially in the religious vocabulary with a number of instances of compounding (3%) adaptations (10%) and borrowings (29%), these two being predominant in the geographical terminology. Moreover, our data highlight a dense network of influences from the classical languages into English and between vernacular languages, through the sustained contacts of English and Continental reformers and translators: 39% of the words come from Latin or Greek, 17% from Romance languages (mainly French but also Italian and Spanish), 3% from German and 8% from other languages. This emphasises that the vocabulary of the church, of religion and of the peoples and nations of the world was discussed, re-codified, and significantly enriched

2021 Jan Kochanowski University Press. All rights reserved.

throughout this period, during which we see provenance, ethnicity and belief as overlapping notions and a source for terminological creativity, as well as confusion. From today's perspective, we can observe that 42% of the words are now marked as obsolete and/or rare, 10% as historical and/or archaic and 1% as poetic, which can be explained by the circumstantial nature of the religious terminology regarding the debates and controversies of the time and the historical distance between early modern Britain and present-day English-speaking countries. The words still in common use are mainly those that refer to entities that have undergone little or no change in identity (e.g. *Lutheranism*, *Calvinist*, *Sorbonne*, *piazza*), or those that have gained ground amongst competing variants (e.g. *Asian, Sodom*), or those that have been reframed as references that are still significant on historical grounds (e.g. *Etruscan*, *Bohemian*).

As mentioned at the onset, this is a pilot study devised to test the validity of our methodological approach. The next steps in our research will be to expand and update the sources of our database by refining the selection criteria and to investigate further semantic domains emerging from the data.

REFERENCES

Sources

Early English Books Online (*EEBO*)
https://proquest.libguides.com/eebopqp, accessed February 2021
Historical Thesaurus of the Oxford English Dictionary Online (*HTOED*)
https://www.oed.com/thesaurus, accessed February 2021
Oxford Dictionary of National Biography (*ODNB*)
https://www.oxforddnb.com/, accessed February 2021
Oxford English Dictionary Online (*OED*)
http://www.oed.com, accessed February 2021

Special studies

Ahmad, K. – M. Rogers
2001 "Corpus linguistics and terminology extraction". In: S.E. Wright – G. Budin (eds.) *Handbook of Terminology Management.* Vol. 2: *Application-Oriented Terminology Management*. Amsterdam: John Benjamins, 725-760.
Baker, P.
2006 *Using Corpora in Discourse Analysis.* London: Bloomsbury.

2021 Jan Kochanowski University Press. All rights reserved.

Barber, C.
1997 *Early Modern English*. Edinburgh: Edinburgh University Press.
Biber, D.
2009 "Corpus-based and corpus-driven analyses of language variation and use". In: B. Heine – H. Narrog (eds.) *The Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press, 159-191.
Boukhaled, M.A. – B. Fagard – T. Poibeau
2019 "The dynamics of semantic change: A corpus-based analysis". In: J. van den Herik – A.P. Rocha – L. Steels (eds.) *Agents and Artificial Intelligence. 11th International Conference, ICAART 2019, Prague, Czech Republic, February 19-21, 2019, Revised Selected Papers.* Cham: Springer, 1-15.
Brewer, C.
2010 "The use of literary quotations in the *Oxford English Dictionary*", *The Review of English Studies* 61 (248), 93-125.
2013 "*OED Online* re-launched: Distinguishing old scholarship from new", *Dictionaries: Journal of the Dictionary Society of North America* 34, 101-126.
Cabré Castellví, M. T.
1998 *Terminology: Theory, Methods and Applications*. Amsterdam: John Benjamins.
Cheung Salisbury, M.
2009 "Sarum use. The ancient customs of Salisbury", *The Journal of Ecclesiastical History* 60 (3), 582.
Coleman, J.
2013 "Using dictionary evidence to evaluate authors' lexis: John Bunyan and the *Oxford English Dictionary*", *Journal of the Dictionary Society of North America* 34, 66-100.
Considine, J.
2009 "Literary classics in *OED* quotation evidence", *The Review of English Studies* 60 (246), 620-38.
Conway, S.
2009 "Christians, Catholics, Protestants: The religious links of Britain and Ireland with continental Europe, c.1689-1800", *The English historical review* 124 (509), 833-862.
Durkin, P.
2014 *Borrowed Words: A History of Loanwords in English*. Oxford: Oxford University Press.
Gamper, J. – O. Stock
1998 "Corpus-based terminology", *Terminology* 5 (2), 147-159.
Goodland, G.
2013 "Reading Early Modern literature through *OED3*", *English Text Construction* 6 (1), 17-39.
Gotti, M.
2005 *Investigating Specialized Discourse*. Bern: Peter Lang.

2021 Jan Kochanowski University Press. All rights reserved.

Gotti, M. – S. Šarčević
2006 *Insights into Specialized Translation*. Bern: Peter Lang.
Görlach, M.
1991 *Introduction to Early Modern English*. Cambridge: Cambridge University Press.
Hadni, M. – A. Lachkar – S.A. Ouatik
2014 "Multi-word term extraction based on new hybrid approach for Arabic language", *Computer Science & Information Technology* 4, 109-120.
Hammond, M.
2020 *Python for Linguists*. Cambridge: Cambridge University Press.
Hanks, P.
2012 "The corpus revolution in lexicography", *International Journal of Lexicography* 25 (4), 398-436.
Hughes, G.
1988 *Words in Time: A Social History of English Vocabulary*. Oxford; New York: Basil Blackwell.
Isozaki, H.
2001 "Japanese named entity recognition based on a simple rule generator and decision tree learning". In: B.L. Webber (ed.) *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*. Stroudsburg: Association for Computational Linguistics, 314-321.
Jang, H. – Y. Leong – B. Yoon
2021 "TechWord: Development of a technology lexical database for structuring textual technology information based on natural language processing", *Expert Systems With Applications* 164, 114042.
Jones, H.L. (ed.)
1978 [1917] *Strabo: Geography.* Book VI. Cambridge, Mass.: Harvard University Press.
Kay, C. – K. Allan
2015 *English Historical Semantics*. Edinburgh: Edinburgh University Press.
Kilgarriff, A. et al.
2004 "Itri-04-08 the Sketch Engine", *Information Technology* 1, 105-116.
2014 "The Sketch Engine: Ten years on", *Lexicography* 1, 7-36.
Ma, L.
2011 "Clarification on linguistic applications of Fuzzy Set Theory to natural language analysis". In: *Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), 26-28 July 2011, Shanghai, China*. Institute of Electrical and Electronics Engineers, IEEE Circuits and Systems Society, 811-815.
Mei, L. et al.
2016 "A novel unsupervised method for new word extraction", *Science China Information Sciences* 59, 92102.
Mills, D.
2003 *A Dictionary of British Place-Names.* Oxford: Oxford University Press.

2021 Jan Kochanowski University Press. All rights reserved.

Nevalainen, T.
2000 "Early Modern English lexis and semantics". In: R. Lass (ed.) *The Cambridge History of the English Language. V*ol. 3: *1476-1776*. Cambridge: Cambridge University Press, 332-458.
Pantel, P. – D. Lin
2001 "A statistical corpus-based term extractor". In: E. Stroulia – S. Matwin (eds.) *Advances in Artificial Intelligence. 14th Biennial Conference of the Canadian Society for Computational Studies of Intelligence, AI 2001 Ottawa, Canada, June 7-9, 2001. Proceedings.* Berlin; Heidelberg: Springer, 36-46.
Pecina, P.
2010 "Lexical association measures and collocation extraction", *Language Resources and Evaluation* 44 (1/2), 137-158.
Schäfer, J.
1980 *Documentation in the O.E.D.: Shakespeare and Nashe as Test Cases*. Oxford: Clarendon Press.
Stanković, R. et al.
2016 "Rule-based automatic multi-word term extraction and lemmatization". In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016).* European Language Resources Association (ELRA), 507-514.
Tognini-Bonelli, E.
1996 "Towards translation equivalence from a corpus linguistics perspective", *International Journal of Lexicography* 9 (3), 197-217.
Weikum, G. et al.
2012 "Big Data methods for Computational Linguistics", *IEEE Data Engineering Bulletin* 35 (3), 46-64.
Wright, S.E. – G. Budin (eds.)
2001 *Handbook of Terminology Management.* Vol. 2: *Application-Oriented Terminology Management*. Amsterdam: John Benjamins.
Xiao, R.
2008 "Theory-driven corpus research: Using corpora to inform aspect theory". In: A. Lüdeling – M. Kytö (eds.) *Corpus Linguistics: An International Handbook*. Vol. 2. Berlin: Mouton de Gruyter, 987-1008.

Address: Angela Andreani, Università degli Studi di Milano, Department of Foreign Languages and Literatures, Piazza S. Alessandro, 1, 20123 Milan, Italy.
ORCID code: orcid.org/0000-0002-5331-8915

Address: Daniel Russo, University of Insubria, Department of Human Sciences and Territorial Innovation (DiSUIT), via Dunant 7, 21100 Varese, Italy.
ORCID code: orcid.org/0000-0002-6164-1384

2021 Jan Kochanowski University Press. All rights reserved.