

Andrea Bölcskei

✉ bolcskei.andrea@kre.hu

🆔 <https://orcid.org/0000-0003-2628-5297>

🏠 Károli Gáspár University of the Reformed Church
in Hungary

🌐 Budapest, Hungary

🔗 <https://doi.org/10.4467/K7478.47/22.23.17724>

The Use of Terms in the Proceedings of ICOS Congresses

Abstract

In the modern theory of Terminology, it is claimed that terminology work is descriptive in approach, exploring actual term use in specialized discourse. Terminological units are examined from cognitive, linguistic and socio-communicative points of view. This paper examines the use of terms presented in the “ICOS List of Key Onomastic Terms” in the volumes of the proceedings of the latest ICOS congresses with the help of modern corpus linguistic tools. Based on the text corpora provided in the conference volumes, the present enquiry focuses on features such as the frequency of terms and distributional patterns, performing a collocation analysis and a search for synonyms and definitions with respect to established ICOS key terms. By extracting term candidates from the texts and comparing their features with those of the already accepted onomastic terms, the paper also comments on terms recommended recently for acceptance by the Terminology Group and makes suggestions on adding some further expressions to the next edition of the ICOS key terms list. The paper explains and illustrates through relevant examples how the latest ideas, trends and corpus linguistic applications in Terminology can be adopted in terminology work on onomastic terms, for the benefit of experts in several distinct branches of science in general and of the community of name scholars in particular.

Keywords

onomastic terms, corpus linguistic tools, ICOS, theory of Terminology, conference volumes

1. Introduction: Theories of Terminology

Terminology is a relatively new discipline; still, there have been a lot of changes in this field of science since its beginnings in the early 1930s. The classical approach known as the General Theory of Terminology, released by Eugen Wüster in his seminal doctoral dissertation (1931), required subject-field experts to compile, systematize and standardize previously miscellaneous and diverse terminologies, in order to enhance the efficiency of specialized discourse and improve knowledge transfer at an international level. Wüster's students and followers, especially Helmut Felber and the members of the Vienna School took steps towards involving several ideas of contemporary linguistics in Terminology, when developing the Extended General Theory of Terminology. Still, on epistemological grounds, they also emphasized the importance of the onomasiological (i.e., the priority of concepts over terms) and prescriptive (i.e., normative) considerations in terminology management, as propagated by their master. Many principles of the General and the Extended General Theories survive today in the activities of the Infoterm and TermNet networks (Wüster, 1931, 1979; Felber & Budin, 1989; Cabré, 2003; Fóris, 2021).

A paradigm shift took place in the 1990s, when a group of linguists turned towards Terminology. Language experts were eager to describe the actual linguistic behaviour of terms in real-life professional communication, in the firm conviction that terms, as language items in general, are inherently diverse and resist “simplification” and “unification”. The Communicative Theory of Terminology, as presented by Teresa Cabré in several of her writings, compares the features of specialized languages to those of the standard language variant (cf. Cabré, 1999, pp. 160–193). In this theory, multifaceted terminological units are examined from cognitive (i.e., the knowledge component: conceptual mappings, logical relations), linguistic (i.e., the language component: linguistic description) and socio-communicative (i.e., the pragmatic component: communicative function, variants, socio-cultural background) points of view to explore their peculiarities (cf. Cabré, 2003, p. 186).

The Sociocognitive Theory of Terminology, elaborated by Rita Temmerman (2000), accepts terms as units of knowledge, and brings their dynamic nature to the fore. Frame-Based Terminology, as explained by Pamela Faber (2012) and the members of her research group, focuses on the description of prototypical

domain-specific events. These events activate templates for the processes, actions, entities, and participants involved. Topic-related specialized knowledge is conveyed by established semantic frames. Ontoterminology, invented by Christopher Roche (2012) and his colleagues, defines concepts and their relations in a formal language, easily used by computers, while terms and their relations are given in natural human languages. As a last step, the conceptual and lexical units of knowledge representation and their networks are combined into a single system in ontoterminological databases (Temmerman, 2000; Faber, 2012; Roche, 2012).

Newer trends, under the influence of the recent theories in Linguistics, advocate the adoption of a descriptive approach to Terminology; accept the relevance of the onomasiological (i.e., from concept to term) model in terminology management, but also stress the practical use of the semasiological (i.e., from term to concept) approach in terminography; and propose the harmonization of terms and terminological systems as an effective working method in the field instead of standardization. Reconciling distinct (systems of) terms in harmonization processes enhances multilingualism in specialized discourse and protects the individual's language rights. In the view of linguists, corpus-based term analyses are an inevitable step in the understanding of the characteristics of professional communication. Furthermore, terms, definitions, sample sentences and style and usage notes are best extracted from authentic specialized texts (if necessary, in different languages), written by native-speaker subject-field experts (Cabr e, 1999, pp. 160–193; 2003; Temmerman, 2000; Faber, 2012; Roche, 2012).

The representatives of the General and the Extended General Theories of Terminology believed that concepts are universal, static, clear-cut, independent, and are in hierarchical relations within their (sub)systems. The representatives of the linguistic terminological trends, however, recognized that concepts are culture- and language-dependent, non-static, with blurred boundaries, and are in both hierarchical and non-hierarchical relations with one another in their (sub)systems. The concept-centred trends in Terminology before the 1990s tried to produce unambiguous, univocal, monosemic terms by singling out one of the competing terms for general use. They focused almost exclusively on the level of lexemes and defined isolated terms, and typically described the synchronic use of terms in written specialized discourse. The more term-centred trends since the 1990s have demonstrated that terms can be synonymous, homonymous and polysemic in professional

communication. Metaphors and connotations also play an important role in creating new specialized meanings, while term variants are used to indicate the users' social, professional and power status. Furthermore, not only lexemes, but also phrases, collocations, morphology as well as syntax, that is, all levels of linguistic expression should be examined, when looking for terminological units. Finally, synchronic and diachronic observations (on term formation, term changes, unstable term systems, etc.) are equally relevant when terms are concerned (Cabré, 1999, pp. 160–193; 2003; Temmerman, 2000; Faber, 2012; Roche, 2012; Fóris, 2021).

2. Computational and manual terminology work

Computational Linguistics and Terminology can mutually contribute to each other's development. Software applications and tools – by providing fully or semi-automatic solutions – help terminologists build and manage large domain-specific corpora; to extract terms, semantic and syntactic information as well as definitions from computer-processed professional texts and store them in databases; and also, to map, display and extend conceptual relations in appropriate ontological formats. The more precisely described the characteristic linguistic features of professional texts are, the more effective corpus building web crawlers and language query systems can be developed. The latest trends in Terminology place great emphasis on exploring knowledge representation and try to give detailed descriptions of the conceptual and linguistic units and their relations in a given domain, which contributes to successful (semi-)automatic text and discourse analyses in Computational Linguistics.

Although manual terminology work is undoubtedly more accurate than computational term extraction and description, it is also more time-consuming and less cost-efficient. Moreover, computational term processing can easily deal with enormous amounts of data that could not be processed by humans. Thus, combining computational and manual term analyses can be a solid basis for terminology work of high quality. Checking the computational outcome and eliminating its errors manually (by terminologists in close cooperation

with subject-field experts) in the validation phase of terminology work can provide optimal and satisfactory results.

3. Term use in ICOS proceedings

In the present paper, I will check the proceedings of the last three ICOS congresses with respect to (3.1) the frequency of the terms listed in the document entitled “ICOS List of Key Onomastic Terms” (see below); (3.2) the use of onomastic terms containing *onym*; (3.3) some keywords and term candidates extracted automatically from the texts. The survey was completed with Sketch Engine, a leading corpus manager and text analysis software in lexicography. The sets of volumes of the proceedings are those of the Barcelona (2011), Glasgow (2014) and Debrecen (2017) congresses. Sketch Engine created three corpora from the pdf versions of the congress proceedings, which I conveniently named the Barcelona (= B), Glasgow (= G) and Debrecen (= D) corpus. According to Sketch Engine, the size of the three corpora is as follows: B: 1,638,575; G: 477,079; D: 550,265 words. These corpora include all the papers in the volumes, regardless of their languages, as non-English (i.e., German, French, Spanish, Catalan) abstracts and papers in several cases make use of English onomastic terms. Results and observations in sections 3.1 and 3.2 are based on these three corpora.

However, monolingual corpora should be compared to identify keywords and term candidates. In such a case, a corpus of domain-specific texts is compared to a corpus of general texts to find lexemes, phrases appearing in special patterns in the former source, and thus presumably acting as keywords or terms in the domain-specific texts. For this purpose, I selected the English papers from each proceedings volume, compiled them into three monolingual corpora with Sketch Engine, and indexed them as B(en), G(en), D(en), respectively. Their size as accounted by the application is: B(en): 631,414; G(en): 357,875; D(en): 453,263 words. These three English corpora were compared to English Web 2020 (enTenTen20), a 38,149,437,411-word corpus of contemporary general texts from the web, which is provided as one of the possible reference corpora in Sketch Engine. Results and observations in section 3.3 are based on these English corpora.

3.1. The frequency of key onomastic terms

The “ICOS List of Key Onomastic Terms”, published on the ICOS website, enumerates 70 terms (ICOS Terminology Group, n.d.). These terms were considered the most relevant in Onomastics by the ICOS Terminology Group in 2010. The master list was prepared in English and has already been translated into several languages.¹

The total number of occurrences of the 70 key onomastic terms in the proceedings of the last three ICOS congresses paints the following picture (from the lowest to the highest absolute frequency): *denominatum* (B: 0, G: 0, D: 0 = 0); *proprialisation* (B: 0, G: 0, D: 0 = 0); *cryptonym* (B: 0, G: 1, D: 0 = 1); *psychoonomastics* (B: 1, G: 0, D: 0 = 1); *nesonym* (B: 3, G: 0, D: 0 = 3); *allonym* (B: 3, G: 0, D: 1 = 4); *socioonomastics* (B: 3, G: 1, D: 2 = 6); *inhabitant name* (B: 2, G: 3, D: 2 = 7); *metronym* (B: 0, G: 2, D: 6 = 8); *minor name* (B: 4, G: 4, D: 0 = 8); *deonymisation* (B: 5, G: 0, D: 5 = 10); *onymy* (B: 4, G: 0, D: 6 = 10); *deonym* (B: 8, G: 1, D: 2 = 11); *macrotoponym* (B: 6, G: 2, D: 3 = 11); *onymisation* (B: 8, G: 1, D: 2 = 11); *theonym* (B: 1, G: 9, D: 1 = 11); *hagionym* (B: 7, G: 3, D: 2 = 12); *zonym* (B: 10, G: 2, D: 4 = 16); *choronym* (B: 7, G: 2, D: 9 = 18); *ergonym* (B: 11, G: 6, D: 2 = 19); *anthroponomasticon* (B: 0, G: 5, D: 16 = 21); *odonym* (B: 11, G: 6, D: 4 = 21); *hodonym* (B: 12, G: 8, D: 7 = 27); *product name* (B: 20, G: 4, D: 5 = 29); *toponomasticon* (B: 0, G: 1, D: 28 = 29); *eponym* (B: 13, G: 15, D: 3 = 31); *anthroponomastics* (B: 7, G: 27, D: 1 = 35); *appellativisation* (B: 3, G: 0, D: 40 = 43); *animal name* (B: 19, G: 20, D: 16 = 55); *last name* (B: 31, G: 2, D: 23 = 56); *Christian name* (B: 12, G: 13, D: 35 = 60); *field name* (B: 7, G: 29, D: 24 = 60); *oronym* (B: 18, G: 3, D: 40 = 61); *hypocoristic* (B: 9, G: 13, D: 43 = 65); *toponomastics* (B: 17, G: 48, D: 0 = 65); *patronym* (B: 20, G: 6, D: 40 = 66); *pseudonym* (B: 53, G: 10, D: 4 = 67); *oikonym* (B: 10, G: 5, D: 63 = 78); *name bearer* (B: 13, G: 22, D: 54 = 89); *microtoponym* (B: 33, G: 25, D: 32 = 90); *onym* (B: 45, G: 14, D: 35 = 94); *onomastician* (B: 34, G: 17, D: 53 = 104); *forename* (B: 81, G: 19, D: 14 = 114); *ethnonym* (B: 17, G: 50, D: 64 = 131); *namegiver* (B: 35, G: 32, D: 77 = 144); *anthroponymy* (B: 74, G: 45, D: 35 = 154); *hydronym* (B: 40, G: 25, D: 96 = 161); *by-name* (B: 68, G: 67, D: 92 = 227); *endonym* (B: 130, G: 75, D: 30 = 235); *onomasticon* (B: 139, G: 46, D: 57 = 242); *brand name* (B: 228, G: 10, D: 37 = 275); *namegiving* (B: 75, G: 50, D: 177 = 302); *geographical name* (B: 274, G: 65, D: 58 = 397); *street name* (B: 241, G: 59, D: 112 = 412); *exonym* (B: 291, G: 140,

¹ <https://icosweb.net/publications/onomastic-terminology>

D: 23 = 454); *settlement name* (B: 80, G: 38, D: 336 = 454); *first name* (B: 184, G: 158, D: 126 = 468); *family name* (B: 225, G: 132, D: 164 = 521); *anthroponym* (B: 178, G: 104, D: 402 = 684); *toponymy* (B: 341, G: 161, D: 212 = 714); *given name* (B: 255, G: 185, D: 427 = 867); *naming* (B: 308, G: 252, D: 373 = 933); *personal name* (B: 328, G: 218, D: 470 = 1,016); *onomastics* (B: 441, G: 222, D: 413 = 1,076); *proper name* (B: 560, G: 289, D: 403 = 1,252); *nickname* (B: 403, G: 489, D: 609 = 1,501); *surname* (B: 609, G: 810, D: 814 = 2,233); *toponym* (B: 1,144, G: 385, D: 1,008 = 2,537); *place name* (B: 1,323, G: 712, D: 1,020 = 3,055); *name* (B: 11,263, G: 8,429, D: 10,699 = 30,391).

When judging how well-established a term is in specialized discourse, the total number of occurrences is only one of the indicators that should be considered. The distribution of the term is also a relevant factor. For example, the term *nesonym* happens to appear three times in two papers of the Barcelona corpus; the term *deonymisation* (or *deonymization*) can be found ten times in three papers of the Barcelona and Debrecen corpora. The use of these terms can be considered restricted even in onomastic contexts. Their relative frequency in the ICOS proceedings is B: 1.4; and B: 2.33, D: 6.86 per million tokens, respectively.

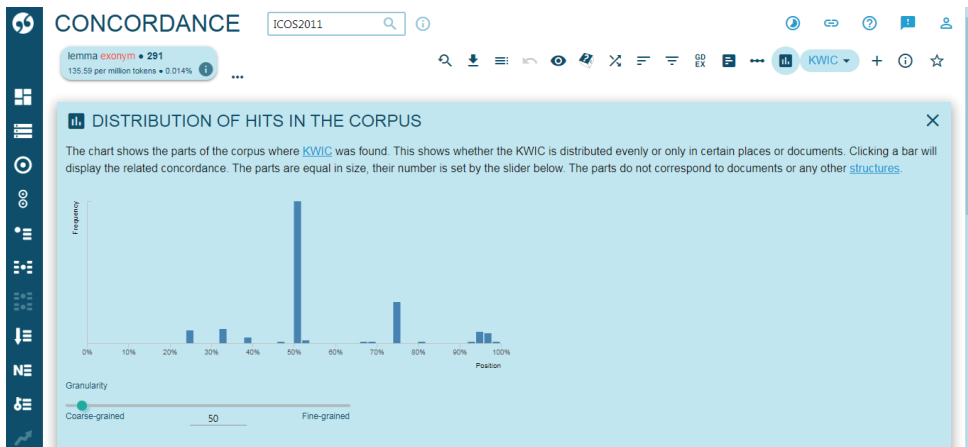


Figure 1. The distribution of the term *exonym* in the Barcelona corpus

Source: Sketch Engine (screenshot).

In comparison, the term *exonym* is used 454 times in 25 papers and occurs in all three corpora. Its relative frequency in the ICOS proceedings is B: 135.59, G: 223.09, D: 31.54 per million tokens. For further comparison, in the English Web 2020 corpus, which consists of several text types on various

topics, the term *exonym*, not surprisingly, is a much rarer, but acknowledged expression, resulting in 1,759 hits out of the approximately 40 billion words, its relative frequency is 0.04 per million tokens. Thus, the lexeme *exonym* can clearly be accepted as a well-established term in onomastic discourse, based on its distributional patterns.

3.2. A case study: Onomastic terms containing *onym*

By searching for lemmas containing *onym* in the ICOS proceedings, words considered to be candidates for “characteristic onomastic terms” can be identified. In this case study, two restrictions were made. First, ordinary words (such as *anonymous*, etc.) and generally used linguistic terms including *onym* (such as *synonym*, *hyperonym*, *metonymy*, *metaphonymy*, *auto-antonyms*, *contronymy*, *meronymy*, etc.) were excluded. Second, fictitious terms used merely to illustrate the possible consequences of the unreasonable adoption of the quasi-suffix *onym* in onomastic terminology were also disregarded (e.g., *genonym*, *anchistonym*, *genethlonym*, *rheitronym*, *krenonym*, etc.). In this way, 161 term candidates were identified in the three corpora, out of which 32 items are included in the current ICOS term list.

94 terms containing *onym* appear in one of the three corpora (NB not necessarily once), out of which the meanings of the 16 bold items can easily be deduced from the current ICOS term list by non-experts as well: *abionym*, *aegistoponym*, *agnonym*, *agronym*, *agrotoponym*, *anonym*, *anoikonym*, *anoikonymy*, *anonymonymous*, *anthroponymicon*, *anthroponymism*, **anthroponymist**, **anthroponymization**, *aoronym*, *appellonym*, *appellonymy*, *astronymy*, *avonymic*, *bionym*, *connotonym*, *connotonymisation*, *connotonymy*, **cryptonym**, **de-anthroponymic**, *dromonym*, *econym*, *econymy*, *ecotoponym*, **eponymics**, **eponymization**, **eponymy**, **ergonymy**, *ethnotheonym*, *ethnotheonymy*, *eventonym*, *exo-ethnonym*, *gastronym*, *geortonym*, *geotoponym*, *glottonym*, *gyneconymic*, *helonym*, **hodonymy**, *ideonymic*, **interonymisation**, *irconym*, *kyrionym*, *literaronym*, *logonym*, *meso-toponym*, **metaonymic**, **metronymy**, *mythonym*, *nanotoponym*, *nanotoponymy*, *neonym*, **nesonym**, **odonymy**, *oikodomonym*, *orohodonym*, *orotoponym*, *paleotoponymy*, *papponymy*, *poetonymology*, *poetonymy*, *politonym*, *polyonym*, *polyonymy*, *potamonym*, *pragmatonym*, *propriumy*, *pseudoandronym*, *pseudonymity*, *pseudonymization*, *pseudotoponym*, *remionym*, *spelenonymy*, **semi-onym**, *shibbolethnonym*, *skionym*, *socioanthroponymy*,

*socioorohodonym, symbantonym, symphytotoponym, synphytonym, teknonym, theotoponym, toponymism, **toponymized**, toponymsociological, toponymity, trawleronym, urbanhydronym, zootoponym.*

31 *onym*-terms occur in two of the three corpora, out of which (the stems of) the 16 bold items are included in the current ICOS term list: ***allonym, andronym, anthropotoponym, chrononym, cosmonym, deonymisation, exonymisation, geonym, hagiotoptonym, isonymy, limnonym, macrotoponymy, matronymic, metronym, oikonymy, onym-naming, onym-building, onym-making, onymy, oronymy, patronymy, phaenomenonym, phytotoponym, poetonym, theonymy, toponymics, toponymisation, transonymic, transonymisation, urbanonym, urbanonymy.***

36 terms including *onym* can be found in all three corpora, out of which the 31 bold items are surely comprehensible for those who are familiar with the current ICOS term list: ***anthroponym, anthroponymy, astronom, charactonym, choronym, chrematonym, chrematonymy, demonym, deonym, endonym, eponym, ergonym, ethnonym, exonym, hagianym, hodonym, hydronym, hydronymy, macrotoponym, microtoponym, microtoponymy, odonym, oikonym, onym, onymisation, oronym, patronym, phytonym, pseudonym, theonym, toponym, toponymicon, toponymist, toponymy, zoonym, zoonymy.***

Onym-terms with more than 50 occurrences from the latter two groups are mostly included in the current ICOS term list: *anthroponym, anthroponymy, endonym, ethnonym, exonym, hydronym, microtoponym, oikonym, onym, oronym, patronym, pseudonym, toponym, toponymy*. The two exceptions are *charactonym* (appearing in all the three corpora: 63 occurrences, used in 4 papers by 2 authors) and *urbanonym* (appearing in two corpora: 126 occurrences, used in 12 papers by 15 [co-]authors). In the ICOS proceedings, *urbanonym* has twice as many occurrences as *charactonym*; in the English Web 2020 corpus, however, *charactonym* proves to be significantly more frequent and widespread in use than *urbanonym* (68 vs 5 occurrences, respectively). Based on their frequency and distribution in specialized and general texts, these two terms are worth considering as new entries in the next edition of the ICOS term list.

The frequency of *onym*-terms quoted above from the ICOS proceedings was also observed in the English Web 2020 corpus. Special attention was paid to terms whose exact meanings cannot be deduced from the current ICOS term list. Such terms with the highest number of occurrences might be interesting for the general public; consequently, their appearance in the next edition

of the ICOS term list would also be justifiable. These terms include *demonym* (1,722 occurrences), *isonym(y)* (48 occurrences), *geonym* (16 occurrences), *phytonym* (10 occurrences), *andronym* (7 occurrences), *cosmonym* (6 occurrences), *astronym* (5 occurrences).

The Wikipedia entry for the suffix *-onym* (“-onym”, 2021) lists and defines 118 terms ending in *-onym* (15 of which are not relevant in the present survey: *contronym* or *antagonym* or *autoantonym*, *homonym*, *meronym*, *metonym*, *paronym*, *phantonym*, *retronym*, *synonym*, *textonym*, and terms used in nomenclature systems: *autonym*, *basionym*, *chresonym*, *heterochresonym*, *orthochresonym*, *taxonym*). 55 terms in the Wikipedia entry also appear in (one of) the ICOS proceedings; 48 of them, however, cannot be found in the volumes of the last three ICOS proceedings, which – even if they are intelligible – questions their acceptability as true onomastic terms (i.e., terms used commonly by subject-field experts in specialized discourse). These “illusory” *onym*-terms are *agoronym*, *anacronym*, *anepronym*, *apronym*, *apronym*, *asteroidonym*, *astionym*, *autoethnonym*, *autoglossonym* or *autoglottonym*, *autolinguonym*, *backronym*, *caconym*, *capitonym*, *cometonym*, *comonym*, *drymonym*, *ecclesionym*, *endochoronym*, *endoethnonym*, *endolinguonym*, *endotoponym*, *ethnochoronym*, *ethnohydronym*, *ethnotoponym*, *exochoronym*, *exolinguonym*, *exotoponym*, *gamonym*, *glacionym*, *gynonym*, *hypocoronym* or *hypocorism* or *hypocoristic*, *isonym*, *linguonym*, *meteoronym*, *mononym*, *necronym*, *numeronym*, *oceanonym*, *paedonymic*, *pelagonym*, *phaleronym*, *planetonym*, *prosoponym*, *speleonym*, *tautonym*, *theronym*, *topoanthroponym*, *topoethnonym*. Seemingly, whoever compiled the list strongly supported the adoption of the prefixes *auto-*, *endo-*, *ethno-*, *exo-*, etc. as term elements, which in fact is not very typical of specialized language use by onomasticians.

The *onym*-terms that appear both in the volumes of the ICOS proceedings and in the Wikipedia entry on *-onym* but which are missing from the current ICOS term list are very often compounds or derivations from onomastic terms which are defined in the key terms list (e.g., *hagiotoponym*, *oikodomonym*, *transonymic*). In other cases, they are spelling, or formal variants of terms included in the ICOS term list, for example, *econym* (see the entry *oikonym* in the ICOS term list), *matronym* (see the entry *metronym* in the ICOS term list). Some of them are culture-dependent terms, for example, *chrononym* = “a proper name of a time period, like the *Bronze Age*, or the *Middle Ages*” (as defined in the relevant Wikipedia entry). However, it should be noted that, for instance, the Hungarian equivalents of these *age*-expressions are not considered proper

names. Thus, chrononyms are only elements of the onymic systems in some languages. Many of the *onym*-terms not included in the ICOS term list are individual innovations (e.g., *gastronym*, *nanotoponym*); or the results of the necessity of adopting suitable terms for new concepts (e.g., *geonym*).

Through comparing the frequency of the internationally intelligible synthetic *onym*-terms with that of their traditional English analytic equivalents in onomastic specialized discourse and in general web texts, some interesting trends can be observed. For example, the term *hipponym* appears in few cases in onomastic papers and is practically unknown in general web texts (B: *hipponym* 3; English Web 2020: *hipponym* 1), in which its traditional analytic equivalent, *horse name* is used instead (English Web 2020: *horse name* 762), a term hardly used in the conference proceedings (B: *horse's name* 1; D: *horse's name* 1).

Similarly, the traditional English analytic term *house name* is significantly more frequent in general web texts as well as in the volumes of the proceedings than the synthetic term *oikonym* in its strict sense ('house name') (English Web 2020: *house name* 5,812, *oikonym* 1; B: *house name* 7, *oikonym* 0; G: *house name* 65, *oikonym* 1; D: *house name* 4, *oikonym* 0). Interestingly enough, *oikonym* is used more often in a general sense ('place name', 'settlement name') in the three onomastic corpora (B: *oikonym* 15; G: *oikonym* 6; D: *oikonym* 70). The term *oikonym* in the latter sense appears mostly in papers by German, Lithuanian, Latvian, Romanian, Hungarian scholars, and is often used by the onomasticians of Slavic languages in all three corpora, reflecting the specific discourse practices of (some) Central and Eastern European subject-field experts.

A somewhat similar trend can be discovered in a comparison of the use of the terms *toponym* and *place name*. Although *place name* displays far more occurrences than *toponym* in general web texts and a slightly higher frequency in the Glasgow corpus (edited by native English speakers) (English Web 2020: *toponym* 7,675, *place name* 52,271, *place-name* 17,252, *placename* 9,396; G: *toponym* 385, *place name* 618, *place-name* 94), its dominance is less significant in the Barcelona corpus (B: *toponym* 1,144, *place name* 611, *place-name* 669, *place-name* 43), and in the Debrecen corpus the two terms are more or less level (D: *toponym* 1,008, *place name* 345, *place-name* 632, *placename* 43).

3.3. Some extracted keywords and term candidates

With respect to the keywords and term candidates extracted automatically by Sketch Engine, one can realize that – apart from proper onomastic terms (see below) – the search returns results featuring some specific, but from the current perspective non-relevant types of terms, words and collocations, as they appear in patterns characteristic of strictly domain-specific terms in specialized texts. These include phonetic terms (e.g., *voiceless velar stop, palatal sonorant, approximant consonant*), orthographic terms (e.g., *grapheme, digraph, cyrillic script, initial capital letter, romanization*), general linguistic terms (e.g., *appellative, polysemy, lemma, grapheme, geographical distribution, folk etymology, word formation, calque, mental lexicon, informant*), identifications of languages (e.g., *source language, receptor language, recipient language, official language, target language, mother tongue, Old Norse, Spanish language, native language*), sources of names (e.g., *land register, town court record book, official gazetteer, tax schedule, epigraph, street map, death registry, marriage record, census book*), potential denotata of names (e.g., *meeting-place, farmstead, hundredal system, residential complex, administrative unit*), references to unique onomastic phenomena (e.g., *hyphenated double-barrelled personal name, carnival place nickname, Hebrew theophoric name, Scottish namescape*), word-for-word translations from mother tongue into English (e.g., *name body, toponym, name community, name creation, bare toponym*), branches of science relevant to onomastics (e.g., *cartography, historical geography, human geography, linguistic anthropology, settlement history*) and geographical terms (e.g., *natural landscape, geographical feature, low land, topographical feature, karst region, hydrographic feature, undersea feature*).

The statistics below demonstrate how reliable term extraction by Sketch Engine and how well-edited the document “ICOS List of Key Onomastic Terms” is. Out of the 70 items in the “ICOS List of Key Onomastic Terms” 53 (75.71%) were identified as terms by Sketch Engine. Seven (10%) of the missed 17 items are either generally used name-related expressions (i.e., *given name, last name*), or derivations from *onym* or *proprial*, acknowledged as keywords in the automated term extraction processes (i.e., *deonym, deonymisation, onymisation, onymy, proprialisation*). Thus 10 terms in the list were not recognized automatically (14.29%), nine of which are in fact used in some of the volumes (*allonym, cryptonym, hagnonym, inhabitant name, macrotoponym, minor name, nesonym, psychoonomastics, zonym*). The 10th term, *denominatum* cannot be

found in any of the volumes; instead, authors prefer the term *denotatum* (B: 23, G: 9, D: 8 occurrences). Additionally, although the term *allonym* appears three times in the Barcelona corpus, the expression *name variant* seems to be more accepted in this sense by onomasticians (B: 9, G: 9, D: 11 occurrences).

Sometimes definitions provided in the ICOS term list can be improved on the basis of the actual term use in the volumes of the three proceedings. For instance, in the ICOS list the informal name of a place is labelled as a *by-name*, while *nicknames* are exclusively connected to people. In practice, a study devoted to Dutch carnival *place nicknames* in the Glasgow corpus illustrates the well-established use of the term *nickname* in connection with locations as well (cf. G vol. 5, pp. 124–131). The identification “product and company names (ergonyms)” in the Debrecen corpus suggests a broad meaning of the term *ergonym* (defined as “name of a product or a brand” in the ICOS term list), similar to *chrematonym* in some languages (D OU vol. 13, p. 315; see also the entry *ergonym* in the ICOS term list).

Apart from the terms listed in the ICOS document, some additional strictly onomastic keywords and term candidates are suggested by Sketch Engine. These expressions can be categorized as special types of personal names (e.g., *apotropaic name*, *baptismal name*, *derisive name*, *theophoric name*); subtypes of place names (e.g., *farm name*, *valley name*, *fjord name*, *urbanonym*, *urban toponym*); umbrella terms (e.g., *astronym*, *geonym*, *commercial names*, *commemorative names*); distinct types of names (e.g., *phytonym*, *charactonym* ~ *characternym*, *hotel name*, *station name*, *ship name*); linguistically based onomastic expressions (e.g., *namespace*, *onomastic landscape*, *toponymic landscape*, *onomastic community*, *onymic system*, *onomastic geography*, *toponymic stock*, *toponymicon*, *toponymist*); derivatives of the term *onymisation* (e.g., *toponymisation*, *connotonymisation* ~ *connotonymization*, *anthroponymization*, *eponymization*); elements of names (e.g., *specific element*, *distinctive addition*, *generic term* ~ *generic appellative* ~ *geographic[al] appellative* ~ *geographic[al] common noun/name/word*, *locational suffix*, *patronymic suffix*). In rare cases, the definitions of some of these terms are also provided in the observed texts: “Geonym is defined as a geographical name or name of a geographical feature, while geoname is an informal alternative for geographical name” (B, p. 1975); “The process of the connotative transfer is called *connotonymisation* or *connotonymization* – the network of associations” (B 86); “*eponymization* or name transfer” (D OU vol. 12, p. 206).

4. Computer-assisted terminology, terminological entries

As seen above, automatic corpus building and corpus analysis can help, but cannot substitute human work in terminology. It is people who set up the web crawler to build domain-specific corpora, prepare the collected texts for processing, then assess, sort out and validate automatically gained data. An “ideal” terminological entry is supposed to comprise the following types of information in connection with each term: ID number; head term; source of the head term; domain; definition; source of the definition; context(s) and source(s) of the context(s); illustrative examples; origin of the head term; grammatical category of the head term; variants; synonyms; hyperonym; hyponym(s) (if available); foreign language equivalent(s) and source(s) of foreign language equivalent(s); notes on term use, applicability and frequency; bibliography; experts’ comments, discussion; validity; author of the record/validation; date of record/validation; etc. In practice, out of all these pieces of information it is enough to focus on the items relevant to the actual enquiry, task or project, to support time and cost efficiency.

Terminologists claim that preparing a proper terminological entry with all the necessary information in appropriate forms from the results of (semi-) automatic corpus analysis usually takes a lot of time and effort. In general, 40 terms can be processed per day (approx. 12 minutes/term), if all data needed are easily available, and cooperation with subject-field experts is unimpeded. The practical use of this activity, however, is extensive. Terminological entries can be imported to CAT-tools, termbases of translation memories (e.g., SDL Trados MultiTerm). Thus, precision, quality and consistency can be improved in translation services; the comprehensibility, integrity and credibility of translated texts, professional documentation or specialized discourses could also be enhanced.

5. Conclusion: The practical use of an onomastic terminological database

A multilingual, online onomastic terminological database could effectively ensure consistent term use in many fields of life. Such a database could serve as a source of knowledge in public and higher education, contribute to quality assurance in translation and language services, and consolidate term use in law, geoinformatic devices, mass media and public administration. It would also help with giving and receiving internationally intelligible information. It could be relied on in research projects, when onomastic and interdisciplinary enquiries are carried out (e.g., in dialectology, sociolinguistics, historical linguistics, contact linguistics, pragmatics, lexicography, geography, history, archaeology, psychology, law, theology). It would disseminate authentic information about the functions of names in discourses, the role of names in preserving culture and shaping identity to the general public. For optimal results, establishing an onomastic terminological database should be accomplished in conformity with other national and international terminological works.

Acknowledgements

Work on this paper was supported by a research grant obtained from the Faculty of Humanities and Social Sciences, Károli Gáspár University of the Reformed Church in Hungary (“The Hungarian Language and Professional Culture: Hungarian Linguistic Research in the Fields of Content Development and Technical Documentation”, project no. 20632B800).

References

- B = Tort i Donada, J., & Montagut i Montagut, M. (Eds.). (2014). *Els noms en la vida quotidiana. Actes del XXIV Congrés Internacional d'ICOS sobre Ciències Onomàstiques / Names in Daily Life: Proceedings of the XXIV ICOS International Congress of Onomastic Sciences*. Barcelona: Generalitat de Catalunya, Departament de Cultura.
- Cabré, M. T. C. (1999). *Terminology: Theory, Methods and Applications*. Amsterdam, Philadelphia: John Benjamins.

- Cabré, M. T. C. (2003). Theories of terminology: Their description, prescription and explanation. *Terminology*, 9(2), 163–199.
- D = Ainiala, T., Aldrin, E., & Győrffy, E. (Eds.). (2018). *Onomastica Uralica*, 10; Coates, R., & Reszegi, K. (Eds.). (2018). *Onomastica Uralica*, 11; Bölcskei, A., Felecan, O., & Štěpán, P. (Eds.). (2018). *Onomastica Uralica*, 12; Slíz, M., & Walkowiak, J. B. (Eds.). (2018). *Onomastica Uralica*, 13; Farkas, T. & Nyström, S. (Eds.). (2018). *Onomastica Uralica*, 14. Debrecen–Helsinki: Debrecen University Press.
- Faber, P. (Ed.). (2012). *A Cognitive Linguistics View of Terminology and Specialized Language*. Berlin–Boston: De Gruyter Mouton.
- Felber, H., & Budin, G. (1989). *Terminologie in Theorie und Praxis*. Tübingen: Narr.
- Fóris, Á. (2021). *Translation and Terminology: Theory and Practice in Hungary*. Budapest: KRE, L'Harmattan.
- G = Hough, C., & Izdebska, D. (Eds.). (2016). *Names and Their Environment: Proceedings of the 25th International Congress of Onomastic Sciences, Glasgow, 25–29 August 2014*. Glasgow: University of Glasgow.
- ICOS Terminology Group. (n.d.). *List of Key Onomastic Terms*. Retrieved November 8, 2021, from <https://icosweb.net/wp/wp-content/uploads/2019/05/ICOS-Terms-en.pdf>
- onym. (2021, August 20). In *Wikipedia*. <https://en.wikipedia.org/w/index.php?title=-onym-&oldid=1039704701>
- Roche, Ch. (2012). Ontoterminology: How to unify terminology and ontology into a single paradigm. In N. Calzolari, K. Choukri, Th. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC-2012)* (pp. 2626–2630). Paris: ELRA.
- Sketch Engine*. *Learn how language works*. Retrieved August 12–15, 2021, from <https://www.sketchengine.eu>
- Temmerman, R. (2000). *Towards New Ways of Terminology Description: The Sociocognitive Approach*. Amsterdam, Philadelphia: John Benjamins.
- Wüster, E. (1931). *Internationale Sprachnormung in der Technik, besonders in der Elektrotechnik. Die nationale Sprachnormung und ihre Verallgemeinerung*. Berlin: VDI.
- Wüster, E. (1979). *Einführung in die Allgemeine Terminologielehre und Terminologische Lexikographie* (Vols. 1–2). Wien: Springer.