Paulina Judycka
ORCID: 0000-0002-0609-092X
University of Szczecin, Szczecin

Linguistische Treffen in Wrocław, Vol. 24, 2023 (II)
ISSN: 2084–3062, e-ISSN: 2657–5647
https://doi.org/10.23817/lingtreff.24-20
S. 291–301

Comparing Multi-Word Term Candidates in Business Valuation Standards with the Use of Sketch Engine and Power Query Functionalities[1]

The purpose of the study was twofold: to examine terminological similarities and differences between three samples of specialized texts representing the same genre – business appraisal standards and to present a method extending the Keyword and term extraction function of Sketch Engine, a tool in corpus linguistics. The standards selected for the analysis were ASA Business Valuation Standards 2009, IBA Professional Standards 2015 and NACVA Professional Standards 2017.
The character of the study was synchronic – the texts selected for the analysis were standards in their latest possible versions and thus applicable for use at around the same time.
The study was carried out in two steps. First, text were compiled into separate corpora and automatically POS-tagged with the use of Sketch Engine corpus software. Then, the multi-word terms were automatically extracted from each text, based on their relative frequency against the preloaded reference corpus English Web 2020, providing the total number of 737 terms in ASA, 373 terms in IBA and 367 terms in NACVA. Next, the data were compiled using Excel spreadsheet functionalities. The results showed a considerable terminological overlap between the NACVA and the IBA standard (366 terms occurred in both standards) as well as a significantly different pool of terms in the ASA standard (with 219 terms occurring in all three standards and the same number shared by ASA and IBA and by ASA and NACVA respectively). A tentative conclusion would therefore be that much greater similarity is to be expected between the IBA and NACVA standards also on other levels of description.

Keywords: term extraction, corpus linguistics, domain-specific texts, Sketch Engine

Vergleich von Begriffskandidaten mit mehreren Wörtern in Unternehmensbewertungsstandards mit Verwendung von Sketch Engine- und Power Query-Funktionen

Die Studie verfolgte zwei Ziele: die Untersuchung terminologischer Ähnlichkeiten und Unterschiede zwischen drei Stichproben von Fachtexten, die dasselbe Genre repräsentieren – Standards für Unternehmensbewertungen – und die Vorstellung einer Methode zur Erweiterung der Funktion zur Extraktion von Schlüsselwörtern und Begriffen von Sketch Engine, einem Werkzeug der Korpuslinguistik. Die für die Analyse ausgewählten Standards waren ASA Business Valuation Standards 2009, IBA Professional Standards 2015 und NACVA Professional Standards 2017.
Der Charakter der Studie war synchron – bei den für die Analyse ausgewählten Texten handelte es sich um Standards in ihren letztmöglichen Versionen und somit um Texte, die ungefähr zur gleichen Zeit verwendet werden konnten.
Die Studie wurde in zwei Schritten durchgeführt. Zunächst wurden die Texte zu separaten Korpora zusammengestellt und mit Hilfe der Korpussoftware Sketch Engine automatisch mit POS-Tags

[1] I would like to express my special thanks to Assoc. Prof. Jacek Cypryjański, Prof. of the University of Szczecin, for helping me explore the Power Query features and functionalities.

annotiert. Dann wurden die Mehrwortbegriffe automatisch aus jedem Text extrahiert, und zwar auf der Grundlage ihrer relativen Häufigkeit im Vergleich zum vorgeladenen Referenzkorpus English Web 2020, was die Gesamtzahl von 737 Begriffen in ASA, 373 Begriffen in IBA und 367 Begriffen in NACVA ergab. Anschließend wurden die Daten mit Hilfe von Excel-Tabellenfunktionen zusammengestellt. Die Ergebnisse zeigten eine beträchtliche terminologische Überschneidung zwischen der NACVA- und der IBA-Norm (366 Begriffe kamen in beiden Normen vor) sowie einen signifikant unterschiedlichen Bestand an Begriffen in der ASA-Norm (219 Begriffe kommen in allen drei Normen vor, ebenso viele in ASA und IBA bzw. in ASA und NACVA). Eine vorläufige Schlussfolgerung wäre daher, dass auch auf anderen Beschreibungsebenen eine viel größere Ähnlichkeit zwischen den IBA- und NACVA-Standards zu erwarten ist.

Schlüsselwörter: Begriffsextraktion, Korpuslinguistik, domänenspezifische Texte, Sketch Engine

Author: Paulina Judycka, University of Szczecin, ul. Mickiewicza 64, 71-101 Szczecin, Poland, e-mail: paulina.judycka@usz.edu.pl

Received: 1.2.2023 **Accepted**: 9.5.2023

1. Introduction

Establishing the base of common terms is one of the vital problems when performing a comparative analysis of domain-specific texts. Researchers in this area are in need of methods and tools that would facilitate as well as precipitate the selection of terms for further analysis. The aim of this paper is to show how combining functionalities of two different tools can allow for a swift comparison of terminology databases of several similar texts within a specific discipline. It will also present the results of a terminological analysis involving texts (corpora) representative of the domain of business valuation performed with the application of the functionalities in question.

2. Theoretical background

The term selection methods applied in corpus linguistics include those involving automatic term extraction, which makes use of statistical properties of words, such as their frequency of occurrence. The main assumption in this approach is that words appearing with certain frequencies in a given text can be considered more relevant for its content – and are thus called **keywords** (cf. Kageura/Marshman 2020: 63). According to Kilgarriff, the examination of keywords for comparative analysis brings best results in corpora which "are very well matched in all regards except the one in question" (Kilgarriff 2009: 1). The author claims that comparing keyword lists obtained from such corpora makes it possible for one particular dimension or difference between texts – for example a difference of genre or of domain – to stand out clearly (cf. Kilgarriff 2009: 1).

In Kilgariff's argument, the corpora to be compared are the focus corpus (fc) and the reference corpus (rc), that is the corpus from which and against which the keywords are identified. The statistics described by Kilgariff to measure how many times a given word is more common in the focus corpus than in the reference corpus, with the size of both corpora taken into account (cf. ibid.) that is keyness[2], identifies the core vocabulary of a general purpose text (cf. Kageura/Marshman 2020: 66). However, since, as pointed out by Kageura and Marshman, terms behave like keywords as far as their distributional aspects are concerned – they occur with high frequencies in domain-specific texts and with frequencies that are considerably higher in corpora related to a given domain than in corpora representing other domains or general language (cf. Kageura/Marshman 2020: 63) – a statistics based on the relative frequency of occurrence can also be used to express termhood – "the degree of representation of the domain-specific concepts" (Kageura/Marshman 2020: 63).

There are, however, at least two restrictions for the application of this statistics-based method. The first one, which is more of a precondition, is that the text corpus[3] used for the extraction of keywords or key phrases has already been identified as domain-specific, that is pertaining to or representative of the specialized domain in question (cf. Kageura/Marshman 2020: 63, Heylen/De Hertog 2015: 205, Dubuc/Lauriston 1997: 81, Wright 1997: 13). This corresponds with the semasiological approach to terminology, making *term* the starting point in research (cf. Santos/Costa 2015: 154) as a unit of discourse which reflects the knowledge of a given community of specialists about a particular subject (cf. Cabré Castellví 2003: 182, Santos/Costa 2015: 158) and relying on texts as (authentic) products of language of this community (cf. Sager 1997: 25, Cabré Castellví 2003: 183, Wright 1997: 13) that provide "living proof that the term is used in the field of reference" (Dubuc/Lauriston 1997: 81).

The second restriction involves the fact that keywords and terms are based on different theoretical grounds – while the former are considered to be attributes of texts, the latter are seen as attributes of domains (cf. Kageura/Marshman 2020: 66, Culpeper/Demmen 2015: 90) and "semantically defined" (Heylen/De Hertog 2015: 203). This means that for a lexical item to be called a term sheer occurrence in a domain-related text, although indicative (cf. Dubuc/Lauriston 1997: 84, Wright 1997: 13), is not sufficient – the word or a phrase must also be assigned to a domain-specific **concept**, that is it must designate an object belonging to a subject field (domain)[4], a particular field of activity or a sphere of reality (cf. Dubuc/Lauriston 1997: 80, Wright 1997: 13, Heylen/De Hertog 2015: 203, Lee 2001: 52, Steen 1999: 113, Kageura 2015: 48, Cabré Castellví 2003: 183). Not all of the particularly frequent words or phrases from a given corpus will fulfill this condition – some will simply occur more frequently in particular texts than in others, indicating tendencies or domain-characteristic features of usage, but

[2] Kilgariff does not actually use the word "keyness" in his article.

[3] In this article a text corpus representing a given domain is understood as a single (authentic) text or a collection of texts and in this sense the words "text" and "corpus" can be used interchangeably.

[4] Lee uses "domains" and "subject fields" interchangeably and relates them to what can also be called "context", that is "categories covering the major spheres of social life" (Lee 2001: 51).

without any relation to actual specific concepts[5]. For this reason, the identification of terms based on statistical methods alone constitutes just a preliminary, pre-semantic selection stage – and in fact it can only identify what is called term candidates (cf. Santos/Costa 2015: 169, Dubuc/Lauriston 1997: 84, Wright 1997: 19), since ultimately the list resulting from the automatic extraction needs to be validated by relevant domain experts (cf. Heylen/De Hertog 2015: 203, Santos/Costa 2015: 154).

With two above mentioned restrictions considered, "sheer frequency of use" (Dubuc/Lauriston 1997: 84) becomes an acceptable measure to indicate the degree to which a term candidate can be associated with a given subject field. Kageura and Marshman even state straightforwardly that "the difference between occurrences of term candidates in the target domain and those in other domains or general corpora gives a good indicator of termhood" (Kageura/Marshman 2020: 63).

The fact that term candidates linked to any domain can be reliably identified against a general corpus brings an interesting methodological solution for terminological analyses performed on specialized texts representing the same subject field. It makes the measure of frequency more flexible, which means that it is possible to apply the all-but-one approach proposed by Kilgariff in a manner that is reverse in order. Thus, it is possible to first extract term candidates from a number of similar domain-specific texts against the same general corpus and then perform the actual comparative analysis on term candidates identified this way.

The question is, which criteria of similarity besides the already mentioned condition of belonging the same domain should be adopted when constructing corpora (collecting texts) for the analysis. An easily distinguishable criterion would be that of a shared text category, where the most commonly exploited category in research and often adopted in corpus building seems to be that of a genre (cf. Lee 2001: 37, Davies 2015: 11). According to Lee "*genre* is the level of text categorization which is theoretically and pedagogically most useful and most practical to work with" (Lee 2001: 37), which is probably because most texts are representatives of some generic category. Another advantage of the genre category is that it is often reflexive (cf. Sinclair et al. 1996: 8), that is self-labelling[6], and yet another one – that it does not require any additional analyses to be performed (whereas a typology based on text types proposed e.g. by Biber, does).

Biber observes, however, that the folk-typology of text genre is not necessarily consistent in terms of linguistic patterns – texts that are counted to represent a particular

[5] This is the case with a lexical bundle (Biber's term) used when a witness is sworn before testifying at court in the United States: "Do you solemnly swear that you will tell the truth, the whole truth, and nothing but the truth?". The automatic term extraction from a corpus consisting of court proceedings involving testimonies will most probably show very high frequencies of such phrases as "solemnly swear" or "the whole truth", even though they can hardly be called legal terms, as they do not refer to any specifically legal concepts.

[6] This feature of genre is particularly useful when compiling text for the analysis and in case when the compiler has to defend or justify their choice.

genre such as a novel or a newspaper article may have quite different linguistic characteristics, whereas texts from different genres can be close to identical in form (cf. Biber 1989: 6). This is because the typology into genres is not based on linguistic differences – genres are "socially constituted, functional categories of text" (Lee 2001: 47), defined or classified on the basis of criteria that are external to text (cf. Biber 1989: 6, Sinclair et al. 1996: 6), that is on such "systematic nonlinguistic criteria" (Biber 1989: 39) as the intended audience, the purpose for which or a communicative situation in which a text is created (cf. Biber 1989: 13, Lee 2001: 38, Sinclair et al. 1996: 6). The categorization of texts into genres "refers to a conventional, culturally recognized grouping of texts based on properties other than lexical or grammatical co-occurrence features" (Lee 2001: 38).

Still, as Biber also points out, a frequent co-occurrence of sets of linguistic features can be indicative of the shared situational, social, and cognitive functions (cf. Biber 1989: 7) and the groupings of certain linguistic features tend to be present more often in genres sharing multiple functions or purposes than in others[7] (cf. ibid: 26). Therefore, even though linguistic similarities between texts should not be assumed a priori based on their generic features alone, certain genres – such, for example that serve exactly the same purpose and address the same (type of) audience as well as cover the same topic – can ultimately be expected to share certain linguistic characteristics.

There also seem to be certain doubts (or perhaps, the lack of consensus) among the authors as to the interrelation between genres and domains (cf. Oakey 2002: 115). Some authors do not seem to perceive genre as a text category to be specifically connected with any topic. Such a view is clearly supported by Biber, who states that a genre such as press reportage can include a report on sports, but also culture or finance and academic prose can be about such distant disciplines as humanities or engineering (cf. Biber 1989: 13). Other authors link the concept of a domain with the generic description of a text. For instance, according to Steen, domain is one of the attributes of a given genre, next to the medium, content, form, function, type and language[8] (cf. Steen 1999: 112). Lee defines genre as a category characterized simultaneously at several levels or in several dimensions, including "situated linguistic patterns

[7] The results of Biber's multidimensional analysis of text types have shown that texts in a corpus that are grouped in clusters based on the similarity of a set of linguistic features tend to represent certain genres more often (in a larger percentage) than the others.

[8] Steen explains in his work how he understands each of the attributes. By language he means the aspects of language analyzed by rhetoric, stylistics or sociolinguistics and involving "the study of diverging patterns of selection and combination of language items". By type (of discourse) he means the (rhetorical) classes of narrative, argumentation, description and exposition. The function (of discourse) is perceived in terms of its main intention, which can be informative, persuasive and instructive. The generic form is the superstructure in van Dijk's sense, or it can be another text pattern such as the problem-solution structure. Content of discourse is associated with its topics or themes and "the medium involves linguistic and nonlinguistic aspects of the material means by which a message is transmitted".

(register), functional co-occurrences of linguistic features (text types), or subject fields (domain), and […] text-structural/discoursal features" (Lee 2001: 52). Therefore, texts that are considered similar according to one author may not meet the requirement of similarity in the view of the other.

3. Analysis

In order to reconcile the theoretical assumptions mentioned above, the texts selected for comparative analysis involving similar corpora were business valuation standards developed by three American organizations associating members among professional appraisers and valuators:

- ASA Business Valuation Standards 2009, developed by the American Society of Appraisers and effective as of November 2009;
- IBA Professional Standards 2015, developed by the Institute of Business Appraisers and effective for arrangements accepted on and after August 1, 2015;
- NACVA Professional Standards 2017, developed by the National Association of Certified Valuators and Analysts and effective for arrangements accepted on and after June 1, 2017.

The preliminary assumption as to the results of the study was that texts sharing multiple features, that is: representing the same, quite specific domain (business valuation, a subdiscipline of the subject field economics and finance) and genre (a standard), aimed at a similar type of audience (business valuators), created or approved by the same type of author (organizations associating members among professional appraisers and valuators), serving the exact same purpose (to provide guidance to the associated members as to the practice, principles and requirements connected with the business appraisal/valuation procedure as well as to commit the members to comply with the instructions) and addressing the same topic (description of the correct business valuation procedure) will be relatively similar as to the lists of their term candidates.

Each of the selected texts was collected in machine-readable format (an active pdf file) and, since term candidates were to be extracted separately from each standard, each text was methodologically considered as a single, independent corpus. The sample, even though small, met the requirements of a corpus definition proposed by McEnery, Xiao and Tono, where "a corpus is a collection of (1) *machine-readable* (2) *authentic* texts (including transcripts of spoken data) which is (3) *sampled* to be (4) *representative* of a particular language or language variety" (McEnery/Xiao/Tono 2006: 5). The authors suggest a less radical treatment of the quality of balance when it comes to specialized corpora[9], claiming that "specialized corpora serve a different yet important purpose from the balanced corpora" (ibid.) and maintain that it is the

[9] According to the authors, specialized corpora are "domain (e.g. medicine or law) or genre (e.g. newspaper text or academic prose) specific" (McEnery et al. 15).

research question which in fact defines representativeness (cf. McEnery/Xiao/Tono 2006: 18). The question of sampling in case of a domain-specific corpus it treated similarly by Heylen and De Hertog, who emphasize that balanced sampling is needed when the domain that the collected texts represent is considered "a domain at large" and not when the sample consists of a finite set of documents (cf. Heylen/De Hertog 2015: 205).

The character of the study was synchronic. Therefore, the texts selected for the analysis were standards in their latest possible versions and thus it was assumed that, even if issued on different dates, they were applicable for use at around the same time.

The study consisted in comparing and contrasting multi-word term candidates (automatically extracted multi-word terms, before validation) in the aforementioned standards. It was carried out in a number of steps. First, selected texts (standards) were compiled into separate corpora (each text, in the format of a pdf file, was uploaded separately and thus formed a one-document corpus) and automatically POS-tagged with the use of Sketch Engine corpus software. Then, the multi-word terms were automatically extracted for each text/corpus with the use of "Keywords and term extraction"[10] functionality. The term extraction tool in Sketch Engine is based on the formula of simple maths[11], which means that the keyness score, a statistics showing the relative frequency of a word, that is how many times a single word or a multi-word expression is more frequent in one corpus against another is calculated as the normalized (per million) frequency of the word in the focus corpus plus a smoothing parameter N (default value N=1) divided by the normalized (per million) frequency of the word in the reference corpus plus a smoothing parameter N (default value N=1). To proceed with the extraction, a reference corpus was needed. The reference corpus for the purpose of the study was the web-based preloaded corpus English Web 2020 consisting of 36 billion words.

4. Results

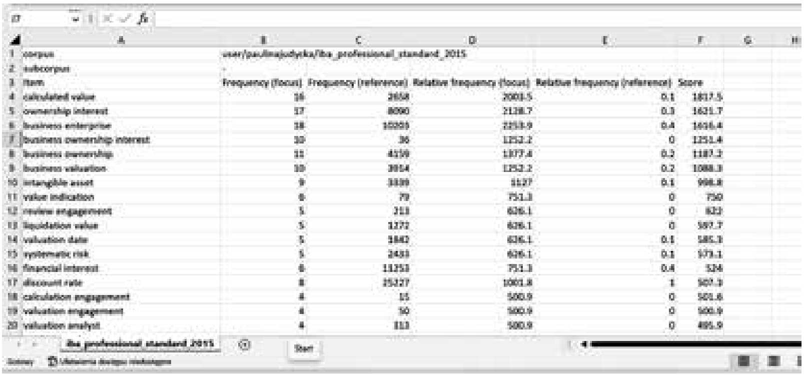
As a result of the extraction, each corpus (that is each document or each standard) produced a file with extracted term candidates. All files were downloaded in cvs format. Fig. 1 below, with the view of the opened cvs file containing the data extracted for the IBA 2015 standard, shows the type of data obtained.

As can be noted, the result of the Sketch Engine "Keyword and term extraction" function is a single file (in cvs or xlsx or xml format) containing term candidates from a corpus (which can be a single document), together with their frequencies (total and relative frequency in both focus and reference corpus) and the keyness score. There are

[10] Description of the extraction method at: https://www.sketchengine.eu/guide/keywords-and-term-extraction, date of access: 27.4.2021.

[11] Description of the simple maths at: https://www.sketchengine.eu/documentation/simple-maths, date of access: 27.4.2021.

no other functionalities built in the Sketch Engine tool that would allow for operations on extracted data or files. “The only way to get the other statistics is to download the results with frequencies and use a spreadsheet to calculate the statistics outside Sketch Engine”[12]. Therefore, in order to compare term candidates across different documents (in this case: standards) it was necessary to make use of extra tools. The tools used for the study were Power Query and Power Pivot, built-in as MS Excel functionalities.



	A	B	C	D	E	F
1	corpus	user/paulinajudycka/iba_professional_standard_2015				
2	subcorpus	-				
3	item	Frequency (focus)	Frequency (reference)	Relative frequency (focus)	Relative frequency (reference)	Score
4	calculated value	16	2658	2003.5	0.1	1817.5
5	ownership interest	17	8090	2128.7	0.3	1621.7
6	business enterprise	18	10205	2253.9	0.4	1616.4
7	business ownership interest	10	36	1252.2	0	1251.4
8	business ownership	11	4159	1377.4	0.2	1187.2
9	business valuation	10	3954	1252.2	0.2	1088.3
10	intangible asset	9	3339	1127	0.1	998.8
11	value indication	6	79	751.3	0	750
12	review engagement	5	213	626.1	0	622
13	liquidation value	5	1272	626.1	0	597.7
14	valuation date	5	1842	626.1	0.1	585.3
15	systematic risk	5	2433	626.1	0.1	573.1
16	financial interest	6	11253	751.3	0.4	524
17	discount rate	8	25227	1001.8	1	507.3
18	calculation engagement	4	15	500.9	0	501.6
19	valuation engagement	4	50	500.9	0	500.9
20	valuation analyst	4	313	500.9	0	495.9

Fig. 1. The data obtained by extracting multi-word term candidates from IBA 2015 standard with Sketch Engine

The application of queries created a common pool of term candidates in a single excel file, with their instances marked for each of the standards separately. As a result, it was possible not only to calculate the total number of term candidates in all three documents (with those appearing in more than one standard counted only once), but also to identify exactly which terms occurred in which standards.

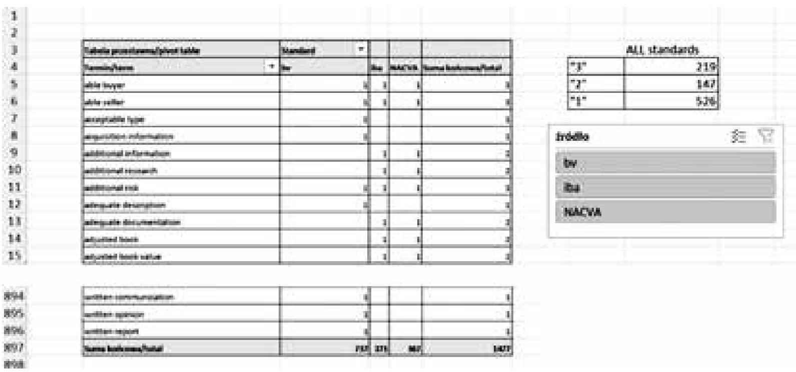


Fig. 2. The calculation of term candidates in ASA, IBA and NACVA business valuation standards after the application of Power Query

As shown in Fig. 2, the total number of term candidates was 1477, out of which 737 were present in ASA Business Valuation Standards 2009, 373 in IBA Professional Standards 2015 and 367 in NACVA Professional Standards 2017.

[12] A reply provided on the Boot Camp forum on December 21, 2021.

The number of term candidates which occurred in all three standards was 219, there were 147 term candidates present in two out of three standards and 526 of them appeared in only one standard.

Also, as can be seen in Fig. 3, the results showed a considerable terminological overlap between the NACVA and the IBA standard – 366 terms occurred in both standards, with 373 term candidates in IBA and 367 term candidates in NACVA in total.

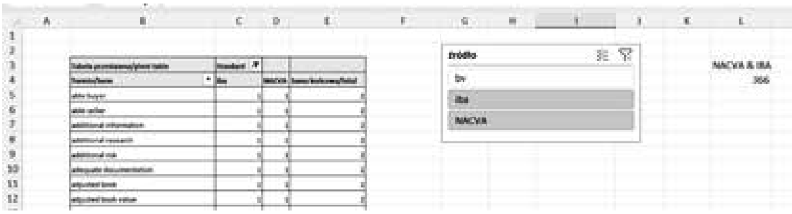


Fig. 3. The number term candidates present in IBA and NACVA business valuation standards

The pool of term candidates shared by ASA and IBA was significantly smaller – it amounted to 219 instances (Fig. 4), that it is 30 % of all term candidates present in ASA (737). The number of term candidates which appeared both in the ASA and NACVA standard was equally 219 (Fig. 5).

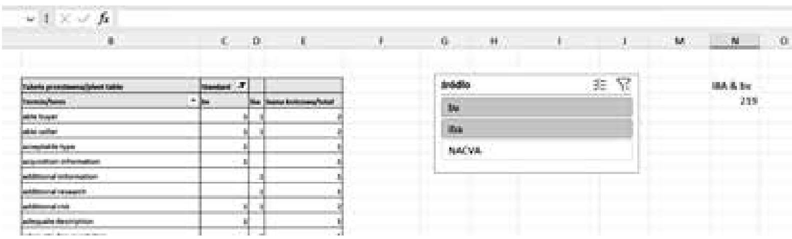


Fig. 4. The number of term candidates present in ASA and IBA business valuation standards

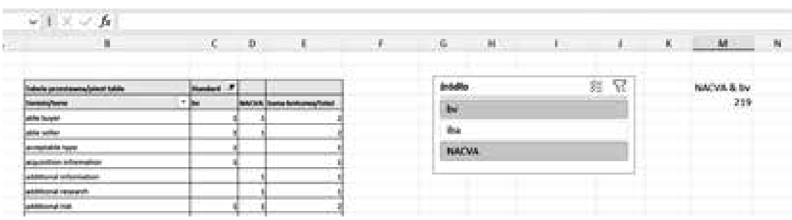


Fig. 5. The number of term candidates present in ASA and NACVA business valuation standards

5. Conclusions

The results showed a much greater disparity between the pool of term candidates extracted from the ASA standard and the that of the remaining two standards. Terminological consistency between the IBA and NACVA standards was almost at the level of 100 %, whereas less than one third of ASA pre-selected terms were shared by

IBA and NACVA respectively. Also, 219 term candidates that occurred in all three standards were exactly the same number as the pool shared by ASA and either the IBA or the NACVA standard, which means that it was in the ASA standard where a set of different term candidates was to be found.

These results led to a tentative conclusion that much greater similarity was to be expected between the IBA and NACVA standards also on other levels of description. This conclusion was later confirmed, since a closer examination of the content of each of the documents revealed that the two standards were practically the same on the level of their superstructures and even included exact wording of the corresponding text sections. According to information published by NACVA in 2022, it was because since 2011 both standards had been developed by co-operating teams of specialists, since NACVA was eventually to take over all IBA operations.[13]

The study also showed how the joint application of Sketch Engine and Power Query functionalities helped establish a common pool of pre-selected terms which served for the analysis as well as to calculate the subsets of extracted term candidates across different domain-specific documents. The method itself can be further applied to compare terminology in a number of documents or (sub)sets of corpora created via corpus linguistics tools.

References

American Society Of Appraisers. *ASA Business Valuation Standards*. Herndon: American Society of Appraisers (2009). https://www.appraisers.org/about/standards-ethics-and-policies. 7.6.2020.

Biber, Douglas. "A Typology of English Texts." *Linguistics*, 27 (1989): 3–43. Print.

Cabré Castellví, Maria Teresa. "Theories of Terminology: Their Description, Prescription and Explanation." *Terminology*, vol. 9, no. 2 (2003): 163–99. Print.

Culpeper, Jonathan and Jane Demmen. "Keywords." *The Cambridge Handbook of English Corpus Linguistics*. Ed. Douglas Biber and Randi Reppen. Cambridge: Cambridge University Press, 2015, 90–105. Print.

Davies, Mark. "Corpora: An Introduction." *The Cambridge Handbook of English Corpus Linguistics*. Ed. Douglas Biber and Randi Reppen. Cambridge: Cambridge University Press, 2015, 11–31. Print.

Dubuc, Robert and Andy Lauriston. "Concept Description: Terms and Contexts." *Handbook of Terminology Management. Volume 1: Basic Aspects of Terminology Management*. Ed. Sue Ellen Wright and Gerhard Budin. Amsterdam, Philadelphia: John Benjamins Publishing Company, 1997, 80–88. Print.

Heylen, Kris and Dirk De Hertog. "Automatic Term Extraction." *Handbook of Terminology. Volume 1*. Ed. Hendrik Kockaert and Frieda Steurs. Amsterdam, Philadelphia: John Benjamins Publishing Company, 2015, 203–221. Print.

Institute Of Business Appraisers. *Professional Standards*. Salt Lake City: Institute of Business Appraisers (2015). http://web.nacva.com/TL-Website/Files/IBA/IBA_Professional_Standards_Incl_Review_Stnds_Effective_8-1-15_Final.pdf. 9.6.2020.

[13] See https://www.nacva.com/go-iba, date of access: 25.8.2022.

Kageura, Kyo. "Terminology and Lexicography." *Handbook of Terminology. Volume 1*. Ed. Hendrik Kockaert and Frieda Steuers. Amsterdam, Philadelphia: John Benjamins Publishing Company, 2015, 45–59. Print.

Kageura, Kyo, and Elisabeth Marshman. "Terminology Extraction and Management. "*The Routledge Handbook of Translation and Terminology*. Ed. Minako O'Hagan. London, New York: Routledge, Taylor & Francis Group, 2020, 61–77. Print.

Kilgarriff, Adam. "Simple Maths for Keywords." *Proceedings of the Corpus Linguistics Conference.* Liverpool: Univerity of Liverpool (2009): 1–6. https://www.sketchengine.eu/wp-content/uploads/2015/04/2009-Simple-maths-for-keywords.pdf. 12.1.2023.

Lee, David Y. "Genres, Registers, Text Types, Domains, and Styles: Clarifying the Conceptsand Navigating a Path through the BNC Jungle." *Language Learning & Technology*, vol. 5, Sept. (2001): 37–72. https://www.lltjournal.org/item/10125-44565. 19.10.2022.

Mcenery, Tony, Richard Xiao and Yukio Tono. *Corpus-Based Language Studies: An Advanced Resource Book*. London, New York: Routledge, 2006. Print.

National association of certified valuators and analysts. *Professional Standards*. Salt Lake City: NACVA (2017). http://web.nacva.com/TL-Website/PDF/NACVA_Professional_Standards_Incl_Review_Stnds_Effective_6-1-17_Final.pdf. 9.6.2020.

Oakey, David. "Formulaic Language in English Academic Writing: A Corpus-Based Study of the Formal and Functional Variation of a Lexical Phrase in Different Academic Disciplines." *Using Corpora to Explore Linguistic Variation*. Ed. Randi Reppen, Susan M. Fitzmaurice and Douglas Biber. Amsterdam, Philadelphia: John Benjamins Publishing Company, 2002, 115–132. Print.

Sager, Juan. "Concept Representation: Term Formation." *Handbook of Terminology Management. Volume 1: Basic Aspects of Terminology Management*. Ed. Sue Ellen Wright and Gerhard Budin. Amsterdam, Philadelphia: John Benjamins Publishing Company, 1997, 25–41. Print.

Santos, Claudia and Rute Costa. "Domain Specificity: Semasiological and Onomasiological Knowledge Representation." *Handbook of Terminology. Volume 1. Ed.* Hendrik Kockaert and Frieda Steurs, Amsterdam, Philadelphia: John Benjamins Publishing Company, 2015, 153–179. Print.

Sinclair, John et al. EAGLES Preliminary Recommendations on Text Typology, 1996. http://www.ilc.cnr.it/EAGLES96/spokentx/node1.html#SECTION00010000000000000000. 9.10.2022.

Steen, Gerard. "Genres of Discourse and the Definition of Literature." *Discourse Processes*, vol. 28, no. 2, 1999: 9–20, doi:10.1080/01638539909545075, https://www.researchgate.net/publication/254302202_Genres_of_discourse_and_the_definition_of_literature. 19.10.2022.

Wright, Sue Ellen. "Term Selection: The Initial Phase of Terminology Management." *Handbook of Terminology Management. Volume 1: Basic Aspects of Terminology Management*. Ed. Sue Ellen Wright and Gerhard Budin. Amsterdam, Philadelphia: John Benjamins Publishing Company, 1997, 13–23. Print.

ZITIERNACHWEIS:

Judycka, Paulina. „Comparing Multi-Word Term Candidates in Business Valuation Standards with the Use of Sketch Engine and Power Query Functionalities", *Linguistische Treffen in Wrocław* 24, 2023 (II): 291–301. DOI: 10.23817/lingtreff.24-20.