Journal of Mechatronics, Automation and Identification Technology
Vol. 11, No. 1, pp. 12–17, 2026.

Use of Sketch Engine in Linguistic and Literary Analysis

Darko KOVAČEVIĆ

Faculty of Electrical Engineering, University of East Sarajevo, East Sarajevo, Bosnia and Herzegovina
darko.kovacevic@etf.ues.rs.ba

***Abstract*—This paper examines the applicability of Sketch Engine, a web-based corpus analysis platform, across two distinct but complementary domains: linguistic analysis and corpus stylistic analysis of literary texts. The study begins with an overview of corpus linguistics, stylistics, and corpus stylistics as theoretical frameworks, followed by a description of Sketch Engine's principal analytical tools and its position relative to other available corpus analysis software. A structured literature review is provided for both domains of application. The central part of the paper presents a comprehensive account of the ways in which Sketch Engine's tools—including Wordlist, Concordance, Word Sketch, Word Sketch Difference, Thesaurus, N-grams, and Keywords—can be employed to uncover lexical, grammatical, phraseological, and stylistic patterns in language corpora. Examples of their application are discussed, with attention to methodological implications and analytical insights. The paper concludes with reflections on the scope and versatility of Sketch Engine as a resource for corpus-driven linguistic and literary research, and identifies directions for future investigation.**

***Keywords*—Sketch Engine, corpus linguistics, corpus stylistics, linguistic analysis, literature, stylistic analysis.**

I. Introduction

The study of language through large, digitized collections of authentic texts—commonly known as corpus linguistics—has profoundly transformed the methods available for linguistic research and language description. Over the past three decades, the development of increasingly powerful corpus analysis tools has made it possible to investigate lexical, grammatical, and stylistic phenomena with a degree of empirical precision that was not previously achievable. Among the platforms that have emerged in this field, Sketch Engine has established itself as one of the most versatile and widely used web-based environments for corpus analysis, offering a comprehensive suite of tools that support both the management and the systematic investigation of language corpora.

The applicability of Sketch Engine extends across different areas of language study. On the one hand, it has proven useful in the analysis of specialized registers, where the identification of domain-specific vocabulary, grammatical patterns, and multi-word expressions is of practical relevance for language teaching, lexicography, and terminology work. On the other hand, it has demonstrated considerable potential in corpus stylistics—a discipline that brings together literary studies and corpus linguistics in order to examine the linguistic features of literary texts in a systematic and data-driven manner.

This paper aims to provide a unified account of how Sketch Engine can be deployed across these two areas of application. It begins by establishing the relevant theoretical frameworks, covering corpus linguistics, stylistics, and corpus stylistics, before presenting an overview of available corpus analysis software and a detailed description of Sketch Engine's principal tools and functionalities. A literature review surveys existing research in both domains. The central sections discuss, in a generalized way, how Sketch Engine's analytical tools can be applied to investigate key features of language and style in different types of corpora. The paper concludes with a synthesis of the main findings and a discussion of future research directions.

II. Theoretical Background

A. Corpus Linguistics

One of the defining components of the scientific study of language includes an account of how language works [1]. In this context, corpus linguistics has emerged as a research method that draws on authentic language examples, collected and organized into corpora—that is, searchable collections of data [2]. More formally, a corpus has been defined as a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research [3], [4], [5]. Corpus-based analytical methodologies involve the systematic exploration of principled, computerized collections of real-world texts in order to examine patterns of language use. Corpus linguistics is distinguished by its empirical orientation, analyzing typically large text collections through automatic and interactive data retrieval techniques, and integrating quantitative and qualitative approaches [6]. By relying on corpus evidence with minimal prior assumptions, researchers are able to observe lexico-grammatical patterns and other linguistic phenomena that may not have been previously identified or adequately described [7]. With the development of information technologies, various forms of corpus analysis tools have emerged alongside increasingly easy access to large numbers of electronic texts, making corpus knowledge and skills—including familiarity with available query tools and corpora, their capabilities, and

their methods of use—progressively more important for researchers, lexicographers, terminologists, translators, and language instructors [8].

B. Stylistics and Corpus Stylistics

Stylistics has been variously defined in the literature. At its core, it concerns the linguistic study of style in language and the manner in which style is influenced by particular non-linguistic variables [9]. It has also been described as the study of style in text, which considers the language choices that text producers make over other possible choices and how these relate to the artistic functions of the texts [10], as well as the study of language in literature, which offers insight into the encoding of literary effects in language [11]. For the purposes of this paper, stylistics may be understood as the linguistic analysis of literary texts whose goal is to decode literary meanings and structural features by identifying linguistic patterns and their functions, where style implies the lexical and grammatical patterns in a text that contribute to its meaning [12]. Stylistics is grounded in the assumption that meaning in language is a linguistic phenomenon and can be accessed through systematic analysis. Corpus stylistics emerges at the intersection of corpus linguistics and literary studies, merging a range of methods from the plain use of digitized literary texts to the deployment of statistical analyses and contrastive or comparative corpora [13]. It has been understood as a way of bringing the study of language and literature closer together by using innovative descriptive tools that fit into linguistic frameworks while also leaving room to account for individual qualities of texts and thereby link up with literary interpretation [14]. Corpus stylistics enables the analysis of extended texts and, through the use of software tools to identify recurring linguistic patterns in the analyzed data, offers detailed and methodologically neutral insights that do not depend on prior interpretations, genre expectations, or critical reception, thereby making it possible to uncover meanings that have previously remained inaccessible to approaches based on intuition alone [12].

III. Overview of Corpus Analysis Software

The key characteristic shared by corpus analysis software is the capability to store, annotate, and analyze large collections of electronic texts, enabling data-driven descriptions of language use. Available tools differ in scope, functionality, and analytical orientation. WordSmith Tools (https://www.lexically.net/wordsmith) is one of the earliest and most influential corpus analysis platforms. Designed as a desktop-based suite, it popularized core corpus techniques such as wordlists, concordances, and keyword analysis, while its intuitive interface made corpus methods accessible to users with limited technical background. AntConc (https://www.laurenceanthony.net/software/antconc) is a free, platform-independent concordance tool offering functionalities similar to WordSmith, including concordance generation, extraction of multi-word expressions (n-grams), collocation analysis, and keyword lists. LancsBox (https://lancsbox.lancs.ac.uk) offers advanced visualization features, including dispersion plots and collocation networks, which facilitate the interpretation of frequency and co-occurrence data by enabling the identification of distributional patterns that might otherwise be obscured in numerical output. Wmatrix (https://ucrel.lancs.ac.uk/wmatrix) extends the scope of corpus analysis through semantic tagging, allowing users to examine semantic domains and conceptual patterning alongside lexical and grammatical features. As corpus linguistics expanded toward larger datasets and more complex analyses, Sketch Engine emerged as a comprehensive, web-based corpus management and analysis platform.

IV. Sketch Engine

Sketch Engine (https://www.sketchengine.eu/) is a web-based corpus analysis platform that brings together a suite of online tools supporting corpus creation, management, annotation, and analysis within a single interface [15]. It also maintains an extensive collection of pre-built corpora immediately available for analysis—currently hosting 846 pre-built corpora covering 109 languages, of which 103 are in English—along with options for users to upload, compile, and manage their own corpora. During corpus compilation, Sketch Engine performs automated part-of-speech (POS) tagging, annotating each token with a tag carrying information about the part of speech and often morphological and grammatical information such as number, gender, case, and tense. Sketch Engine provides a wide range of analytical tools for generating frequency statistics, identifying co-occurrence and collocational patterns, visualizing contrasts between corpora, and conducting comparative, contrastive, and multilingual analyses [15]. Its interface is organized around a central Dashboard, which serves as the main access point for analytical tools and corpus management functions. The principal tools include the following: Word Sketch generates grammaticalized collocation profiles showing typical syntactic relations of a lexical item. Word Sketch Difference enables direct comparison of the collocational behavior of two lexical items. The Concordance tool produces keyword-in-context (KWIC) lines, facilitating detailed qualitative analysis of word usage in authentic textual environments. Wordlist generates frequency-based inventories of words or lemmas, providing an overview of lexical distribution within a corpus. The N-grams tool identifies recurrent multi-word sequences and formulaic expressions. Keywords highlights statistically significant lexical items by contrasting a focus corpus with a reference corpus. The Thesaurus tool identifies semantically related words based on distributional similarity, and Parallel Concordance enables translation-oriented analysis through aligned bilingual corpora. Additional tools include Trends for diachronic analysis, Text Type Analysis for statistical information about corpus composition, and OneClick Dictionary for semi-automatic creation of dictionary entries.

Access to Sketch Engine operates through a subscription-based model. New users are provided with a 30-day free trial granting full access to core analytical tools and to both pre-built and user-created corpora, with no

payment details required. Continued use after the trial requires a paid subscription, available on monthly, quarterly, or annual terms, with pricing options tailored to individuals, students, educators, academic researchers, and institutions. Sketch Engine also provides extensive user support through detailed video tutorials and comprehensive written documentation.

V. Literature Review

A. Sketch Engine in Linguistic Analysis

Sketch Engine has been employed in a wide range of linguistic research studies. In [16], the authors use the British National Corpus and Sketch Engine to examine usage differences between two synonymous verbs, investigating their collocates as well as the syntactic patterns they typically occupy and their functional implications. In [17], the representation of men and women in the British National Corpus is discussed through focusing on the collocational and grammatical behavior of the corresponding nouns, revealing systematic differences in the ways these terms pattern in different grammatical relations. The use of Sketch Engine in constructing a computational lexicon for the English language is discussed in [18], evaluating its effectiveness by analyzing large-scale general and domain-specific corpora to extract lexical information, including word frequency, grammatical structures, and semantic relationships. The features of Sketch Engine supporting bilingual lexicography and the creation of bilingual learner's dictionaries are summarized in [19]. The Sketch Engine Thesaurus tool has been evaluated on a subset of analogy queries [20], and the automatic definition-finding mechanism implemented within Sketch Engine has been described in [21], exploiting complex pattern-matching queries in CQL and the indexing mechanism of word sketches. Reference [22] presents the role of the Sketch Engine Concordance tool in the development of a trilingual lexicographic resource based on linguistic corpora. References [23] and [24] use Sketch Engine to investigate near-synonymous adverbs and the countable usages of nouns normally considered uncountable. In [25] Sketch Engine is used to examine lexical bundles in corpora of English and Indonesian research articles on literature and linguistics. Finally, [26] discusses how Sketch Engine can be used for the analysis of a Technical English corpus based on IT conference papers.

B. Sketch Engine in Literary Analysis

The application of Sketch Engine to corpus stylistic analysis of literary texts is a more specialized area, with a smaller but growing body of research. In [27], the author investigates the possibility of obtaining data useful for the interpretation of literary characters through computer-assisted analysis grounded in narratology, using Sketch Engine as the analytical platform. Reference [28] presents a stylometric study of fiction through a corpus-based approach using Sketch Engine, applying quantitative and statistical processing to a text corpus drawn from a series of science fiction novels. In [29], lexical patterning in selected Shakespearean plays is investigated through a customized Sketch Engine corpus, focusing on key lexical items to analyze frequency, collocational patterns, parts of speech, and n-grams, with the findings revealing how repeated exposure to specific lexical items contributes to meaning-making and thematic development. Reference [30] demonstrates the use of the Russian National Corpus and Sketch Engine for extracting and semantically processing proper names in connection with a lexicographic project on military vocabulary in a major Russian novel. In [31], it is demonstrated how a pedagogically oriented literary corpus can be constructed and analyzed using Sketch Engine tools for the purpose of Arabic language instruction. Finally, [32] gives a comprehensive overview of the aspects in which different Sketch Engine tools can be used for corpus stylistic analysis, on the example of a corpus made of seven Cormac McCarthy's novels.

VI. Applying Sketch Engine Tools in Linguistic and Literary Analysis

A. Vocabulary Profiling and Style: The Keywords Tool

The Keywords tool performs a comparison between a focus corpus and a reference corpus, identifying what is lexically unique or overrepresented in the former relative to the latter. Under its Advanced tab, parameters such as the reference corpus, minimum and maximum frequency, the maximum number of items, and the Focus on setting can be adjusted. The Focus on parameter is particularly significant: when set to rare, the tool prioritizes words that are infrequent in the reference corpus, highlighting those which are characteristic or specialized in the focus corpus; when set to common, it emphasizes words that are more broadly shared. The output is divided into keywords (single-word items) and terms (multi-word items), both linked directly to concordance lists in the respective corpora.

When applied to specific corpora, the Keywords tool can be expected to surface a vocabulary profile reflecting the characteristic features of specialized discourse. High-frequency keywords will typically include nouns central to the field in question and verbs associated with the dominant modes of activity within that register, while the terminology present is likely to reflect the broader communicative purposes of the genre. The multi-word term list will generally consist of established collocations and noun compounds, illustrating how lexical precision and high information density are achieved in specialized registers.

In the case of literary corpora, the Keywords tool is likely to reveal a vocabulary profile that differs markedly from that of domain-specific or general-language reference corpora. Depending on the author, period, or genre under analysis, the results may foreground colloquial or non-standard language, lexis associated with particular thematic or semantic domains, or concentrations of vocabulary linked to specific narrative settings and strategies. Whatever the specific content, keyword clusters of this kind provide a data-grounded entry point into the stylistic signature of the work or author under examination, and can direct the analyst toward more focused investigations using other tools..

B. *Frequency and Lexical Distribution: The Wordlist Tool*

The Wordlist tool generates frequency-ranked inventories of words or lemmas within a corpus, organized by part of speech category, and provides statistical information on the presence of each word category within the corpus. The resulting lists offer a first-level quantitative overview of lexical composition and frequency distribution, making it possible to distinguish general from specialized lexis, identify dominant word categories, and reveal patterns of lexical preference and repetition.

When applied to specific language corpora, the Wordlist tool wih assist in revealing the patterns of lexical distribution characteristic of specialized registers. High-frequency nouns will typically be abstract and domain-specific, while general, non-specific vocabulary is likely to be underrepresented. Frequent verbs will generally reflect the dominant modes of activity and reasoning within the register, and the overall distribution of word categories can provide an initial indication of the communicative purposes and stylistic conventions of the corpus under analysis.

Regarding literature corpora, Wordlist analysis will emphasise vocabulary categories that reflect the narrative and stylistic orientation of the texts. The relative prominence of concrete versus abstract nouns, or of dynamic versus stative verbs, may offer initial indications of an author's characteristic perspective and narrative strategies. The distribution of nouns across semantic categories—whether they cluster around human figures, spatial settings, temporal experience, or other domains—can further reveal dimensions of an author's literary world when examined across a substantial corpus, and can direct subsequent analysis toward more targeted investigations..

C. *Contextual Analysis: The Concordance Tool*

The Concordance tool produces aligned keyword-in-context (KWIC) lines, presenting instances of a search term together with the surrounding textual context. Searches can be performed at the level of individual word forms, lemmas, phrases, or character strings, using either the Basic or Advanced interface. Crucially, the tool also supports Corpus Query Language (CQL) queries, which allow users to specify highly complex search criteria—for example, targeting specific morphological or syntactic configurations—that cannot be handled through the standard interface. CQL queries can be constructed manually or with the assistance of an integrated AI Search function.

In language analysis, the Concordance tool serves as an effective means of retrieving and examining the frequency and contextual distribution of specific grammatical structures. Patterns can be targeted through appropriately formulated CQL queries, yielding both quantitative counts and qualitative contextual data that contribute to a systematic characterization of the grammatical conventions governing the writing.

In literary stylistic analysis, the Concordance tool enables close, corpus-wide examination of how individual lexical items operate across the full extent of a text or collection of texts. By surveying all occurrences of a given word in their immediate contexts, the analyst can identify recurring collocational patterns, track figurative uses and semantic shifts, and map the distribution of a term across different narrative situations. Examining all instances of a word through the concordance allows these distinct functions to be distinguished and compared in a principled way, grounding interpretive conclusions in the full range of textual evidence rather than in a selection shaped by prior expectations.

D. *Collocational Behavior: Word Sketch and Word Sketch Difference*

The Word Sketch tool generates one-page summaries of the typical grammatical and collocational behavior of a given word, organizing results into grammatical relation categories such as subject of, object of, modified by, and others relevant to the specific search item. For most corpora, results are available for nouns, adjectives, verbs, and adverbs. The tool allows the analyst to examine, at a glance, the principal collocates of a lexical item organized by grammatical function, facilitating the identification of conventional usage patterns, terminological structures, and register-specific meanings. Graphical visualization of the results is also available.

Word Sketch Difference extends this functionality by enabling direct comparisons between two lemmas, two-word forms of the same lemma, or the occurrences of the same lemma in two different subcorpora. This comparative functionality is particularly valuable for distinguishing near-synonymous items, investigating register variation, or comparing authorial usage across different periods or genres. In language analysis, it can reveal subtle differences in the collocational behavior of terms that appear functionally equivalent, while in literary analysis it can support the investigation of how a writer's use of a given lexical item varies across different works, narrative contexts, or textual segments.

E. *Multi-Word Expressions: The N-grams Tool*

The N-grams tool identifies recurrent multi-word sequences in a corpus, with configurable parameters for n-gram length (from two to six tokens), minimum and maximum frequency, and additional criteria including regular expressions. N-grams essentially present multi-word expressions and formulaic sequences, and their analysis is a bridge between quantitative frequency analysis and qualitative interpretation: by identifying recurring phraseological units, the tool reveals grammatical, discourse, and stylistic patterns that would not be apparent from single-word frequency data alone.

In the analysis of language corpora, n-gram analysis typically surfaces formulaic expressions used to organize discourse, signal evaluation, or structure argumentation. The distribution of such metadiscursive and logical connective expressions reflects adherence to specific and characteristic writing conventions.

In literary stylistic analysis, n-gram results reveal recurring syntactic and narrative patterns characteristic of a given author's prose. Phrasal units related to observation, perception, and physical action—verb-object sequences denoting typical narrative events, locative expressions

indicating spatial positioning, temporal markers regulating narrative pacing—may all emerge as recurrent n-grams, providing quantitative evidence for various stylistic feature. Dialogue-related sequences and patterns of interpersonal interaction can similarly be identified and analyzed in terms of their frequency and distribution across a corpus.

F. Semantic Similarity: The Thesaurus Tool

The Thesaurus tool in Sketch Engine generates lists of synonyms and semantically related words on the basis of their distributional similarity within the selected corpus. The underlying assumption is that words that appear in similar contexts are likely to be semantically related—a principle derived from distributional semantics. Results are available for nouns, adjectives, verbs, and adverbs in most corpora, and visualization options including bubble charts and word clouds are provided.

In linguistic analysis, the Thesaurus is particularly useful for investigating the semantic field of key terms in a corpus, identifying synonyms and near-synonyms, and examining whether distributional evidence in a specialized corpus corresponds to the semantic relations posited in general-language dictionaries and thesauri. In other words, the terms whose general-language synonymy relationships are well established may cluster differently when analyzed on the basis of authentic usage in domain-specific contexts, providing evidence for terminological precision or semantic shift. In literary corpora, the Thesaurus can reveal unexpected associative networks around key lexical items, supporting the analysis of thematic and symbolic patterning.

VII. Discussion

Sketch Engine's analytical toolkit provides robust and versatile support for corpus-driven inquiry across for both linguistic and literary analysis. The platform enables a movement away from intuition-based description toward insights grounded in authentic, large-scale textual data. The tools are mutually reinforcing: Keywords and Wordlist provide macro-level overviews of vocabulary and frequency; Concordance enables micro-level contextual examination; Word Sketch and Word Sketch Difference reveal collocational and grammatical behavior; N-grams uncover phraseological and discourse conventions; and the Thesaurus supports semantic investigation. Moreover, all tools are interconnected within the Sketch Engine interface, so that any item appearing in one set of results can be immediately examined through the lens of another tool.

In language analysis, the platform's analytical capabilities support the systematic characterization of register-specific features. These findings have both theoretical relevance and applied value for language teaching, curriculum development, lexicography, and terminology work. The GDEX functionality further enhances the pedagogical utility of the platform by enabling the automatic selection of illustrative examples suitable for instructional use.

In corpus stylistic analysis, Sketch Engine offers a means of investigating the linguistic features of literary texts with the empirical precision that large corpora make possible, without displacing the interpretive dimension of literary analysis. By identifying characteristic lexical clusters, collocational patterns, grammatical tendencies, and formulaic expressions across a corpus of literary works, the analyst can obtain data-grounded evidence for stylistic features that can then be integrated into broader interpretive accounts. The platform is equally suited to the comparative analysis of multiple authors or works, the diachronic investigation of a single author's stylistic development, and genre- or register-sensitive analyses within a single authorial corpus.

The integration of CQL queries, AI-assisted search, and the interconnected nature of the tools significantly extends the scope and flexibility of possible analyses beyond what is achievable through standard interface options alone. At the same time, the availability of extensive pre-built corpora facilitates comparative and reference-based analyses without requiring users to compile their own data from scratch.

VIII. Conclusions

This paper has presented a comprehensive overview of how Sketch Engine can be applied across linguistic and literary analysis, covering both the theoretical foundations of corpus linguistics and corpus stylistics and the practical capabilities of the platform's principal analytical tools. The discussion has shown that Sketch Engine's tools—Wordlist, Concordance, Word Sketch, Word Sketch Difference, Thesaurus, N-grams, and Keywords—can be deployed in a coordinated and complementary manner to investigate a wide range of lexical, grammatical, phraseological, and stylistic phenomena in language corpora of different types.

The platform proves particularly effective when the goal is to move beyond impression-based characterizations of language use toward systematic, evidence-driven descriptions grounded in authentic textual data. In language analysis, it supports the identification of the features that define specialized registers; in corpus stylistics, it provides the quantitative and qualitative data needed to characterize authorial style and to uncover patterns that would remain inaccessible to close reading alone.

Several directions for future research can be identified on the basis of the present study. Sketch Engine's comparative tools—Keywords, Word Sketch Difference, and the subcorpus functionality—offer significant potential for contrastive analyses of multiple authors, genres, or language varieties, allowing stylistic and linguistic differences to be examined systematically. Diachronic analyses investigating stylistic development across an author's career, or language change across different periods, represent another promising area. The platform also offers a basis for corpus-based discourse analysis, contributing to the investigation of how discourse shapes and reflects social relations, ideology, and power. Finally, the pedagogical applications of Sketch Engine, particularly the GDEX feature and the ability to extract representative examples of language use, merit further attention in the context of language teaching and materials development.

REFERENCES

[1] E. Csomay and W. J. Crawford, Doing Corpus Linguistics. New York: Routledge, 2024.

[2] C. F. Meyer, English Corpus Linguistics. Cambridge: Cambridge University Press, 2023.

[3] D. McIntyre and B. Walker, Corpus Stylistics: Theory and Practice. Edinburgh: Edinburgh University Press, 2019.

[4] T. McEnery and A. Wilson, Corpus Linguistics: An Introduction, 2nd ed. Edinburgh: Edinburgh University Press, 2001.

[5] J. Sinclair, "Corpus and text – basic principles," in Developing Linguistic Corpora: A Guide to Good Practice, M. Wynne, Ed. Oxford: Oxbow Books, 2005, pp. 1–16.

[6] E. Le Foll, Textbook English: A Multi-Dimensional Approach, Studies in Corpus Linguistics. Amsterdam: John Benjamins, 2024.

[7] S. Hunston, Corpora in Applied Linguistics, The Cambridge Applied Linguistics Series. Cambridge: Cambridge University Press, 2002.

[8] M. Kunilovskaya and M. Koviazina, "Sketch Engine: A toolbox for linguistic discovery," Journal of Linguistics / Jazykovedný časopis, vol. 68, 2017, pp. 503–507.

[9] D. McIntyre and B. Walker, Corpus Stylistics: Theory and Practice. Edinburgh: Edinburgh University Press, 2019.

[10] J. Lugea and B. Walker, Stylistics: Text, Cognition and Corpora, 1st ed. Cham, Switzerland: Springer International Publishing, 2023.

[11] M. Toolan, Language in Literature: An Introduction to Stylistics. London: Arnold, 1998.

[12] [B. Fischer-Starcke, Corpus Linguistics in Literary Analysis: Jane Austen and Her Contemporaries, 1st ed. London: Continuum, 2010.

[13] [M. Wynne, "Stylistics: corpus approaches," in The Encyclopedia of Language and Linguistics, K. Brown, Ed. Oxford: Elsevier, 2006, pp. 223–226.

[14] M. Mahlberg, "Corpus stylistics: bridging the gap between linguistic and literary studies," in Text, Discourse, and Corpora, M. Hoey et al., Eds. London: Continuum, 2007, pp. 219–246.

[15] A. Kilgarriff, V. Baisa, J. Bušta, M. Jakubíček, V. Kovář, J. Michelfeit, P. Rychlý, and V. Suchomel, "The Sketch Engine: Ten years on," Lexicography, vol. 1, 2014, pp. 7–36.

[16] C. Hu and Y. Bei, "Using Sketch Engine to investigate synonymous verbs," International Journal of English Linguistics, vol. 5, 2015, pp. 29–41.

[17] M. Pearce, "Investigating the collocational behaviour of MAN and WOMAN in the BNC using Sketch Engine," Corpora, vol. 3, 2008.

[18] [M. Mobark, "Building a computational lexicon for English using Sketch Engine: A pilot study," 2025. [Online]. Available: https://www.academia.edu/128813152/, Accessed: 23 April 2026.

[19] V. Kovář, V. Baisa, and M. Jakubíček, "Sketch Engine for bilingual lexicography," International Journal of Lexicography, vol. 29(3), 2016, pp. 339–352.

[20] P. Rychlý, "Evaluation of the Sketch Engine thesaurus on analogy queries," RASLAN, vol. 1, 2016, pp. 147–151.

[21] [V. Kovář, M. Močiariková, and P. Rychlý, "Finding definitions in large corpora with Sketch Engine," in Proc. 10th Int. Conf. Language Resources and Evaluation (LREC 2016), Portorož, Slovenia, 2016, pp. 391–394.

[22] R. Chishman, A. Nardes dos Santos, D. Spader de Souza, and J. Padilha, "The relevance of the Sketch Engine software to build field–football expressions dictionary," Revista de Estudos da Linguagem, vol. 23, 2015, pp. 769–796.

[23] P. Biasi, "Using Sketch Engine: An analysis of five adverbs," 2018. [Online]. Available: https://www.researchgate.net/publication/350213126, Accessed: 23 April 2026.

[24] P. Biasi, "Using Sketch Engine: An analysis of nouns normally considered uncountable," 2021. [Online]. Available: https://www.researchgate.net/publication/350441983, Accessed: 23 April 2026.

[25] S. Yuliawati, D. Ekawati, and R. Mawarrani, "Investigating lexical bundles in the corpora of English and Indonesian research articles with the Sketch Engine," Jurnal Sosioteknologi, vol. 20, 2021, pp. 188–200.

[26] D. Kovačević, "Analysis of a Technical English Corpus with Sketch Engine," 2026 25th International Symposium INFOTEH-JAHORINA (INFOTEH), East Sarajevo, Bosnia and Herzegovina, 2026, pp. 1-6.

[27] L. Ninčević, "Analiza likova hrvatske književnosti pomoću računalnih alata," [sic] – A Journal of Literature, Culture and Literary Translation, 2023. Available: https://www.sic-journal.org/Article/Index/726. Accessed: 23 April 2026.

[28] O. S. Taran, O. S. Palchevska, A. A. Luchyk, V. V. Shabunina, and O. V. Labenko, "Stylometric study of fiction using Sketch Engine," in Proc. Digital Humanities Workshop (DHW 2021), Kyiv, Ukraine, Dec. 23, 2021, pp. 1–7.

[29] U.-e. Naqvi and Prof. Bhatti, "Lexical patterning in Shakespearean tragedies: A corpus-based analysis using Sketch Engine through the lens of lexical priming theory," Contemporary Journal of Social Science Review, vol. 3, 2025, pp. 1360–1375.

[30] Н. В. Козловская, Е. В. Сергеева, Е. Е. Силантьев и О. С. Бахтина, "Имена собственные в учебном проекте словаря военной лексики к роману Л. Н. Толстого 'Война и мир'," Вестник Череповецкого государственного университета, № 6 (129), 2025, pp. 92–106.

[31] H. Maachi, H. Khamar, and L. Abubkr, "Using a pedagogic corpus and Sketch Engine tools to analyze a literary text in an Arabic language classroom," International Journal of Arabic-English Studies, vol. 25, 2025, pp. 273–294.

[32] D. Kovačević, "Corpus Stylistic Analysis with Sketch Engine," 2026 25th International Symposium INFOTEH-JAHORINA (INFOTEH), East Sarajevo, Bosnia and Herzegovina, 2026, pp. 1-6.