



Article

The Role of Corpus Linguistics in The Terminological Analysis of Botanical Terms

Murodova Shaxnoza Choriyeвна^{*1}

1. Independent Researcher, Kokand University

* Correspondence: Shakhnozamurodova0804@gmail.com

Abstract: It studies the application of corpus linguistics to make a glimpse of botanical items particularly in English – Uzbek translation. The conventional methods of automatic terminological analysis usually do not yield insights into contextualized and functionalized term usage, providing incomplete information for practical translation and understanding of knowledge, especially in scientifically rich domains such as botany. By focusing on large, annotated corpora of real text, corpus linguistics offers a powerful tool for exploring terms in their natural linguistic context. Based on a broad range of corpus-based analysis methods, such as frequency, collocational and morpho-semantic analyses, this study analyses similarities and differences in the usage, particulars synonyms and translation equivalents for botanical terms in both languages.

We conclude that the results underscore the role of corpus linguistics in identifying semantic change, polysemy, and translation problems (such as the case of "soybean", "alfalfa", and "hemp" terms). These terms differ in morphology and semantics in English and Uzbek, so they need context-sensitive translation. Notably, it also monitors the emergence of neologism in plant science, such as "bio-cotton" and "CRISPR crops" and identifies the implication for accurate translation strategies. Using corpora including COCA and Sketch Engine, the research standardizes botanical lexicon and improves its use in scientific research and lexicography.

Keywords: Corpus linguistics, grammatical, semantic, contextual, statistical, translation equivalents, synonymy, statistical approach, neologisms.

Citation: Choriyeвна, M. S. The Role Of Corpus Linguistics In The Terminological Analysis Of Botanical Terms. Central Asian Journal of Literature, Philosophy, and Culture 2025, 6(4), 966-970.

Received: 13th Sept 2025

Revised: 20th Sept 2025

Accepted: 05th Oct 2025

Published: 21th Oct 2025



Copyright: © 2025 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>)

1. Introduction

The corpus Linguistics has assumed an ever more central role in modern linguistics-its focus is the study of linguistic units in their natural habitat, that is, the analysis of or large scale, systematic assembled text corpora. Such measures are essential in highly technical fields such as botany, where terms can have a highly specific and context-sensitive meaning. However, in the botany sciences the precise use and translation of terminology is crucial to clear scientific communication, teaching, and research. Although a central aspect of traditional terminological analysis, subtle nuances of contextual varieties typically go unconsidered, resulting in semantic ambiguities and errant translations [1]. Therefore, the correlation of corpus linguistics with the analysis of terminology is becoming increasingly popular in the study of botanical terms in English and Uzbek. Corpus Linguistics, Terminology, and the notion of Translation Equivalence as key concepts of this study According to the Oxford English Dictionary, terminology is the vocabulary specific to a particular field of study and, therefore, its meaning must be analyzed as well as the context where they are used [2]. Frequency, collocation, and semantic context are the three key terms that provide the methodological framework for such an analysis and that derive from the field of quantitative corpus linguistics. While

several works have examined similar corpus-based approaches to terminological analysis, including in multidisciplinary context, few of these methods generalise to botanical terminology on a comparative basis across more than one language [3]. This research gap is due to the lack of corpus-based approaches in solving research problems and unification of terminological lexicons for botanical sciences in non-English oriented sciences like Uzbek.

Through the use of frequency, collocational, concordance and morpho-semantic analysis, and based on corpus-based analytical methods, this study investigates the formation, semantic properties and translation equivalents of botanical terms in English and Uzbek [4]. This expectation is that using these methods will result in a deeper understanding of botanical terminology, highlight recurring problems with translation, and provide insight into what rendering of terms may be more uniform. The results will help to enhance translation quality, advance scientific lexicography, and provide applicable tools both for the translation of botanical terms and their teaching. In short, this research strives to bring theoretical terminological studies closer to practice to stimulate and achieve more efficient communication in botany and related fields [5].

2. Materials and Methods

The present study is a corpus-based analysis of some aspects of botanical term formation, semantics and translation equivalents in English and Uzbek. This study uses big, manually labeled aligned corpora of scientific papers and botanical dictionaries data and uses frequency analysis to determine important keywords, collocational analysis to analyze word partnerships, and concordance analysis to observe words in their contexts [6]. In addition, Morphosemantics is used for the syntagmatic and paradigmatic analysis of words, for example, processes of composition and affixation. The work also attempts to deal with the translation problems where the English botanical words have been compared with Uzbek and the perfection of the semantic gaps, synonymy and polysemy has been noted. Its purpose is to clarify ambiguities in translation and, help to standardize the use of botanical terminology among languages. All these techniques allow the research to present a context-sensitive means of translating and interpreting botanical terms that will be beneficial for scientific lexicography and a basis for terminology development [7].

3. Results and Discussion

Relationship Between Corpus Linguistics and Terminology: Terminology refers to the system of specialized lexical units used within a particular field of science, technology, or profession. Traditional approaches (reference works, explanatory dictionaries, scientific articles) are insufficient in studying terms, because:

1. The contextual usage of terms is not always reflected in theoretical sources;
2. Translation challenges are often solved on the basis of real texts;
3. Modern vocabulary is constantly updated with new terms.

Therefore, corpus linguistics makes it possible to analyze terms in their actual linguistic environment, i.e., in practical usage [8].

Advantages of Corpus Linguistics in Terminological Analysis. Corpus-based terminological analysis stands out for the following aspects:

Statistical methodology Through the number of times terms are used, we can ascertain their degree of activity and centrality in the field. For example: COCA: cotton: > 25,000 flax: ~1,800 So cotton is much more important+ or high-frequency

Contexts analysis Similar words (collocations) and grammatical structures are recognized. Such as: root in root vegetable, deep root, root cause and they have little different meanings.

When the same concept is expressed in multiple term(s), this is known as Synonymy and Variability, which helps in translation and standardization.

For example: yam vs. sweet potato – a semantic distinction, often confused.

Translation Solutions Parallel corpora (e.g. English-Uzbek, English-Russian) enable us to find actual translation equivalents and semantic gaps.

For example, gourd is a very general classification in English, but it might translate as qovoq, luffa, or poliz ekini in Uzbek, according to context [9].

Corpus of neologism monitoring, usually around biotechnology and agriculture. Ex: GM crop "occurrence frequency has risen significantly from 2000.

Neologism Monitoring Corpora track the emergence of new terms, especially in biotechnology and agriculture. *Example: GM crop* has seen a marked frequency increase since 2000.

Corpus Linguistics in Terminological Lexicography

The advantages of corpus-based terminological dictionaries:

1. Terms are presented with their real context of usage;
2. Each entry includes collocations and translation options;
3. Synonyms and terminological shifts are identified;
4. Registers across scientific, journalistic, and applied contexts are distinguished.

This table presents English botanical terms, their common collocations, and their corresponding Uzbek translation options. It highlights key terms related to legumes, such as soybean, chickpea, and lentil, along with their usage in agricultural contexts, see Table 1.

Table 1. Legumes (Dukkakli don ekinlari)

No	English Term	Collocations	Uzbek Translation Options
1	Soybean	soybean intercropping, soybean yield, soybean aphid infestation	soya ekini bilan qo'sh ekin tizimi, soya hosili, soya o'rgimchigi zarari
2	Chickpea	chickpea cultivation, chickpea protein content	no'xat yetishtirish, no'xat tarkibidagi protein
3	Lentil	lentil pulse, red lentil production	loviya yasmiqi, qizil loviya hosili

This table provides English terms related to forage crops, such as alfalfa, clover, and ryegrass, along with their common collocations and corresponding Uzbek translation options. It emphasizes their usage in agriculture, particularly in relation to forage and yield (Table 2).

Table 2. Forage Crops (Yem-xashak o'tlar)

No	English Term	Collocations	Uzbek Translation Options
1	Alfalfa	alfalfa forage, alfalfa yield	luserna yem ekini, luserna hosili
2	Clover	red clover hay, clover forage quality	qizil bedana o'ti, bedana yem sifati
3	Ryegrass	perennial ryegrass pasture, ryegrass seed	doimiy roj o'ti yaylov, roj urug'i

Corpus-Based Analytical Methods

1. **Frequency Analysis** – Identifies the most active terms (*potato* 85,000+ vs. *taro* 4,000 occurrences).
2. **Collocational Analysis** – Reveals semantic fields (*cotton yield*, *cotton fiber*, *cotton field*).
3. **Concordance Analysis** – Studies surrounding context; e.g., *carrot* in literal vs. metaphorical uses.
4. **Synonymy Analysis** – Distinguishes lexical variants (*sweet potato* vs. *yam*).
5. **Translation Equivalents** – Established via parallel corpora (*gourd* with 5 Uzbek renderings).
6. **Neologism Tracking** – Identifies newly emerging terms (*CRISPR crops*, *bio-cotton*).
7. **Morpho-Semantic Analysis** – Examines term structure (*sunflower* = *sun* + *flower*).
8. **Semantic Group Comparison** – Differentiates root crops, tubers, oil plants, fiber plants, etc.

Plant Groups in Corpus-Based Study

1. **Tubers:** potato, yam, cassava, taro, jicama, etc. (often generalized as *root vegetables* in English; Uzbek equivalents may require descriptive translation).
2. **Root Crops:** carrot, beet, radish, turnip, etc. (*carrot* being the most frequent).
3. **Melons & Gourds:** melon, watermelon, cucumber, pumpkin, squash, luffa, etc. (*melon's* equivalent in Uzbek depends on context: *qovun*, *poliz ekini*).
4. **Fiber Crops:** cotton, flax, hemp, jute, ramie, etc. (*cotton* may mean both *paxta* and raw fiber).
5. **Oil Crops:** sunflower, soybean, sesame, peanut, olive, jojoba, etc. (*jojoba* appears mostly in cosmetic contexts).
6. **Narcotic Plants:** opium poppy, cannabis, coca, peyote, ayahuasca, etc. (mostly require descriptive translation in Uzbek).

Corpus linguistics provides:

- a. Statistical and contextual precision in terminological studies;
- b. Identification of synonymy, polysemy, and translation equivalents;
- c. Support for terminological standardization and lexicography;
- d. A dynamic alternative to static dictionaries by capturing real usage.

In rapidly evolving and multilingual fields such as botany and agriculture, corpus-based approaches ensure accuracy, contextual relevance, and scientific consistency in terminology [10].

In this article, the terminology related to plant groups (legume-cereal crops, forage grasses, tuberous and root crops, melon crops, fiber, oil, and narcotic plants) in English and Uzbek is analyzed in terms of their formation, semantic and lexical features, as well as translation peculiarities. In addition, by analyzing plant science terminology based on modern corpus linguistics, the contextual, functional, and morphological aspects of the field's language have been clarified.

1. The results of lexical and semantic analysis show that the plant group terms in English and Uzbek are formed using particular ways and methods of different parts of speech, which belong to the morphological methods (composition, affixation, conversion) and the specific semantic tools (metonymy, metaphor, generalization and narrowing) [11]. Especially because terms like soybean, alfalfa, taro, and hemp are not always formed the same way in all languages and therefore highlight cases of synonymy and polysemy in translation.
2. Translation problems have shown that many specialized terms need to be clarified depending on the context [12]. For example, terms like *soybean intercropping*, *clover nitrogen fixation*, and *opium poppy* may cause semantic errors in direct translation. This demonstrates the necessity of a contextual approach in the language of the field.
3. Corpus linguistics-based analytical methods – frequency analysis, collocational analysis, concordance analysis, synonymy, morpho-semantic, and translational equivalence analysis – have been used to identify the actual usage forms, contextual registers, and translation equivalents of the terms [13]. Data obtained through corpus platforms such as *Sketch Engine*, *AntConc*, and *COCA* serve as a key source for enhancing scientific precision, standardization, and dictionary compilation in the field's terminology.
4. Within plant science terminology, semantic differentiation exists inside the field itself; for example: *cotton* and *flax* – both are fiber crops, yet their usage scope, collocations, and semantic roles differ [14][15]. This aspect requires precision in Uzbek translations as well, especially from a professional perspective in the field.
5. Through a corpus-based approach, neologisms and elements of dynamic lexis have also been analyzed. For instance, terms such as *bio-cotton*, *CRISPR crops*, and *drought-resistant varieties* are actively used in modern plant science discourse, and selecting appropriate equivalents in translation is necessary.

4. Conclusion

So, this study successfully shows the power of the corpus linguistics the analyzing the vocabulary of plants and provides extremely valuable information about plants terms in terms of their meaning, morphology and use in the context in English and Uzbek languages. The results show some important challenges for translation, where synonymy, polysemy and equivalence of terms across languages come into play and call for context-sensitive translation strategies. The study emphasizes the need for a corpus-based approach for the standardization of botanical terminology which is expected to bring improvements in accuracy at least in the realm of scientific lexicon and translation practices. In addition, the research adds nuance to the well-documented process of linguistic change in fast-changing areas such as botany, especially with regard to novel concepts and terms, examples here being "bio-cotton" and "CRISPR crops". Based on the results, similar research could use corpus linguistics to study other professional fields, for example, environmental science and biotechnology, to provide additional insights into term usage and translation on a transnational and multilanguage basis.

REFERENCES

- [1] McEnery, T., & Hardie, A., *Corpus Linguistics: Method, Theory and Practice*, Cambridge University Press, 2012.
- [2] Sinclair, J., *Trust the Text: Language, Corpus and Discourse*, Routledge, 2004.
- [3] Cabré, M. T., *Terminology: Theory, Methods and Applications*, John Benjamins Publishing, 1999.
- [4] Davies, M., *The Corpus of Contemporary American English (COCA)*, 2008. [Online]. Available: <https://www.english-corpora.org/coca/>
- [5] Stubbs, M., *Words and Phrases: Corpus Studies of Lexical Semantics*, Blackwell, 2001.
- [6] *The British National Corpus (BNC)*, version 3, Oxford University Computing Services, 2014.
- [7] *UCREL Multilingual Parallel Corpus (Lancaster University)*, Agricultural Texts Subset, 2021.
- [8] Davies, M., *The Corpus of Historical American English (COHA)*, Brigham Young University, 2010.
- [9] Giller, K. E., *Nitrogen Fixation in Tropical Cropping Systems*, CABI Publishing, pp. 129–131, 2001.
- [10] Sketch Engine, "cotton" collocations (Brown corpus, enTenTen20), 2023.
- [11] *AGRICOLA corpus*, USDA National Agricultural Library, Morphosemantic analysis list, 2021.
- [12] Kilgarrieff, A., & Grefenstette, G., Introduction to the Special Issue on Corpus Linguistics, *Journal of Language and Technology*, vol. 2, no. 1, pp. 1-6, 2003.
- [13] Biber, D., Conrad, S., & Reppen, R., *Corpus Linguistics: Investigating Language Structure and Use*, Cambridge University Press, 1998.
- [14] O'Keeffe, A., McCarthy, M., & Carter, R., *From Corpus to Classroom: Language Use and Language Teaching*, Cambridge University Press, 2007.
- [15] Louw, B., Contextual Prosody and the Corpus-based Study of Lexical Semantics, *International Journal of Corpus Linguistics*, vol. 6, no. 2, pp. 161-187, 2001.