

LINGUISTICS AND VIETNAMESE LINGUISTICS

EXPLORING SEMANTIC RELATIONSHIPS IN THE SEMANTIC TAGSET SYSTEMS FOR ENGLISH NOUNS

PHAN THI MY TRANG* - DANG NGOC LE** -
DINH DIEN*** - TRAN THI MINH PHUONG****

ABSTRACT: Semantic tagset systems are fundamental to computational semantics research. Each semantic tagset system contains different parts of speech, and these parts of speech have various semantic relationships. In this research, we investigated the semantic relationships of English nouns through three prominent semantic tagset systems: LLOCE, UCREL, and WordNet. Using a descriptive research method as well as a comparative and contrastive method, the study explored the underlying semantic relationships of these three systems, highlighting their similarities and differences. The findings help identify the strengths and limitations of each system, providing a better understanding of the semantic relationships of nouns in English semantic tagset systems, and consequently propose similar models for the semantic relationships of Vietnamese nouns. These discoveries will provide a solid theoretical foundation for constructing semantic tagset systems for Vietnamese nouns, contributing to practical applications in automatic semantic processing as well as in the teaching and research of lexical semantics.

KEY WORDS: semantic relationships; semantic tagset systems; LLOCE; UCREL; WordNet.

RECEIVED: February, 18th, 2025.

EDITED: June, 17th, 2025

1. Introduction

Several semantic tagset systems have been developed and categorized based on semantic domains worldwide, including the Longman Lexicon of Contemporary English (LLOCE) (McArthur, 1988), the UCREL Semantic Analysis System (USAS) (Rayson et al., 2004), and WordNet (Miller, 1990). Each of these systems employs its own unique approach and structure to represent semantic relationships. The LLOCE, derived from the Longman Dictionary of Contemporary English (LDOCE), is a lexicon organized based on semantic principles. Unlike traditional dictionaries that list words alphabetically, the LLOCE groups related words by meaning into thematic sets (McArthur, 1988). USAS, a multilingual automatic semantic analysis system developed at Lancaster University, is recognized for its ability to classify and label lexicon within texts. The UCREL semantic tagset, originally inspired by LLOCE, was first developed for English and later extended to 12 other languages. There is currently no semantic tagset system for Vietnamese (Piao et al., 2016). WordNet, a famous lexical database developed by Princeton University, organizes words into synonym sets (synsets) and provides detailed information on semantic relationships. With over 117,000 synset classes (Fellbaum, 2005), WordNet is now extensively used in linguistic research and computational applications.

Research on the semantic relationships of nouns in these representative semantic tagset systems not only enhances our understanding of how nouns are organized and categorized based on meaning, but also opens up the opportunity to develop a similar semantic tagset system for Vietnamese. At present, there is no specific research addressing the semantic relationships between nouns in these systems. Therefore, the primary objective of this study is to explore the semantic relationships of English nouns in three representative semantic tagset systems: LLOCE, UCREL, and WordNet. The study used a descriptive research method to analyze each system individually and a comparative and

**Ph.D. student - University of Social Sciences and Humanities, Vietnam National University Ho Chi Minh City; Saigon Technology University; Email: mytrang779@gmail.com*

***Assoc.Prof.Dr; Nguyen Tat Thanh University; Email: ngocle.dang1309@gmail.com*

****Assoc.Prof.Dr; University of Science, Vietnam National University Ho Chi Minh City; Email: ddien@fit.hcmus.edu.vn*

*****Dr; University of Social Sciences and Humanities, Vietnam National University Ho Chi Minh City; Email: minhphuongtrn@hcmussh.edu.vn*

contrastive method to highlight the similarities and differences among the systems, while also identifying their respective strengths and weaknesses. The research results show that WordNet has a more detailed lexical structure and more diverse semantic relationships compared to the other two systems. Based on these insights, the study proposes using semantic relationships such as synonymy, polysemy, hypernymy, and hyponymy as the foundation for building a semantic tagset system for Vietnamese nouns.

2. Literature review

The development of thesauruses can be traced back to the philosophical ideas of Leibniz, Bishop Wilkins, and other scholars in the 17th century, which laid the foundation for logical empiricism and influenced Russell and later Peter Mark Roget (Lyons, 1995). Roget's "Thesaurus of English Words and Phrases", published in 1852, marked a significant milestone in lexicography by organizing lexicon thematically rather than alphabetically, inspiring subsequent linguistic and thesaurus research.

One of the modern advancements in semantic tagset systems is the Longman Lexicon of Contemporary English (LLOCE), developed from the Longman Dictionary of Contemporary English (LDOCE) by a team of linguists and lexicographers led by Tom McArthur. First published in 1981, LLOCE organizes vocabulary thematically. The lexicon was expanded with each subsequent edition, with the 9th edition released in 1988 (McArthur, 1988). This was followed by bilingual versions, including English-Vietnamese and English-Chinese dictionaries (McArthur, 1997). LLOCE contains 16,000 headwords and 25,000 meanings, systematically arranged into 14 major codes, 127 group codes, and 2,441 set codes.

The UCREL Semantic Analysis System (USAS) is an automatic software tool for assigning semantic tags to English, initially based on LLOCE (Rayson et al., 2004). USAS has a multi-layered structure with 21 major discourse fields, 232 category labels, organized into a maximum of three layers. These categories can be further subdivided for more detailed classification in certain cases. Initially developed for tagging semantics in English, USAS has since expanded to 12 other languages, including Arabic, Chinese, Czech, Dutch, Finnish, Italian, Malay, Portuguese, Russian, Spanish, Urdu, and Welsh (Piao et al., 2016). Currently, USAS does not have a semantic tagging system for Vietnamese.

WordNet is an English lexical database developed at Princeton University in 1986 by a group of scientists, led by psychologist George Armitage Miller. WordNet organizes the English conceptual system into groups of words that have semantic relationships with each other. Nouns, verbs, adjectives, and adverbs in English are grouped into synsets, with each synset representing a distinct concept. These sets are linked through conceptual-semantic and lexical relations. The semantic relationships between words in WordNet include synonymy, antonymy, hyponymy, hypernymy, meronymy, holonymy, entailment, and troponymy. WordNet differs from a thesaurus because it links specific meanings of words and assigns labels to the semantic relationships between them, whereas thesaurus entries do not follow any explicit pattern other than similarity in meaning. WordNet currently contains 117,000 synsets (Fellbaum, 2005), divided into 25 basic concepts, which are further reduced to 11 primitive concepts (Miller et al., 1990).

So far, there has been no detailed research on the semantic relationships of the three systems mentioned above. Most studies focused on exploring the importance of lexical semantic resource systems in language analysis and natural language processing. Archer et al. (2004) used the UCREL (USAS) semantic annotation scheme to analyze semantic fields in lexicography. The paper compared the UCREL semantic analysis system with two dictionary classification systems, LLOCE and Collins English Dictionary (CED), testing the mapping capability of the UCREL labels with the Collins English Dictionary and WordNet. This study guided the process of label mapping, bringing lexicographers closer to using semantic fields as the organizing principle in dictionaries. Piao et al. (2016) expanded the multilingual scope by developing a large-scale multilingual semantic lexicon

system for 12 languages, using the Lancaster semantic classification to build a multilingual lexical knowledge base for the automatic UCREL semantic annotation system (USAS). The study highlighted the importance of vocabulary coverage for the quality and performance of lexical annotation tools and corpora, with some languages achieving over 90% vocabulary coverage, while others required further expansion.

Overall, the studies mentioned above emphasize the importance of semantic tagset system resources in language analysis and natural language processing, opening up significant potential for practical applications. They provided a scientific foundation and highlighted the feasibility of constructing semantic tagset systems and developing semantic annotation tools for new languages. Therefore, investigating semantic relationships in the semantic tagset system for English nouns is necessary in order to propose a similar model for semantic relationships in Vietnamese, contributing to practical applications in fields such as lexical semantics, lexicography, corpus linguistics, and computational linguistics.

3. Theoretical Background

3.1. Conceptual Hierarchies

Conceptual hierarchies organize knowledge based on hierarchical relationships between concepts, systematically classifying information from general to specific. There are two main types of conceptual hierarchies: *type-of hierarchies* and *part-whole hierarchies*.

A *type-of hierarchy* organizes items based on common properties or characteristics, following the principle of class inclusion. For example, the class “animals” consists of the sub-classes “mammals”, “birds”, and “reptiles”. At the next level, the class “mammals” includes “dogs”, “cats”, “cows”, “lions”, “elephants”, and “mice.” Further, the class “dogs” includes “terriers”, “bulldogs”, “Alsatisans”, “poodles”, and other dog breeds (Ungerer & Schmid, 2006) (see Figure 1).

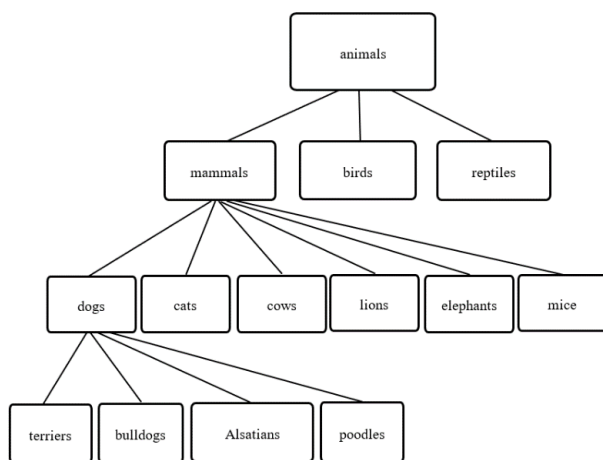


Figure 1. Illustration of a branch in the hierarchical tree structure of conceptual classes

On the other hand, a *part-whole hierarchy* describes the relationship between parts that make up a whole, based on spatial proximity (i.e., closeness) and the continuity of individual objects in the real world (Cruse 1986:178, as cited in Ungerer & Schmid, 2006). For example, a finger is a part of a hand, a hand is a part of an arm, and an arm is a part of the body.

Type-of hierarchies utilize basic categories, hyponyms, and hypernyms, which are commonly used in scientific classification. In contrast, part-whole hierarchies focus on folk classification and are observable in everyday language (Ungerer & Schmid, 2006).

3.2. Semantic Relationships for Nouns

Semantic relationships within a lexical system refer to the connections between words based on their meanings. There are various types of lexical semantic relationships, and this research focuses on the relationships of synonymy, polysemy, hypernymy, and hyponymy within the scope of English nouns.

3.2.1. Synonymy

Synonymy refers to the relationship between words that share similar meanings in certain contexts. According to Geeraerts, there are two types of synonymy: *absolute synonymy* and *partial synonymy* (Geeraerts, 2010). Absolute synonyms are words that have exactly the same meaning and can be substituted for each other in all relevant contexts without changing the overall meaning of the sentence (Geeraerts, 2010). However, in practice, absolute synonyms are very rare, as most synonyms differ in at least one characteristic or aspect of meaning. Partial synonyms are words that share similarities in meaning, but may have slight differences in meaning in specific contexts (Geeraerts, 2010). For example, both “house” and “home” refer to a place of residence, but “house” typically emphasizes the physical structure, while “home” brings to mind a sense of warmth and belonging. The degree of synonymy can be classified on a scale from absolute synonymy to non-synonymy, with synonyms not only showing considerable overlap in meaning but also reflecting minimal potential differences (Cruse, 1986, p.266).

3.2.2. Polysemy

A polysemous word is a single lexical form that is linked to two or more related meanings, which are connected by a common concept (Vicente & Falkum, 2017). A typical example is the word “bank” in English, which can refer to a financial institution, the side of a river, a reserve of assets, a money box, a pile of earth, or a slope. These meanings are often related to the concept of containment or boundaries. The phenomenon of polysemy occurs because language has a limited number of lexical items, while the concepts in the real world are diverse and infinite.

3.2.3. Hypernymy and Hyponymy

Hypernymy and hyponymy are terms that describe the semantic inclusion relationship between words, where a hypernym is a higher-level term that has a more general meaning and includes hyponyms as its subcategories. For example, “flower” is a hypernym of “rose”. On the contrary, a hyponym has a more specific meaning and its parent is the hypernym. For instance, “rose” is a hyponym of “flower”. Co-hyponyms are words that refer to the same level of specificity under a shared hypernym. For example, “tulip”, “daisy”, and “rose” are co-hyponyms of “flower”. This hierarchical structure of hypernyms and hyponyms forms a “taxonomy” (Geeraerts, 2010). Hypernymy and hyponymy are particularly significant in computer science, especially in the fields of natural language processing and computational linguistics.

4. Methodology

In order to study the semantic relationships of nouns across semantic tagset systems such as LLOCE, UCREL, and WordNet, we conducted data collection using a descriptive research method, along with a comparative and contrastive method.

4.1. Data Collection

This study collected data from three semantic tagset systems: the Longman Lexicon of Contemporary English (LLOCE) (McArthur, 1988), the UCREL Semantic Analysis System (USAS) (Rayson et al., 2004), and WordNet (Miller, 1990). The data collection process began by using the LLOCE dictionary to look up sets of labels and related semantic information, accessing the USAS database to gather semantic information from labeled and used materials, and retrieving data from WordNet to compile a corpus of English nouns along with their corresponding semantic relationships.

4.2. Descriptive Research Method

This study focused on investigating and describing the semantic relationships of nouns, including synonymy, polysemy, hyponymy, and hypernymy, across three semantic tagset systems: LLOCE, UCREL, and WordNet.

4.3. Comparative and Contrastive Method

The comparative and contrastive method was used to identify similarities and differences in the semantic relationships across the three semantic tagset systems: LLOCE, UCREL, and WordNet. This approach aimed to uncover the strengths and limitations of each semantic tagset system and propose semantic relationships for the development of a semantic tagset system for Vietnamese nouns.

5. Results and Discussion

5.1. Semantic Relationships of Nouns in three English Semantic Tagset Systems

In this section, we focus on analyzing the semantic relationships of nouns in three English semantic tagset systems (LLOCE, UCREL, WordNet). Each of these systems makes distinct contributions to the understanding of semantic relationships.

5.1.1. Semantic Relationships of Nouns in LLOCE

5.1.1.1. Synonymy

The LLOCE semantic tagset system does include synonymy, though it is used infrequently and is not common. Synonymy appears at the lowest level of the system, corresponding to the practical meanings of words. For example, the word group under label C13, which refers to “father and mother”, includes synonyms for “father” such as “dad”, “daddy”, “pa”, “papa”, “poppa”, “pop”, and synonyms for “mother” such as “mum”, “mummy”, “mom”, “mommy”, “ma”, “mam”, “mammy”, “mama”, “momma” (McArthur, 1988, p.83) (see Figure 2). This is a rare example of synonymy in LLOCE.

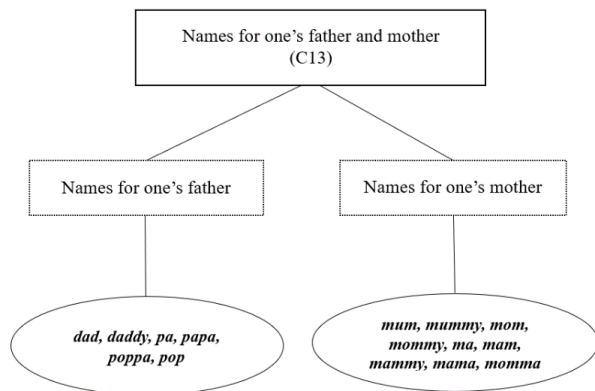


Figure 2. Synonyms for “Names for one’s father and mother” in LLOCE

5.1.1.2. Polysemy

LLOCE organizes the polysemy relationship through cross-references, linking different meanings of the same word without repeating the old information. LLOCE presents polysemy relationships in detail by categorizing all polysemous words in the Index alphabetically, with specific codes to help users easily look up meanings (McArthur, 1988, p.787-910). For example, under the word “bank”, the codes “money (n) J104” and “river (n) L99” are listed (McArthur, 1988, p.794). Using these codes, users can easily refer to the content and find the different meanings of the word “bank” in the LLOCE dictionary (see Figure 3).

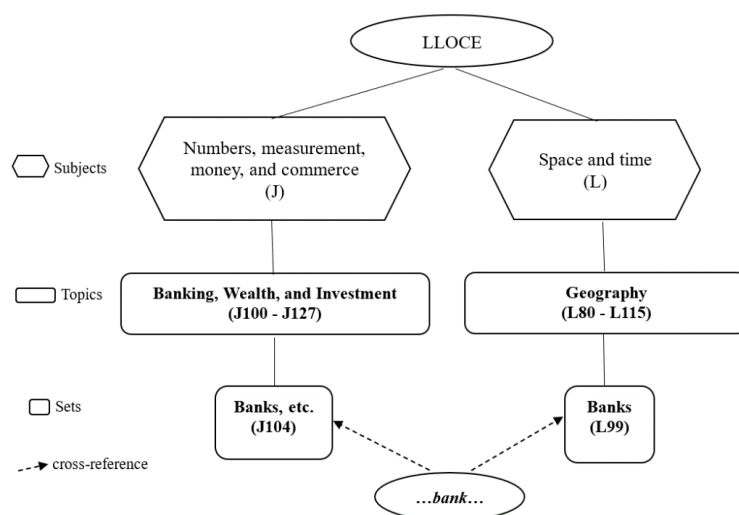


Figure 3. Semantic subjects, topics, sets, and cross-references of the word “bank” in LLOCE

5.1.1.3. Hypernymy and Hyponymy

LLOCE organizes vocabulary into topics, with each topic divided into groups, and each group further subdivided into classes (meanings) according to a hierarchical structure of class inclusion. The hypernym class includes the items in the hyponym class, helping to systematically organize and classify vocabulary from general to specific information. For example, under the code D37 for “rooms in houses”, “rooms in houses” is the hypernym that includes hyponyms such as specific types of rooms: “living room”, “lounge”, “front room”, “study”, “workroom”, “bedroom”, and “dining room” (McArthur, 1988, p.178). These specific room types are co-hyponyms of “rooms in houses”. LLOCE presents the hypernym and hyponym relationships in detail. However, the hierarchical system remains relatively simple, with only three levels (see Table 1).

Table 1. Hierarchical System, Hypernym and Hyponym Relations in LLOCE for “Rooms in Houses”

Code Type	Code	Description
Major code	D	Buildings, houses, the home, clothes, belongings, and personal care
Group codes	D20-43	Parts of houses
Set code	D37	Rooms in houses
Actual word senses		- living room - lounge - front room - study - workroom - bedroom - dining room - ...

5.1.2. Semantic Relationships of Nouns in UCREL

The UCREL semantic analysis system is an automated software tool that assigns labels to every word in a text. Initially, the label set in this system was built based on LLOCE, but it was later modified due to issues encountered in the tagging process during the research (Archer et al., 2024). This system inherits semantic relationships from LLOCE, such as synonymy, polysemy, hypernymy, and hyponymy. However, unlike LLOCE, the UCREL semantic analysis system mapped its labels to synonym sets in WordNet to clarify the meanings of words that had not been previously classified, and then assigned labels to them (Archer et al., 2024).

5.1.3. Semantic Relationships of Nouns in WordNet

5.1.3.1. Synonymy

In the WordNet semantic tagset system, synonymy plays a central role, represented in detail through synonym sets (synsets). These synsets are interconnected through conceptual-semantic and lexical relations (Fellbaum, 2005). The analysis of synonyms in WordNet demonstrates a rich and diverse network of words with similar meanings. For example, synset ID (10007601) includes {dad, dada, daddy, pa, papa, pappa, pop}, all referring to informal terms for “father” and can be used interchangeably in equivalent contexts. Similarly, synset ID (10297825) includes {ma, mama, mamma, mom, momma, mommy, mamma, mum, mummy}, which are used to refer to informal terms for “mother” (see Figure 4).

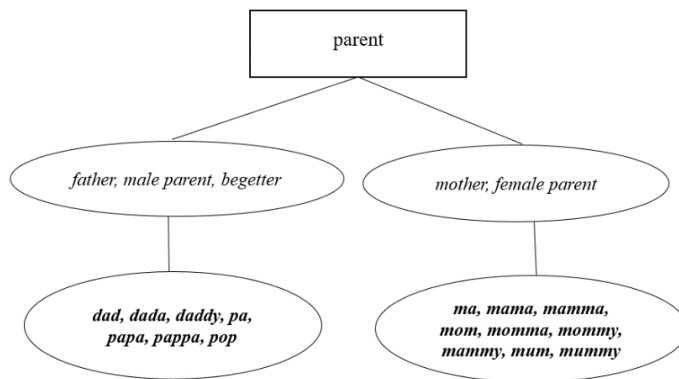


Figure 4. Synonyms for “father” and “mother” in WordNet

5.1.3.2. Polysemy

Polysemy is the phenomenon in which a word has more than one meaning. In English, many words exhibit polysemy. WordNet helps users easily search for and analyze polysemous words. When sending a query (search word) to the Princeton WordNet system, the result often shows more than one synset, indicating the variety of meanings in different contexts. For example, a search for the word “bank” in WordNet returns 10 distinct meanings (see Table 2). These meanings of the noun “bank” can refer to a slope of land, a financial institution, or a container for holding money, among others.

Table 2. Polysemy of the noun “bank” in WordNet (Fellbaum, 2005)

Polysemy	BANK (noun)
No.	“S:” = Show Synset (semantic) relations
1	S: (n) bank (sloping land (especially the slope beside a body of water))
2	S: (n) <i>depository financial institution</i> , bank , <i>banking concern</i> , <i>banking company</i> (a financial institution that accepts deposits and channels the money into lending activities)
3	S: (n) bank (a long ridge or pile)
4	S: (n) bank (an arrangement of similar objects in a row or in tiers)
5	S: (n) bank (a supply or stock held in reserve for future use (especially in emergencies))
6	S: (n) bank (the funds held by a gambling house or the dealer in some gambling games)
7	S: (n) bank , <i>cant</i> , <i>camber</i> (a slope in the turn of a road or track; the outside is higher than the inside in order to reduce the effects of centrifugal force)
8	S: (n) <i>savings bank</i> , <i>coin bank</i> , <i>money box</i> , bank (a container (usually with a slot in the top) for keeping money at home)
9	S: (n) bank , <i>bank building</i> (a building in which the business of banking transacted)
10	S: (n) bank (a flight maneuver; aircraft tips laterally about its longitudinal axis (especially in turning))

5.1.3.3. Hypernymy and Hyponymy

In WordNet, hypernym and hyponym relationships are organized in a tree structure, which helps systematize and simplify the process of looking up concepts. Each word is arranged in a branching tree starting from a root concept, with higher levels containing attributes of lower levels to avoid redundancy. For example, the hypernym class “ungulate” (hoofed animal) includes the hyponym class “odd-toed ungulate”, which in turn includes the class “equid”, and finally the class “horse”. Moving from the bottom up, the class “horse” is a hyponym of “equid”, “equid” is a hyponym of “odd-toed ungulate”, and “odd-toed ungulate” is a hyponym of “ungulate” (see Figure 5).

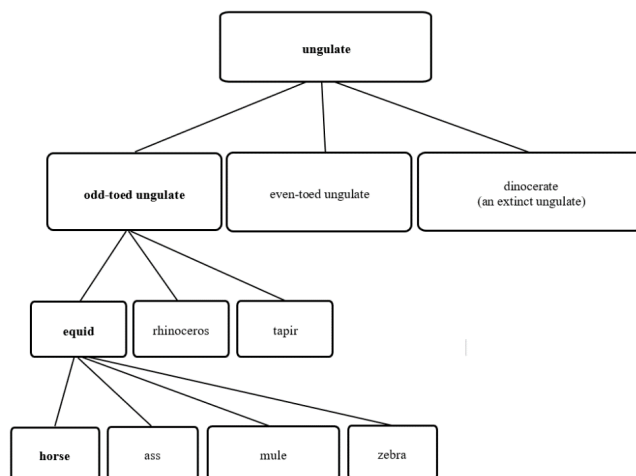


Figure 5. Branch of the hierarchical tree structure showing hypernym and hyponym relationships in WordNet

WordNet does not limit the number of levels, but lexical systems used in everyday life typically do not go deeper than ten levels. The deeper levels lead to the “Tops” file, which contains more technical concepts. For example, starting a query with the word “pony”, we get: pony @→ horse @→ equid @→ odd-toed ungulate @→ ungulate (ungulate) @→ mammal @→ vertebrate @→ animal; if continuing into the “Tops” file, it proceeds to concepts like: animal @→ organism @→ living thing @→ {object, physical object} @→ physical entity @→ entity. This chain illustrates the hypernym and hyponym relationships between concepts, starting from “pony” and ending with “entity”. WordNet provides detailed hypernym and hyponym relationships with an undefined number of levels, and the depth depends on the specific topic.

5.2. Similarities and Differences between the Semantic Relationships in LLOCE, UCREL, and WordNet

All three systems - LLOCE, UCREL, and WordNet - organize and link vocabulary based on meaning through basic semantic relationships such as synonymy, polysemy, hypernymy, and hyponymy. However, each system has a different approach and organization (see Table 3).

Table 3. Comparison of Semantic Relationships in LLOCE, UCREL, and WordNet

Semantic Relationship	LLOCE	UCREL	WordNet
Synonymy	Not emphasized	Inherited from LLOCE; expanded through mapping with WordNet	Playing a central role; represented through synonym sets (synsets)
Polysemy	Presented in detail; organized through cross-referencing (linking meanings of the same word without repeating old information)	Inherited from LLOCE	Easily searchable with multiple synsets for each word
Hypernymy and hyponymy	Simple, consisting of three levels: topic - group - class; actual words appear only at the lowest level	Inherited from LLOCE	Tree structure with an undefined number of levels; depth varies by topic; deeper levels contain technical concepts

LLOCE focuses on organizing vocabulary by topic and does not place much emphasis on synonym relationships. This system provides a detailed analysis of polysemy, using cross-references to connect a word to multiple meanings that represent different characteristics and attributes, and organizing polysemous words in the Index. The hierarchical system of LLOCE is quite simple, consisting of only three levels (topic - group - class), which makes it difficult to identify semantic relationships between the levels (Dinh Dien, 2018). The number of labels (semantic classes) is not large, and actual words are entirely located at the lowest level. With its simple and accessible vocabulary system, LLOCE is a useful tool for English learners to look up and expand their vocabulary within topic-based groups.

UCREL, inheriting from LLOCE, expands synonym relationships through mapping with WordNet and focuses on automatic semantic tagging and analysis of every word in a text. This system plays an important role in extracting semantic annotations for words, sentences, paragraphs, and texts. However, UCREL may present challenges for users in terms of accessibility and understanding the semantic relationships.

Meanwhile, WordNet emphasizes semantic relationships and organizes its vocabulary based on conceptual meanings. It provides synset IDs for each set of words and allows users to easily search for polysemous words. Hypernym and hyponym relationships in WordNet are organized in a tree

structure, with an undefined number of levels, and the depth varies depending on the topic. However, the deepest levels in WordNet often contain technical concepts, rather than words commonly used in everyday life. WordNet generally works well with specific and concrete nouns, but for abstract concepts, its structure is sometimes not optimally organized to fully reflect their meanings. In terms of lexical organization based on semantic relationships, WordNet offers a more detailed system with a broader variety of semantic relations compared to LLOCE and UCREL. WordNet plays an important role in applications such as information retrieval on semantic relationships, semantic analysis, and supporting machine translation.

5.3. Suggestions for Semantic Relationships in Developing a Semantic Tagset System for Vietnamese Nouns

Modern linguistics acknowledges that the lexicon-semantics is a system (Le Quang Thiem, 2018). In the process of constructing a lexical-semantic tagset system, classification and organization according to a hierarchical structure is crucial, with the fundamental components being basic categories, hyponyms, and hypernyms. The lexical-semantic system is a dynamic one, characterized by its vastness and extreme complexity (Le Quang Thiem, 2018). Language constantly evolves under the influence of social, historical, and cultural factors. Therefore, when developing a semantic tagset system, it is essential to view lexical meaning as part of a dynamic system rather than treating it as static or rigid in classification and arrangement. The lexical-semantic system operates at multiple levels, with large systems covering semantic fields corresponding to various thematic domains. These semantic fields should not only consider synonymy but also pay attention to polysemy, using both literal and derived meanings as the basis for grouping meanings. Moreover, it is necessary to avoid limiting exploration to thematic domains alone. When constructing the system, attention must be given to expressive functions and cognitive factors (Le Quang Thiem, 2018).

Constructing a semantic tagset system that aligns with the Vietnamese language and culture is a complex and labor-intensive task, requiring extensive research and long-term development. The authors initially propose focusing on concrete nouns and utilizing semantic relationships such as synonymy, polysemy, hypernymy, and hyponymy. The linguistic data should be collected from Vietnamese dictionaries and actual language usage by Vietnamese speakers. To ensure comprehensiveness and practicality, the hierarchical structure of the initial semantic tagset system should consist of approximately 4–6 levels, organized top-down according to conceptual hierarchy. This structure reflects how humans mentally associate words, starting with broad parent concepts that branch into more specific child concepts, and further divide into detailed sub-concepts. This approach creates a network connecting vocabulary in terms of semantics.

6. Conclusion

This study examined the semantic relationships of English nouns across three representative semantic tagset systems: LLOCE, UCREL, and WordNet. Using a descriptive research method and a comparative and contrastive method, the research analyzed semantic relationships such as synonymy, polysemy, hypernymy, and hyponymy within each system. The findings reveal that LLOCE focuses on organizing vocabulary thematically, providing detailed presentations of polysemy and hypernym-hyponym relationships. However, its hierarchical structure is relatively simple and shallow, with limited emphasis on synonymy. UCREL, inherits from LLOCE, focusing on automatic semantic analysis and mapping labels to WordNet to expand synonymy relationships. WordNet organizes words based on conceptual meaning, detailing various semantic relationships. These findings offer a solid theoretical foundation for developing a similar semantic tagset system for Vietnamese nouns. Such a system could support applications in natural language processing, including semantic distance measurement, information extraction, and the research of lexical semantics.

REFERENCES

Vietnamese

1. Đinh Điền (2018), *Sách chuyên khảo ngôn ngữ học ngữ liệu*. Nxb Đại học Quốc gia Thành phố Hồ Chí Minh.
2. Lê Quang Thiêm (2018), *Sự phát triển nghĩa từ vựng tiếng Việt*. Nxb Đại học Quốc gia Hà Nội.

English

3. Archer, D. E., Rayson, P., Piao, S., & McEnery, A. M. (2004), Comparing the UCREL semantic annotation scheme with lexicographical taxonomies. In *Proceedings of the EURALEX-2004 Conference* (pp. 817-827). European Association for Lexicography.
4. Cruse, D. A. (1986), *Lexical semantics*. Cambridge University Press.
5. Geeraerts, D. (2010), *Theories of lexical semantics*. Oxford: Oxford University Press.
6. Lyons, J. (1995), *Linguistic semantics: An introduction*. Cambridge University Press.
7. McArthur, T. (1988), *Longman Lexicon of Contemporary English*. London: Longman.
8. McArthur, T. (1997), *Longman Lexicon of Contemporary English (English - Chinese)*. Longman Asia Limited and Shanghai Foreign Language Education Press.
9. McArthur, T. (1997), *Longman Lexicon of Contemporary English (English - Vietnamese)*. Vietnam Education Publishing House Limited Company.
10. Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990), Introduction to WordNet: An on-line lexical database. *International journal of lexicography*, 3(4), 235-244.
11. Piao, S. S., et al. (2016), Lexical coverage evaluation of large-scale multilingual semantic lexicons for twelve languages. Available from http://www.lrec-conf.org/proceedings/lrec2016/pdf/257_Paper.pdf
12. Fellbaum, C. (2005), WordNet and wordnets. In: Brown, Keith et al. (eds.), *Encyclopedia of Language and Linguistics, Second Edition*, Oxford: Elsevier, 665-670. Available from <https://wordnet.princeton.edu/>
13. Rayson, P., Archer, D., Piao, S., & McEnery, A. M. (2004), The UCREL semantic analysis system. Available from https://www.lancaster.ac.uk/staff/rayson/publications/usas_lrec04ws.pdf
14. Ungerer, F., & Schmid, H. J. (2006), *An introduction to cognitive linguistics*. Pearson Longman.
15. Vicente, A., & Falkum, I. L. (2017), *Polysemy*. In Oxford research encyclopedia of linguistics.

Khảo sát mối quan hệ ngữ nghĩa trên các hệ thống nhãn ngữ nghĩa danh từ Tiếng Anh

Tóm tắt: Hệ thống nhãn ngữ nghĩa là nền tảng cho các nghiên cứu ngữ nghĩa học tính toán. Mỗi hệ thống nhãn ngữ nghĩa chứa đựng các từ loại khác nhau và các từ loại đó có những mối quan hệ ngữ nghĩa khác nhau. Trong nghiên cứu này, chúng tôi khảo sát các mối quan hệ ngữ nghĩa của danh từ tiếng Anh qua ba hệ thống nhãn ngữ nghĩa tiêu biểu: LLOCE, UCREL và WordNet. Bằng phương pháp mô tả và so sánh đối chiếu, nghiên cứu khám phá các mối quan hệ ngữ nghĩa làm nền tảng cho ba hệ thống nhãn ngữ nghĩa nói trên, đồng thời làm rõ điểm tương đồng và dị biệt giữa chúng. Kết quả nghiên cứu giúp xác định các ưu điểm và hạn chế riêng của từng hệ thống, hiểu rõ hơn về các mối quan hệ ngữ nghĩa của danh từ trong các hệ thống nhãn ngữ nghĩa tiếng Anh và từ đó đề xuất mô hình tương tự cho các mối quan hệ ngữ nghĩa của danh từ trong tiếng Việt. Những phát hiện này sẽ cung cấp cơ sở lý thuyết vững chắc cho việc xây dựng các hệ thống nhãn ngữ nghĩa danh từ tiếng Việt, góp phần vào ứng dụng thực tế trong xử lý ngữ nghĩa tự động và trong giảng dạy-nghiên cứu ngữ nghĩa học từ vựng.

Từ khóa: quan hệ ngữ nghĩa; hệ thống nhãn ngữ nghĩa; LLOCE, UCREL; WordNet.