

УДК 81'33:811.161.2'373

DOI <https://doi.org/10.32447/2663-340X-2024-16.6>

УКРАЇНСЬКИЙ ВОРДНЕТ ЯК ІНСТРУМЕНТ ІНТЕГРАЦІЇ УКРАЇНСЬКОЇ МОВИ У СВІТОВИЙ МОВНИЙ ПРОСТІР

Вакуленко Максим Олегович

доктор філологічних наук, науковий співробітник
лабораторії мовних технологій факультету інформатики
Дармиштадтського університету прикладних наук
Шефферштрассе, 3, Дармиштадт, Німеччина
<https://orcid.org/0000-0003-0772-7950>

У цій статті розглянуто основні проблеми, пов'язані з процесами створення, розширення та редагування Українету – українського ворднету. Як представник родини ворднетів, Українет – це сучасний віртуальний мовний ресурс, який містить поняття (концепти), виражені рядами лексичних синонімів (синсетів) разом з їхніми означеннями, прикладами вживання та зазначенням лексико-семантичних відношень між поняттями – гіперонімія / гіпонімія, меронімія / голонімія, антонімія. Такі ресурси забезпечують важливу інформацію для різноманітних процесів опрацювання природної мови – машинного перекладу, перефразування, семантичного аналізу тексту тощо. Реальні й близькі перспективи включення української мови в коло мов Євросоюзу робить такий ресурс украї важливим. Тож створення Українету, яке стало можливим завдяки співпраці українських і німецьких фахівців, поставило Україну в один ряд із передовими в технологічному сенсі державами. Особливу увагу було приділено узгодженню лексикографічної інформації, що походить із різних джерел – таких як традиційні словники, Відкритий англійський ворднет (Open English WordNet), польський ворднет та інші ворднети. Було проаналізовано різні випадки синонімії – зокрема, лексичної (як різновиду формальної) та змістової синонімії – а також гіпо-гіперонімії, наявні у Відкритому англійському ворднеті та в звичайних термінологічних фізичних і фонетичних словниках. Було зіставлено також означення термінів у цих джерелах. Було виявлено, що традиційні словники не містять усіх необхідних гіпо-гіперонімічних відношень, а означення слів у ворднетах часто є неточними або неповними. Тому ворднети потребують істотного лексикографічного доопрацювання. Ця процедура передбачає включення додаткових лексичних одиниць для заповнення семантичних лакун, уточнення означень, а також введення нових понять і семантичних зв'язків, зокрема гіпо- та гіперонімічних, які пов'язані з уточненими означеннями. Ворднети є майбутнім лексикографії.

Ключові слова: ворднет, концепт, мовні ресурси, цифрова лексикографія, семантичні відношення.

Постановка проблеми в загальному вигляді та обґрунтування її актуальності. Швидкий розвиток ворднетів (від англ. *word net*, букв. *словосіть*) спричинив справжній прорив у новітній лексикографії, зробивши доступними для опрацювання та використання величезні масиви лінгвістичних ресурсів (Fellbaum, 1998; Bond et al., 2016; Siegel, Vakulenko, and Baum, 2023). Ворднети містять поняття (концепти), виражені рядами лексичних синонімів (синсетів) разом з їхніми означеннями, прикладами вживання та зазначенням лексико-семантичних відношень між поняттями – гіперонімія / гіпонімія, меронімія / голонімія, антонімія. Загалом існує два підходи до побудови ворднетів: **розширення** (англ. *expansion*) та **злиття** (англ. *merge*), де перший базується на перекладі наявних в іншому ворднеті синсетів, а другий передбачає побудову семантичних структур у цільовій мові з подальшим зв'язуванням концептів у різних мовах за допомогою так званих **міжмовних індексів** (англ. *interlingual indices*,

ILIs) (Bond et al., 2016, сс. 50–51). Отже, розробка ворднетів на основі злиття та комбінованого підходів (поєднання розширення та злиття) потребує узгодженого використання лексикографічної інформації, наявної у ворднетах, головним чином у Відкритому англійському ворднеті (<https://en-word.net/>) – англ. *Open English Wordnet, OEWN*) – і в традиційних словниках, які містять лексичні одиниці та їхні означення.

Аналіз останніх досліджень і публікацій. Першу версію української словосіті – *Ukrainet 1.0* – було створено на основі злиття (Siegel, Vakulenko, and Baum, 2023) з використанням «Тлумачного словника з фізики» (Вакуленко, Вакуленко, 2008) та «Словника фізичних термінів-синонімів» (Вакуленко, Вакуленко, 2017). Досягнення цієї мети поставило Україну поруч із провідними технологічними державами стосовно семантичних мовних ресурсів для автоматичного опрацювання мови. Зазначимо, що попередні спроби створити

український ворднет (Кульчицький, Романюк, Харів, 2010; Anisimov et al., 2013) виявилися безуспішними – очевидно, через брак кваліфікованих лексикографів. У ході подальшої роботи над Українетом було додано близько 750 означень понять і 700 прикладів використання, а також доповнено цей ресурс перекладом польського ворднету (Rudnicka et al., 2019) українською мовою. Означення взято зі «Словника української мови в 20-ти томах» (Русанівський, 2010), «Тлумачного словника з фізики» (Вакуленко, Вакуленко, 2008) та Відкритого англійського ворднету. Поточна версія Українету містить понад 15 тисяч лексичних гасел, виражених простими лексемами та , які об'єднані в понад 12 тисяч синонімічних рядів.

На рис. 1 зображено синсет Українету, який містить лексему «пастишок», із зазначенням його лексико-семантичних зв'язків з іншими синонімічними рядами. На рис. 2 зображено лексему «вага» разом зі синсетами, в які вона входить.

Лінгвістичні дані, отримані з різних джерел, становлять «особливий виклик для анотаторів» (Rudnicka et al., 2019, с. 310), потребуючи копіткого доопрацювання та корегування. Найбільш проблематичними тут є лексичні одиниці, що входять у відповідні лексико-семантичні відношення, та їхні означення. Для зв'язування ворднетів найважливішими відношеннями є синонімія та гіперонімія, де найбільшу кількість еквівалентів у синсетах спричиняє абсо-

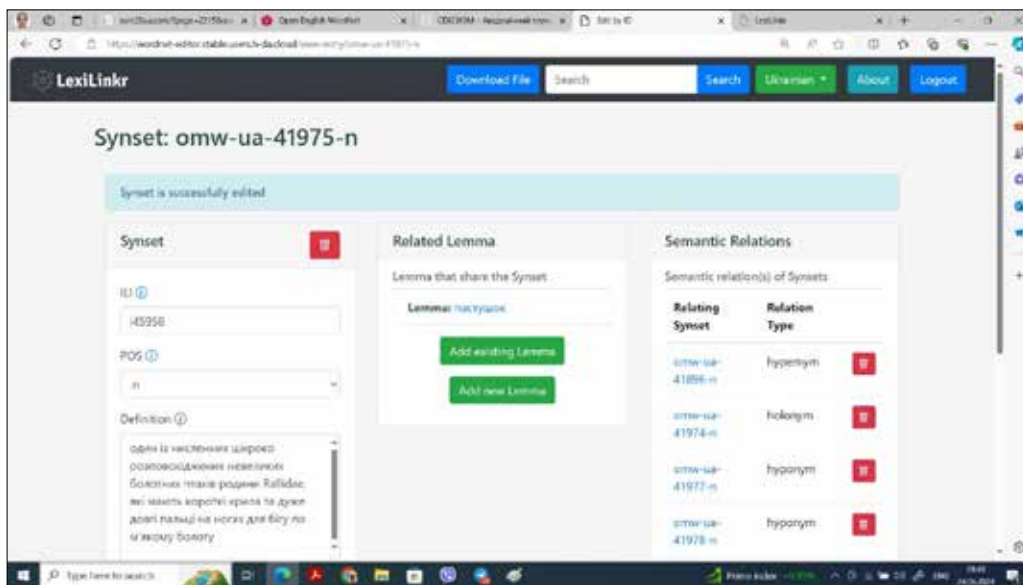


Рис. 1. Синсет OMW-ua-41975-n в Українеті, який містить лексему «пастишок»

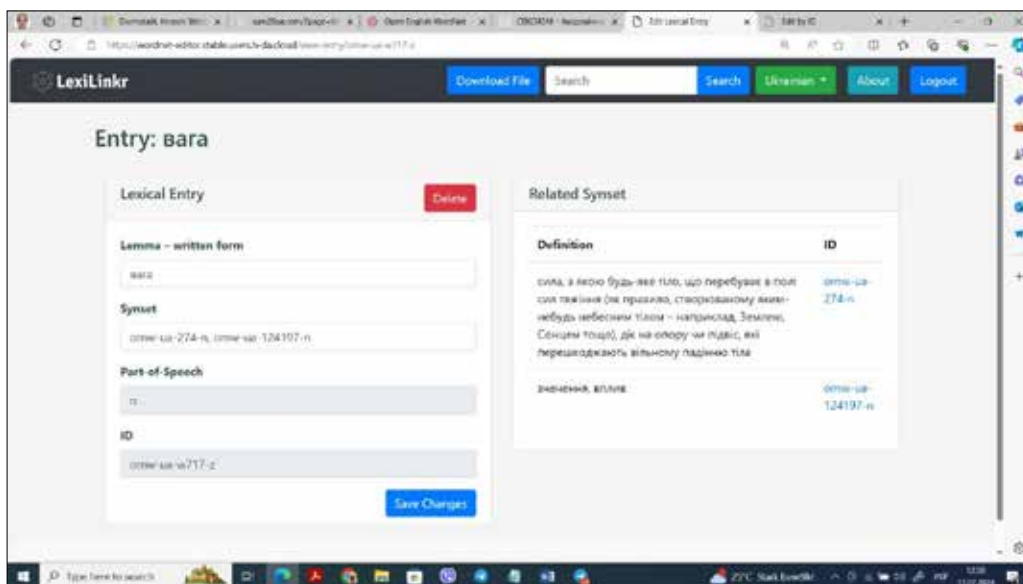


Рис. 2. Лексема Українету «вага» з відповідними синсетами означеннями

лютна синонімія (Rudnicka et al., 2019). Отже, синонімія відіграє ключову роль у створенні ворднетів.

Істотною проблемою розвитку ворднетів, зокрема Українету, є координація та корегування лексикографічних даних, які походять із різних джерел.

Формулювання мети і завдань статті.

Мета роботи полягає в тому, щоб дослідити суперечності в лексикографічній інформації для розвитку Українету, яка походить із інших ворднетів та традиційних словників, і визначити шляхи усунення цих суперечностей.

Завдання роботи полягають у тому, щоб:

– виявити й проаналізувати приклади синонімії в термінологічних словниках; порівняти семантику відповідних гіперонімів;

– порівняти означення у Відкритому англійському ворднеті з означеннями з традиційних словників;

– сформулювати рекомендації щодо використання у ворднетах лексикографічної інформації, яка надходить із різних джерел.

Виклад основного матеріалу дослідження.

Синонімія та гіпо-гіперонімія. Постання лексичної синонімії (Вакуленко 2023, сс. 88–90) зумовлене певним ступенем семантичної близькості понять, гіпоніми до яких мають тотожні значення:

– англ. *Doppler shift* ‘доплерівський зсув’ – *Doppler displacement* ‘доплерівське зміщення’ – *Doppler frequency shift* ‘доплерівський зсув частоти’ (Sube, Eisenreich, 1973, с. 375);

– англ. *initiation of the discharge* ‘ініціація розряду’ – *ignition of the discharge* ‘запалювання розряду’ (Sube, Eisenreich, 1973, с. 726);

– англ. *level of significance* ‘рівень значущості’ – *significance level* ‘значущісний рівень’ – *error probability* ‘імовірність похибки’ (Sube, Eisenreich, 1973, с. 838);

– англ. *level of significance* ‘рівень значущості’ – *confidence level* ‘довірчий рівень’ – *degree of confidence* ‘ступінь довіри’ – *confidence probability* ‘довірча ймовірність’ – *confidence coefficient* ‘довірчий коефіцієнт’ – *error probability* ‘імовірність похибки’ (Sube, Eisenreich, 1973, с. 838);

– англ. *contact clearance* ‘контактний зазор’ – *contact gap* ‘контактний проміжок’ – *contact spacing* ‘контактна відстань’ (Greulich, Meenenga, 1997, с. 594);

– англ. *monitor* ‘монітор’ – *monitoring instrument* ‘моніторинговий пристрій’ – *survey instrument* ‘оглядовий пристрій’ – *survey meter*

‘оглядовий вимірювач’ (Greulich, Meenenga, 1997, с. 594);

– англ. *sighting apparatus* ‘прицільний апарат’ – *sighting device* ‘прицільний пристрій’ (Greulich, Meenenga, 1997, с. 594);

– англ. *magnetometer* ‘магнітометр’ – *magnetic meter* ‘магнітний вимірювач’ – *magnetic-field meter* ‘вимірювач магнітного поля’ – *magnetic detector* ‘магнітний детектор’ – *gaussmeter* ‘гаусметр’ (Вакуленко, Вакуленко, 2008, с. 296–297);

– англ. *diminuendo accent* – *falling accent* – *diminuendo stress* – *falling stress* ‘спадний наголос’ (Vakulenko, 2018, с. 53);

– нім. *Polschuhabstand* *m* ‘полюсна відстань’ – *Luftspaltslänge* *f* ‘довжина повітряного проміжку’ – *Luftspaltsbreite* *f* ‘ширина повітряного проміжку’ (Sube, Eisenreich, 1973, с. 570);

– нім. *Visiervorrichtung* *f* ‘візирувальний пристрій’ – *Zielgerät* *n* ‘прицільний пристрій’ (Greulich, Meenenga, 1997, с. 594);

– нім. *statistische Sicherheit* ‘статистична достовірність’ – *Sicherheitsschwelle* *f* ‘безпечний поріг’ – *Sicherheitsgrad* *m* ‘довірчий ступінь’ – *Vertrauenskoeffizient* *m* ‘довірчий коефіцієнт’ – *Konfidenzkoeffizient* *m* ‘довірчий коефіцієнт’ – *Konfidenzniveau* *n* – ‘довірчий рівень’ – *Konfidenzwahrscheinlichkeit* *f* ‘довірча ймовірність’ – *Bedeutsamkeitstufe* *f* ‘значущий рівень’ (Sube, Eisenreich, 1973, с. 838);

– нім. *spiegelnder Körper* ‘відбивальне тіло’ – *weisses Objekt* ‘білий об’єкт’ (Greulich, Meenenga, 1997, с. 306);

– *оптичний ізомер* – *оптичний антипод* (Вакуленко, Вакуленко, 2008, с. 165);

– *світлопроменевий осцилограф* – *світлопроменевий прилад* (Вакуленко, Вакуленко, 2008, с. 165);

– *аргумент* – *незалежна змінна* (Вакуленко, Вакуленко, 2017, с. 13; Vakulenko, 2018, с. 41);

– *апроксимант* – *безтертьовий континуант* (Vakulenko, 2018, с. 53).

У цих прикладах ми бачимо чітку семантичну кореляцію між головними термінами, включаючи гіперонімію / гіпонімію (*shift* – *frequency shift*, *instrument* – *monitor*, *instrument* – *meter*, *weight* – *density*, *meter* – *detector*, *apparatus* – *device*, *прилад* – *осцилограф*, *змінна* – *аргумент*, *апроксимант* – *континуант*) і синонімію (*shift* – *displacement*, *level* – *degree*, *accent* – *stress*). Звідси видно, що необхідною умовою синонімії двох лексем є наявність у відповідних понять гіпонімів, які входять в один синсет. Отже, еквівалентність гіпонімів зумовлена синонімією їхніх гіперонімів. Як можна бачити з синонімічної пари *доплерівський зсув* – *дон-*

плерівський **зсув частоти**, протилежне правильне тоді, коли гіперонім (*зсув*) має додаткове значення, тотожне значенню його гіпоніма (*зсув частоти*).

Граматична варіативність також часто стає причиною синонімії, насамперед для словосполучень. Прикладом цього є такий ряд синонімів: *unshared electron pair* ‘неподілена електронна пара’ – *electron lone pair* ‘електронна самотня пара’ – *lone pair* ‘самотня пара’ – *lone-pair electrons* ‘самотня пара електронів’ (Вакуленко, Вакуленко, 2008, с. 394). Тут елемент *pair*, який є гіперонімом до *electron lone pair*, можна розглядати як гіперонім до лексеми *electrons*, яка є гіперонімом до *lone-pair electrons*.

Дещо інакшою є ситуація зі **змістовою синонімією**, яка виникає, коли різні терміни відображають різні аспекти поняття або різні підходи до його вивчення (Вакуленко 2023, с. 91; Vakulenko, 2018, с. 41). Цей різновид синонімії характерний і для термінів-словосполучень, тобто коли деякі семи їхніх гіперонімів нейтралізуються. У традиційних словниках відповідні головні терміни для членів синсету проявляють або незначну семантичну подібність, або складні та непрямі семантичні зв'язки:

– англ. *piezo effect* ‘п'єзоефект’ – *piezoelectric effect* ‘п'єзоелектричний ефект’ – *polarization of distortion* ‘поляризація дисторсії’ (Greulich, Meenenga, 1997, с. 594);

– нім. *schwarzer Körper* ‘чорне тіло’ – *schwarzer Strahler* ‘чорний випромінювач’ (Greulich, Meenenga, 1997, с. 22);

– *матриця густини* – *статистична матриця* – *статистичний оператор* (Вакуленко, Вакуленко, 2008, с. 305);

– *енергія Гіббса* – *термодинамічний потенціал Гіббса* – *ізобарно-ізотермічний потенціал* – *вільна ентальпія* (Вакуленко, Вакуленко, 2008, с. 166);

– *стала Планка* – *квант дії* (Вакуленко, Вакуленко, 2008, с. 608);

– *ізотопний індикатор* – *індикатор* – *індикатор ізотопів* – *мітка* – *мічений атом* (Вакуленко, Вакуленко, 2008, с. 225);

– *точка замерзання* – *температура замерзання* (Вакуленко, Вакуленко, 2017, с. 181; Vakulenko, 2018, с. 43).

Такі терміни ввійшли в традиційні словники завдяки їх практичному використанню, що з погляду термінологічної теорії є результатом застосування статистичного методу в симптоматичному сенсі (Вакуленко 2023, с. 123–135; Vakulenko, 2018, с. 14–19). Проте поява синонімії між словосполученнями передбачає синонімію

або гіпо- гіперонімію головних термінів, які описують відповідні поняття. Іншими словами, концептуальний підхід, на якому базується розробка ворднетів, вимагає введення додаткових значень для цих гніздових термінів, які не були явно відображені у звичайних словниках. Такий розвиток надбань лексикографічної практики свідчить про застосування аналітичного методу науки термінології (Вакуленко 2023, с. 123–135; Vakulenko, 2018, с. 14–19).

Існування подібних корельованих і недостатньо уточнених понять було помічено і при безпосередній роботі над Українетом. Зокрема, синсет **omw-1354-n** містить лексеми «НВЧ-генератор» і «лампа зворотної хвилі», які, згідно з «Тлумачним словником із фізики» (Вакуленко, Вакуленко, 2008), належать до гнізд різних термінів: «генератор» (**omw-124123-n**) і «лампа» (**omw-1617-n**) відповідно. Синсет **omw-689-n** (i61958) містить лексеми «гомоморфізм» і «відображення гомоморфне», які, згідно з «Тлумачним словником із фізики» (Вакуленко, Вакуленко, 2008), також входять у різні термінні гнізда: «схожість» (**omw-100-n**) і «відображення» (**omw-54804-n**). У таких випадках ми додаємо відсутні значення до відповідних гніздоутворювальних термінів.

Словникові означення та означення Відкритого англійського ворднету. Ступінь семантичної схожості між гіперонімами до вищезгаданих синсетів обчислити нелегко, оскільки вона дуже залежить від відповідних означень (Vakulenko, 2019). Наприклад, Відкритий англійський ворднет трактує *енергію* як “a thermodynamic quantity equivalent to the capacity of a physical system to do work” ‘термодинамічна величина, еквівалентна спроможності фізичної системи виконувати роботу’, таким чином пов’язуючи це поняття з гіперонімом «величина» (<https://en-word.net/lemma/energy>). Своєю чергою, в «Тлумачному словнику з фізики» (Вакуленко, Вакуленко, 2008) *енергія* означається як «загальна міра різноманітних процесів і типів взаємодії, що дозволяє вимірювати всі види фізичних форм руху та взаємодії», що відносить це поняття до гіпероніма «міра». Ми рекомендуємо використовувати друге. Щобільше, деякі означення слів, доступні у Відкритому англійському ворднеті, не можуть коректно передавати важливі особливості відповідних понять, зокрема поняття з такими міжмовними індексами (ILI).

Fluorescence ‘флуоресценція’ (i97800): “light emitted during absorption of radiation of some other (invisible) wavelength” ‘світло, що випромінюється під час поглинання

випромінювання іншої (невидимої) довжини хвилі». Це означення не виділяє основних характеристик флуоресценції і тому неповне. У «Тлумачному словнику з фізики» (Вакуленко, Вакуленко, 2008) подано означення, яке відносить флуоресценцію до люмінесценції та описує її істотні ознаки і тому може бути рекомендоване: «короткочасна люмінесценція з часом загасання 10^{-8} – 10^{-9} сек».

Luminescence 'люмінесценція' (i97901): "light not due to incandescence; occurs at low temperatures" 'світло не внаслідок розжарювання; відбувається при низьких температурах'. Цей опис не стосується люмінесценції. «Словник сучасної англійської мови Лонгмана» (Longman, 2003) також не підкреслює основні властивості люмінесценції, означуючи це явище як "a soft shining light" 'м'яке сяюче світло'. «Тлумачний словник із фізики» (Вакуленко, Вакуленко, 2008) трактує люмінесценцію як «випромінювання, що являє собою надлишок над тепловим випромінюванням тіла і продовжується протягом часу, який значно перевищує період світлових коливань», що ми рекомендуємо для ворднетів.

Metal 'метал' (i113678): "a mixture containing two or more metallic elements or metallic and nonmetallic elements usually fused together or dissolving into each other when molten" 'суміш, що містить два або більше металічних елементів або металічні та неметалічні елементи, які зазвичай злиті разом або розчиняються один в одному під час розплавлення'. Це означення некоректне, оскільки не розрізняє сплави та метали. «Словник сучасної англійської мови Лонгмана» (Longman, 2003) пропонує більш релевантне, але дещо спрощене означення на основі характеристик металу: "a hard, usually shiny substance such as iron, gold, or steel" 'тверда, зазвичай блискуча речовина, така як залізо, золото або сталь'. Більш чітке означення є в «Тлумачному словнику з фізики» (Вакуленко, Вакуленко, 2008): «конденсований стан речовини (тверде тіло, рідина), побудований з атомів у хімічному розумінні, тобто з атомів, які легко віддають електрони в процесі хімічних реакцій» – яке ми й рекомендуємо.

Paramagnetism 'парамагнетизм' (i97922): "materials like aluminum or platinum become magnetized in a magnetic field but it disappears when the field is removed" 'такі матеріали, як алюміній або платина, намагнічуються в магнітному полі, але воно зникає, коли поле прибирають'. Це означення є некоректним, оскільки парамагнетизм є властивістю, а не матеріалом. «Тлумачний словник із фізики»

(Вакуленко, Вакуленко, 2008) містить означення, яке слід використовувати замість цього: «властивість речовин (парамагнетиків) намагнічуватися в напрямку зовнішнього магнітного поля».

Resistance 'опір', *impedance* 'імпеданс', *ohmic resistance* 'омічний опір', *resistivity* 'опірність', *electric resistance* 'електричний опір', *electrical resistance* 'електричний опір' (i98048): "a material's opposition to the flow of electric current; measured in ohms" 'протидія матеріалу потоку електричного струму; вимірюється в Омах'. Це означення не розрізняє сталий і змінний струми, що має тут вирішальне значення. Насправді імпеданс хоч і є аналогом опору, але відповідає іншому поняттю. «Словник сучасної англійської мови Лонгмана» (Longman, 2003) означає імпеданс як "a measure of the power of a piece of electrical equipment to stop the flow of an alternating current" 'міру потужності елемента електричного обладнання, необхідної для зупинки потоку змінного струму'. «Тлумачний словник із фізики» (Вакуленко, Вакуленко, 2008) дає більш точне означення імпедансу: «аналог електричного опору для гармонічних процесів» – яке ми рекомендуємо.

Spectrometry 'спектрометрія', *spectroscopy* 'спектроскопія' (i38839): "the use of spectrometers to analyze spectra" 'використання спектрометрів для аналізу спектрів'. Це означення є неповним і не розрізняє спектрометрію та спектроскопію, які не є тотожними поняттями. «Тлумачний словник із фізики» (Вакуленко, Вакуленко, 2008) означає спектрометрію як «сукупність методів і теорія вимірювання спектрів електромагнітного випромінювання та вивчення спектральних властивостей речовин і твердих тіл в оптичному діапазоні довжин хвиль» і пропонує два значення для поняття спектроскопія: 1 «область фізики, присвячена дослідженню розподілу інтенсивності електромагнітного випромінювання за довжинами хвиль або частотами»; 2 «дослідження розкладів у спектр». Тож ці поняття слід розглядати у ворднетах відповідним чином.

Запропоновані поправки до означень спричинять відповідно зміни у відношеннях гіпо- та гіперонімії, які мають коректно відображатися у ворднетах.

Обговорення. Глибинна причина невідповідностей між поняттями та їхніми словесними позначеннями полягає в явищі асиметричного дуалізму мовного знака, яке є фундаментальною мовною властивістю. Асиметричний дуалізм мовного знака означає, що значення слова (*signifié*) і його словесне вира-

ження (*signifiant*) перебувають у нестійкій рівновазі та демонструють динамічну поведінку: перше має тенденцію виражатися іншими словоформами, таким чином породжуючи синонімію, а друге має тенденцію набувати додаткових значень (див. Karcevskij, 1929). Цю безперервну появу додаткових значень неможливо відстежити у традиційній лексикографії належним чином.

Крістіан Фелльбаум зауважує, що у ворднетах існують члени синсету, які не можуть бути замінені в усіх контекстах (Fellbaum, 2006, с. 665). Проте якщо розглядати синонімію як явище, яке формується на більш високому, контекстуальному рівні, то можна прийти до висновку, що лексичні одиниці, які неможливо замінити у всіх контекстах, повинні належати до різних синсетів. Тому варто дотриму-

ватися означення синонімів, яке передбачає їхню взаємозамінність у контексті (Вакуленко, 2023, с. 82).

Висновки та рекомендації для подальших досліджень. У цій статті ми проаналізували деякі проблеми розвитку українського ворднету, пов'язані з корегуванням лексем та їхніх означень, узятих із різних джерел. Було показано, що ворднети, які є результатом «злиття» та комбінованих підходів, як-от Українет, потребують подальшого узгодження лексикографічної інформації, зокрема введення додаткових лексичних одиниць для понять з відсутніми назвами, редагування означень, а також уведення нових понять і семантичних структур (переважно гіпо- та гіперонімії) відповідно до змінених означень.

Ворднети є майбутнім лексикографії.

ЛІТЕРАТУРА

1. Fellbaum C. WordNet: An Electronic Lexical Database. Cambridge, MA: MIT Press, 1998. 422 p.
2. Bond F., Vossen P., McCrae J., Fellbaum C. CILI: the Collaborative Interlingual Index. *8th Global WordNet Conference (GWC): proceedings* (Bucharest, Romania, 27–30 January, 2016). Editors: Christiane Fellbaum, Piek Vossen, Verginica Barbu Mititelu, Corina Forascu. Bucharest, Romania, 2016. P. 50–57. URL: <https://aclanthology.org/2016.gwc-1.9>.
3. Siegel M., Vakulenko M., Baum J. Towards Ukrainian WordNet: Incorporation of an Existing Thesaurus in the Domain of Physics. *19th Conference on Natural Language Processing (KONVENS 2023): proceedings* (Ingolstadt, Germany, 19–21 September 2023). Ingolstadt, Germany, 2023. P. 121–126. URL: <https://aclanthology.org/2023.konvens-main.12/>.
4. Вакуленко М. О., Вакуленко О. В. Глумачний словник із фізики: [6644 статті]. Київ: ВПЦ «Київський університет», 2008. 767 с.
5. Вакуленко М. О., Вакуленко О. В. Словник фізичних термінів-синонімів. Київ: ВПЦ «Київський університет», 2017. 191 с.
6. Кульчицький І. М., Романюк А. Б., Харів, К. Б. Розроблення Wordnet-подібного словника української мови. *Вісник Національного університету Львівська політехніка. Інформаційні системи та мережі*. 2010. № 673. С. 306–318.
7. Anisimov A., Marchenko O., Nikonenko A., Porkhun O., Taranukha, V. (2013). Ukrainian wordnet: creation and filling. *Flexible Query Answering Systems (FQAS 2013): proceedings of the 10th International Conference* (Granada, Spain, 18–20 September, 2013). P. 649–660. Granada, Spain, 2013. Springer.
8. Rudnicka E., Piasecki M., Bond F., Grabowski Ł., Piotrowski T. Sense Equivalence in plWordNet to Princeton WordNet Mapping. *International Journal of Lexicography*. 2019. Vol. 32 (3). P. 296–325. DOI: 10.1093/ijl/ecz004.
9. Словник української мови: в 20 томах. / ред. Віталій Русанівський. Київ: Наукова думка, 2010.
10. Вакуленко М. О. Сучасна українська термінологія: методологія, кодифікація, лексикографічна практика. Рукопис. Дис. на здобуття наук. ступеня док. філол. наук: спец. 10.02.01 – українська мова. Київський національний університет імені Тараса Шевченка, Міністерство освіти і науки України. Київ, 2023. 445 с. DOI: 10.13140/RG.2.2.34645.81122.
11. Sube R., Eisenreich G. Wörterbuch Physik. A-M. Englisch-Deutsch-Französisch-Russisch. Zürich und Frankfurt am Main: Verlag Harri Deutsch, 1973. 1020 S.
12. Greulich W., Meenenga D. Wörterbuch der Physik: Englisch-Deutsch/Deutsch-Englisch. Heidelberg; Berlin: Spektrum, Akad. Verlag, 1997. 840 S.
13. Vakulenko M. From Terminology-Vocabulary to Terminology-Science: A Ukrainian Trend. Beau Bassin: LAP, 2018. 120 p.
14. Vakulenko M. Calculation of Semantic Distances between Words: From Synonymy to Antonymy. *Journal of Quantitative Linguistics*. 2019. Volume 26(2). P. 116–128. URL: <https://doi.org/10.1080/09296174.2018.1452524>.
15. Longman Dictionary of Contemporary English / director Della Summers. Harlow, England: Pearson Education Limited, 2003. 1951 p.

16. Karcevskij S. Du dualisme asymétrique du signe linguistique [On the asymmetric dualism of the linguistic sign]. *Travaux du Cercle linguistique de Prague*. 1929. Volume 1. P. 88–92. URL: <https://www.degruyter.com/document/doi/10.7560/780439-004/html>.

17. Fellbaum C. WordNet(s). *Encyclopedia of Language & Linguistics, Second Edition* / Editor-in-Chief Keith Brown. 2006. Volume 13. P. 665–670. Oxford: Elsevier.

REFERENCES

1. Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. MIT Press.
2. Bond, F., Vossen, P., McCrae, J., & Fellbaum, C. (2016). CILI: the Collaborative Interlingual Index. *Proceedings of the 8th Global WordNet Conference (GWC)*, 50–57. Bucharest, Romania, 27–30 January, 2016. Editors: Christiane Fellbaum, Piek Vossen, Verginica Barbu Mititelu, Corina Forascu. URL: <https://aclanthology.org/2016.gwc-1.9>.
3. Siegel, M., Vakulenko, M., & Baum, J. (2023). Towards Ukrainian WordNet: Incorporation of an Existing Thesaurus in the Domain of Physics. *Proceedings of the 19th Conference on Natural Language Processing (KONVENS 2023)*, 121–126. 19–21 September 2023, Ingolstadt, Germany. URL: <https://aclanthology.org/2023.konvens-main.12/>.
4. Vakulenko, M. O., & Vakulenko, O. V. (2008). *Tlumachnyj slovnyk iz fizyky: [6644 staty]* (Explanatory dictionary on physics: [6644 articles]). Kyjiv: Vydavnycho-poligrafichnyj centr "Kyjivskij universytet". 767 p. [in Ukrainian].
5. Vakulenko, M. O., & Vakulenko, O. V. (2017). *Slovnyk fizychnykh terminiv-synonimiv* [Dictionary of physical synonymous terms]. Kyjiv: Vydavnycho-poligrafichnyj centr "Kyjivskij universytet". 191 p. [in Ukrainian].
6. Kuljchyc'kij, I. M., Romanjuk, A. B., & Khariv, K. B. (2010). Rozroblennja WordNet-podibnogo slovnyka ukrajins'kogoji movy [Development of a WordNet-like dictionary for the Ukrainian language]. *Visnyk Nacional'nogho universytetu "L'vivs'ka politekhnika"*. *Informacijni systemy ta merezhi* 673, 306–318 [in Ukrainian].
7. Anisimov, A., Marchenko, O., Nikonenko, A., Porkhun, O., & Taranukha, V. (2013). Ukrainian wordnet: creation and filling. *Proceedings of the 10th International Conference "Flexible Query Answering Systems" (FQAS 2013)*, 649–660. Granada, Spain, September 18–20, 2013. Springer.
8. Rudnicka, E., Piasecki, M., Bond, F., Grabowski, Ł., & Piotrowski, T. (2019). Sense Equivalence in plWordNet to Princeton WordNet Mapping. *International Journal of Lexicography* 32 (3), 296–325. DOI: 10.1093/ijl/ecz004.
9. Rusanivskij, V. (Ed.). *Slovnyk ukrajins'kogoji movy v 20 tomakh* [Dictionary of Ukrainian language in 20 volumes]. Kyjiv, 2010: Naukova dumka. [in Ukrainian].
10. Vakulenko M. O. *Modern Ukrainian Terminology: Methodology, Codification, and Lexicographic Practice*. Manuscript. Dissertation for a Doctoral Degree in Philology, Specialty 10.02.01 – Ukrainian Language. Kyiv National University after Taras Shevchenko. Kyjiv, 2023. 445 p. [in Ukrainian]. DOI: 10.13140/RG.2.2.34645.81122..
11. Sube, R., & Eisenreich, G. (1973). *Wörterbuch Physik. A-M. Englisch-Deutsch-Französisch-Russisch*. Verlag Harri Deutsch – Zürich und Frankfurt am Main.
12. Greulich, W., & Meenenga, D. (1997). *Wörterbuch der Physik: Englisch-Deutsch/Deutsch-Englisch*. Heidelberg; Berlin: Spektrum, Akad. Verlag.
13. Vakulenko, M. (2018). *From Terminology-Vocabulary to Terminology-Science: A Ukrainian Trend*. Beau Bassin: LAP. 120 p.
14. Vakulenko, M. (2019). Calculation of Semantic Distances between Words: From Synonymy to Antonymy. *Journal of Quantitative Linguistics* 26(2), 116–128. URL: <https://doi.org/10.1080/09296174.2018.1452524>.
15. *Longman Dictionary of Contemporary English*. (2003) / [director, Della Summers]. Pearson Education Limited.
16. Karcevskij, S. (1929). Du dualisme asymétrique du signe linguistique [On the asymmetric dualism of the linguistic sign]. *Travaux du Cercle linguistique de Prague* 1, 88–92. URL: <https://www.degruyter.com/document/doi/10.7560/780439-004/html>.
17. Fellbaum, C. (2006). *WordNet(s)*. In: Keith Brown, (Editor-in-Chief) *Encyclopedia of Language & Linguistics, Second Edition, Volume 13*, pp. 665–670. Oxford: Elsevier.

UKRAINIAN WORDNET AS A TOOL OF INTEGRATION THE UKRAINIAN LANGUAGE INTO THE GLOBAL LANGUAGE SPACE

Vakulenko Maksym Oleghovych

*Doctor of philological sciences, Researcher of the Faculty of Informatics,
Darmstadt University at Applied Sciences
3, Schoefferstrasse Str., Darmstadt, Germany*

This paper addresses several key challenges encountered in the ongoing development of the Ukrainian wordnet, known as Ukrajinet, focusing on integrating and aligning vocabularies derived from various sources. These sources include traditional dictionaries, the Open English WordNet, and other wordnets from different languages. Ukrajinet, like other wordnets, serves as a linguistic resource that organises concepts into a series of lexical synonyms known as synsets. These synsets are provided with their definitions, examples of usage, and detailed mapping of lexical-semantic relations, such as hypernymy (generalisation) and hyponymy (specification), metonymy (part-whole relations), holonymy (whole-part relations), and antonymy (oppositeness). These relations are critical for enhancing the capabilities of natural language processing. The paper explicitly examines various synonymy forms and the relations of hyponymy and hypernymy as they are presented in the Open English Wordnet and conventional terminological physical and phonetic dictionaries. A distinction is made between lexical synonymy (synonymy based on formal similarity of terms) and contextual synonymy (synonymy based on conceptual similarity). A detailed analysis of term definitions across different sources revealed that traditional dictionaries often lack the hierarchical relations of hyponyms and hypernyms, essential for the structured representation of knowledge in wordnets. Furthermore, wordnets frequently contain inaccuracies or incomplete information, particularly in the definitions of concepts. The findings underscore the need for lexicographic post-processing in creating and refining wordnets. This procedure involves several key actions: introducing missing lexical units for concepts not captured in the initial datasets, revising existing definitions to enhance clarity and accuracy, and incorporating new conceptual entities and semantic structures. These tasks are especially important in managing relations of hypo- and hypernymy, ensuring that they align with the updated definitions. This paper posits that wordnets represent the future of lexicography.

Key words: wordnet, concept, linguistic resource, digital lexicography, semantic relations.