https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2102 (Project)

# Lexicographic Post-processing and Refinement Issues in Wordnets: The Ukrainian Wordnet Experience

Maksym Vakulenko, *Darmstadt University of Applied Sciences, Darmstadt, Germany*
*(maxvakul@yahoo.com)*
*(https://orcid.org/0000-0003-0772-7950)*

**Abstract:** This paper considers challenges arising within the development of the Ukrainian wordnet, specifically focusing on the integration and alignment of vocabularies originating from diverse sources, such as conventional dictionaries, the Open English Wordnet, and other wordnets. Various cases of synonymy and hyper-hyponymy presented in the Open English Wordnet and in conventional terminological and phonetic dictionaries were analyzed, as well as the definitions of terms in these sources were compared. It was found that traditional dictionaries do not contain all the necessary hyper- and hyponymic relations, and definitions in wordnets are often inaccurate or incomplete. It is shown that creating wordnets based on the "merge" and combined approaches, exemplified by Ukrajinet, necessitates specific lexicographic post-processing. This procedure involves several crucial steps, including incorporating additional lexical units for lacking concepts, refining definitions, and introducing novel concepts and semantic structures, particularly those associated with hyper- and hyponymy, aligned with revised definitions.

**Keywords:** WORDNETS, UKRAJINET, CONCEPT, DEFINITION, LEXICAL-SEMANTIC RELATIONS, SYNONYMY, HYPER-HYPONYMY, LEXICOGRAPHIC DATA

**Opsomming: Leksikografiese postverwerking en verfyningskwessies in woordnette: Die Oekraïense woordnet-ervaring.** Hierdie bydrae beskou uitdagings wat binne die ontwikkeling van die Oekraïense woordnet (Ukrajinet) ontstaan, en fokus spesifiek op die integrasie en belyning van woordeskat wat afkomstig is van diverse bronne, soos konvensionele woordeboeke, die Open English Wordnet en ander woordnette. Verskeie gevalle van sinonimie en hiperohiponimie wat in die Open English Wordnet en in konvensionele terminologiese en fonetiese woordeboeke aangebied word, is geanaliseer. Die definisies van terme in hierdie bronne is ook met mekaar vergelyk. Daar is bevind dat tradisionele woordeboeke nie al die nodige hipero- en hiponimiese verwantskappe bevat nie, en definisies in woordnette is dikwels onakkuraat of onvolledig. Daar word aangetoon dat die skep van woordnette gebaseer op die "samesmeltings-" en gekombineerde benaderings, soos geïllustreer deur Ukrajinet, spesifieke leksikografiese postverwerking noodsaak. Hierdie prosedure behels verskeie belangrike stappe, insluitend die inkorporering van bykomende leksikale eenhede vir ontbrekende konsepte, die verfyning van definisies en die bywerking van nuwe konsepte en semantiese strukture, veral dié wat met hipero- en hiponimie geassosieer word sodat dit met hersiene definisies strook.

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2102 (Project)



**Sleutelwoorde:** WOORDNETTE, UKRAJINET, KONSEP, DEFINISIE, LEKSIKAAL-SEMANTIESE VERHOUDINGS, SINONIMIE, HIPERO-HIPONIMIE, LEKSIKOGRAFIESE DATA

## 1. Introduction

The rapid development of linguistic resources such as wordnets (Fellbaum 1998; Bond et al. 2016; McCrae et al. 2020; Siegel and Bond 2021) caused a breakthrough in lexicography, providing multilingual lexical databases containing definitions, use examples, lexical-semantic relations (LSRs), etc. Wordnets are useful in various natural language processing tasks, such as machine translation and semantic text analysis. In general, there are two approaches to developing wordnets: "expansion" and "merge", where the first is based on translations of existing lemma forms and the second foresees building semantic structures of the language in question (synonymy, hyper-/hyponymy, mero-/holonymy, and antonymy) with subsequent linking of concepts through interlingual indices — ILIs (Bond et al. 2016: 50-51). Then the development of wordnets based on the "merge" and combined ("expansion" and "merge") approaches necessitates the joined use of lexicographic information available in existing wordnets, mainly Open English Wordnet (OEWN)[1], and traditional dictionaries containing lexical units and their definitions.

In particular, the first version of the Ukrainian wordnet, Ukrajinet 1.0[2], was created based on the merge approach, making use of the "Explanatory Dictionary on Physics" (TSF)[3] and the "Dictionary of Physical Synonymous Terms" (SFS)[4]. As the follow-up of this project, the 750 concept definitions and 700 examples of use have been added so far, and this resource was supplemented with a translation of OEWN into Ukrainian. The definitions were taken from the "Dictionary of Ukrainian Language in 20 volumes" (SUM)[5], TSF, and OEWN. The current version of Ukrajinet contains 153 683 lexical entries (single and multi-word lexemes) combined in 124 214 synsets.

The linguistic data coming from different sources pose "a particular challenge for annotators" (Rudnicka et al. 2019: 310), requiring tedious refinement and adjustment. Extremely difficult is to find corresponding concepts in different languages where some of them may be absent, and others may differ in their scope that requires thorough analyzing definitions and use examples. For example, the concept i29195 "put" — 'cause to be in a certain state; cause to be in a certain relation' — corresponds to several Ukrainian concepts, in particular: 1 *викликати [в когось настрій/стан]* 'put [someone in a mood/state]'; 2 *оформляти* 'put [something] in form'; 3 *упорядковувати* 'put [something] in order'.

The most problematic are here the lexical units that enter the corresponding LSRs, and their definitions. For linking wordnets, the most important relations are synonymy and hypernymy, where the greatest number of equivalents across synsets comes from absolute synonymy (Rudnicka et al. 2019). So, synonymy plays a fundamental role in creating wordnets. It is remarkable that if synonymy is defined on the basis of hypernymy, treating hypernymy as the fundamental relation, the resulting synsets would differ (Halas Popović 2023).

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2102 (Project)



Synonymy is widespread in common language and also in various terminologies (Vakulenko 2018; Bokshanj 2021). However, this concept is treated differently in different sources. For example, the Longman dictionary[6] defines a synonym as "a word with the same meaning as another word in the same language" (Longman: 1685), where only lexical synonymy is considered and no partial synonymy and cross-language synonymy are recognized.

In contrast, Löbner understands synonymy in a way that:

"Two expressions are synonymous if they have the same meaning" (Löbner 2013: 203). This is an interesting approach where the synonymy is taken more generally, involving expressions instead of a single word and depending on their meanings.

The more general definition is based on the notion of a concept and suggests that: "Synonyms are the units of a language system (morphemes, lexemes, syntactic constructions) that within usually one linguistic level represent the same concept, and that can be substituted in the context without distorting the general sense" (Vakulenko 2018: 36). A similar approach is adopted for wordnets (Fellbaum 2006; Rudnicka et al. 2019: 299) where synonymy is considered in a narrower sense as restricted to the lexical level.

Another approach to synonymy that is aimed at meeting the lexicographer's requirements, derives synonymy from the so-called constitutive relations such as hypernymy/hyponymy, meronymy/holonymy, and antonymy (Maziarz et al. 2013; see also Rudnicka et al. 2019: 300).

The theory of synonymy in the context of terminological research was further developed in Vakulenko (2018: 40-43) where it was proposed to discern formal synonymy (covering lexical and grammatical synonymy) and contentual one. In particular, terminological dictionaries in the fields of physics (WPEDFR[7]; WPED[8]; TSF; SFS) and phonetics (Yurtbaşı 2017a; 2017b; Vakulenko 2018: 49-55) contain various kinds of full synonyms resulting from these types of synonymy. In what follows, these notions will be used.

A large problem in developing wordnets, particularly Ukrajinet, is the coordination and adjustment of lexicographic data coming from different sources. This problem will be addressed here. Synonymy and hyper-/hyponymy in wordnets will be considered and traditional dictionaries in connection with the existence of "hidden" senses not explicitly displayed in the last sources. A comparative analysis will be provided of definitions available in OEWN in comparison with those coming from academic dictionaries.

The article is organized as follows. In Section 2, we describe the method and material used for this research. Section 3 is devoted to the study of synonymy and hypernymy/hyponymy as they appear in traditional terminological dictionaries. In Section 4, we carry out a comparative analysis of term definitions available in OEWN and traditional dictionaries. Section 5 provides a discussion of related issues and future plans. The concluding remarks are presented in Section 6.

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2102 (Project)



## 2. Methods and material

The statistical (in the symptomatic sense) and analytical methods of terminology science (see Vakulenko 2018) are applied to carry out the comparative analysis of terminological units and their definitions coming from conventional dictionaries in physics and phonetics (WPEDFR; WPED; TSF; Yurtbaşı 2017a; 2017b), with those of OEWN. The statistical method determines in this case the availability of terms and their definitions, as well as the use examples, while the analytical method reveals scientific quality of these. The study is restricted to English, German, and Ukrainian lexical units, and uses standard conventions for traditional dictionaries. Within this lexicographic treatment, the units appearing in the same terminological article, are considered full synonyms, and the nouns in the word combinations which are specified by preceding adjectives (modifiers), are considered hypernyms. For example, *shift* is a hypernym for *Doppler shift*. As for definitions, the term to which the described term is referred is regarded as its hypernym.

## 3. Synonymy and hypernymy

Approximately 10% of concepts available in the physical dictionaries (WPEDFR; WPED; TSF) have synonyms. The rise of lexical synonymy is conditioned by a certain degree of semantic similarity of concepts whose hyponyms have identical meanings:

— Doppler **shift** – Doppler **displacement** – Doppler **frequency shift** (WPEDFR: 375);
— Ger. Polschuh**abstand** m – Luftspalts**länge** f – Luftspalts**breite** f (WPEDFR: 570);
— **initiation** of the discharge – **ignition** of the discharge (WPEDFR: 726);
— level of significance – significance **level** – error **probability** (WPEDFR: 838);
— **level** of significance – confidence **level** – **degree** of confidence – confidence **probability** – confidence **coefficient** (WPEDFR: 838);
— Ger. statistische Sicherheit – Sicherheits**schwelle** f – Sicherheits**grad** m – Vertrauens**koeffizient** m – Konfidenzkoeffizient m – Konfidenz**niveau** n – Konfidenz**wahrscheinlichkeit** f – Bedeutsamkeit**stufe** f (WPEDFR: 838);
— contact **clearance** – contact **gap** – contact **spacing** (WPED: 594);
— **monitor** – monitoring instrument – survey **instrument** – survey **meter** (WPED: 594);
— Ger. spiegelnder **Körper** – weisses **Objekt** (WPED: 306);
— sighting **apparatus** – sighting **device** (WPED: 594);
— Ger. Visier**vorrichtung** f – Ziel**gerät** n (WPED: 594);
— ***density*** *– specific* ***weight****, optical* ***isomer*** *– optical* ***antipode*** (TSF: 165);
— *light-beam* ***oscillograph*** *– light-beam* ***instrument*** (TSF: 165);
— ***argument*** *– independent* ***variable****,* ***momentum*** *–* ***quantity*** *of motion* (Vakulenko 2018: 41);

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2102 (Project)



— *magnetometer – magnetic(-field)* ***meter*** *– magnetic* ***detector*** *– gaussmeter* (TSF: 296-297);
— ***approximant*** *– frictionless* ***continuant*** (Vakulenko 2018: 53);
— *diminuendo* ***accent*** *– falling* ***accent*** *– diminuendo* ***stress*** *– falling* ***stress*** (Vakulenko 2018: 53).

In these examples, there is a clear semantic correlation between the head terms, including hypernymy/hyponymy (*shift* **–** *frequency shift*, *instrument – monitor*, *instrument – meter*, *weight – density*, *instrument – oscillograph*, *variable – argument*, *quantity – momentum*, *meter – detector*, *continuant* **–** *approximant*) and synonymy (*shift* **–** *displacement*, *level – degree*, *accent* **–** *stress*). It can be seen from this that a necessary condition for the synonymy of two lexemes is the presence of hyponyms in the corresponding concepts that are part of the same synset. So, the equivalence of hyponyms is caused by synonymy of their hypernyms. As can be seen from the synonymous pair *Doppler* ***shift*** *– Doppler* ***frequency shift***, the opposite is correct only when the hypernym (*shift*) has an additional meaning identical to the meaning of its hyponym (*frequency shift*).

Grammatical variability also often causes synonymy, primarily for compound terms. This is exemplified by the following series of synonyms: *unshared electron pair –* ***electron lone pair*** *– lone pair –* ***lone-pair electrons*** (TSF: 394). Here the element *pair* that is a hypernym to *electron lone pair*, can be considered as a hypernym to the unit *electrons* that is a hypernym to *lone-pair electrons*.

The situation is somewhat different for the **contentual synonymy** that arises when different terms reflect different aspects of the concept or different approaches to its study (Vakulenko 2018: 41). This kind of synonymy is also typical for composed terms, i.e. when some semes of their hypernyms are neutralized. In traditional dictionaries, the corresponding head terms for synset members often display either little semantic similarity or complicated and unstraightforward semantic relations:

piezo **effect –** piezoelectric **effect – polarization** of distortion (WPED: 594);
Ger. *schwarzer* ***Körper*** *– schwarzer* ***Strahler*** 'blackbody radiator' (WPED: 22);
*density matrix – statistical* ***matrix*** *– statistical* ***operator*** (TSF: 305);
*Gibbs* ***energy*** *– Gibbs thermodynamic* ***potential*** *– isobaric-isothermal* ***potential*** *– free* ***enthalpy*** (TSF: 166);
*thermal electron* ***emission*** *– Richardson* ***effect*** *– filament emission* (TSF: 165);
*Planck* ***constant*** *–* ***quantum*** *of action* (TSF: 608);
*isotopic* ***indicator*** *–* ***tracer*** *– tracer* ***isotope*** *–* ***label*** *– label(l)ed* ***atom*** (TSF: 225);
*freezing* ***point*** *– freezing* ***temperature***; *dielectric loss* ***coefficient*** *– dielectric loss* ***factor*** *– dielectric loss* ***index*** *– dielectric power factor – dielectric dissipation factor – dielectric loss* ***tangent*** (Vakulenko 2018: 43).

Historically, these terms entered the conventional dictionaries due to their practical use that from the standpoint of terminological theory is the result of the statistical method in the symptomatic sense (Vakulenko 2018: 14-19). However, the

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2102 (Project)



appearance of synonymy between composed terms assumes synonymy or hypernymy between head terms describing corresponding concepts. In other words, the conceptual approach underlying the development of wordnets necessitates the introduction of extra meanings of these head terms that have not been explicitly displayed in conventional dictionaries. Such augmentation of the human lexicographic practice manifests the application of the analytical method of terminology science (Vakulenko 2018: 14-19).

While the relative number of such correlated and under-refined concepts is an open question, their existence was already observed while working directly on Ukrajinet.

In particular, the synset **omw-1354-n** contains lexemes "НВЧ-генератор" 'microwave oscillator' and "лампа зворотної хвилі" 'backward-wave tube' which, according to the TSF, belong to the nests of different terms: "генератор" 'oscillator' (**omw-124123-n**) and "лампа" 'tube' (**omw-1617-n**), respectively. The synset **omw-689-n** (i61958) contains lexemes "гомоморфізм" 'homomorphism' and "відображення гомоморфне" 'homomorphic mapping' which, according to the TSF, are also included in different term nests: "подібність" 'similarity' (**omw-100-n**) and "відображення" 'mapping' (**omw-54804-n**). In such cases, lacking senses are added to the corresponding nesting terms.

## 4. Dictionary definitions vs OEWN definitions

The degree of semantic similarity between hypernyms to the abovementioned synsets cannot be easily calculated, as it strongly depends on the corresponding definitions (see Vakulenko 2019). For example, the Open English Wordnet treats energy as "a thermodynamic quantity equivalent to the capacity of a physical system to do work", connecting this concept with the hypernym "quantity" (https://en-word.net/lemma/energy). In turn, the TSF defines energy as "a general measure of various processes and types of interaction that allows measuring all kinds of physical forms of movement and interaction", relating this concept to the hypernym "measure". Using the latter is recommended, because some definitions of physical and technical terms available in the OEWN, fail to correctly convey significant features of the defined concepts such as those with the following ILIs.

### *Incorrect OEWN definitions*

**Determinant** (i80590): 'a square matrix used to solve simultaneous equations'. This definition is completely incorrect. It is recommended to use the latter formulation from the TSF that is freely available and defines this concept as 'a polynomial, each term of which is a product of matrix elements, and only one coefficient is included in the product from each row and each column of the matrix' ['поліном, кожен член якого є добутком елементів матриці, причому з кожного рядка і кожного стовпця матриці в добуток входить лише один співмножник'].

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2102 (Project)



**Field, field of force, force field** (i97792): 'the space around a radiating body within which its electromagnetic oscillations can exert force on another similar body not in contact with it'. This definition is not correct, because the force field is not about a body radiating electromagnetic oscillation. The Longman (which is freely available) defines this concept better as 'the area in which the natural force is felt or has an effect'. The TSF treats this concept more comprehensively, as 'a part of space (limited or unlimited), at each point of which a force with a certain numerical value and direction acts on a material particle located there, which depends only on the coordinates of this point' ['частина простору (обмежена чи необмежена), у кожній точці якої на розміщену там матеріальну частинку діє певна за числовим значенням і напрямком сила, що залежить тільки від координат цієї точки']. The last definition is recommended.

**Luminescence** (i97901): 'light not due to incandescence; occurs at low temperatures'. This description is not relevant to luminescence. However, the Longman also does not highlight essential properties of luminescence, defining this phenomenon as 'a soft shining light'. The TSF treats luminescence as 'radiation, which is an excess over the thermal radiation of the body and continues for a time that significantly exceeds the period of light oscillations' ['випромінювання, що являє собою надлишок над тепловим випромінюванням тіла і продовжується протягом часу, який значно перевищує період світлових коливань'] that should be used in wordnets.

**Paramagnetism** (i97922): 'materials like aluminum or platinum become magnetized in a magnetic field but it disappears when the field is removed'. This definition is incorrect, because paramagnetism is a property, not a material. The TSF contains a definition that should be used instead: 'the property of substances (paramagnets) to be magnetized in the direction of an external magnetic field' ['властивість речовин (парамагнетиків) намагнічуватися в напрямку зовнішнього магнітного поля'].

**Vibration, oscillation** (i75292): 'a regular periodic variation in value about a mean'. These terms are not synonyms, as can be seen from the Longman that proposes two senses for vibration(s): 1 'a continuous slight shaking movement'; 2 'vibes'; and four senses for oscillation(s): 1 'frequent changes between two extreme amounts of limits'; 2 'frequent changes between one feeling or attitude and another'; 3 'a regular movement of something from side to side'; 4 'a regular change in direction of an electrical current'. This treatment is much more relevant and should be applied in wordnets.

### *Inexact OEWN definitions*

**Elasticity** (i63383): 'the tendency of a body to return to its original shape after it has been stretched or compressed'. This definition uses common language and lacks scientific exactness. The Longman suggests a somewhat better definition: 'the ability of something to stretch and go back to its usual length or size'. The

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2102 (Project)



TSF contains an even more exact formulation: 'the ability of a material or product to undergo more or less significant springy reversible deformations without destruction' ['спроможність матеріалу або виробу зазнавати більш або менш значних пружних зворотливих деформацій без руйнування']. This last definition should be used in wordnets.

**Flexure, flexion, flection** (i110688): 'the state of being flexed (as of a joint)'. This definition refers the concept to the corresponding verb and is, therefore, not sufficient. The TSF defines this concept as 'a type of deformation characterized by a change in the curvature of the axis (bar, balk, rod) or the middle surface (plates, shells) under the influence of external force or temperature' ['вид деформації, що характеризується зміною кривизни осі (бруса, балки, стержня) або серединної поверхні (пластинки, оболонки) під дією зовнішньої сили або температури'] that is much more informative and can be recommended.

**Fluorescence** (i97800): 'light emitted during absorption of radiation of some other (invisible) wavelength'. This definition does not emphasize essential features of fluorescence and refers to light which is correct but not complete. The TSF gives a definition that refers fluorescence to luminescence and describes its essential features and, therefore, may be recommended: 'short-term luminescence with decay time of $10^{-8}$–$10^{-9}$ sec' ['короткочасна люмінесценція з часом загасання $10^{-8}$–$10^{-9}$ сек'].

**Loupe** (i55838): 'small magnifying glass (usually set in an eyepiece) used by jewelers and horologists'. This definition is too narrow. The TSF gives a more general and comprehensive definition, which is recommended: 'an optical system consisting of a lens or several lenses, designed to observe small objects located at a finite distance, under an increased angle of view' ['оптична система, що складається з лінзи або кількох лінз, призначена для спостереження дрібних предметів, розташованих на скінченній відстані, під збільшеним кутом зору'].

**Resistance, impedance, ohmic resistance, resistivity, electric resistance, electrical resistance** (i98048): 'a material's opposition to the flow of electric current; measured in ohms'. This definition makes no distinction between direct and alternating currents, which is crucial here. In fact, impedance is analogical to resistance but represents a different concept. The Longman defines impedance as 'a measure of the power of a piece of electrical equipment to stop the flow of an alternating current'. The TSF gives a more precise definition of impedance: 'an analogue of electrical resistance for harmonic processes' ['аналог електричного опору для гармонічних процесів'], which is recommended.

**Selectivity** (i64341): 'the property of being selective'. The TSF gives another definition used in optics that should be also present in wordnets: 'the ability of a spectral device to distinguish narrow spectral intervals from a continuous spectrum' ['спроможність спектрального прилада виділяти вузькі спектральні інтервали з суцільного спектру'].

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2102 (Project)



**Sensing element, sensor, detector** (i52820): 'any device that receives a signal or stimulus (as heat or pressure or light or motion etc.) and responds to it in a distinctive manner'. This formulation is too broad, and the term "detector" does not have this sense. The Longman defines sensor similarly but incompletely, as 'a piece of equipment used for discovering the presence of light, heat, movement etc.', and detector differently, as 'a machine or piece of equipment that finds or measures something'. The TSF defines this concept as 'a converter of a controlled or regulated value into an output signal, convenient for remote transmission and further processing' ['перетворювач контрольованої чи регульованої величини у вихідний сигнал, зручний для дистанційної передачі та подальшої обробки'], and detector as 1 'a device for detecting certain particles or radiations' ['прилад для виявлення тих чи інших частинок або випромінювань'] and 2 'a device for converting modulated high-frequency oscillations into oscillations with a modulation frequency' ['прилад для перетворення модульованих високочастотних коливань у коливання з частотою модуляції']. These definitions are recommended for the use in wordnets.

**Set, place, pose, position, put, lay** (i29197): 'put into a certain place or abstract location'. This definition is not well-formed because it repeats the defined concept ("put"). The Longman avoids this inconsistency: 'to move something to a particular place or position, especially using your hands'. This definition should be recommended for wordnets.

**Undulation, wave** (i75294): '(physics) a movement up and down or back and forth'. This definition is too simplified to convey notable information. The Longman proposes a more relevant but too narrow definition: 'a line of raised water that moves across the surface of the sea'. The more strict definition is available in the "Explanatory dictionary of physics": 'a change in some set of physical quantities (fields), capable of moving (spreading), moving away from the place of its origin, or oscillating within limited regions of space' ['зміна деякої сукупності фізичних величин (полів), здатна переміщуватися (поширюватися), віддаляючись від місця її виникнення, або коливатися всередині обмежених областей простору'] which we recommend.

### *Incomplete OEWN definitions*

**Flint** (i115216): 'a hard kind of stone; a form of silica more opaque than chalcedony'. The Longman distinguishes two meanings: 1 'a type of smooth hard stone that makes a small flame when you hit it with steel'; 2 'a piece of this stone or a small piece of metal that makes a small flame when you hit it with steel' and the TSF adds a sense 'optical glass that contains lead oxide and thus has a refractive index and its dispersion greater than, for example, crown' ['оптичне скло, яке містить окис свинцю і завдяки цьому має показник заламу та його дисперсію більші, ніж, наприклад, крон']. The wordnet definitions should include all these meanings.

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2102 (Project)



**Graphite, black lead, plumbago** (i114791): 'used as a lubricant and as a moderator in nuclear reactors'. This is not a complete definition, but rather a part of it. Another and essential part is available in the "Longman Dictionary of Contemporary English": 'a soft black substance that is a kind of carbon, used in pencils, paints, and electrical equipment'. The combined formulation for graphite is recommended: 'a soft black substance that is a kind of carbon, used in pencils, paints, and electrical equipment, and that serves as a lubricant and as a moderator in nuclear reactors'.

**Median** (i68434): 'the number midway between the two middle numbers in a series containing an even or odd number of items'. This definition is too complicated and encompasses only one sense of this polysemant. In comparison, the Longman distinguishes three senses of this noun: 1 'a narrow area of land that separates the two sides of a big road in order to keep traffic travelling in different directions apart'; 2 'the middle number of measurement in a set of numbers or measurements that have been arranged in order'; 3 'a line passing from one of the points of a triangle to the centre of the opposite side'. The TSF discerns also three senses of this term: 1 'the line that bisects the opposite side of the polygon' ['лінія, що ділить протилежну сторону багатокутника навпіл'], which is absent in OEWN and is more general than sense 3 of Longman; 2 'a point with the smallest sum of distances from given points' ['точка з найменшою сумою відстаней від заданих точок'], which is absent in OEWN and Longman; 3 'the value in the middle of a certain set of values' ['серединне значення певного набору величин'], which correlates with the OEWN and Longman definitions but is more comprehensive and clear. It is recommended to use the definitions of median given in the "Explanatory dictionary of physics".

**Oscillator** (i56791): 'generator that produces sonic oscillations or alternating current'. This treatment is too narrow. The Longman contains a more exact formulation: 'a machine that produces electrical oscillations', whereas the TSF adds two more definitions describing two other meanings of this term: 1 'a system (or a material point) carrying out oscillating periodic motion near the position of stable equilibrium' ['система (або матеріальна точка), що здійснює коливальний періодичний рух біля положення стійкої рівноваги']; 2 'exciter (source) of oscillations' ['збуджувач (джерело) коливань']. The last three definitions should be recognized also in wordnets.

**Photometry** (i40656): 'measurement of the properties of light (especially luminous intensity)'. There is also another meaning presented in the "Explanatory dictionary of physics": 'a section of physical optics dedicated to the measurement of electromagnetic radiation in the optical range' ['розділ фізичної оптики, присвячений вимірюванню електромагнітного випромінювання оптичного діапазону'] — that should be included in wordnets.

**Radiometer** (i57900): 'meter to detect and measure radiant energy (electromagnetic or acoustic)'. The TSF three more definitions that should be present in

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2102 (Project)



wordnets: 1 'a device for measuring the energy of electromagnetic radiation, based on its thermal action' ['прилад для вимірювання енергії електромагнітного випромінювання, заснований на його тепловій дії']; 2 'a radio telescope receiver' ['приймальний пристрій радіотелескопа']; 3 'a device for measuring the activity of radioactive sources' ['прилад для вимірювання активності радіоактивних джерел'].

**Spectrometry, spectroscopy** (i38839): 'the use of spectroscopes to analyze spectra'. This definition is incomplete and makes no difference between different notions of spectrometry and spectroscopy. The TSF defines spectrometry as 'a set of methods and theory of measuring the spectra of electromagnetic radiation and studying the spectral properties of substances and solids in the optical range of wavelengths' ['сукупність методів і теорія вимірювання спектрів електромагнітного випромінювання та вивчення спектральних властивостей речовин і тіл в оптичному діапазоні довжин хвиль'] and proposes two senses for spectroscopy: 1 'a field of physics devoted to the study of the distribution of the intensity of electromagnetic radiation by wavelengths or frequencies' ['область фізики, присвячена дослідженню розподілу інтенсивності електромагнітного випромінювання за довжинами хвиль або частотами']; 2 'study of decompositions into the spectrum' ['дослідження розкладів у спектр']. These two different notions should be treated in wordnets accordingly.

**Transparency, transparence** (i98144): 'permitting the free passage of electromagnetic radiation'. The Longman does not have an entry "transparence" and proposes three senses for "transparency": 1 'a sheet of plastic or a piece of photographic film through which light can be shone to show a picture on a large screen'; 2 'the quality of glass, plastic etc. that makes it possible for you to see through it'; 3 'the quality of being easy to understand or know about'. Here, the second sense gives qualitative description of this concept. In turn, the TSF presents more strict definition: 'a value that shows how much of the radiation flux (or for visible light – the luminous flux) falling on a certain surface passes through a layer of unit thickness without changing its direction' ['величина, що показує, яка частина потоку випромінювання (або для видимого світла – світлового потоку), що падає на деяку поверхню, проходить без зміни напрямку через шар одиничної товщини'], which is recommended.

The proposed definition amendments will, accordingly, cause the changes in hypo-hypernymy relations that should be correctly displayed in wordnets.

The relative number of such inconsistencies is small because the total number of analyzed synsets in OEWN is 3 220. However, each third definition therein requires minor refinement.

The described procedure may be executed both automatically — through hypothetical corresponding tool based on the findings of (Vakulenko 2019; 2021*)* — or manually. Human validation by an experienced lexicographer is required.

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2102 (Project)



## 5. Discussion

The analyzed problems are typical for wordnets created on the basis of the "merge" and "combined" approaches. The "expansion" approach alone does not have these inconsistencies because the resulting data in this case come from only one source — plWordNet (Rudnicka et al. 2019), which is linked to OEWN. However, its disadvantage is that in most cases, the resulting synsets and their semantic structures do not correspond to those inherent to the target language, not to mention individual lexical units, whose machine translation is often incorrect today, especially with regard to the Ukrainian language. This results in the need for tedious post-processing that cannot yet be performed by a machine.

The profound reason for these inconsistencies between concepts and their verbal designations lies in asymmetrical dualism of the linguistic sign that is the fundamental property of language. Asymmetrical dualism of the linguistic sign means that the word meaning (signifié) and its written form (signifiant) are set in an unstable balance and display dynamic behavior. The first is inclined to be expressed by other word forms, thus giving rise to synonymy, while the second tends to acquire additional meanings (see Karcevskij 1929). This continuous emergence of additional meanings cannot be properly tracked by conventional lexicography methods.

Fellbaum (2006: 665) makes a reserving remark about the interchangeability of synset members in different contexts, indicating that there exist synset members that cannot be substituted in all contexts. However, if synonymy is considered as a phenomenon formed at a higher, contextual level, it may be concluded that lexical units that cannot be substituted in all contexts, should belong to different synsets. From this standpoint, a revised definition of synonyms strictly requiring their interchangeability in all contexts should be adhered to.

Regarding different definitions of the same lexeme, a reasonable question may be asked: how crucial is the given form of a definition for a semantic load of the defined unit? In other words, how do the differences in definitions available in different sources affect the sense of the lexeme in question?

A metric method to measure semantic distances between lexemes was applied to build semantic fields using a recursive procedure that represents a lexeme's meaning by its semes taken from a dictionary definition (Vakulenko 2021). It was observed that two types of lexicographic effects appear in the Merriam-Webster dictionary[9]: the horizontal crossing of different semantic trees when some semes belonging to different trees, are the same at certain semantic depth; and the vertical looping (appearance of lexicographic hypercycles) when the same seme appears at some depth of the given semantic tree. A similar lexicographic effect was reported earlier in Shyrokov (2018: 12-18) on the material of the "Dictionary of Ukrainian Language in 20 Volumes" where the following semantic hypercycle was found: ***prylad*** 'instrument' – *prystrij* 'device' (hypernym) – *obladnannja* 'equipment' (hypernym) **–** ***prylad*** 'instrument' (meronym). Such hypercycles appear in sufficiently large dictionaries and manifest the rise

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2102 (Project)



of semantic structures that are conditioned by vertical and horizontal semantic correlations between concepts. The existence of these structures indicates that the resulting semantic fields display a certain level of stability against different formulations in definitions.

The outcomes of the recent research on concept ontology (Bense 2021) suggest that complex concepts can finally be decomposed into a significantly smaller number of elementary concepts displaying more semantic clarity, which are the same across analyzed languages. This fact also supports the assumption that the semantic load of a given lexeme is less dependent on the exact formulation of the corresponding concept.

## 6. Prospects of future research

Some related questions that should be examined in more detail in future investigations are briefly discussed here.

1. The notion of "meaning dimensions" (Löbner 2013) has a major sense in semantics because it sets the weight of any individual meaning. The further development if this idea leads to the conclusion that any sense of a linguistic unit can be presented in the form of its semantic field represented by a sum of elementary meanings taken with corresponding weight coefficients (see Vakulenko 2021). In other words, a proper account of "meaning dimensions" requires basing on their weight coefficients which determine the importance of these dimensions.

   The definition of synonyms by Löbner (op. cit.) uses a vague concept of "sameness" and tempts a logical question of what senses are considered the "same", or to what extent of precision the lexical units must have the "same" meaning, and what the tools to measure this "sameness" could be. This fuzzy definition makes the synsets themselves ambiguous and unclear. On the other hand, a mathematically strict approach to calculate semantic distances between lexical units based on their definitions (Vakulenko 2019) may resolve this issue. The possible use of underlying mathematics to determine and measure synonymy would be promising for further development of wordnets.

2. The semantic equivalence of the compound synonyms ***electron lone pair*** and ***lone-pair electrons*** (TSF: 394) is due to syntactic reasons. This kind of similarity may be displayed also by corresponding contextualized embeddings as described in (Liu et al. 2020; Lee 2021; Loureiro et al. 2021). However, the method described in Section 3 to refine relevant semantics seems to be much more straightforward and clearer.

3. The development of semantic structures obtained within the "merge" approach, requires assignment of semantic relations not displayed in the lexicographic data sources. For example, "A Multilingual Dictionary on Phonetics and Phonology" (Yurtbaşı 2017a; 2017b; Vakulenko 2018: 49-55) and "A Concise

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2102 (Project)



Dictionary of Synonyms of the Ukrainian Language"[10] contain synsets but have no information about hyper-/hyponymy, mero-/holonymy, antonymy, etc. In this case, tools for automatically assigning LSRs are useful, such as the tool for building hypernymy described in Romanyshyn et al. 2024.

However, such a semantic hierarchy is disrupted in multiword expressions where the overall meaning cannot be inferred from the meanings of individual components (Voula 2023).

4. AI technology has been successfully applied to the task of writing dictionary definitions (De Schryver 2023: 381) and text generation, including Ukrainian (*Ukrainian Text Generator*). This makes it possible to expect that this technology will be able to solve also the tasks described above.

## 7. Conclusion

This study has demonstrated that the development of the Ukrainian Wordnet (Ukrajinet) under the "merge" and "combined" approaches inevitably exposes inconsistencies between lexicographic sources and wordnet structures, especially regarding synonymy and hyper-/hyponymy. What is new and significant here is the identification of systematic mismatches between traditional dictionary definitions and their Open English Wordnet counterparts, and the formulation of a reproducible procedure for their correction. By applying methods of terminology science and comparative lexicography, this work clarified how missing or imprecise definitions distort conceptual hierarchies and proposed a structured post-processing framework that restores their integrity.

These findings contribute to the broader field of computational lexicography by highlighting that semantic refinement is not merely an editorial task but an essential step toward ensuring conceptual accuracy and interoperability of multilingual lexical databases. The recognition that synonymy and hyper-/hyponymy relations often depend on subtle shifts in definitions provides a pathway for developing more semantically consistent ontologies. In practical terms, the proposed procedures can inform the semi-automatic refinement of lexical databases, combining algorithmic processing with expert validation.

However, several limitations remain. The current refinement was performed on a subset of data and focused primarily on the domains of physics and phonetics. Extending this methodology to other domains will require additional testing and refinement of the semantic alignment rules. Moreover, while partial automation is achievable, full automation of lexicographic post-processing still depends on advances in AI-based semantic modeling.

The practical implications of this research lie in improving the quality of multilingual wordnets, supporting tasks such as machine translation, semantic search, and ontology construction. Future research should focus on implementing computational tools capable of identifying definitional inconsistencies automatically, experimenting with weighted semantic fields, and integrating contextual embeddings to enhance synonymy detection. Further theoretical work

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2102 (Project)



on semantic metrics and concept ontologies may also provide mathematical foundations for refining lexical relations.

Ultimately, this study underscores that the future of lexicography lies in the synergy of human expertise and computational precision. Wordnets are not static repositories but evolving models of linguistic and conceptual knowledge. The refinement procedures outlined here contribute to moving closer to a truly interoperable and semantically transparent lexicon — a cornerstone for intelligent language technologies of the future.

## 8. Acknowledgments

The author is grateful to Prof. Melanie Siegel for useful comments and suggestions that helped to improve the content of this article.

## 9. Endnotes

1. https://en-word.net/
2. https://ukrajinet.yoda.media.h-da.de/
   https://github.com/hdaSprachtechnologie/ukrajinet
   https://code.fbi.h-da.de/ukrajinet/ukrajinet-dev/-/blob/main/20230728_merged_ukrajinet.xml
3. Vakulenko, M.O and O.V. Vakulenko. 2008. *Tlumachnyj slovnyk iz fizyky* [Explanatory Dictionary on Physics]. Kyjiv: VPC "Kyjivsjkyj universytet". 767 p.
4. Vakulenko, M.O. and O.V. Vakulenko. 2017. *Slovnyk fizychnykh terminiv-synonimiv* [Dictionary of Physical Synonymous Terms]. Kyjiv: VPC "Kyjivsjkyj universytet". 191 p.
5. Rusanivsjkyj, Vitalij (Ed.). 2010–. *Slovnyk ukrajinsjkoji movy v 20 tomakh* [Dictionary of Ukrainian Language in 20 Volumes]. Kyjiv: Naukova dumka.
6. *Longman Dictionary of Contemporary English*. 2003. [Della Summers (ED.).] Harlow, Essex: Pearson Education.
7. Sube, Ralf and Günther Eisenreich. 1973. *Wörterbuch Physik. A–M. Englisch–Deutsch-Französisch–Russisch*. Zürich/Frankfurt am Main: Verlag Harri Deutsch.
8. Greulich, Walter and Dirk Meenenga. 1997. *Wörterbuch der Physik: Englisch–Deutsch/Deutsch–Englisch*. Heidelberg/Berlin: Spektrum, Akademischer Verlag.
9. *Dictionary by Merriam-Webster*. 2023.
   https://www.merriam-webster.com/
10. Derkach, Pylyp. 1960. *Korotkyj slovnyk synonimiv ukrajinsjkoji movy* [A Concise Dictionary of Synonyms of the Ukrainian Language]. Kyjiv: Radjansjka shkola.
    https://chtyvo.org.ua/authors/Derkach_Pylyp_Maksymovych/Korotkyi_slovnyk_synonimiv_ukrainskoi_movy/

## References

**Bense, Hermann.** 2021. A Formal Ontology of Concepts for Deep Semantic Searching. *KONT-21 Knowledge — Ontology — Theory, 8th International Conference, 8–12 November 2021, Novosibirsk.* https://www.ontology4.us/download/dot/ontopage/sub_Bens2021d.pdf

**Bokshanj, Ghalyna.** 2021. Synonimija v terminosystemi anghlomovnykh naukovykh tekstiv ekonomichnogho sprjamuvannja [Synonymy in Terminology System of English-language Scientific Texts in the Field of Economics]. *Pivdennyj arkhiv (filologhichni nauky)* 86: 54-57 (in Ukrainian). DOI: 10.32999/ksu2663-2691/2021-86-8

**Bond, Francis, Piek Vossen, John P. McCrae and Christiane Fellbaum.** 2016. CILI: the Collaborative Interlingual Index. Fellbaum, Christiane, Piek Vossen, Verginica Barbu Mititelu and Corina Forascu. 2016. *Proceedings of the 8th Global WordNet Conference (GWC),* Bucharest, Romania, 27–30 January 2016: 50-57. Global Wordnet Association. https://aclanthology.org/2016.gwc-1.9

**De Schryver, Giles-Maurice.** 2023. Generative AI and Lexicography: The Current State of the Art Using ChatGPT. *International Journal of Lexicography* 36(4): 355-387. DOI: https://doi.org/10.1093/ijl/ecad021

**Fellbaum, C.** 1998. *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.

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

**Halas Popović, Ana.** 2023. Synonymy from a Prototype Theory Perspective and its Symbiosis with Polysemy: Towards a New Dictionary of Synonyms. *Lexikos* 33: 1-23. DOI: https://doi.org/10.5788/33-1-1773

**Karcevskij, Sergej.** 1929. Du dualisme asymétrique du sign linguistique [On the Asymmetric Dualism of the Linguistic Sign]. *Travaux du Cercle linguistique de Prague* 1: 88-92. https://www.degruyter.com/document/doi/10.7560/780439-004/html

**Lee, Younghoon.** 2021. Systematic Homonym Detection and Replacement Based on Contextual Word Embedding. *Neural Processing Letters* 53: 17-36.

**Liu, Qi, Matt J. Kushner and Phil Blunsom.** 2020. A Survey on Contextual Embeddings. ArXiv.2003. 07278v2 [cs.CL]. DOI: https://doi.org/10.48550zrXiv.2003.07278

**Loureiro, Daniel, Kiamehr Rezaee, Mohammad Taher Pilehvar and Jose Camacho-Collados.** 2021. Analysis and Evaluation of Language Models for Word Sense Disambiguation. *Computational Linguistics* 47(2): 387-443. DOI: https://doi.org/10.1162/coli_a_00405

**Löbner, Sebastian.** 2013. *Understanding Semantics.* Second Edition. New York: Routledge.

**Maziarz, M., M. Piasecki and S. Szpakowicz.** 2013. The Chicken-and-Egg Problem in Wordnet Design: Synonymy, Synsets and Constitutive Relations. *Language Resources and Evaluation* 47(3): 769-796.

**McCrae, John Philip, Alexandre Rademaker, Ewa Rudnicka and Francis Bond.** 2020. English Wordnet 2020: Improving and Extending a Wordnet for English Using an Open-source Methodology. Declerk, Thierry, Itziar Gonzalez-Dios, German Rigau. 2020. *Proceedings of the LREC 2020 Workshop on Multimodal WordNets (MMW2020), Marseille, France, May 2020*: 14-19. The European Language Resources Association (ELRA). https://aclanthology.org/2020.mmw-1.3/

**Romanyshyn, Nataliia, Dmytro Chaplynskyi and Mariana Romanyshyn.** 2024. Automated Extraction of Hypo-Hypernym Relations for the Ukrainian WordNet. *The Third Ukrainian Natural Language Processing Workshop (UNLP) @LREC-COLING-2024, 25 May 2024:* 51-60. *ELRA Language Resource Association.*

https://lexikos.journals.ac.za; https://doi.org/10.5788/35-2-2102 (Project)



**Rudnicka, Ewa, Maciej Piasecki, Francis Bond, Łukasz Grabowski and Tadeusz Piotrowski.** 2019. Sense Equivalence in plWordNet to Princeton WordNet Mapping. *International Journal of Lexicography* 32(3): 296-325.
DOI: https://doi.org/10.1093/ijl/ecz004

**Shyrokov, V.A. (Ed.).** 2018. *Linghvistychno-informacijni studiji. Praci Ukrajinsjkogho movno-informacijnogho fondu NAN Ukrajiny: u 5 tomakh, tom 3. Tlumachna leksykoghrafija, knygha 2. Systemna semantyka tlumachnykh slovnykiv* [Linguistic Information Studies. Works of Ukrainian Lingua-Information Fund of National Academy of Sciences of Ukraine: in 5 Volumes, Vol. 3. Explanatory Lexicography, Book 2. Systemic Semantics of Explanatory Dictionaries] (in Ukrainian). Kyjiv: Ukrajinsjkyj movno-informacijnyj fond.

**Siegel, Melanie and Francis Bond.** 2021. OdeNet: Compiling a German Wordnet from Other Resources. Vossen, Piek and Christiane Fellbaum (Eds.). 2021. *Proceedings of the 11th Global Wordnet Conference (GWC 2021), Potchefstroom, South Africa, 18–21 January 2021*: 192-198.
https://www.aclweb.org/anthology/2021.gwc-1.22

***Ukrainian Text Generator***. Linguanex.
https://lingvanex.com/services/ukrainian-text-generator/

**Vakulenko, Maksym.** 2018. *From Terminology-Vocabulary to Terminology-Science: A Ukrainian Trend.* Beau Bassin: LAP.

**Vakulenko, Maksym.** 2019. Calculation of Semantic Distances between Words: From Synonymy to Antonymy. *Journal of Quantitative Linguistics* 26(2): 116-128. (Published online: 28 May 2018.)
https://doi.org/10.1080/09296174.2018.1452524

**Vakulenko, Maksym.** 2021. From Semantic Metrics to Semantic Fields. *Proceedings of the 2021 IEEE 16th International Conference on Computer Science and Information Technologies (CSIT), 22–25 September 2021, Lviv, Ukraine: 44-47.*
DOI: 10.1109/CSIT52700.2021.9648675

**Voula, Giouli.** 2023. A Model for Representing the Semantics of MWEs: From Lexical Semantics to the Semantic Annotation of Complex Predicates. *Frontiers in Artificial Intelligence* 6.
DOI: https://doi.org/10.3389/frai.2023.802218

**Yurtbaşı, Metin.** 2017a. The French Phonological Corpus with Equivalents in English, German and Turkish, as Part of a New Multilingual Dictionary Project. *International Journal of Arts, Humanities, and Social Sciences* 2(11): 27-77.
http://www.ijahss.com/vol2-issue11.html

**Yurtbaşı, Metin.** 2017b. A Multilingual Dictionary also Containing Qazaq Terminology in Phonetics and Phonology. *International Journal of Arts, Humanities, and Social Sciences* 2(11): 78-103.
http://www.ijahss.com/vol2-issue11.html