Ученые записки Таврического национального университета им. В.И. Вернадского
Серия "Филология". Том 20 (59). 2007 г. №4. С.57-61.

УДК 81'374.221.823:004.912

ФОРМУВАННЯ ЕЛЕКТРОННОГО СЛОВНИКА СИНОНІМІВ УКРАЇНСЬКОЇ МОВИ

Устимець О.В.

Український мовно-інформаційний фонд НАН України, м. Київ, Україна

В статті подаються принципи побудови Електронного словника синонімів української мови (ЕСС). Продемонстровано використання даного словника у функції основної лінгвістичної бази при формуванні української зони автоматичного багатомовного перекладного словника як частини інтегрованої лексикографічної системи, що розробляється в Українському мовно-інформаційному фонді. Основними параметрами при формуванні синонімічних рядів (синсетів) ЕСС є семантична схожість слів та взаємозамінність їх у тексті

Ключові слова: *синоніми, взаємозамінність, семантична подібність, автоматичний багатомовний перекладний словник, перекладний еквівалент*

Дослідження синонімії проводиться нами в рамках проекту „Електронний словник синонімів української мови (ЕСС)", який, в свою чергу, виступає в функції основної лінгвістичної бази, що використовується при формуванні української зони автоматичного багатомовного перекладного словника (АБПС), як частини інтегрованої лексикографічної системи, що розробляється в Українському мовно-інформаційному фонді. Мета статті – продемонструвати використання даного словника у функції основної лінгвістичної бази при формуванні української зони автоматичного багатомовного перекладного словника

Постановка проблеми. Зупинимося коротко на принципах побудови АБПС з використанням ЕСС. З метою розмежування полісемічних та синонімічних відношень перекладні одиниці (ПО) в даному словнику представлені семантичними групами, що утворюються словами в їхніх конкретних значеннях з відповідними їм синонімами, якими ці значення можуть передаватися в конкретній мові, тобто синсетами.

Визначаємо синсет як групу семантично схожих слів, що мають однаковий лексико-граматичний статус і при перекладі можуть бути взаємозамінними. Перекладні еквіваленти (ПЕ) однієї перекладної одиниці (представлені синсетами) мають однакові номери в усіх мовних зонах словника і є кодами їхніх загальних значень. Це робить можливим переклад за словником у будь-якому напрямку [1, с. 14].

Приклад перекладних еквівалентів в словнику для українського багатозначного слова „закон" у російській, англійській, німецькій, іспанській та французькій зонах, що входять у синсети 126, 127, 129:

126 – постанова державної влади: *закон* (укр.), *закон* (рос.), *low* (англ.), *gezetz* (нім.), *loi* (франц.);

127 – об'єктивно існуючий зв'язок між предметами, явищами:

закономірність 1 закономерность1 regularity1 Gesetzmabigkeit 1 regularite1
закон 2 закон 2 Gezetz 2 logique2
логіка 3 логика 3 logic 2 Logic 3
129 – те, чим керуються у праці, житті:
правило 1 правило 1 regulations 1 Gegel 1 regle 1
закон 2 закон 2 rule 2 Grundsatz 2 reglement 2
припис 3 канон 3 principle 3 Maxime 3 principe 3
принцип 4 норма 4 norm 4 Pranzip 4 norme 4
канон 5 предписание 5 standart 5 Norm 5
норма 6 Vorschrift 6
Bestimmung 7

Елементи синсета ієрархічно розташовані за ступенем близькості їхніх значень до загального значення синсета. Стилістично маркованим елементам синсета (наприклад, таким, що мають позначку в словнику „*розм.*", „*рідко*" тощо) надають найвищі ранги. Морфолого-фонетичним варіантам додається спеціальний ідентифікатор; наприклад, у варіанті даного словника перед рангом такого варіанта ставиться цифра 5. За збігом другої цифри з рангом іншого слова і відбувається ідентифікація „абсолютних синонімів". Так, наприклад, синсет зі значенням "*мати обсяг, достатній для розташування кого-, чого-небудь*" набуває з урахуванням рангів такого вигляду: {вміщати 1, уміщати 51, вміщувати 2, уміщувати 52, містити 3}.

Вибір варіантів перекладу здійснюється у такий спосіб: у *зоні словника, яка* відповідає мові, з якої робиться переклад слова, визначаються групи, що містять у собі дане слово. Головний варіант перекладу шукатиметься в тій групі, в якій перекладна одиниця має ранг 1 (як правило, ним є слово з найбільш нейтрально вираженим значенням у синсеті).

У наведеному вище *прикладі* перекладу українського слова „закон" головним варіантом перекладу для всіх мов будуть ПЕ зі синсета 126. В обраній групі головним варіантом буде визначено ПО з рангом 1. За цим же принципом вибираються з усіх інших синсетів решта варіантів перекладу. Так, варіантами перекладу, наприклад, на російську мову будуть *закономірність* (з синсета 127), *правило* (з синсета 129). У випадку графічно тотожних варіантів один з них замінюється синонімом з рангом 2.

При використанні АБПС в *автономному режимі людина-перекладач*, навіть якщо її знання мови, якою робиться переклад, обмежуються лише загальними уявленнями про її структурно-граматичні властивості, може однозначно визначити адекватний перекладний еквівалент. Ним буде слово із синсета, номер *якого збігається з номером* синсета у вихідній мові, до якого віднесено ПО. При машинному перекладі однозначний вибір ПЕ з усіх запропонованих варіантів можливий лише на етапі післяредагування.

Орієнтація на розв'язання означених прикладних завдань зумовила *певне* переосмислення засад, на яких формувався базовий Словник синонімів української мови (ССУМ) [2].

Насамперед це стосується інтерпретації основних параметрів, за якими слова об'єднуються в синсети, – семантичної схожості синонімів та їхньої взаємозамінності в тексті, які в нашій роботі інтерпретуються через вимогу до синонімічного відношення бути рефлексивним і транзитивним. З ознаки рефлексивності випливає неможливість об'єднання в один синсет слів на позначення родо-видових понять, а також слів, які є синонімами до домінанти *синонімічної групи, але не є такими* один до одного. В

базовому Словнику синонімів такі слова часто подаються в одній словниковій статті через крапку з комою. Так, наприклад, синонімічна група прикметника **червоний**: **червоний** (кольору крові), **кров'яний, кривавий, калиновий; бордовий, бордо, вишневий, червлений, гранатовий** *діал.* (темно-червоний); **малиновий** (кольору ягоди малини – темно-червоний); **рубіновий** (кольору рубіну); **маковий** *розм.* (кольору червоної квітки маку); **червіньковий** (кольору червоно-коричневої глини – червіньки); **кумачевий** (кольору кумачу – яскраво-червоний); **кораловий** (кольору коралів – яскраво-червоний); **кармінний, карміновий, шарлаховий** (яскраво-червоний); **червоногарячий, полум'яний, полум'янистий, жаркий, пломенистий** *поет.* (кольору вогню – яскраво-червоний); **рожевий** (світло-червоний); **фрез** (світло-червоний з бузковим відтінком).

Прикладом невизнання синонімічних груп базового словника через визначення між їхніми елементами родо-видових відношень можуть бути словникові групи: *{**білила, блейвес** – свинцеві білила}*, *{**верблюд, дромедар** – верблюд одногорбий}*, *{**тріска, лабардан** – просолена й пров'ялена без кісточок}*, *{**хата, глинянка** – хата, зроблена з глини, **саманка** – хата, побудована з саману}* тощо. Такі синсети перебудовуються або взагалі вилучаються зі словника, якщо вони складаються лише з двох елементів.

Як зазначалося раніше, базову основу укладання Електронного словника синонімів становить Словник синонімів української мови. В електронному представленні ССУМ – це словникові статті, що являють собою синонімічні ряди, в яких слова в межах одного ряду мають схоже значення та об'єднані спільною домінантою.

Розкриття значень і взаємозв'язків членів синонімічного ряду здійснюється за допомогою тлумачень, пояснень. У ССУМ у більшості синонімічних рядів після домінанти наводиться стисле тлумачення, що містить спільну семантичну характеристику всіх членів ряду. У багатьох випадках синонімічний ряд з однією домінантою розбивається на окремі групи слів (іноді одну) і записуються вони через крапку з комою (;). Як правило, до всієї групи або до окремого слова дається уточнення тлумачення.

Виділені синсети стають об'єктом аналізу щодо семантичної схожості їхніх елементів. У результаті детального дослідження встановлено, що такий розподіл синонімів за допомогою крапки з комою в ССУМ не є послідовним і чітким. Як показав аналіз, саме через крапку з комою подано слова з відношенням рід-вид (це видно з представленого вище синсету з домінантою ***червоний***), а також слова, що є синонімами лише до домінанти і між собою, але не є такими по відношенню до інших елементів, записаних через крапку з комою (наприклад: **абстрактний** *(який ґрунтується на теоретичних, відірваних від досвіду міркуваннях, позбавлений конкретності)* – **загальний, умоглядний, ідеальний;** і – **платонічний** *(духовний, не пов'язаний з реальними цілями)* і, навпаки, не зрозуміло чому через крапку з комою подано синоніми: **квапити** *(спонукати когось до швидкого виконання)* та **підганяти** *(примушувати або заохочувати робити що-небудь швидше)* Це відповідно означає, що практично більша частина синонімічних рядів перебудовується.

Процедурним визначенням семантичної схожості синонімів є встановлення схожості семантичних компонентів в їхніх тлумаченнях. Аналіз і порівняння значень слів-претендентів на роль синонімів здійснюється за допомогою порівняння тлумачень елементів синсета, поданих в Тлумачному словнику української мови.

За основу приймаємо традиційний погляд на синоніми і вважаємо, що їхня головна властивість – наявність у значеннях всіх відповідних лексем достатньо великої частини, що співпадає. Тобто формальними критеріями зіставлення тлумачень елементів синсета є збіг їхніх семантичних компонентів. Так, наприклад, синсет слова „буквар"

буквар *– книжка для початкового навчання грамоти,*

азбука *– книжка для початкового навчання грамоти: буквар;*

абетка *– книжка для початкового навчання грамоти, буквар;*

граматка *– те саме, що буквар.*

Як бачимо, загальне значення синсета є еталоном зіставлення для інших його членів. Формальними критеріями такого зіставлення є збіг у слів ***буквар, азбука, абетка*** семантичних компонентів: *„книжка для початкового навчання грамоти"*; а також подання тлумачення через синонім, наприклад: ***азбука*** *–... буквар;* або через мітку *„те саме, що"*(у випадку тлумачення синоніма ***граматка)***. Отже, повне співпадання тлумачення даних слів є підставою для того, щоб визначити їх синонімами.

Як було зазначено в роботі, встановлення можливості словникових синонімів бути взаємозамінними здійснюється шляхом аналізу їхнього контекстного оточення в текстах Українського національного лінгвістичного корпусу обсягом 42 млн. словоживань. Для виділення контекстів використовується компонент програмного інструментарію семантичної розмітки текстів, розроблений в УМІФі для здійснення автоматичного маркування синонімів, який видає конкорданси синонімів довжиною в одне речення в контексті.

Вимога взаємозамінності диктує вилучення з синсетів тих лексем, лексична сполучуваність яких характеризуються дуже вузьким колом контекстів. Наприклад, з дієслівного синсета „повертати, звертати, вернути, брати, забирати" зі значенням за ССУМ „змінювати напрямок свого руху, повертати" дві останніх лексеми вилучаються, оскільки лексема *брати* в цьому значенні можлива лише в контекстах „брати вліво, вправо", а *забрати* – „забрати вліво, вправо, вгору" (пор.: повертати вліво, вправо, з дороги, на стежку, до лісу, в напрямку, від дуба...) [3, с. 56].

Висновки. З огляду на обрані параметри, в процесі роботи над побудовою синсетів ЕСС відбувалися зміни в наступних напрямках: а) в наш словник включались нові синонімічні ряди, що утворювалися в результаті руйнування існуючих в Базовому словнику з огляду на недопустимість таких варіантів (наприклад, синсет з домінантою **убити** містить близько 30 синонімів з відношенням рід-вид); б) в інших випадках проводилось розширення синсета шляхом включення нових слів: наприклад, коли тлумачення слова подається через синонім, але він не входить в даний синсет:

Чорнобиль *1 – різновид полину, нехвощ,* ***чорнобильник;***

Хвощ *2 – те саме, що чорнобиль*

В синсет включено слово ***чорнобильник:***

Чорнобиль *1 – різновид полину, нехвощ,* ***чорнобильник;***

Хвощ *2 – те саме, що чорнобиль;*

Чорнобильник *3 – те саме, що чорнобиль*

Із синсета, заданого Базовим словником, виключалися деякі претенденти-синоніми, якщо їх значення не співвідносились з основним значенням синсета (наприклад: із синсета {свідок, посвідник, посвідчувач, понятий} - *„особа, яку викликають до суду для посвідчення відомих їй обставин справи"* було виключено слово ***понятий*** в значенні *„особа, яку залучають органи влади як свідка при обшукові, описові майна і т. ін."*;

г) видалялися діалектні слова, які не фіксуються Тлумачним словником (оселок, жертвенний, перегудка, щепки, дриготіти, поладнувати і т. ін.).

Список літератури

1 Грязнухіна Т.О. Система багатомовного машинного перекладу// Мовознавство. – 2001. - №5. – С.14-25.
2. Словник синонімів української мови: В 2 т. – К., 1999-2000.
3. Широков В.А. Семантичні стани мовних одиниць та їх застосування в когнітивній лексикографії // Мовознавство. – 2005 – №3-4. - С. 56.

Устимец Е.В. **ФОРМИРОВАНИЕ ЭЛЕКТРОННОГО СЛОВАРЯ СИНОНИМОВ УКРАИНСКОГО ЯЗЫКА**

В статье поданы принципы построения Электронного словаря синонимов украинского языка (ЭСС). Продемонстрировано использование данного словаря в функции основной лингвистической базы при формировании украинской зоны автоматического многоязыкового переводного словаря как части интегрированной лексикографической системы, которая разрабатывается в Украинском языково-информационном фонде Основными параметрами при формировании синонимических рядов (синсетов) ЭСС избрано семантическое сходство слов и взаимозаменяемость их в тексте.

Ключевые слова: *синонимы, взаимозаменяемость, семантическое сходство, автоматический переводной многоязыковой словарь, переводной эквивалент*

Ustymets O.V. **CONSTRUCTION OF THE ELECTRONIC DICTIONARY OF SYNONYMS OF THE UKRAINIAN LANGUAGE**

In this paper the principles of construction of the electronic dictionary of synonyms of the Ukrainian language are described. The use of the dictionary in the function of the main linguistic base in forming the Ukrainian zone of the automatic multilanguage translation dictionary, as part of the integrated lexicographic system, which is developed in the Ukrainian Language-Information Fund. The main purpose of this paper is to show that the basic parameters of formation of synonymic rows are semantic similarity and interchangeability of these words

Key words: *synonyms, interchangeability, semantic similarity, automatic multilanguage translation dictionary, translation equivalent*

Поступила до редакції 05.03.2007 р.