

ВИЗНАЧЕННЯ СЕМАНТИЧНОЇ КОРЕЛЯЦІЇ УГРУПОВАНЬ ЗНАЧЕНЬ НЕОДНОЗНАЧНИХ СЛІВ УКРАЇНСЬКОЇ ТА АНГЛІЙСЬКОЇ МОВИ

Солдатова Л.П. (Київ)

У міжсупільних комунікативних процесах важливою тенденцією стає якість міжмовних трансформацій інформації. Проблема співвідношення автентичної інформаційної наповненості значень слів особливо гостро виявляється для неоднозначних слів, тому що з'являється проблема розмежування цих неоднозначностей за інформаційними значеннями всередині кожної мови.

Мета роботи полягає в удосконаленні технології коректного та однозначного міжмовного перекладу інформації.

Актуальність досліджень виявляється при автоматизованих перекладах. Наявні алгоритми та програмне забезпечення міжмовних перекладів не дають необхідного автентичного перекладу, а іноді призводять до абсурдних ситуацій через проблемність слів з кількома інформаційними наповненнями.

У даній роботі розглядається можливість знаходження **вектора інформаційної відповідності (ВІВ)** значень неоднозначного слова однієї мови з можливими інформаційними значеннями цього слова в іншій мові. Основою для вирішення цього завдання є інформаційний аналіз порівнюваних слів у двох мовах та визначення ступеня кореляції інформації по суті у цих значеннях.

Вирішення цього завдання проводиться у 2 етапи: наведення тлумачення кожного значення та розмежування їх за омонімічними групами та полісемами всередині кожної такої групи слів мови, що зіставляються за суттю інформації, яка міститься у центральних омонімах цих слів; проведення кореляційного аналізу між угрупованням розмежованих значень у двох мовах та визначення ВІВ.

Без обмеження загальності досліджень використаємо результати роботи [2] та розглянемо найскладнішу частину виконаних кореляційних розмежувань, а саме ПЛЮ (повних лексичних омонімів) та полісемів іменникової частини лексики.

Поняття „інформація” (інформаційний масив) як лінгвістичний об'єкт, використано поряд з таким лінгвістичним об'єктом, як „слово”. Усі відомі визначення останнього можна класифікувати як два комплекси: до першого відносяться визначення, які розглядають „слово” акцентовано як елемент системи мови, а до другого відносяться визначення, які розглядають „слово” акцентовано як знак (код, позначуване) елементів дійсності. Аналіз визначень першого комплексу показує, що у системі мови слово характеризується такими основними ознаками, як відносна самостійність і відносна міцність. Другий комплекс визначень характеризується тим, що в їх основі є те, для чого слова і потрібні – як позначувані субстанцій та елементів дійсності, де поняття „дійсність” трактується як єдність сутності

внутрішнього для людини буття та сутності зовнішнього буття, процесів і результатів функціонування свідомості мозку. Наявні у літературі визначення „слова” в основному представляють характеристики (архітектури) форми подання інформації про значення (інформи) і представляють співвідношення цієї форми у складі таких форм як ідіоми, речення. Але ці характеристики форми накладають собою обмеження і на суть вміщеної у ній інформації. Основною узагальнювальною ознакою „слова” (як словесного знака) у відомих визначеннях є *функція репрезентації*.

Поняття „*інформація*” - *це визначений об’єм узгоджених понять, знань і даних, які відображають властивості та стани об’єктів і явищ об’єктивного світу в природній, соціальній та штучно утворених системах (у тому числі і абстрактних), у часових і просторових координатах і відображених та створених процесом мислення у людських суб’єктивних свідомостях.*

Основною функцією поняття „*інформація*” є репрезентація або уявлення дійсності визначеного об’єму. Те саме є основною ознакою „слова” як Знака. Але поняття „*інформація*” не обмежене формою репрезентації, тоді як „слово” має такі обмеження (відносна самостійність і відносна міцність). Як репрезентатор дійсності „*інформація*” є поняттям більш загальним, ніж слово, речення і т. ін., тобто „*інформація*” є загальним родоводом для „слова”, „речення”. *Інформація* – поняття категоріальне, через нього можна виконувати багато визначень у комунікативних аспектах життєдіяльності людей і соціумів. Наприклад, визначення „слова” згідно з теорією визначень може синтезуватися як знак певного інформаційного масиву, який має відносну самостійність і відносну міцність.

Поняття „слова” визначається через поняття „*інформація*” як його родовід за основною ознакою. Саме за цією ознакою заміна одного поняття іншим правомірна за законами формальної логіки як заміна часткового загальним. Це дає підстави для визначення *омонімів* як об’єктів інформації, вміщених в форму морфем, слів, ідіом, синтаксичних конструкцій і т. ін., в яких співвідношення форми і змісту такі, що при більшому ступені тотожності форми існують такі відмінності у змісті, які являють собою різні інформаційні наповненості за суттю буття, явищ, об’єктів. **ПЛО (повні лексичні омоніми)** – це інформаційні масиви (об’єкти), які мають тотожний акустично-графічний код (АГК) у всіх своїх формах і мають повну відмінність в інформаційній наповненості по суті інформації, яка передається чи зберігається у конкретний часовий період.

На основі запропонованих визначень „*інформація*” та „*інформаційні масиви*” представлено визначення *полісемів* як інформаційних масивів, які мають тотожний акустично-графічний код вираження своєї форми у всіх

граматичних формах, належать до однієї частини мови і характеризуються існуванням об'єму інформаційного пересічення, в якому є спільний семантичний компонент, основне інформаційне значення якого повністю або частково переноситься на можливі інформаційні масиви (лексико-семантичні варіанти) даного слова або словосполучення у конкретний часовий період існування.

Згідно з цими визначеннями ПЛО та полісемів і проведеним аналізом існуючих визначень логічним критерієм розмежування ПЛО та полісемів має бути наявність для полісемів чи відсутність для омонімів кореляції інформації за суттю (об'єму перетину інформації за суттю), яка міститься в інформаційних масивах (значеннях слів), які підлягають розмежуванню – **інформаційно-кореляційний критерій розмежування (ІККР)**.

Основою **ІККР** є аналіз можливих кореляцій інформаційних масивів. Тому основною складовою для проведення кореляційних досліджень є сама технологія формування уявлення, презентації інформації у масиві та її характеристик у формі, зручній для кореляції. Це є важливим завданням для всієї лінгвістики взагалі і для аналізу значень слів та словосполучень. У роботі зроблена спроба створити модель формування інформації та визначено форму та структуру формування інформації для її диференціювання та кореляції.

Основними вихідними принципами і передумовами для створення такої структури є: психофізіологічність і функціональність мови; значення слів, **ЛСВ** (лексико-семантичні варіанти) тотожно уявляються як ідіоматичні масиви, а їх позначення – як *знаки* і *символи* для функціонування мозку; унормована характеристична структурність формування презентації, аналоги якої мають позитивний досвід подання інформації в інших науках і сферах життя людей.

Створена структура **ТФЗС** (**тлумачна формула значення слова**) відповідає таким вимогам: має єдину базову **характеристичну структуру** визначення інформації за суттю (**ХС**); містить функціональну суть у загальній системі людських знань, спільне, відмінне, зв'язки, в тому числі логічні та причинно-наслідкові; повністю та однозначно ідентифікує суть інформації в масиві.

Вся описова частина **ТФЗС** виконується на спеціальній мові-ідентифікаторі, створення якої є дуже важливим інформаційно-лінгвістичним завданням, та не містить ускладнення сприйняття суті значення.

Базова **ХС** визначень інформації за суттю, як варіант, може бути створена, виходячи з визначення поняття „**інформація**” та основного інформаційного значення за суттю, під яким розуміють функціональне або спостережувальне відображення та вираження станів буття, але з можливим

функціональним впливом на життя. Це дає можливість визначити такі основні елементи *ХС* (ознаки певної інформації): сутність явища в середовищі його існування; функціональна суть об'єкта в середовищі його існування; статична суть об'єкта (явища): компонентний склад, речовинний склад, форма й об'єм, розташування у просторі та ін.; зв'язки з іншими; динамічна суть об'єкта (явища): динамічні характеристики, рух відносно системи існування, відносний рух всередині об'єкту (явища); дії на нього, дії його на ін.; характеристики станів систем існування.

Опис інформації за цими характеристичними складовими повинен проводитися, за Баллі Ш., особливою „інтелектуальною мовою-ідентифікатором”. Основними вимогами до необхідної нам мови-ідентифікатора повинна бути визначена базова система понять і термінів – лексичних одиниць (однозначних іменників, дієслів і т. ін.); визначена базова структура словосполучень із сукупністю синтагм та більш складних синтаксичних конструкцій; визначена унормована стилістика; всі базові складові мови-ідентифікатора повинні бути загальнозживаними в усіх сферах життєдіяльності людей (наука, мистецтво, тощо) і бути, по можливості, близькими до категорій.

Таким чином, сформовані *ТФЗС* будуть достатніми для проведення кореляції між ними. За допомогою простих логічних операцій порівняння алгоритм детектування кореляції складається з порівняння за відповідними тотожними елементами безлічі інформації.

Набір основних дій *базового алгоритму* для використання *ІККР* неоднозначностей складається з формування інформаційних масивів *ТФЗС* у формі роздиференційованих *позицій порівняння* (п.п.) за визначеними вище умовами з однаковими ознаками; із здійснення порівняння інформаційних масивів за відповідними позиціями порівняння; із здійснення порівняння результату попередньої дії з визначеними „*Договірними умовами*” (*ДУ*).

ДУ – це норма (співвіднесеність), за якою визначають успішність тієї чи іншої події, успішність розв'язання поставленого завдання за визначених умов. *ДУ* повинні давати однозначність у прийнятті рішень у процесі кореляції. Відповідно до базового алгоритму використання *ІККР* неоднозначностей першою дією є формування – презентація інформаційних масивів *АГК* у вигляді *ТФЗС* у формі роздиференційованих позицій порівняння за однаковими характеристичними ознаками – елементами *ХС* інформації.

Продемонструємо визначену технологію для пари слів “голова” (укр.) – “head” (англ.), які за інформаційною суттю центрального омоніма словників [1, 3] корелюють.

З більш, ніж 14 *позицій порівняння* (п.п.) за *ХС ТФЗС*, ми наведемо тільки аналітичний опис суті явища (п.п. 1.1) та аналітичний опис функціональної суті (п.п. 1.2).

В українській мові слово “голова” використовується у 10 значеннях. Англійська мова демонструє більше 66 варіантів значення. Але для демонстрації технології формування *BIB* ми використаємо значення за 3 корелюючими омонімами. Для автоматизованих перекладів важливіше розподілити угруповання значень слів на більшу кількість омонімів у синхронічному аспекті вживання, тому помилка в міжполісемному *BIB* значно менше спотворює переклад, ніж помилка у міжомонічному *BIB*.

Голова – 1 – основна частина живої істоти (п.п. 1.1); призначено для функціонування живого організму (п.п.1.2); **Голова – 2** – основна частина живої істоти – людини як орган мислення, пам’яті, розуму (п.п. 1.1); призначено для оптимального функціонування живого організму в цілому (п.п.1.2); **Голова – 3** – основна особа (персона) соціуму – лідер, який керує колективом людей (п.п.1.1); призначено та використовується для оптимального керування групою людей (п.п.1.2); **Голова – 4** – основна особа (персона) соціуму – керівник зборів людей (п.п.1.1); призначено та використовується для оптимального функціонування зборів людей (п.п.1.2); **Голова – 5** – основна особа у будь-якій справі – головний спеціаліст (п.п.1.1); призначено та використовується для оптимального керування справою (п.п.1.2.); **Голова – 6** – основне у будь-чому (п.п.1.1); призначено та використовується для здійснення основної функції чого-небудь (п.п.1.2); **Голова – 7** – верхня передня (за положенням) частина будь-чого (п.п.1.1); призначено та використовується як основна частина будь-чого (п.п.1.2); **Голова – 8** – передня частина колони, загону та т.ін. (п.п.1.1); призначено та використовується для оптимального руху колони, загону і т.ін. (п.п.1.2).

Результати розмежування угруповання цих значень за технологією ІККР: Голова –1 – Голова¹₁; Голова –2 – Голова²₁; Голова –3 – Голова¹₂; Голова –4 – Голова²₂; Голова –5 – Голова³₂; Голова –6 – Голова¹₃; Голова –7 – Голова²₃; Голова –8 – Голова³₃, де верхній індекс – позначає полісеми даного омоніма, нижній індекс – позначає омоніми.

Продемонструємо позиції порівняння для “Head” за відповідними омонімічними групами: **Head – 1** – основна частина живої істоти (п.п. 1.1); призначено для функціонування живого організму (п.п.1.2); **Head – 2** – основна частина живої істоти – людини як орган мислення, пам’яті, розуму (п.п. 1.1); призначено для функціонування живого організму в цілому (п.п.1.2); **Head – 25a** – персона соціуму, яка керує будь-якою справою (п.п.1.1); призначено для оптимального керування справою (п.п.1.2); **Head – 26a** – основна особа у будь-якій справі (п.п.1.1); призначено для

оптимального керування справою (п.п.1.2); **Head – 8** – верхня, передня за положенням частина чого-небудь (п.п.1.1); призначено та використовується для називання основної частини чого-небудь (п.п.1.2); **Head – 18a** – передня частина, фронт колони, армії (п.п.1.1); призначено та використовується для оптимального руху колони та ін. (п.п.1.2).

Результати розмежування: Head – 1 - Head¹₁; Head – 2 – Head²₁; Head – 25a – Head¹₂; Head – 26a – Head²₂; Head – 8 – Head¹₃; Head – 18a – Head²₃

За результатами продемонстрованих значень **VIB** будуть такими: [Голова¹₁ – Head¹₁], [Голова²₁ – Head²₁], [Голова¹₂ – Head¹₂], [Голова³₂ – Head²₂], [Голова²₃ – Head¹₃], [Голова³₃ – Head²₃]. Для значень Голова²₂ та Голова¹₃ в угрупованні значень Head немає кореляючих значень.

За результатами дослідження можна зробити висновок, що незважаючи на те, що центральні омонімічні групи корелюють за суттю інформації, інші значення в угрупованні можуть не корелювати між собою (наприклад, head of a ship, boat відповідає інформаційній наповненості в українській мові – ніс корабля) для **VIB** цих значень необхідно шукати кореляцію значень від інших слів мови.

Продемонстрована технологія дозволяє однозначно та повно визначити **VIB** значень неоднозначних слів мови апіорі відповідних за центральними омонімами у різних мовах; дозволяє визначити поле припущень **VIB** як всередині слів, що відповідають центральним омонімам, так і за значеннями інших слів; знання **VIB** дозволяє створити алгоритми та програми для автентичного міжмовного автоматизованого перекладу; отримані результати розмежування за значеннями можуть з успіхом використовуватися у лексикології, лексикографії та інших дослідженнях.

Література:

1. Словник української мови в 11 томах / Ред кол. Білодід І.К. та інш. – К.: Наукова думка, 1970 – 1980. Т.2. – С. 153– 155
2. Солдатова Л.П. Використання інформаційно-кореляційного критерію в практиці розмежування неоднозначностей // Актуальні проблеми української лінгвістики: теорія і практика. – Випуск 10. – К.: Київський національний університет імені Тараса Шевченка, 2004, С.146–155.
3. The Oxford English Dictionary (Vol. 12). Oxford University Press, Ely House, London W, 1970. Vol. 5, P. 140–146.