УДК 81’255:004 Філологічні науки, Перекладознавство

Моренцова Алла Володимирівна
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»
(Київ, Україна)

# УСУНЕННЯ ЛЕКСИЧНОЇ БАГАТОЗНАЧНОСТІ ПРИ МАШИННОМУ ПЕРЕКЛАДІ: ВІД ТЕРМІНОЛОГІЧНИХ СЛОВНИКІВ ДО ОНТОЛОГІЇ ПРЕДМЕТНОЇ ОБЛАСТІ

***Анотація:*** *Розглянута концепція побудови онтології предметної області на основі термінологічних словників. Представлені основні етапи й шаги цього процесу. Запропонований підхід дозволяє суттєво розширити множину предметних областей, у яких можливе застосування онтологічних моделей для розв'язку задачі усунення лексичної багатозначності при машинному перекладі.*
***Ключові слова:*** *машинний переклад, лексична багатозначність, термінологічний словник, тезаурус, таксономія, онтологія, онтологічна модель*

*Моренцова Алла Владимировна*
*Национальный технический университет Украины «Киевский политехнический институт имени Игоря Сикорского»*
*(Киев, Україна)*

# *УСТРАНЕНИЕ ЛЕКСИЧЕСКОЙ МНОГОЗНАЧНОСТИ ПРИ МАШИННОМ ПЕРЕВОДЕ: ОТ ТЕРМИНОЛОГИЧЕСКИХ СЛОВАРЕЙ К ОНТОЛОГИИ ПРЕДМЕТНОЙ ОБЛАСТИ*

***Аннотация:*** *Рассмотрена концепция построения онтологии предметной области на основе терминологических словарей. Представлены основные этапы и шаги этого процесса. Предложенный подход позволяет существенно расширить множество предметных областей, в которых возможно*

*применение онтологических моделей для решения задачи устранения лексической многозначности при машинном переводе.*

***Ключевые слова:*** *машинный перевод, лексическая многозначность, терминологический словарь, тезаурус, таксономия, онтология, онтологическая модель*

*Morentsova Alla*
*National Technical University of Ukraine*
*«Igor Sikorsky Kyiv Polytechnic Institute»*
*(Kyiv, Ukraine)*

# *ELIMINATION OF LEXICAL AMBIGUITY IN MACHINE TRANSLATION: FROM TERMINOLOGICAL DICTIONARIES TO THE DOMAIN ONTOLOGY*

***Abstract:*** *The concept of creation of domain ontology on the basis of terminological dictionaries is considered. The main stages and steps of this process are presented. The suggested approach allows to expand significantly a set of domains in which application of ontological models will make it possible to solve the problem of eliminating lexical ambiguity in machine translation.*

***Keywords:*** *machine translation, lexical ambiguity, terminological dictionary, thesaurus, taxonomy, ontology, ontological model.*

Незважаючи на постійний розвиток і удосконалення методів та систем машинного перекладу, на сьогоднішній день у машинному перекладі (МП) однією з головних проблем залишається ***лексична багатозначність*** — неоднозначність змісту слова, коли у нього може бути більше ніж одне значення. Розв'язання лексичної багатозначності — це встановлення значення слова в деякому контексті [1]. Автором даної статті було запропоновано [2-3] використовувати онтологію предметної області (ПрО), до якої належить текст, що перекладається, як один з дієвих інструментів вирішення цієї проблеми. При цьому виникає задача побудови такої онтології на основі термінологічного

словника.

Термінологічні питання з'являються при перекладі наукового й науково-технічного тексту, що відноситься до деякої області знання. Терміни в спеціальному тексті так само, як і звичайні слова, можуть бути багатозначні, виступаючи в ПрО як назви різних речей і понять залежно від контексту. Це полісемантична властивість терміна й умовою правильного перекладу, тобто вибору потрібного перекладу з множини пропонованих для терміна оригіналу, є правильне розуміння, про що йде мова в контексті. При перекладі науково-технічних текстів в основному користуються готовими термінами, що вже існують у відповідній області наукової літератури. Як правило ці терміни представлені в термінологічних словниках.

Проблема вирішення лексичної багатозначності пов'язана з лексичним значенням. А завдання її автоматичного вирішення було вперше сформульовано при створенні систем машинного перекладу.

Реалізацію того або іншого значення слова здійснює контекст або ситуація, загальна тематика мовлення. Точно так само, як контекст обумовлює конкретне значення багатозначного слова, у певних умовах він може створювати семантичну дифузність, тобто сумісність окремих лексичних значень, коли їх розмежування не здійснюється. Деякі значення проявляються тільки в комбінації з визначальним словом, а в деяких комбінаціях значення багатозначного слова представлене як фразеологічно зв'язане.

Не тільки лексична сполучуваність і словотворчі особливості характеризують різні значення слів, але також у ряді випадків і особливості граматичної сполучуваності.

Для перекладу багатозначних слів також використовуються контекстологічні словники, словникові статті яких фактично являють собою алгоритми запиту до контексту на наявність або відсутність контекстних визначників значення. Для кожного багатозначного слова вказується його пріоритетний переказний еквівалент, специфічний для розглянутої ПрО. Реалізація системи такого словника на комп'ютері в якості онтології ПрО,

представленої у вигляді онтологічної моделі дозволяє ефективно розв'язати проблему лексичної багатозначності при машинному перекладі.

Наявні на цей час розробки предметно-орієнтованих онтологій для різних природничих наук (біології, хімії, фізики й ін.) вимагають єдиного способу звертання до загальних і спеціальних знань при аналізі тексту. Використання для цього онтологічної моделі ПрО, яка по суті є розширеним графічним відображенням концептуального словника, дозволяє розв'язати це завдання Елементами онтології є концепти й концептуальні відносини між ними. При цьому онтологія допускає не тільки строгі концепти, але й одиниці, які можуть стати поняттями й концептами.

За одним із поширених визначень ***онтологія*** – це формальне представлення знання, яке включає поняття (такі як об'єкти, процеси і т.д.) у ПрО й деякі відношення між ними. Онтологію можна розглядати як базу знань спеціального виду з семантичною інформацією про визначену предметну область. Модель онтології ПрО містить визначені концепти (поняття, класи), властивості концептів (атибути, ролі), відношення між концептами (залежності, функції) і обмеження на використання, які визначаються аксіомами. Формальна модель онтології представляється у вигляді трійки множин O={T, R, F}, де T – множина понять ПрО; R – множина відносин між ними; F – множина функцій інтерпретації понять і відносин. Фундаментальні поняття визначеної ПрО відповідають класам онтології.

На багато рішень, що стосуються моделювання, впливає знання того, для чого буде використовуватися онтологія й наскільки детальною або загальною вона буде. Модель, як правило, конкретизується в залежності від призначення і сфери застосування онтології. Для обробки інформації на природній мові, зокрема при машинному перекладі, необхідно застосування спеціалізованих онтологій. Основне їхнє призначення — забезпечити зв'язок між фрагментами тексту природньою мовою й поняттями ПрО (наприклад, класами або екземплярами онтології) [4]. Поняття в онтології повинні бути близькі до об'єктів (фізичних або логічних) і відносинам в ПрО, що цікавить нас. Як правило, це іменники (об'єкти) або дієслова (відносини) у реченнях, які

описують нашу предметну область. Онтології можуть використовуватися в якості джерела знання для систем МП. З доступом до великої бази знань система може розв'язати багато двозначностей (особливо лексичних) самостійно.

Як показує досвід автора, для значної кількості лінгвістів онтологія починається і закінчується використанням онтологічних словників. Дійсно, як зазначено в [5], сучасні онтологічні словники містять не лише конкретні значення слів, а й лексичні (антонімія, слова-відношення, номіналізація та ін.) та семантичні (гіперонімія/гіпонімія, меронімія/голонімія та ін.) зв’язки між ними, що дає змогу використовувати їх для усунення багатозначності слів на основі цих зв’язків. Однак, онтологічні словники не описують в повній мірі мовні конструкції для вираження семантичних зв’язків між поняттями. Тому автор пропонує при застосуванні онтологій в машинному перекладі використовувати онтологічну модель ПрО, представлену в графічному вигляді (як приклад, ER-модель Чена) і дещо розширену під поставлену задачу.

При ***побудові онтології на основі словника*** ми проходимо чотири основні етапи, на кожному з яких одержуємо свій проміжний результат. На першому етапі будується ***керований словник***, що містить список явно заданих і ретельно відібраних термінів (слів, фраз або нотацій) ПрО. Усі терміни повинні мати однозначне й ненадлишкове тлумачення (визначення). Керовані словники фіксують можливі варіанти вибору значень і зменшують невизначеність і неоднозначність, властиву природній мові.

На другому етапі створюється ***таксономія*** – предметна класифікація, яка групує терміни у вигляді керованих словників і впорядковує ці словники у вигляді ієрархічних структур. Математично таксономією є деревоподібна структура класифікацій певного набору об'єктів.

На третьому етапі будується тезаурус. З погляду лінгвістики тезаурус – це множина значеннєвих одиниць деякої мови із заданій на ній системою семантичних відносин. Тезаурус фактично визначає семантику мови (національної й/або мови конкретної науки).

Наприклад, лінгвістичний, тезаурус містить:

1) морфологічні й синтаксичні властивості (частина мови, рід, відмінювання, корінь, словоформи в різних відмінках, родах і числах);

2) семантику (значення, синоніми, антоніми, гіпероніми, гіпоніми);

3) родинні слова;

4) фразеологізми й стійкі комбінації;

Тезаурус є розширенням таксономії і є одним з діючих інструментів для опису окремих ПрО. Він також дозволяє зв'язувати дві й більше таксономій. Це основа для вироблення єдиної, погодженої, нормативної, що однозначно розуміється, повної й несуперечливої термінології, використовуваної всіма, хто має відношення до ПрО. Це ж – засіб, призначений для класифікації, структурування, систематизації, моделювання і додання змісту поняттям і зв'язкам, що відносяться до ПрО. Також це ще й засіб для вичерпного опису інформаційної моделі ПрО (на основі якої будується онтологічна модель), включаючи й її семантику, самостійно, тобто не залежно від завдань (оперативних знань), які будуть вирішуватися з використанням цієї моделі.

На основі тезауруса на четвертому етапі будується онтологія ПрО. У такий спосіб увесь процес створення онтології ПрО, починаючи від словника, включає:

1) виявлення й чітке визначення ***понять (концептів)***. Це основа для створення керованих словників;

2) виявлення й чітке визначення множини ***властивостей*** (атрибутів, характеристик), що характеризують кожне поняття. Обов'язковим є виділення множини властивостей, що ідентифікують поняття;

3) виявлення й чітке визначення ***родовидових залежностей*** (зв'язків) між поняттями й, тим самим, завдання таксономій на множині понять;

4) виявлення й чітке визначення інших ***довільних бінарних зв'язків*** між поняттями із вказівкою до якого типу зв'язків вони відносяться (частина/ціле, агрегація, асоціація, причина/наслідок і т.д.). Тим самим будується тезаурус.

5) виявлення й чітке визначення ***аксіом*** (правил, прикладних обмежень), що характеризують поглиблену семантику понять, атрибутів і зв'язків. Тим

самим будується онтологія ПрО, яка використовується як інструмент для вирішення задачі усунення лексичної багатозначності.

Узагальнюючи сказане вище, відзначимо, що на першому етапі побудови онтології фактично будується глосарій термінів, що включає всі терміни (концепти та їх екземпляри, атрибути, дії й т.п.), важливі для ПрО, й їхні природномовні описи. Потім будуються дерева класифікації концептів. Таким чином, ідентифікуються основні таксономії предметної області, а кожна таксономія, згідно з розглянутою методологією, дає в остаточному підсумку онтологію. Далі будуються діаграми бінарних відносин, які фіксують відносини між концептами в рамках онтології.

Для машинного представлення онтології в системах машинного перекладу, як відзначено в [6], додатково створюються:

1) словник концептів, що містить усі концепти предметної області, екземпляри таких концептів, атрибути екземплярів, відносини, джерелом яких є концепт, а також синоніми й акроніми концепту;

2) таблиця бінарних відносин для кожного відношення, вихідний концепт якого втримується в класифікаційнім дереві. Для кожного відношення фіксується його ім'я, імена концепту-джерела й цільового концепту, інверсне відношення та інші характеристики;

3) таблиця атрибутів екземпляра для кожного екземпляра зі словника концептів. Вона містить: ім'я атрибута, тип значення, значення «за замовчуванням», атрибути, які можуть бути виведені з використанням даного атрибута, формула або правило для виведення й ін;

4) таблиця атрибутів класу для кожного класу зі словника концептів з аналогічними характеристиками.

5) таблиця екземплярів для кожного входу в словник концептів. Тут специфікується ім'я екземпляра, його атрибути і їхні значення.

Таким чином побудова онтологічної моделі ПрО на основі термінологічного словника надає можливість подальшого використання цієї моделі при вирішенні залачі усунення лексичної багатозначності при машинному перекладі науково-технічних текстів.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ ТА ЛІТЕРАТУРИ:


1. *Agirre E., Edmonds P. G.* Word Sense Disambiguation: Algorithms and Applications. — U.K.:Springer, 2007. — 380 c.
2. *Моренцова А. В.* Застосування онтологічних моделей предметної області для усунення лексичної багатозначності при машинному перекладі // Актуальные научные исследования в современном мире – Переяслав–Хмельницкий, 2018. – Вып. 4(36), ч. 7. – С. 65–71.
3. *Моренцов Є. І., Моренцова А. В.* Використання онтології предметної області для усунення неоднозначностей при комп'ютерному перекладі технічних текстів. Актуальні питання сучасної науки: III Міжнародна науково–практична інтернет–конференція: тези доповідей, Дніпро, 30 січня 2018 р. – Ч. 1. – Дніпро: НБК, 2018 – С. 82–88.
4. *Лесько О.Н., Рогушина Ю.В.* Использование онтологии предметной области для снятия омонимии в естественно–языкових текстах // Проблеми програмування. — 2017. — №2. — С. 61–71.
5. *Лозинська О.В., Давидов М.В.* Математична модель граматично–доповненої онтології // Вісник НТУ “ХПІ». — 2015. — №11(1120). — С.102–107.
6. *Fernandez, M.; Gomez–Perez, A.; Juristo, N.* METHONTOLOGY: From Ontological Art Towards Ontological Engineering. Workshop on Ontological Engineering. Spring Symposium Series. AAAI97 Stanford, USA