УДК 801.009

А.О. Рось, д-р техн. наук, проф.,
О.Ю. Ніколаєвський, здобувач,
КНУ імені Тараса Шевченка

ВИМОГИ ДО ФОРМАЛІЗАЦІЇ СЕМАНТИЧНОЇ ІНФОРМАЦІЇ В СИСТЕМІ МАШИННОГО ПЕРЕКЛАДУ

У статті запропоновано підхід до семантичного аналізу природно-мовного тексту на основі побудови розподілених семантичних моделей на різних рівнях представлення тексту. Текст розглядається як взаємодія трьох систем: семіотичної, мовної, системи знань про навколишній світ. Визначені й обґрунтовані мінімальні одиниці семантики для кожної системи. Показана практична реалізація запропонованого підходу для побудови систем машинного перекладу.

Ключові слова: семантичний аналіз, семіотична система, мовна система, семантичні одиниці, природно-мовний текст, система знань про світ, система машинного перекладу.

Актуальність дослідження. На сьогодні побудова формальної моделі семантики тексту є самою слабкою ланкою в системах автоматичного опрацювання природно-мовного тексту (ПМТ). І хоча розробки в цьому напрямі велися тривалий час і продовжуються, зараз якихось усталених універсальних методів аналізу змісту тексту немає. Ю. Марчук дослідження в галузі формалізації семантики умовно поділяє на два напрями [1]:

1) дослідження, що проводяться на дедуктивному абстрактно-теоретичному рівні, мета яких встановити співвідношення між семантикою і семіотикою, з одного боку, семантикою і синтактикою та прагматикою, з іншого; побудувати моделі розуміння ПМТ взагалі й у зв'язку з процесом комунікації;

2) дослідження індуктивного емпіричного характеру, його мета – розв'язання конкретних прикладних проблем: машинного перекладу, автоматичного інформаційного пошуку, реферування тощо.

Другий напрям носить фрагментарний характер, методи, що розробляють під конкретну систему, не придатні для розв'язання інших задач. Перший напрям має фундаментальний характер, і результати, отримані в межах такого дослідження, дозволили б позбавити недоліків прикладний напрям досліджень. Аналіз теоретичних напрацювань в рамках першого напряму показав, що однією з причин досить скромних результатів є різне бачення дослідників на проблему обґрунтування і вибір одиниці змісту тексту. Вирішення цієї проблеми дозволить значно просунутися у галузі штучного інтелекту.

Постановка проблеми дослідження. Для вибору й обґрунтування семантичних одиниць ПМТ проведено порівняльний аналіз відомих моделей семантики, які використовуються сьогодні в системах автоматичної обробки тексту.

У відомих моделях семантики англійської мови, зокрема, в моделі «концептуальних залежностей» Р. Шенка [2] в якості мінімальної одиниці змісту пропонується поняття, що відповідає слову або словосполученню в тексті. В моделі «семантик переваги» Уілкса [3] аналіз тексту починається з рівня твердження, що відповідає простому ядерному реченню. Існуючі моделі розуміння ПМТ з точки зору розглядуваних семантичних одиниць тексту показують, що розпізнавання смислу в кращому випадку починається з морфологічного рівня мовної системи: модель "Смисл↔Текст" І. Мельчука [4], в якій в якості мінімальної одиниці змісту пропонується морфема.

Всі перераховані моделі ігнорують текст як знакову систему. В практичному плані аналіз знакового рівня організації ПМТ обмежується виділенням синтаксичних розділових знаків, абревіатур, скорочень тощо. В той же час, аналіз текстів реальної складності показав, що вже на рівні знакової організації тексту людина використовує описові можливості семіотичної системи для кодування знань про фрагменти реального світу. Так, використання лапок (кінотеатр *"Салют"*) свідчить, що лексему в лапках не можна розглядати в значенні, поданому в словнику. Власні назви можуть збігатися з написанням загально вживаних слів, але при цьому мати інший зміст (депутат *Хмара*, прем'єр-міністр *Major*, вул. *23 Серпня*).

Пропонований підхід до розробки семантичної моделі ПМТ, ґрунтується на таких концептуальних положеннях:

- вхідний ПМТ - є зв'язний текст (тобто, дискурс);
- зв'язність дискурсу забезпечується графемними засобами оформлення тексту (відношення між заголовками і змістом абзаців тексту тощо), лінгвістичними засобами (граматичними узгодженнями, анафоричними посиланнями тощо) і екстралінгвістичними (часові, причинно-наслідкові відношення тощо);
- всі ці засоби є інструментом кодування знань про світ (предметну область).

Виходячи з цього, семантичний аналіз вхідного ПМТ є розподіленим і здійснюється, починаючи з знакового рівня організації тексту. В якості мінімальної семантичної одиниці цього рівня виступає *графема*. На рівні організації тексту як мовної системи ми розглядаємо морфологічний, синтаксичний та власне семантичний рівень. На морфологічному рівні в якості мінімальної одиниці змісту виступає *морфема*, на синтаксичному – *словосполучення*, на семантичному рівні в якості мінімальної одиниці змісту виступає *слово*, або нерозривне синтаксичне утворення, яке в розроблюваній системі машинного перекладу (СМП) розглядається як одиниця перекладу. На рівні організації тексту як системи відображення знань про фрагменти навколишнього світу семантична складова представлена моделлю знань про предметну область (ПО). В якості мінімальної одиниці змісту виступає *поняття* в системі його парадигматичних і синтагматичних відношень.

Викладення результатів дослідження. Як вже зазначалося, семантична модель тексту є розподіленою і її побудова починається зі знакового рівня організації тексту. Такий підхід зумовлений різноманітністю знакового (графемного) подання лексичних одиниць в тексті, яка визначає їх семантичні функції в тексті. Крім того, для вирішення задач перекладу суттєвим є також визначення структури тексту, для відокремлення службової інформації, виділення абзаців, заголовків тощо. Текст при цьому розглядається як певним чином організована послідовність рядків і графем. Задачею цього рівня розпізнавання є побудова формалізованого подання графемної структури тексту. Вхідними даними графемного аналізу є поточний текстовий файл і апріорні еталонні моделі (рядків і графем). В основу класифікатора графем покладені такі ознаки: тип знаку (цифра, буква, синтак-

сичний знак, службовий знак тощо), розмір (прописна, заголовна), фонетичні ознаки (голосна, приголосна). Кінцевою метою доморфологічного аналізу тексту є побудова графемної структури тексту, яка включає виділення на множині рядків і графем вхідного тексту таких семантично самостійних одиниць тексту: фрагментів, речень, синтагм, лексем; визначення класів перелічених одиниць тексту та встановлення відношень між ними в тексті.

Інформаційне забезпечення семантичної моделі доморфологічного аналізу включає словники, що відбивають екстралінгвістичні знання, необхідні для розпізнавання й вилучення знань про навколишній світ безпосередньо із вхідного тексту. В основу єдиної семантичної параметризації словникових одиниць (для української, російської та англійської мов) закладені універсальні (енциклопедичні) знання про навколишній світ. Семантичний код – це двопозиційний цифровий код: перша позиція відбиває семантичний тип лексеми, друга – її семантичне значення.

Так виділяються такі семантичні типи лексем (табл.1):

1 – *географічна назва*. Цей клас включає: назви міст, морів, океанів, річок, озер, материків тощо. Необхідність введення словника географічних назв обумовлена тим, що ці назви в тексті подаються без детермінуючих лексем, оскільки позначають загально відомі знання. В цей клас ми не включили назви держав, оскільки для нашої ПО ці назви мають політичний контекст, це власне й обумовило внесення їх до семантичного класу – політична назва;

2 – *історична назва*. Цей клас включає відомі назви історичних подій;

3 – *ім'я*. Необхідність введення даного семантичного типу обумовлена тим, що в англійських текстах мало відомі прізвища подаються разом із іменем. Це дозволяє автоматично ідентифікувати, що це є особа й об'єднати дві лексеми в одне неподільне поняття, крім того, визначення категорії роду для імені дозволяє досягти більшої точності при перекладі з англійської мови;

4 – *установа*. Клас включає відомі назви організацій, установ, видів збройних сил тощо;

5 – *одиниця вимірювання*. Цей клас включає скорочення, що визначають одиниці вимірювання та назви місяців і днів тижня для англійської мови (вони пишуться з прописної літери);

6 – *назва, що не перекладається*. Клас включає назви організацій, установ, прізвищ тощо, які передаються засобами іншої мови виключно заданими правилами транслітерації;

7 – *посада*. Цей клас включає назви посад, які пишуться із заголовної літери;

8 – *політична назва*. Необхідність введення даного семантичного типу обумовлена тим, що назви держав в нашому контексті (ПО: воєнно-політичні тексти) розглядаються як геополітичні об'єкти, а не як географічні назви.

9 – *не визначений семантичний тип*. Даний семантичний тип призначається, коли лексема не підходить не під один із перерахованих семантичних типів.

Друга позиція семантичного коду визначає семантичні характеристики лексеми у зіставленні зі світом. Так, виділяються такі значення семантичних типів лексем:

1 – *час*. Характеристика часу визначає лексему відповідного семантичного типу у часі;

2 – *простір*. Характеристика простору;

3 – *час-простір*. Дана характеристика притаманна деяким складним одиницям вимірювання (наприклад: км./год.);

4 – *кількість*. Характеристика, що відноситься виключно до оцінювання кількості;

5 – *об'єкт*. Характеристика, яка визначає конкретність (предметність) лексеми відповідного семантичного типу;

6 – *особа*. Характеристика, яка визначає людину (посадову особу);

7- 8 – характеристики, які є резервними.

9 – *інше*. Характеристика лексеми відповідного семантичного типу, яка не підпадає під жоден із перерахованих класів.

В таблиці 1 наведено семантичну параметризацію словникових одиниць, які були виявлені при аналізі російських, англійських та українських текстів, спеціальної військової тематики.

Для автоматизації процесу формування словників екстралінгвістичних знань створено АРМ-«ЕКСПЕРТ». Розроблений програмний продукт підтримує англійську, російську та українську мову. Для кожної мови створюється окрема база даних, яка залучається у відповідності із мовою аналізованого тексту. Бази даних незалежно від мови мають єдину уніфіковану семантичну параметризацію, оскільки відбивають однакові фрагменти знань про навколишній світ. Крім того, існують специфічні лексичні одиниці, які є загальноприйнятими в заданій ПО (наприклад: *омбр* – окрема механізована бригада). З цією метою в АРМ-«ЕКСПЕРТ» аналізується показова вибірка текстів заданої тематичної спрямованості і поповнюється база даних відповідної мови.

Лінгвістичний рівень організації тексту представлений морфемним, синтаксичним й власне семантичним аналізом.

Таблиця 1

Приклад семантичної параметризація лексем на рівні організації тексту як знакової системи

Код сем./кл.	**Приклади**	**Інтерпретація**
12!	*Asia, Africa, Europe, Средиземное море, Тихий океан*	Географічна назва: характеризується простором
21!	*WorldWarII*, *Перша світова війна*	Історична подія: характеризується часом
22!/21!	*Брестский мир*	Історична подія: характеризується часом і простором
35!	*Тарас, Martha, Александр*	Ім'я: особа
45!	*Раданаціональноїбезпекиіоборони,theNational SecurityCouncil,*	Установа
51!	*January, Monday, хв.,р.*	Одиниця вимірювання: характеризується часом

52!	*См, кт, mm*	Одиниця вимірювання: характеризується простором
55!	*Омбр*	Одиниця вимірювання: структурний підрозділ
59!	*MHz*	Одиниця вимірювання: характеристика не визначена
65!	*Верховна рада, Дума,*	Власна назва, що транслітерується.
76!	*Президент, Верховний Головнокомандувач*	Посада: особа

Семантична модель морфологічного рівня мовної системи представлена словозмінною та словотвірною моделями. Одиницею змісту на цьому рівні виступає *морфема*. При зазначеному підході словозмінна модель, виконуючи функції автоматичного морфологічного аналізу, в кінцевому результаті містить такі семантичні ознаки, які «істотність» та значення «відмінку» для дієслів, що в практичному плані дозволяє автоматично реалізувати лексико-семантичні валентності дієслівних форм (фактично це аналог семантичних відмінків Ч. Філмора) [5].

Словотвірна модель вхідної мови призначена для розпізнавання «нових» для системи слів, тобто слів, які утворюються за рахунок продуктивних суфіксів і префіксів в мові. Під словотвірним значенням будемо розуміти відношення між двома однокореневими словами, значення одного з яких визначається або через значення іншого (напр., *дом – домик,* «маленький дім», *победить – победитель* «той, хто переміг»), або тотожно значенню іншого у всіх своїх компонентах, крім частинномовного значення (*near – nearly, белый – белизна*).

Задачею побудови семантичної моделі словотворення – визначення значень афіксів, які беруть участь у словотворенні та значення словотвірного форманта. З цією метою формується лінгвістична база даних словотвірних формантів. Слід зазначити, що ми не ставили своїм завданням побудувати повну словотвірну модель мови. В розроблюваній СМП словотвірна модель призначена для підтримки електронних перекладних словників, оптимальних за розміром. Крім того, дослідження такого явища як словотворення у мові дозволяє виявити закономірності і спрогнозувати найбільш вірогідні форманти появи нових слів. З цією метою досліджувалися паперові перекладні словники різних років на предмет появи нових слів у часовому вимірі. Дослідження, зокрема показали, що близько 60% нових слів утворені саме за рахунок афіксального словотворення. Це, на наш погляд, підтверджує висунуту гіпотезу, що фахівець при перекладі користується не тільки безпосередньо перекладними еквівалентами, але й у разі їх відсутності враховує похідність слів.

Семантична модель синтаксичного рівня мовної системи підтримується словниками понять і термінів, що представлені словосполученням та словниками лексико-семантичних валентностей дієслова. На етапі контекстного синтаксичного аналізу програмний модуль інтерпретації виявляє терміни і стійкі словосполучення за словником стійких словосполучень і формує єдине нероздільне поняття, яке розглядається як одна лексема (напр.: *Військовий інститут Київського національного університету імені Тараса Шевченка*). Алгоритм пошуку працює наступним чином: спочатку словосполучення, яке виділилося на етапі контекстного аналізу, приводиться до початкової форми, після чого перевіряється у словнику понять (іменникових синтаксичних конструкцій), якщо словосполучення знайдено, то словосполученню присвоюється значення перекладного відповідника, якщо словосполучення відсутнє у перекладному словнику, то словоформи, що входять до словосполучення повертаються до свого текстового представлення. Інваріанти синтаксичних зв'язків при цьому зберігаються. Крім того, до кожної з виявлених синтаксичних сполук визначається головне слово. За правилами лематизації перевіряємо синтаксичні сполуки на семантичну єдність.

На другому етапі аналізу (побудова дерева синтаксичного підпорядкування) перевіряються у словнику лексико-семантичних валентностей стійкі дієслівні синтаксичні конструкції, які відповідають в мові лексико-семантичним валентностям. Словник дієслівних синтаксичних конструкцій може містити як прості синтаксичні конструкції (*мати значення*), так і складні (*забезпечувати недоторканість від дій або впливу противника*). Словник лексико-семантичних валентностей, з одного боку, дозволяє уникнути хибних синтаксичних конструкцій (з точки зору семантики), з іншого – сформулювати вимоги до довжини представлення стійких дієслівних конструкцій у перекладному словнику.

Власне автоматичний семантичний аналіз в СМП призначений для побудови поняттєвої структури тексту і підтримується словниками тезаурусного типу. На вході автоматичного семантичного аналізу тексту ми маємо дерево синтаксичного підпорядкування, розмічене відповідно до категорій синтаксису (підмет, присудок, означення, додаток тощо). Задача семантичного аналізу – перевести категорії синтаксису в категорії семантики (суб'єкт, об'єкт, відношення, характеристика суб'єкта (об'єкта), характеристика відношення. Для понять, відношень та характеристик тезауруси будуються окремо. Ці словники визначають парадигматичні відношення в мові. Ці відношення на відміну від тезаурусу з ПО, містять знання про загальну картину світу на рівні енциклопедичних знань. Такі знання, як правило, у тексті в явному вигляді не подаються, а проявляються лише наслідки (тобто похідні відношення). Людина при сприйманні ПМТ підсвідомо використовує такий тезаурус і на основі його, власне, і відбувається комунікативний акт (тобто процес розуміння і спілкування між людьми).

В основі тезауруса понять лежать семантичні характеристики іменника, які представлені в табл. 2.

Таблиця 2

Класифікатор понять

№ с/к	Найменування семантичного класу	Приклади
1	Поняття-предмети	Стіл, ножиці
2	Поняття-речовина	Крупа, отруйний газ
3	Поняття-живий організм	Собака, бактерія
4	Поняття-факт	Пожар, спектакль
5	Поняття-якість	Доброта, любов
6	Поняття-стан	Радість, горе
7	Поняття-процес	Розв'язання, збирання
…	…	…

В основі тезауруса відношень лежать семантичні характеристики дієслова, які представлені в табл. 3. В основі тезауруса відношень лежать семантичні характеристики прикметника (прислівника), які представлені в табл. 4.

На останньому етапі семантичного аналізу всі фрагменти тексту об'єднуються в єдину логіко-семантичну структуру, яка фактично відображає синтагматичні відношення (тобто ті відношення, що проявляються безпосередньо в ПМТ). Завдання цього етапу – визначити поняття і відношення у часі і просторі відносно описаного фрагменту навколишнього світу. При цьому, обробка полягає в узагальненні та уніфікації понять, відношень та їх характеристик. Сутність процесу становить виділення головної ядерної структури, тобто структури, яка відбиває, про що йдеться у тексті. Така структура формалізується у вигляді ядерного ланцюга: **S (суб'єкт) → A (дія) → O (об'єкт)**. Така ядерна структура є універсальною і має безліч інтерпретацій («хто про що каже», «хто що робить» тощо), її легко трансформувати до предикатної структури, яка безпосередньо обробляється в системі. Наповнюється ця модель безпосередньо з тексту.

Таблиця 3

Класифікатор відношень

Тип відношення	**Найменування відношення**	**Позначення відношення**
Часові	Бути одночасно	R_{11}
	Бути раніше	R_{12}
Просторові	Знаходитись в ☒-оточенні	R_{21}
	Знаходитись	R_{22}
	Знаходитись позаду	R_{23}
	…	R_{2r}
Динамічні	Рухатися до	R_{31}
	…	R_{3n}
Класифікаційні	Належати до класу	R_{41}
	Мати (властивості)	R_{42}
	…	R_{4k}
Ідентифікуючі	Мати ім'я	$\square$
Прагматичні	Слугувати для	R_{51}
	Мати стан	R_{52}
	Бути перепоною	R_{53}
	…	

Таблиця 4

Класифікатор характеристик

№ с/к	**Найменування семантичного класу**	**Приклади**
1	Характеристика-якість	Добрий, цінний
2	Характеристика-розмір	Великий, малий
3	Характеристика-простір	Далекий, близький
4	Характеристика-час	Старий, сьогодні, вчорашній
5	Характеристика-процесу	Швидко, поволі
6	Характеристика-стан	Радісно, весело
7	Характеристика-колір	Зелений, синій
8	Характеристика-матеріал	Залізний, шкіряний
9	Характеристика-форма	Квадратний, круглий
10	Характеристика-національність	Англійський, китайський
…	…	….

Слід зазначити, що наведені таблиці 2-4 не є повними, оскільки нашою задачею було показати принцип формалізації семантичних ознак для понять, відношень і характеристик. Змістове наповнення може варіюватися від цілей щодо глибині проникнення у семантику.

Для адекватного перекладу важливо зберігати цілісність тексту, оскільки графічно оформлений текст виступає як окрема одиниця змісту. У зв'язаному тексті практично кожне речення інтерпретується реципієнтом відносно інтерпретації інших речень (відносно «знань про мову» – мовна компетенція реципієнта, або відносно екстралінгвістичної ситуації – фахова компетенція реципієнта). Виходячи з цього, на останньому етапі семантичного аналізу ми визначаємо міжфразові логіко-семантичні відношення. Класифікатор логіко-семантичних відношень для міжфразового синтаксису представлений в таблиці 5.

Висновки.

1. Модель семантики в розроблювальній СМП є розподіленою. Це пов'язане з тим, що текст розглядається нами як взаємодія трьох систем: знакової, мовної і системи знань про навколишній світ. Кожна з цих систем має свої (властиві лише їй) одиниці змісту і засоби формалізації семантики.

Таблиця 5

Класифікатор логіко-семантичних відношень міжфразових одиниць тексту

Сем.код.	**Найменування відношення**	**Приклад**
1	доповнення	Це підтверджується шкалою … **Дана** шкала має ….
2	локалізація	Семантика наукової інформації більш мінлива, ніж форма…. **Вона** може навіть відриватися від реальної дійсності…..
3	перефразування	Текст характеризується смисловою єдністю. **Іншими словами** семантика окремих його одиниць …….
4	підтвердження	В наибольшей степени подвержена изменениям оперативная

		информация… **Действительно** энциклопедические знания…
5	каузатив	Текст являє собою складну багаторівневу структуру. **Тому** його доцільно …..
6	протиставлення	Главная задача машинного перевода – передача значения… **Однако** история развития машинного перевода…..
7	уточнення	Інваріантність інформації не є її універсальною властивістю. **Наприклад**, естетична інформація …….
…	…	…

2. Мінімальною одиницею змісту (смислу) на знаковому рівні організації тексту є графема. Етап розпізнавання смислу на знаковому рівні дозволяє розв'язати наступні задачі: сформувати лексичні класи змістовно значущих понять в тексті; сформувати семантично правильні речення в тексті; сформувати змістовно закінчені фрагменти в тексті; визначити відношення між переліченими одиницями тексту, які проявляються на знаковому рівні представлення тексту.

3. Відмінною рисою семантичного аналізу тексту як системи знань про мову є розподілений аналіз морфологічного, синтаксичного і семантичного рівня мови.

Одиницею змісту на морфологічному рівні мовної системи виступає морфема. Семантична модель аналізу морфологічного рівня включає словозмінну і словотвірну моделі мови. Введення словотвірної моделі мови дозволило врахувати закономірності утворення «нових» слів на основі суфіксального і префіксального словотворення.

Одиницею змісту на синтаксичному рівні мовної системи виступає словосполучення. Лінгвістичне забезпечення семантичного на цьому рівні представлено словником понять (стійких іменникових синтаксичних конструкцій) і словником лексико-семантичних валентностей дієслова. Також до цього рівня відносяться і словники фразеологізмів.

Мінімальною одиницею змісту на семантичному рівні мовної системи виступає поняття, максимальною – надфразова конструкція. Семантичний аналіз ПМТ забезпечується словниками тезаурусного типу.

Список використаних джерел

1. Марчук, Ю.Н. Комп'ютерна лінгвістика. – М., 2007.
2. Шенк, Р. Обработка концептуальной информации. – М., 1980. – 360с.
3. Tsujii, J. Machine Translation: Productivity and Conventionality of Language // Current Issues in Linguistic Theory – 136. Recent Advances in Natural Language Processing / Ed. By R. Mitkov et al. Lohn Benjamins Publ. Cj. – Amsterdam; Philadelphia, 1997. – P. 377-392.
4. Мельчук, И.А. Опыттеориилингвистическихмоделей "Смысл-Текст". – М., 1974. – 314 с.
5. Филлмор, Ч. Дело о падеже // Новое в зарубежной лингвистике. – М., 1981. – Вып. 10. – С.350-359.

Надійшла до редколегії 02.02.13

А.А. Рось, д-р техн. наук, проф.,
А.Ю. Николаевский, соискатель,
КНУ имени Тараса Шевченко

ТРЕБОВАНИЯ К ФОРМАЛИЗАЦИИ СЕМАНТИЧЕСКОЙ ИНФОРМАЦИИ В СИСТЕМЕ МАШИННОГО ПЕРЕВОДА

В статье предложен подход к семантическому анализу естественно-языкового текста на основе построения распределенных семантических моделей на разных уровнях представления текста. Текст рассматривается как взаимодействие трех систем: семиотической, языковой, системы знаний о мире. Определены и обоснованы минимальные единицы семантики для каждой системы. Показана практическая реализация предложенного подхода при построении систем машинного перевода.

Ключевые слова: семантический анализ, семиотическая система, языковая система, семантические единицы, естественно-языковой текст, система знаний о мире, система машинного перевода.

A.O. Ros', Doctor, Professor,
O.Y. Nikolaievs'kyi, Candidate for a Degree,
Taras Shevchenko National University of Kyiv

REQUIREMENTS FOR THE FORMALIZATION OF SEMANTIC INFORMATION IN THE MASHINE TRANSLATION SYSTEM

The paper proposes an approach to semantic analysis of natural-language text on the basis of distributed semantic models at different levels of text representation. The text is considered as the interaction of three systems: a semiotic, linguistic, and system of knowledge about the world. The minimal semantic units for each system were defined and justified. A practical implementation of the proposed approach for building machine translation systems has been demonstrated.

Keywords: semantic analysis; semiotic system; language system; semantic units, natural-language text; a system of knowledge about the world; machine translation system.