

6. Теория и практика применения технических средств в обучении иностранным языкам / Под. ред. М.В. Ляховицкого. – К.: Вища школа. Головное изд-во, 1979. – 256с.

7. Hüllen W. Geschriebene Sprache als Lernhilfe im English unterricht/ – “Fremdsprachen. Lehren und Erlernen”, München, 1972.

*Статья посвящена проблеме формирования навыков аудирования. Автор рассматривает принципы использования звукозаписи вовремя формирования навыков аудирования, а также анализирует использование этих принципов в контексте обучения аудированию курсантов (студентов) высших военных учебных заведений. В статье подчеркивается, что проблема восприятия иностранного языка на слух имеет огромное значение с точки зрения методики преподавания иностранного языка.*

*The article deals with the listening comprehension skills forming. The author examines principles of sound recording use while forming of listening comprehension skills. He also analyzes use of these principles in context of listening comprehension skills forming of cadets (students) of higher military educational establishments. The article stresses that the problem of listening comprehension is of particular importance in terms of the methodology of teaching of a foreign language.*

**УДК 811.93 + 003.295**

**Балабін В. В., Ленков С. В., Замаруєва І. В.,  
Пампуха І. В., Литвиненко Л. О., Ніколаєвський О. Ю.**

### **ОБҐРУНТУВАННЯ ВИБОРУ ТЕХНОЛОГІЙ ТА ПРИНЦИПІВ ПОБУДОВИ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ПРИ РЕАЛІЗАЦІЇ ПРОЕКТУ „СИСТЕМА МАШИННОГО ПЕРЕКЛАДУ”**

*У статті розглянуто існуючі підходи до автоматизації перекладу, сформульовано функціональні та технічні вимоги до програмного забезпечення, що використовується для побудови запропонованої моделі перекладу текстів. А також вказані переваги обраних для реалізації технологій.*

**Постановка проблеми.** Вважається, що історія машинного перекладу (МП) почалася з 1954 р., коли співробітники Джортаунського університету США під керівництвом Леона Достера за допомогою ЕОМ переклали 60 фраз російської мови англійською на основі словника в 250 слів. Намагання створити системи автоматизованого перекладу почалися ще наприкінці 60-х років. З того часу намагання створити адекватні системи перекладу, схожі на такі, які робить людина, й досі в повній мірі не реалізовані. Загалом, можемо виділити дві основні проблеми:

- переоцінювався ступінь можливості реалізацій на ЕОМ операцій, що є подібними до елементарних функцій людського інтелекту;
- недооцінювалась складність природної мови (нерегулярної мови, якщо говорити суто математично) як інструмента інтелектуальної діяльності людини і перекладу як одного з найскладніших видів цієї діяльності.

Новому поштовху до розробки систем машинного перекладу сприяли досягнення в галузі штучного інтелекту, зокрема в частині комп’ютерного моделювання процесів «розуміння» природно-мовних текстів (ПМТ). Розглянемо існуючі моделі розуміння ПМТ.

**Аналіз останніх досліджень та публікацій.** Модель "семантики переваг", розроблена під керівництвом Уїлкса в Стенфордському університеті (США), і є розвитком досліджень Кембриджської лінгвістичної школи. Модель базується на таких принципових положеннях [1]:

- існує обмежене число різноманітних висловлювань;
- існують формальні засоби опису семантики висловлювань;

- існує формальний метод поєднання висловлювань у речення та текст.

Текст характеризується такими складовими: смислом (значенням) окремих слів, висловлювань та речень і правилами їх семантичного поєднання. Аналіз тексту проводиться за декілька етапів:

1. Спочатку за допомогою спеціальних маркерів (прийменники, сполучники, синтаксичні знаки тощо) речення розбивається на фрагменти.
2. Кожному слову виділеного фрагменту речення надаються відповідні значення із словника.
3. Кожний виділений фрагмент за певними правилами зіставляється з простими висловлюваннями, що містяться в базі даних. Зразок вважається накладеним (істинним), якщо кожний з його елементів розбивається на елементи одного із значень слів, які входять до фрагменту.
4. Спеціальні правила розширення перетворюють простий зразок у повний шляхом додавання тих слів, які не увійшли до простого зразка.
5. Після отримання повного зразка алгоритм переходить до обробки наступного речення.
6. Коли повні зразки отримані, починає працювати процедура аналізу їх на семантичну сумісність.

Модель працює з цілісними текстами на логіко-семантичних засадах і не передбачає попереднього морфологічного та синтаксичного аналізу тексту. Модель "семантики переваг" має ряд недоліків, які обмежують її реалізацію в системах обробки текстової інформації, що використовуються на практиці, а саме: для отримання повного зразка може підійти декілька простих зразків (у силу мовної омонімії); використання для семантичного аналізу жорстких шаблонів фактично зводить вільні словосполучення до різновиду стійких (цю ідею реалізує система RETRANS), що з одного боку робить аналіз доволі приблизним, а з іншого – не всі випадки семантичних сполучень система може проаналізувати.

Модель "концептуальних залежностей" розроблена під керівництвом Р. Шенка (США). Початкові розробки можна охарактеризувати як розширений семантичний аналіз дієслів у поєднанні із семантичними відмінками Ч. Філмора. Будь-яка дія має словниковий список учасників та їх взаємовідношень. Відповідна інформація надається в термінах простих дій. Процес визначення семантичного значення фрази являє собою виділення учасників і об'єднання індивідуальних схем елементарних дій. Основою формального подання моделі є семантична мережа. Вершини мережі позначені символами семантичних одиниць: PP – характеристика об'єкта, АСТ – назва дії, АА – характеристика дії, State – стан дії, Loc – місто дії, T – час дії. Останні розробки спрямовані на перехід від семантики до прагматики. Фактично це означало перехід від аналізу дієслів (та їх текстового оточення) до аналізу подій. Крім того, дещо змінився підхід до обробки текстів, сутність якого можна сформулювати таким чином: синтаксичні і семантичні структури обробляються одночасно; закономірності синтаксису і семантики проявляються в ході одного процесу; обробка мовних одиниць за своєю сутністю ототожнюється з обробкою пам'яті.

На базі моделі "концептуальної залежності" створено ряд систем: Spinosa-1, Spinosa-2, MARGE, SAM, PAM. Так, система MARGE працює в режимах "перифразування" та "логічних виводів". У першому режимі задача полягає в побудові концептуального представлення за вхідним реченням і генерації перифраз за цим представленням. У другому режимі система будує різні логічні виводи за вхідним реченням. Недоліком моделі є аналіз подій тільки на одному рівні детальності, що не дозволяє описувати складні події через сукупність простих, крім того, не враховується складність опису об'єктів. Прагматичний аналіз зводиться фактично до фахового семантичного поля і не враховує конкретну задачу щодо обробки тексту, за рахунок чого можна побудувати досить велику кількість перифраз. На основі моделі "концептуальної

залежності" в Україні розроблено експериментальну систему КОНТЕНТ (Інститут програмних систем НАН України) для обробки великих обсягів природно-мовної текстової інформації [3].

Модель "СМИСЛ ↔ ТЕКСТ" (розроблена під керівництвом проф. І.А. Мельчука, Московська лінгвістична школа) орієнтована на комп'ютерну обробку інформації і є найбільш відпрацьованою для російської мови [2]. В основу моделі покладені такі принципи положення: синонімія мовних висловів виникає не лише за рахунок лексичних синонімів – це можуть бути синтаксичні варіанти подання одного значення слова, які по-різному орієнтують певну ситуацію відносно її учасників; комунікативна організація тексту містить: розподілення на тему і рему, виділення "старої" і "нової" інформації, виділення більш-менш важливих частин змісту; носіями смислу є не лише слова, а й морфеми.

У моделі було запропоновано мову семантичних множників для формалізації опису сутності і міри семантичних тотожностей лексично різних одиниць. Під множниками в даному випадку розуміються елементарні смисли, різні комбінації яких відповідають складним лексичним значенням слів. Значення слова при цьому не є простою сукупністю множників, а певним чином організована структура. У моделі прийняті такі рівні аналізу тексту: семантичний, синтаксичний, морфологічний. Семантичне подання (СЕМП) містять два компоненти: семантичний граф і відомості про комунікативну організацію смислу. Семантичний граф – це орієнтований граф, вершини якого позначені символами сем (атомів смислу). Семи класифікуються на два типи: предикати та імена об'єктів. Логічні зв'язки та квантори трактуються як предикати.

На базі моделі реалізовано ряд систем, серед яких система діалогової взаємодії "ПОЭТ" та система машинного перекладу "ЭТАП" (під керівництвом Ю.Д. Апресяна, Москва) [4]. Загальним недоліком цих систем є те, що вони не враховують закономірності опису знань на знаковому рівні подання тексту, що у свою чергу призводить до грубих помилок семантичного аналізу тексту. Крім того, у моделі відпрацьований аналіз, а синтез, тобто перетворення внутрішньої структури до текстового подання, заявлений лише концептуально.

Порівняльну характеристику існуючих моделей розуміння ПМТ з точки зору рівнів організації тексту наведено в таблиці 1.

Таблиця 1.

### Порівняльний аналіз моделей розпізнавання текстових об'єктів

№ п/п	Моделі розуміння текстової інформації	Рівні подання тексту				
		Знаковий	Морфологічний	Синтаксичний	Семантичний	Прагматичний
1.	Модель "семантики переваги"	-	-	-	+	+ -
2.	Модель "концептуальних залежностей"	-	-	+	+	+ -
3.	Модель "смысл ↔ текст"	-	+	+	+	+ -
4.	Логіко-семантична модель	+	+	+	+	+

Вибрана логіко-семантична модель забезпечує обробку ПМТ на всіх рівнях і реалізує знання-орієнтований підхід до розробки систем МП, який базується на розпізнаванні, вилученні й обробці знань про світ (предметну галузь), що містяться в ПМТ, запропонований в [5]. Моделювання процесу розуміння зводиться до розробки засобів розпізнавання і формалізації знань, які містяться у вхідному тексті, інтерпретації

цих знань щодо цільової настанови певної прикладної задачі (перекладу, реферування, переказу тощо) та синтезу результату у вихідний текст. Запропонований підхід ґрунтується на таких концептуальних положеннях: вхідний природно-мовний текст (ПМТ) – є зв’язний текст (тобто дискурс); зв’язність дискурсу забезпечується графемними засобами оформлення тексту (наприклад, відношення взаємозв’язку між заголовками фрагментів тексту і змістом абзаців), лінгвістичними засобами (граматичними узгодженнями, анафоричними посиланнями тощо) і екстралінгвістичними (наприклад, часові, причинно-наслідкові зв’язки в заданій предметній галузі); всі ці засоби є інструментом кодування знань про світ; як елементи реального або абстрактного світу виступають його об’єкти, відображені в тексті у формі природно-мовних понять, відношень між ними та мовних характеристик понять і відношень; формування об’єктів за знаннями про світ може бути різним і залежить від цільової настанови прикладної задачі, проблемної галузі та логічної картини світу носія інформації (мови).

**Формулювання цілей статті.** Мета запропонованої статті полягає в обґрунтуванні вибору технологій та принципів побудови програмного забезпечення при реалізації проекту „система машинного перекладу”.

**Виклад основного матеріалу.** Логіко-семантична модель подання знань з предметної галузі, які містяться в ПМТ, фактично є технологічним ядром системи МП. Основними компонентами знань щодо їх формалізованого подання є поняття, відношення між ними, характеристики понять і відношень, а також модальності цих характеристик. Отже, обробка вхідного тексту має бути спрямованою на виявлення (розпізнавання) в тексті основних компонентів знань і встановлення логіко-семантичних відношень між ними з метою формування поняттєвої структури (ПС) змісту вхідного тексту. До формалізованого подання знань висувуються такі вимоги: по-перше, воно має бути подано в такому вигляді, який забезпечить можливість коректної логіко-семантичної обробки знань; по-друге, воно має містити всю необхідну інформацію для вирішення певної задачі з обробки ПМТ, тобто максимально повно на змістовому рівні зберігати текстове представлення елементів знань.

Особливістю ПС є її гібридне подання, яке поєднує в собі властивості семантичних мереж і предикатних моделей (в якості вершин мережі виступають елементарні предикатні формули). З метою уніфікації структурного подання відношень, які в тексті можуть мати різну кількість аргументів, у ПС використовуються тільки одно- та двохмісцеві предикати, для чого розроблено метод декомпозиції  $n$ -місцевих предикатів і предикатів вищих порядків на двохмісцеві предикати першого порядку. Для відображення рольових відношень введено поняття неявного предикату. Під неявним предикатом розуміється відношення, яке не має відповідного лексичного еквіваленту в тексті. Наприклад, у словосполученні *фірма* “Лінгвістика-93” відношення “мати назву” між аргументами “фірма” та “Лінгвістика-93” відображається в тексті пробілом. З метою зберігання виразових засобів природно-мовного текстového представлення введені спеціальні засоби – префікси і постфікси предикатів і понять. Елементарна предикатна формула може містити також квантори єдності ( $\forall$ ) та існування ( $\exists$ ).

Елементарна предикатна формула має вигляд:  $NP_k^q(LX_t^i, MY_g^j)$ ;

де  $N, L, M$  – відповідно префікси предиката та аргументів, які визначають тип семантичного класу;  $P$  – назва семантичного класу відношення;  $X, Y$  – назви семантичних класів понять. Аргументи мають фіксоване положення. Формула інтерпретується в термінах класичного числення предикатів: ”поняття  $X$  знаходиться у відношенні  $P$  до поняття  $Y$ ”. В якості постфіксів виступають верхні та нижні індекси предикату і аргументів. Верхній індекс предикату  $q$  ( $q \in Q$ ) визначає лексико-граматичний спосіб мовного сполучення відношення і понять у тексті. Множина  $Q$  являє собою перелік як мовних одиниць (наприклад, прийменники, частки тощо), так і граматичних ознак (наприклад, відмінок керування між дієсловом і відповідним іменником), які відбивають правила сполучення відношень і понять у тексті. Нижній індекс предиката  $k$  визначає

конкретний лексичний представник для відповідного семантичного класу  $N$ . Верхні індекси аргументів  $i$  та  $j$  ( $i, j \in A$ ) визначають граматичні характеристики понять (наприклад, число, істота, неістота тощо). Множина  $A$  є список граматичних характеристик понять, які виступають в якості аргументів предикатної формули. Нижні індекси аргументів  $t$  і  $g$  ( $t \in L, g \in M$ ) визначають конкретний лексичний представник відповідних семантичних класів. У процесі формально-логічного виведення постфікси ігноруються. Вони є вирішальними на етапі синтезу опису фрагментів ПС природно-мовними засобами. В окремих лексико-семантичних класах відношень виділені лексичні одиниці, які мають значення модальності (*хотіти, вміти, треба, необхідно* тощо).

Уніфікація лексичних представників понять і відношень у рамках певного семантичного класу здійснюється за відношенням “*рід-вид*”. Структура семантичного класу являє собою ієрархічну будову, на верхньому рівні якої знаходяться найбільш загальні поняття (відношення), кожний нижчий рівень представляється поняттями (відношеннями), які конкретизують відповідні поняття (відношення) вищого рівня. Вибір родо-видового відношення для уніфікації понять і відношень з предметної галузі має принципове значення. Як доведено в [95], заміна видових понять (відношень) на родові у вільних словосполученнях не веде до порушення семантичного значення вислову (чого не можна сказати про інші ієрархічні відношення).

Для кожної елементарної предикатної формули будується матриця істинності  $C = \|C_{ij}\|$  розміру  $m \times n$ , яка відбиває можливі значення аргументів для певного предикату. За своєю суттю ця матриця є фрагментом знань про відношення між поняттями в предметній області. Стовпці матриці відбивають допустимі значення аргументу  $X$ , а рядки – аргументу  $Y$ . Елементи матриці  $C_{ij}$  приймають значення: “1” (предикат має значення “істина”, тобто відповідні аргументи  $X$  і  $Y$  є допустимими для предикату), “0” (предикат має значення “хиба”, тобто відповідні аргументи  $X$  і  $Y$  є недопустимими для предикату) і “?” предикат має значення “умовна істина”, тобто допустимість відповідних аргументів  $X$  і  $Y$  має бути довідзначеною, виходячи з аксіоматичної моделі та мовного контексту. Слід зазначити, що матриця істинності визначає лише семантичну коректність певної елементарної предикатної формули. Прагматична оцінка здійснюється на основі знань (як декларативних, так і процедурних) про певну прикладну задачу.

Довічною проблемою розпізнавання є розв’язання полісемії та нових для системи слів значень. Пропоноване подання ПС дозволяє як розпізнавати “нові” слова, так і здійснювати вибір на множині можливих значень слів. Обробка нових (незнайомих системі) слів: якщо слово (лексичний представник поняття або відношення) відсутнє, то на основі матриці істинності можна вибрати з відповідного лексико-семантичного класу представника із загальнішим значенням. Так, наприклад, якщо для англійського словосполучення *go by bicycle* у перекладному словнику відсутня словникова стаття: *bicycle – велосипед*, то за допомогою аксіоматичної моделі, програма інтерпретує переклад як *їхати транспортним засобом*. Якщо програма не розпізнала лексичного представника відношення, то в тексті вибираються всі контекстні зв’язки (тобто способи його вживання з іншими поняттями) і підбирає найбільш вірогідну (адекватну) аксіоматичну модель, потім присвоює ім’я відношення з найбільш загальним значенням. Звичайно, щоб знайти відповідний еквівалент для відношення (поняття), треба мати для нового слова контекст, який визначається двома або більшою кількістю елементарних предикатних формул. Вже при визначенні трьох елементарних предикатних формул алгоритм працює з достовірністю >70%. Отже, при аналізі нових слів програма може порушувати стилістичну цілісність тексту, але зберігає його семантичну цілісність, що є вкрай важливим і задовольняє вимогам аналізу науково-технічних текстів.

У цілому розробка системи та реалізація у вигляді кінцевого програмного забезпечення на основі запропонованої логіко-семантичної моделі спрямована на автоматизацію широкого кола задач інформаційно-аналітичної діяльності (ІАД) фахівців з аналізу різномовних текстових джерел на єдиних математичних та інструментальних засадах.

Метою розробки автоматизованих систем машинного перекладу (АСМП) на базі знання-орієнтовного підходу є не тільки створення системи для адекватного перекладу текстів, а й задоволення сучасних потреб користувачів під час роботи з програмним забезпеченням, а також автоматизація інформаційно-аналітичної діяльності адміністраторів системи, а також фахівців з предметних галузей (лінгвістів), що є необхідним для реалізації знання-орієнтовного підходу під час розробки системи.

Програмне забезпечення (ПЗ) АСМП має підтримувати багатомовний інтерфейс, насамперед, такі мови: українська, російська, англійська. Вхідні ПМТ мають задовольняти вимогам орфографічної, синтаксичної та стилістичної коректності перед передачею їх до обробки системою. Тому необхідно створити інструментарій для автоматичної перевірки вхідних текстів та оптимізації і виявленню можливих помилок вхідних даних. Електронний перекладний словник має бути відкритим щодо його поповнення користувачем. При цьому від користувача не повинно вимагатись спеціальних знань з лінгвістики. Крім того, має бути реалізована можливість швидкого поновлення словників користувачами, а також передача словників користувачів на сервер для перевірки лінгвістами на коректність. У функціональному плані програмне забезпечення має задовольняти такі вимоги:

- можливість обробки єдиними інструментальними засобами різномовної текстової інформації;
- можливість автоматизованої інтеграції та накопичення знань у певній предметній галузі;
- можливість виявлення суперечностей в поточній інформації зі знаннями з предметної області;
- можливість узагальнення знань з певної предметної області;
- можливість автоматизованого підключення розширених модулів з необхідною інформацією під час обробки вхідних текстів, перекладу;
- зручність у користуванні (інтуїтивний інтерфейс);
- можливість нарощування нових задач, які виникають у процесі функціонування систем.

У технологічному плані програмне забезпечення має задовольняти такі вимоги:

- невимогливе до ресурсів робочої станції користувача;
- вимагати якомога меншого встановлення додаткових модулів підтримки, окрім власне самої системи;
- простий механізм інсталяції системи;
- модульна система розробки (декілька програмістів одночасно можуть розробляти різні модулі, але системою передбачається їх об'єднання; можливість розширення новими модулями без переробки системи);
- зручний механізм збереження та доступу до даних (оскільки передбачається робота з великим масивом інформації);
- програмне забезпечення має задовольняти вимогам надійності (тобто, повинне мати засоби захисту від неправильних дій користувача, мати засоби відновлення свого функціонування під час збоїв) та ергономічності;
- ПЗ має працювати під операційною системою **MS Windows, як найбільш розповсюдженого серед користувачів.**

Задоволення сформульованих вимог можливе також при виконанні таких двох принципів:

- відокремлення подання знань про мову від алгоритмів лінгвістичної обробки природно-мовних текстів;
- одночасне виконання етапів лінгвістичних досліджень і розробки програмних компонентів лінгвістичного аналізу.

Зазначимо, що модульна реалізація системи звужує використання об'єктно-орієнтовних технологій та середовищ програмування.

Виходячи зі змісту вимог до ПЗ і спираючись на наведені принципи, пропонуються такі середовища реалізації системи.

Вказаним вимогам відповідають сучасні середовища розробки Delphi, MS VisualStudio, C++Builder. З наведених середовищ для реалізації було обрано MS VisualStudio. Основними чинниками, якими ми керувались при цьому стали:

- підтримка технології розробниками операційної системи Windows, що надає можливість використовувати та інтегрувати останні бібліотеки та підвищувати працездатність та стабільність програми;
- зручне середовище та організація роботи для одночасної роботи декількох програмістів;
- можливість швидкого та зручного доступу та обміну інформацією з Базами Даних;
- можливість створення багатомовного інтерфейсу (зручне введення інформації для різних мов та коректне відображення символів);
- наявність ефективних засобів тестування та перехоплення помилок.

Крім того, важливим фактором є те, що отримана можливість не тільки допомагає поєднати роботи програмістів в межах одного проекту, а й створює паралельно два різні програмні продукти, причому, деякі елементи кожного можливо швидко та зручно залучати у інший. Такий підхід надає можливість створювати паралельно системи навчання, тестування, обробки різномовних джерел даних та інші.

Також керуючись вимогами працездатності системи як під існуючі операційні системи, так і під ті, що будуть розповсюджені у майбутньому, це ПЗ використовує не Windows API, а бібліотеки NetFramework [6]. Такий підхід обумовлює те, що система буде працювати не лише під існуючими системами Windows XP та Windows Vista, а й під новою системою Windows 7, що не можна сказати про системи із використанням технологій Windows API.

Технологічна схема розробки і функціонування АСМП на базі знання-орієнтовного підходу представлена на рис. 1.

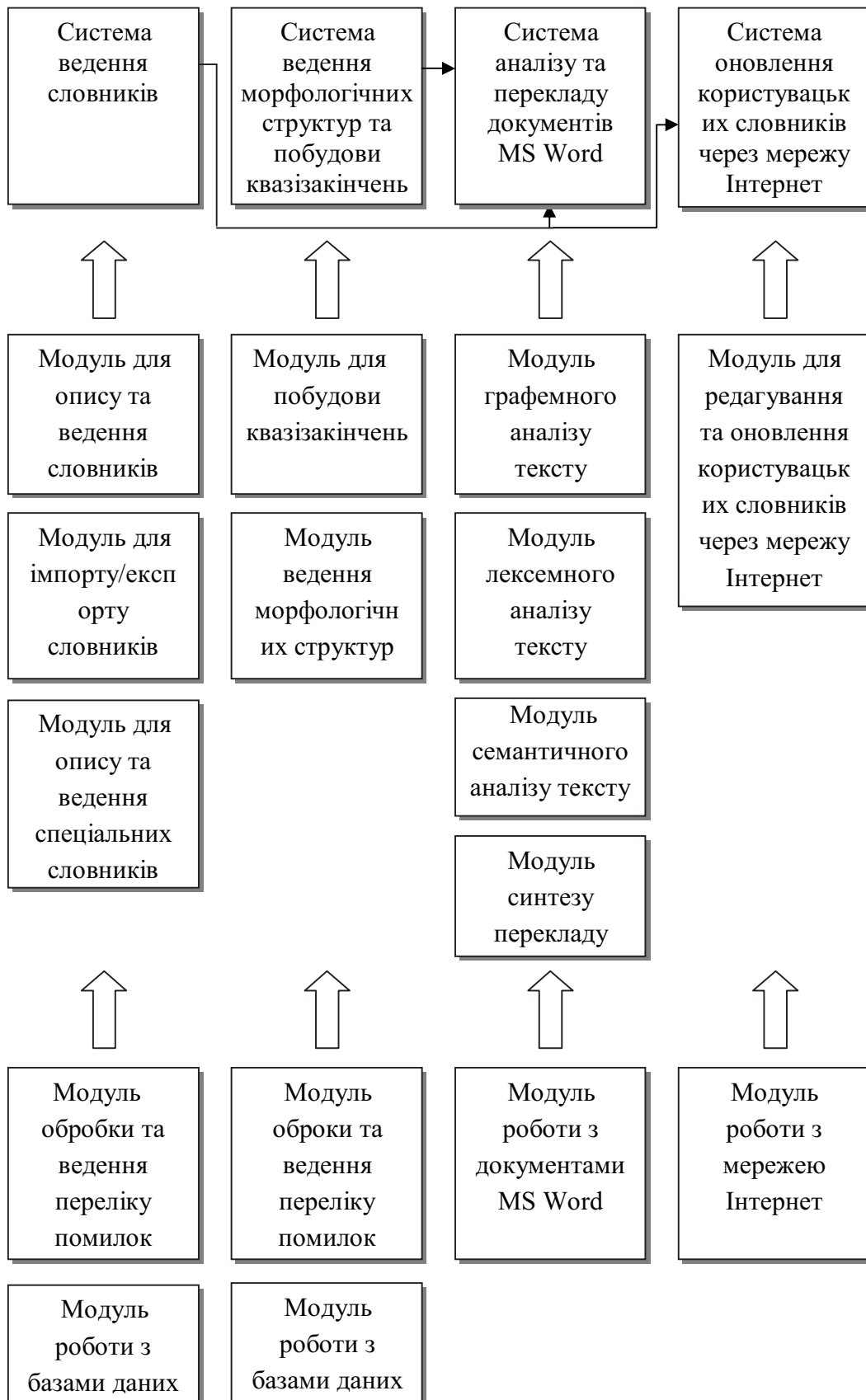


Рис1. Структурна схема АСМП на базі знання-орієнтовного підходу

Представлена структурна схема реалізації АСМП має такі особливості:

- спрямованість технології на автоматизацію вилучення знань про предметну галузь з ПМТ шляхом формування у процесі лінгвістичного аналізу поняттєвої структури тексту;



- комбіноване використання заздалегідь розроблених універсальних компонентів лінгвістичного забезпечення та інструментально-технологічних засобів розробки компонентів ПЗ, частина з яких може бути використаною в якості безпосереднього компонента лінгвістичного забезпечення;
- поєднання на єдиній інструментальній основі етапів лінгвістичних досліджень, відпрацювання правил лінгвістичного аналізу і безпосередньої розробки компонентів ПЗ;
- вилучення програмістів із технологічного циклу створення ПЗ інформаційних систем;
- функціональне розмежування універсальних програмних засобів і відповідних інформаційних компонентів, які забезпечують налаштування підсистем лінгвістичного аналізу на предметну галузь і вхідну мову;
- орієнтація на спрощення роботи фахівців з предметної галузі при створенні й адаптації таких інформаційних компонентів, як словники скорочень, імен, географічних назв, словоформ тощо;
- забезпечення можливості незалежної паралельної (в тому числі щодо розробки програмних компонентів) розробки і відпрацювання всіх інформаційних компонентів.

Використання запропонованої структурної схеми при розробці лінгвістичного забезпечення дозволяє суттєво скоротити час розробки за рахунок використання універсальних програмних компонентів і відокремлювання роботи створення інформаційних компонентів, а також підвищити якість створюваного продукту за рахунок можливості використання інструментально-технологічних засобів розробки і тестування розроблюваних компонентів.

З Рис.1 наочно видно, що ПЗ складаються з таких компонентів:

- система ведення словників;
- система ведення морфологічних структур та побудови квазізакінчень;
- система аналізу та перекладу документів MS Word;
- система оновлення користувацьких словників через мережу Інтернет;

*Модуль ведення морфологічних структур, модуль створення словників квазізакінчень, модуль графемного та лексемного рівнів обробки текстів формують АРМ лінгвіста-розробника, призначений для підтримки функцій:*

- формування правил лінгвістичного аналізу і синтезу ПМТ (лінгвістичної БЗ);
- дослідження правил лінгвістичного аналізу і синтезу ПМТ щодо повноти і суперечливості;
- синтез програмних компонентів лінгвістичного аналізу і синтезу ПМТ;
- формування і ведення словників квазізакінчень;
- формування і ведення лінгвістичної бази даних.

*Модуль опису спеціальних словників, Модуль роботи з мережею Інтернет* призначені для автоматизації функцій:

- формування і ведення перекладних словників;
- формування і ведення бази знань з предметної галузі за ПМТ та в експертному режимі;
- дослідження бази знань з предметної галузі щодо повноти і суперечливості.

*Модуль для редагування та оновлення користувачських словників через мережу Інтернет*

служує для поповнення бази перекладних словників користувачами. Реалізація цього компонента в обраному середовищі (MS VisualStudio) надає такі переваги:

- швидкість обміну інформації та отримання оновлень;
- універсальність щодо типу серверу, який слугує для обміну інформації з користувачем (незалежно UNIX Server або Windows Server чи інші);

– багаторівневий доступ до системи редагування (адміністратор, користувач);

Наслідком залучення такого механізму роботи є те, що власне дійсні користувачі вносять корективи до перекладного словника, а ці корективи одразу можливо внести до системи, що у свою чергу призводить до покращення перекладу.

*Модуль графемного аналізу тексту*

Використовується для попереднього аналізу ПМТ, виявлення можливих помилок в тексті, форматування тексту згідно з встановленими фахівцями-лінгвістами правилами, а також для подальшої передачі на лексемний та семантичний рівні обробки.

*Модуль ведення морфологічних структур, модуль створення словників квазізакінчень*

Використовуються для створення бази знань про мову. Власне з цими компонентами працює фахівець-лінгвіст. Що головне – розроблено алгоритм цифрового представлення морфологічних характеристик для кожного слова, що надає можливість надалі обробляти зручно такі дані на ЕОМ.

*Модуль лексемного аналізу тексту*

Використовується для розбиття поданого ПМТ на структури, які переважно є універсальними для кожної мови, що обумовлює більшу універсальність алгоритму.

Також між деякими структурами встановлюються зв'язки, які визначають певні параметри для структури (слів), та формують основи для подальшої обробки та перекладу. На цьому етапі можуть залучатись спеціальні словники (скорочень, географічних назв, інші)

*Модуль семантичного аналізу*

Призначено для кінцевого формування фраз із визначенням їхнього смислового значення,

Можливе залучення спеціальних словників (словник ідіом, інші).

*Модуль синтезу перекладу текстів*

Власне призначено для отримання перекладу шляхом отримання відповідних значень з перекладного словника згідно з інформацією, яку було передано модулю в результаті роботи Системи аналізу та перекладів документів MS Word.

Крім того, у нижній частині Рис. 1 вказано допоміжні універсальні модулі, які використовує система.

Розроблена структурна схема дозволяє реалізувати знання-орієнтовний підхід у АСМП, а також задовольнити усім вказаним вище вимогам та є масштабованою щодо кількості даних, користувачів, і навантаження на окремі модулі та додавання нових структур. Крім того, обрана мова програмування C# дозволяє перейти в майбутньому на принципово нові технології, які зараз лише розвиваються (наприклад, WPF) щодо роботи з інтерфейсом, локалізацією та програмною реалізацією системи, що надасть можливість роботи за допомогою веб-інтерфейсу. При цьому ядро системи не буде змінено. Таким чином, запропонована модель на декілька років буде відповідати всім сучасним стандартам та вимогам користувачів як щодо якості та надійності, працездатності, так і щодо зручності користування.

Використання запропонованої технологічної схеми при розробці лінгвістичного забезпечення дозволить суттєво скоротити час розробки за рахунок використання універсальних програмних компонентів і розпаралелювання роботи зі створення інформаційних компонентів, а також підвищити якість створюваного продукту за рахунок можливості використання інструментально-технологічних засобів розробки і тестування розроблюваних компонентів.

#### **Література:**

1. Искусственный интеллект. Системы общения и экспертные системы. – Справочник. Кн. 1. /Под ред. Э.В. Попова. – М., 1990. – 315 с.
2. Мельчук И.А. Опыт теории лингвистических моделей "Смысл-Текст". – М., 1974, 314 с.

3. Дерезкий В.А. Об одном подходе к обработке естественно-языковых данных на основе анализа семантических сетей // Праці Першої міжнародної конференції з програмування УкрПРОГ'98. – К.: Кібцентр НАНУ, 2-4 вересня 1998. – С. 405-411.
4. Апресян Ю.Д., Богуславский И.М. и др. Лингвистическое обеспечение в системе автоматического перевода третьего поколения. – М., 1978. – 74 с.
5. Замаруєва І.В. Комп'ютерна модель розуміння природно-мовної текстової інформації // Проблеми програмування. –1999. – №2. – С. 96-102.
6. Кристиан Нейгел, Билл Ивсен, Джей Глинн, Карли Уотсон, Морган Скиннер C# 2008 и платформа. NET 3.5 для профессионалов. Язык программирования C# 3.0 для. NET 3.5 Framework. – М.-СПб.-Киев, «Диалектика», 2007. – 13-92 с.

*В статье рассмотрены существующие подходы к автоматизации перевода, сформулированы функциональные и технические требования к программному обеспечению, которое используется при построении предложенной модели перевода текстов. Также указаны преимущества выбранных для реализации технологий.*

*The article deals with current approaches to the automatic translation process, set up functional and technical requirements for software that is used in the construction of the model is proposed for translation. Also identifies the advantages for the implementation of selected technologies.*

**УДК 37.013.74**

**Марченко Г. В.**

### **ТЕНДЕНЦІЇ РОЗВИТКУ ВИЩОЇ ПЕДАГОГІЧНОЇ ОСВІТИ В КРАЇНАХ ВИШЕГРАДСЬКОЇ ЧЕТВІРКИ (ДРУГА ПОЛОВИНА ХХ – ПОЧАТОК ХХІ СТОЛІТТЯ): СТАН ДОСЛІДЖЕНОСТІ ПРОБЛЕМИ**

*У пропонованій статті автор обґрунтовує актуальність дослідження провідних тенденцій розвитку вищої педагогічної освіти у країнах Вишеградської четвірки в другій половині ХХ – на початку ХХІ століть.*

**Постановка проблеми.** Зміни в економічній, політичній, соціокультурній сферах суспільного життя, що відбуваються сьогодні, глобалізація, інтеграція, інформатизація вимагають відповідного реагування з боку різних держав, і від України у тому числі, для забезпечення економічного процвітання і підвищення рівня добробуту громадян. Визначивши стратегічний напрямок розвитку країни як інтеграцію у світові і європейські економічні структури, орієнтацію і прийняття загальнолюдських цінностей, Україна здійснює численні заходи щодо вирішення означених завдань, чому сприяє посилення демократичних процесів у суспільстві. Одним із найважливіших завдань виступає виховання нової генерації фахівців, здатних до творчого пошуку шляхів розв'язання проблем, які постали перед країною. Якість підготовки професіоналів для різних галузей матеріального і духовного виробництва має забезпечити оновлена система освіти України, і перш за все, система педагогічної освіти, оскільки освіта визнана сьогодні в усьому світі вагомим чинником соціокультурних та економічних змін. Тому думка О. О. Любара та Д. Т. Федоренка про те, що “професійно-педагогічна підготовка вчительських кадрів для національної школи незалежної України повинна здійснюватися паралельно із розбудовою суверенітету нашої молоді держави, а з багатьох позицій – випереджати її”, видається надзвичайно слушною й актуальною.

Одним з головних напрямів реформування системи освіти, як зазначає В. Г. Кремень, виступає “радикальна гуманітаризація”, посилення особистісного компоненту в педагогічній науці; основою освіти ХХІ століття має стати орієнтація на людину, фундаментальні цінності, рішуча демократизація освіти; в основу навчання має