

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
СХІДНОУКРАЇНСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ  
ІМЕНІ ВОЛОДИМИРА ДАЛЯ  
ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ ТА ЕЛЕКТРОНІКИ  
КАФЕДРА ПРОГРАМУВАННЯ ТА МАТЕМАТИКИ

**Пояснювальна записка**  
**до дипломної роботи**  
**бакалавр**  
(освітньо-кваліфікаційний рівень)

**на тему «Комп’ютерна реалізація автоматичної лематизації прикметників  
української мови»**

Виконав: студент 4 курсу, групи ІТ-141  
напряму підготовки 6.040302 „Інформатика”  
спеціальності 7.04030201 „Інформатика”

\_\_\_\_\_ Моторний Є.О.  
(підпис)

Керівник,  
доцент, д.т.н. \_\_\_\_\_ Лифар В.О.  
(підпис)

Рецензент,  
доцент, к.т.н. \_\_\_\_\_ Фесенко Т.М.  
(підпис)

ЛИСТ ПОГОДЖЕННЯ І ОЦІНЮВАННЯ  
бакалаврської роботи студента гр. ІТ-141 Моторного Є.О.

Науковий керівник Лифар В.О., в.о.завідувача каф. ПМ СНУ ім. В. Даля

ПІБ, посада

Оцінка наукового керівника:

Рецензент Фесенко Т.М., доцент каф. ПМ СНУ ім. В. Даля

ПІБ, місто роботи, посада

Оцінка рецензента:

Кінцева оцінка за результатами захисту:

Голова ЕК,  
зав. кафедри ПМ  
д.т.н., доцент

\_\_\_\_\_ підпис    підпис, дата

Лифар В.О.

**СХІДНОУКРАЇНСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ  
ІМЕНІ ВОЛОДИМИРА ДАЛЯ**

Факультет інформаційних технологій та електроніки  
Кафедра програмування та математики  
Освітньо-кваліфікаційний рівень бакалавр  
Напрямок підготовки 6.040302 „Інформатика”  
Спеціальність 7.04030201 „Інформатика”

ЗАТВЕРДЖУЮ  
Завідувач кафедри ПМ,  
д.т.н., доцент  
\_\_\_\_\_ Лифар В.О.  
«\_\_» \_\_\_\_\_ 2018 р.

**З А В Д А Н Н Я  
НА ДИПЛОМНУ РОБОТУ СТУДЕНТУ  
МОТОРНУ ЄВГЕНУ ОЛЕКСАНДРОВИЧУ**

**1. Тема роботи Комп’ютерна реалізація автоматичної лематизації прикметників української мови.**

**керівник роботи Лифар В.О.**

затверджені наказом вищого навчального закладу від “20” лютого 2018 року № 52/48

2. Строк подання студентом роботи 04 червня 2018 р.

3. Вихідні дані до роботи

Об’єктом даної роботи є автоматична обробка природномовних текстів..

3.1 Літературні джерела:

Дарчук Н.П. Комп’ютерна лінгвістика. – К.: Видавничо-поліграфічний центр «Київський університет», 2008. – 351 с.

Зализняк А. А. Грамматический словарь русского языка (словоизменение). 2-е изд., М., "Русский язык", 1980.

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити)

4.1 Вступ

4.2 Аналіз предметної галузі (огляд літератури), з висвітленням наступних питань:

Частина мови і морфологія

Автоматичний морфологічний аналіз, приклади таких систем

4.3 Основна частина, в якій висвітлити:

Теоретичні засади морфологічного аналізу прикметників.

Розробка засобу для автоматичного морфологічного аналізу прикметників

4.4 Висновки

4.4 Перелік використаних джерел

5. Перелік графічного матеріалу немає

6. Дата видачі завдання 12 лютого 2018 року.

## КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів дипломної роботи	Строк виконання етапів роботи	Примітка
1	Одержання завдання на виконання роботи	12.02.18	
2	Укладання і погодження з керівником плану і етапів виконання роботи	20.02.18	
3	Узагальнення даних літературних джерел, укладання розділу «Аналіз предметної галузі»	1.03.18	
4	Аналіз шляхів виконання завдання. Вибір і погодження з керівником оптимального шляху	09.03.18	
3	Аналіз теоретичних засад морфологічного аналізу прикметників	02.04.18	
5	Укладання та тестування програмного продукту	20.04.18	
6	Укладання, оформлення та погодження пояснювальної записки з керівником	10.05.18	
7	Здача готової пояснювальної записки на кафедрі	21.05.18	
8	Укладання доповіді і презентації	01.06.18	

Студент

\_\_\_\_\_

(підпис)

Моторний Є.О.

Керівник роботи

\_\_\_\_\_

(підпис)

Лифар В.О.

## РЕФЕРАТ

Текст — 62 с, таблиць — 1, додатків -2 , джерел — 14.

У ході виконання даної дипломної роботи було проведено дослідження в області автоматичного морфологічного аналізу, вивчені основні особливості обробки українських прикметників.

Мовою Prolog розроблені засоби морфологічного аналізу прикметників, створена база афіксів українських прикметників у формі, придатній для використання в системі автоматичного морфологічного аналізу. Розроблені засоби лематизації прикметників.

Ключові слова: КОМП'ЮТЕРНА ЛІНГВІСТИКА, АВТОМАТИЧНИЙ МОРФОЛОГІЧНИЙ АНАЛІЗ, ПРИКМЕТНИКИ, АФІКСИ, ФЛЕКСІЇ, ЛЕМАТИЗАЦІЯ.

## ЗМІСТ

ВСТУП	12
РОЗДІЛ 1. СИСТЕМИ АВТОМАТИЧНОГО АНАЛІЗУ ТЕКСТУ	13
1.1. З історії комп'ютерної лінгвістики	13
1.2. Частини мови і морфологія	28
1.3. Автоматичний морфологічний аналіз	33
1.4. Приклади систем автоматичного морфологічного аналізу	35
РОЗДІЛ 2. ТЕОРЕТИЧНІ ЗАСАДИ МОРФОЛОГІЧНОГО АНАЛІЗУ ПРИКМЕТНИКІВ	40
2.1. Загальна характеристика прикметника як частини мови	40
2.2. Суфікси прикметників	47
2.3. Номенклатура прикметникових афіксів української мови	48
РОЗДІЛ 3. АВТОМАТИЧНИЙ МОРФОЛОГІЧНИЙ АНАЛІЗ ПРИКМЕТНИКІВ	50
3.1. Представлення афіксів	50
3.2. База прикметників	51
3.3. Автоматизація морфологічного аналізу	53
3.4. Лематизація прикметників	58
ВИСНОВКИ	61
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	62
ДОДАТКИ	64

## ВСТУП

Актуальність. Автоматична обробка текстової інформації довільного спрямування потребує морфологічного аналізу словоформ. Прикметник належить до широко вживаних граматичних класів в українському тексті і тому автоматичний морфологічний аналіз прикметників є необхідного аналізу систем автоматичного опрацювання природної мови.

Об'єкт дослідження: автоматична обробка природномовних текстів.

Предмет дослідження: підсистема морфологічного аналізу прикметників української.

Мета дослідження: розробка інформаційної та програмної складової прикметникової компоненти системи автоматичної обробки текстів.

Задачі дослідження:

1. Ознайомитись з методами комп'ютерної лінгвістики.
2. Обробити первинні дані для створення прикметникової компоненти системи автоматичної морфологічного аналізу україномовних текстів.
3. Розробити елементи системи автоматичного морфологічного аналізу українських прикметників.

## РОЗДІЛ 1. СИСТЕМИ АВТОМАТИЧНОГО АНАЛІЗУ ТЕКСТУ

### 1.1. З історії комп'ютерної лінгвістики

Перша асоціація, що в багатьох викликає словосполучення "штучний інтелект", - автоматичний машинний переклад й "розуміння" пристроєм природномовного тексту. Цю проблематику в останні роки всі частіше відносять до особливої наукової дисципліни за назвою "комп'ютерна лінгвістика", успіхи якої здаються примарними: не бажають комп'ютери розрізняти значеннєві відтінки й видавати гарні літературні переклади.

*Розвиток комп'ютерної лінгвістики.* Комп'ютерна лінгвістика, напрямок у прикладній лінгвістиці, орієнтований на використання комп'ютерних інструментів - програм, комп'ютерних технологій організації й обробки даних - для моделювання функціонування мови в тих або інших умовах, ситуаціях, проблемних сферах і т.д., а також вся сфера застосування комп'ютерних моделей мови в лінгвістиці й суміжних дисциплінах. Властиво, тільки в останньому випадку й мова йде про прикладну лінгвістику в точному значенні, оскільки комп'ютерне моделювання мови може розглядатися і як сфера додатка інформатики й теорії програмування до рішення завдань науки про мову. На практиці, однак, до комп'ютерної лінгвістики відносять практично все, що пов'язане з використанням комп'ютерів у мовознавстві.

Проблему автоматичного перекладу й аналізу тексту природною мовою можна зіставити із проблемою керованого термоядерного синтезу. Світова наука активно працює й над тією, і над іншою від уже піввіку, але, всупереч оптимістичним прогнозам, обидві далекі від рішення. Є, щоправда, маленька відмінність між цими двома фундаментальними технологічними викликами: природних термоядерних реакторів, наскільки відомо, на Землі немає, а от створених природою "пристроїв", що розуміють природну мову, -



мільярди. Гігантський прогрес науки й технологій не дає втратити оптимізм щодо проблеми машинного перекладу.

Комп'ютерна лінгвістика народилася в січні 1954 року, коли в Джорджтаунському університеті (США) був проведений перший у світі публічний експеримент по машинному перекладу. У ті ж часи під керівництвом найбільшого математика й кібернетика Олексія Ляпунова почалися активні роботи з машинного перекладу й у Москві. У створену Ляпуновим групу ввійшли, зокрема, тодішні студенти й аспіранти, майбутні "батьки" вітчизняної комп'ютерної лінгвістики Ігор Мельчук й Ольга Кулагіна.

Втім, основа для успіхів вітчизняних учених заставлялася набагато раніше. В 1920-х роках у Росії велися інтенсивні дослідження з семіотики тексту, у той час як широкомасштабне вивчення семіотики в усім світі встановлюється тільки на початку 1960-х років (що значною мірою пов'язане з роботами творця, що емігрував з Росії в 1920 році, структурної лінгвістики Романа Якобсона 1). Стрімкий прогрес семіотики ставши основою для зближення лінгвістики й математики на ґрунті популярного в 1950-60-і роки математичного структуралізму, у дусі гранично формалізованих математичних структур Бурбаки. Ідеї семіотики поєднували найбільших учених, серед яких математик Володимир Успенський і лінгвіст В'ячеслав Іванов. Успіхи формального підходу до опису мови наочно продемонстрували можливість перетворення чисто гуманітарної науки в логічно строгу дисципліну.

Роботи з кібернетики й, зокрема, по структурній лінгвістиці, розгорнуті наприкінці 50-х років під керівництвом академіка Аксея Берга й член-кореспондента АН СРСР Олексія Ляпунова, вивели вітчизняну науку на передові позиції. Уже на початку 1956 долі в Інституті прикладної математики (ИПМ) ім. М. В. Келдиша заробила перша вітчизняна система машинного перекладу з французької на російську мову. Система ФР-І давала переклад явно більше високої якості, чим в американців. Цікаво, що

математики розглядали алгоритми машинного перекладу як окремі випадки досліджуваних у кібернетику алгоритмів перекодування.

Величезне значення мало вироблене Олексієм Ляпуновим й Ольгою Кулагіной теоретико-множинне подання граматичних категорій мови. Ці результати були отримані незалежно від робіт Ноама Хомського (Noam Chomsky), що вважається творцем теорії формальних граматики. Більшу роль зіграли роботи Ігоря Мельчука по автоматичному граматичному аналізу. Узагальнення й систематизація результатів математичної лінгвістики дозволили створити струнку теорію формальних мов. Проблема полягала в одному: апарат теорії формальних мов був дуже гарний для штучних мов, зокрема для мов програмування, але не давав прийнятних результатів для аналізу природної мови й побудови адекватної системи машинного перекладу.

Семіотика - наукова дисципліна, що вивчає загальне в будові й функціонуванні різних знакових систем, що зберігають і переробляють інформацію. Семіотика розглядає як природні знакові системи (комунікація у світі тварин), так і діючі в людському суспільстві (мова, обряди й ін.)

Згодом сам Мельчук визнав, що роботи з автоматичного синтаксичного аналізу російської мови були "абсолютно тупикові". А чи могло бути інакше? Штучні мови спеціально створюються так, щоб було зручно маніпулювати формальними конструкціями, що описують "прозору" семантику, що зводиться, як правило, до конкретних дій. Такі алгоритмічні мови в програмуванні. У природній мові є щось більше - буквальна семантика, з якого й зв'язане розуміння людиною тексту. Розуміння це може бути не таким вуж простим навіть для носія мови й пов'язане з розумовими процесами, а не з формальними правилами опису самої мови. Тому для автоматичного перекладу треба вміти описувати зміст фрази, а не її формальний запис.

Очевидні сьогодні ідеї були вистраждані й обнародовані кілька десятиліть назад Ігорем Мельчуком й Олександром Жолковським. Розробка

ними лексичних функцій, створення теорії "Зміст-Текст" з'явилися колосальним стрибком у побудові строгої концептуальної системи, застосовної для аналізу природної мови.

Мельчук був занадто талановитий і тому, видимо, незручний. Радянська система виштовхнула його із країни. Книги по загальній теорії морфології він закінчував уже в еміграції. П'ятитомна фундаментальна монографія "Курс загальної морфології" вийшла на Заході в 90-х роках. Не вдаючись у подробиці й трохи спрощуючи, суть теорії Мельчука можна звести до побудови формальної семантичної мови, на якому й буде описуватися зміст тексту. Це відкриває величезні можливості як для завдань аналізу текстової інформації, так і для завдань машинного перекладу.

Реалізацією цих можливостей займався колектив під керівництвом академіка Юрія Апресяна, що розробляє систему машинного перекладу ЕТАП. У цей час ці дуже важливі роботи тривають у лабораторії Інституту проблем передачі інформації РАН, очолюваної Іллею Богуславським.

Пітерським математиком Віталієм Тузовим недавно була створена модель природної мови, що використовує спрощений варіант ідей Мельчука. У моделі Тузова зроблена спроба звести завдання витягу інформації з тексту до завдання перетворення тексту на формальній семантичній мові у внутрішнє подання бази знань. Стаття з розповіддю про роботи по реалізації цієї, безумовно, цікавої моделі ("КТ" #450) викликала активну дискусію на сайті журналу. У зв'язку із цим не можу удержатися від одного зауваження: разюче, що багато програмістів, що займаються створенням комерційних продуктів в області обробки текстової інформації, не розуміють, що деякі завдання складання й поповнення словників цілком алгоритмізуєми - як і багато чого іншого в цьому світі...

Шлях ідей до визнання іноді дуже довгий і важкий. Треба було більше тридцяти років, щоб перейти до реальних розробок на основі створеній Валентином Турчиним теорії суперкомпіляції (див. "КТ" #402). Змінилася елементна база, з'явилися зовсім нові технології програмування, і те, що

раніше багато хто зараховували до нереалізованим теоретичним вишукам, стає затребуваним. Можливо, такий же шлях чекає й теорію Мельчука. Напевно, не випадково схожі долі цих учених.

Комп'ютерна лінгвістика стрімко розвивалася в СРСР в 1960-і роки. Однак у наступному десятилітті роботи в області машинного перекладу виявилися під твердим державним контролем. На відміну від атомного проекту (час уже був іншим), цей контроль не сконцентрував сили вітчизняних учених, а навпаки, сприяв припиненню або вповільненню багатьох робіт. Явна "нелюбов" влади того часу до одного з яскравіших представників вітчизняної лінгвістики Юрія Апресяна на довгі роки загальмувала самий цікавий з вітчизняних проектів. Дослідження з машинного перекладу в ІІМ ім. М. В. Келдиша теж практично припинилися.

Ми поки не знаємо, які можливості дасть людству комп'ютерна революція, що насувається нова. Однак можна сподіватися, що комп'ютерна лінгвістика перейде на зовсім нову технологічну базу, основа якої заставляється в наш час.

Як особливий науковий напрямок комп'ютерна лінгвістика оформилася в 1960-і роки. Російський термін "комп'ютерна лінгвістика" є калькою з англійського *computational linguistics*. Оскільки прикметник *computational* по-російському може переводитися і як "обчислювальний", у літературі зустрічається також термін "обчислювальна лінгвістика", однак у вітчизняній науці він здобуває більше вузьке значення, що наближається до поняття "квантитативної лінгвістики". Потік публікацій у цій області дуже великий. Крім тематичних збірників, у США щокварталу виходить журнал "Комп'ютерна лінгвістика". Більшу організаційну й наукову працю проводить Асоціація по комп'ютерній лінгвістиці, що має регіональні структури (зокрема, європейське відділення). Кожні два роки проходять міжнародні конференції по комп'ютерній лінгвістиці - COLING. Відповідна проблематика звичайно буває широко представлена також на різних конференціях зі штучного інтелекту.

*Інструментарій комп'ютерної лінгвістики.* Комп'ютерна лінгвістика як особлива прикладна дисципліна виділяється насамперед по інструменті - тобто по використанню комп'ютерних засобів обробки мовних даних. Оскільки комп'ютерні програми, що моделюють ті або інші аспекти функціонування мови, можуть використати всілякі засоби програмування, то про загальний понятійний апарат комп'ютерної лінгвістики говорити начебто б не доводиться. Однак це не так. Існують загальні принципи комп'ютерного моделювання мислення, які так чи інакше реалізуються в будь-якій комп'ютерній моделі. У їхній основі лежить теорія знань, що спочатку розроблялася в області штучного інтелекту, а надалі стала одним з розділів когнітивної науки. Найважливішими понятійними категоріями комп'ютерна лінгвістика є такі структури знань, як "фрейми" (понятійні, або, як прийнято говорити, концептуальні структури для декларативного подання знань про типізовану тематично єдину ситуацію), "сценарії" (концептуальні структури для процедурного подання знань про стереотипну ситуацію або стереотипне поведіння), "плани" (структури знань, що фіксують подання про можливі дії, що ведуть до досягнення певної мети). Тісно пов'язане з категорією фрейму поняття "сцена". Категорія сцени переважно використовується в літературі по комп'ютерній лінгвістиці як позначення концептуальної структури для декларативного подання актуалізованих у мовному акті й виділених мовними засобами (лексемами, синтаксичними конструкціями, граматичними категоріями й ін.) ситуацій й їхніх частин.

Певним чином організований набір структур знань формує "модель миру" когнітивної системи і її комп'ютерної моделі. У системах штучного інтелекту модель миру утворить особливий блок, у який залежно від обраної архітектури можуть входити загальні знання про світ (у вигляді простих речень типу "узимку холодно" або у вигляді правил продукції "якщо на вулиці йде дощ, те треба надягти плащ або взяти парасольку"), деякі специфічні факти ("Найвища вершина у світі - Еверест"), а також цінності і їхньої ієрархії, іноді виділювані в особливий "аксіологічний блок".

Напрямки комп'ютерної лінгвістики. Сфера КЛ досить різноманітна й включає такі області, як комп'ютерне моделювання спілкування, моделювання структури сюжету, гіпертекстові технології подання тексту, машинний переклад, комп'ютерна лексикографія. У вузькому змісті проблематика КЛ часто зв'язується з міждисциплінарним прикладним напрямком із трохи невдалою назвою "обробка природної мови" (переклад англійського терміна Natural Language Processing). Воно виникло наприкінці 1960-х років і розвивалося в рамках науково-технологічної дисципліни "штучний інтелект". За своєю внутрішньою формою словосполучення "обробка природної мови" охоплює всі області, у яких комп'ютери використовуються для обробки мовних даних. Тім же часом в практиці закріпилося більше вузьке розуміння цього терміна - розробка методів, технологій і конкретних систем, що забезпечують спілкування людини з ЕОМ на природній або обмеженій природній мові.

Бурхливий розвиток напрямку "обробки природної мови" доводиться на 1970-і роки, що було пов'язане з несподіваним експонентним ростом кількості кінцевих користувачів ЕОМ. Оскільки навчання мовам і технології програмування всіх користувачів неможливо, виникла проблема організації взаємодії з комп'ютерними програмами. Рішення цієї проблеми комунікації йшло по двох основних шляхах. У першому випадку вживали з адаптації мов програмування й операційних систем до кінцевого користувача. У результаті з'явилися мови високого рівня типу Visual Basic, а також зручні операційні системи, побудовані в концептуальному просторі звичних людині метафор - ПИСЬМОВИЙ СТИЛ, БІБЛІОТЕКА. Другий шлях - розробка систем, які дозволяли б взаємодіяти з ЕОМ у конкретній проблемній області природною мовою або якомусь його обмеженому варіанті.

Архітектура систем обробки природної мови в загальному випадку включає блок аналізу мовного повідомлення користувача, блок інтерпретації повідомлення, блок породження змісту відповіді й блок синтезу поверхневої

структури висловлення. Особливою частиною системи є діалоговий компонент, у якому зафіксовані стратегії ведення діалогу, умови застосування цих стратегій, способи подолання можливих комунікативних невдач (збоїв у процесі спілкування).

Серед комп'ютерних систем обробки природної мови звичайно виділяються питально-відповідні системи, діалогові системи рішення завдань і системи обробки зв'язних текстів. Споконвічно питально-відповідні системи стали розроблятися як реакція на погану якість кодування запитів при пошуку інформації в інформаційно-пошукових системах. Оскільки проблемна область таких систем була сильно обмежена, це трохи спрощувало алгоритми перекладу запитів у подання формальною мовою й зворотною процедурою перетворення формального подання у висловлення природною мовою. З вітчизняних розробок до програм такого типу ставиться система ПОЕТ, створена колективом дослідників під керівництвом Э.В.Попова. Система обробляє запити російською мовою (з невеликими обмеженнями) і синтезує відповідь. Блок-схема програми припускає проходження всіх етапів аналізу (морфологічного, синтаксичного й семантичного) і відповідного етапів синтезу.

Діалогові системи рішення завдань, на відміну від систем попереднього типу, грають у комунікації активну роль, оскільки їхнє завдання полягає в тім, щоб одержати рішення проблеми на основі тихнув знань, які представлені в ній самої, і тієї інформації, якові можна одержати від користувача. Система містить структури знань, у яких фіксуються типові послідовності дій для рішення завдань у даній проблемній області, а також відомості про необхідні ресурси. Коли користувач ставить запитання або ставить певне завдання, активізується відповідний сценарій. Якщо якісь компоненти сценарію пропущені або відсутні якісь ресурси, система виступає ініціатором комунікації. Так працює, наприклад, система SNUKA, що вирішує завдання планування воєнних операцій.

Системи обробки зв'язних текстів досить різноманітні за структурою. Їхньою загальною рисою можна вважати широке використання технологій подання знань. Функції систем такого роду полягають у розумінні тексту й відповідях на питання про його зміст. Розуміння розглядається не як універсальна категорія, а як процес витягу інформації з тексту, обумовлений конкретним комунікативним наміром. Іншими словами, текст "читається" тільки з установкою на те, що саме потенційний користувач захоче довідатися про нього. Тім самим і системи обробки зв'язних текстів виявляються аж ніяк не універсальними, а проблемно-орієнтованими. Типовими прикладами систем обговорюваного типу можуть служити системи RESEARCHER й TAILOR, що утворюють єдиний програмний комплекс, що дозволяє користувачеві одержати інформацію з рефератів патентів, що описують складні фізичні об'єкти.

*Розробка інформаційно-пошукових систем.* Найважливішим напрямком комп'ютерної лінгвістики є розробка інформаційно-пошукових систем (ІПС). Останні виникли наприкінці 1950-х - початку 1960-х років як відповідь на різке зростання обсягів науково-технічної інформації. По типу збереженої й оброблюваної інформації, а також по особливостях пошуку ІПС розділяються на дві великих групи - документальні й фактографічні. У документальних ІПС зберігаються тексти документів або їхнього опису (реферати, бібліографічні картки й т.д.). Фактографічні ІПС мають праворуч з описом конкретних фактів, причому не обов'язково в текстовій формі. Це можуть бути таблиці, формули й інші види подання даних. Існують і змішані ІПС, що включають як документи, так і фактографічну інформацію. У цей час фактографічні ІПС будуються на основі технологій баз даних (БД). Для забезпечення інформаційного пошуку в ІПС створюються спеціальні інформаційно-пошукові мови, в основі яких лежать інформаційно-пошукові тезауруси. Пошукову-інформаційно-пошукова мова - це формальна мова, призначена для опису окремих аспектів плану змісту документів, що зберігаються в ІПС, і запиту. Процедура опису документа інформаційно-



пошуковою мовою називається індексуванням. У результаті індексування шкірному документу приписується його формальний опис інформаційно-пошуковою мовою - пошуковий образ документа. Аналогічним образом індексується й запит у, якому приписується пошуковий образ запиту й пошукове приписання. Алгоритми інформаційного пошуку засновані на порівнянні пошукового приписання з пошуковим образом запиту. Критерій видачі документа на запит може складатися в повному або частковому збігу пошукового образа документа й пошукового приписання. У ряді випадків користувач має можливість сам сформулювати критерії видачі. Це визначається його інформаційною потребою. В автоматизованих ІПС хаші використовуються дескрипторні пошуково-інформаційні мови. Тематика документа описується сукупністю дескрипторів. Як дескриптори виступають слова, терміни, що позначають прості, досить елементарні категорії й поняття проблемної області. У пошуковий образ документа вводиться стільки дескрипторів, скільки різних тем зачіпається у документі. Кількість дескрипторів не обмежується, що дозволяє описати документ у багатомірній матриці ознак. Часто в дескрипторному інформаційно-пошуковій мові накладають обмеження на сполучуваність дескрипторів. У цьому випадку можна говорити про те, що пошуково-інформаційна мова має синтаксис.

Одна з перших систем, що працювали з дескрипторною мовою, була американська система УНІТЕРМ, створена М.Таубе. Як дескриптори в цій системі функціонували ключові слова документа - унітерми. Особливість цієї ІПС полягає в тім, що споконвічно словник інформаційної мови не задавався, а виникав у процесі індексування документа й запиту. Розвиток сучасних інформаційно-пошукових систем пов'язане з розробкою ІПС безтезаурусного типу. Такі ІПС працюють із користувачем на обмеженій природній мові, а пошук здійснюється по текстах рефератів документів, по їхніх бібліографічних описах, а часто й по самих документах. Для індексування в ІПС безтезаурусного типу використовуються слова й словосполучення природної мови.

*Створення гіпертекстових систем.* До області комп'ютерної лінгвістики деякою мірою можуть бути віднесені роботи в області створення гіпертекстових систем, розглянутих як особливий спосіб організації тексту й навіть як принципово новий вид тексту, протипоставлений по багатьом своїх властивостях звичайному тексту, сформованому в гутенберговській традиції друкарства. Ідея гіпертексту зв'язується з ім'ям Ванневару Буша - радника президента Ф.Рузвельта по науці. В.Буш теоретично обґрунтував проект технічної системи "Мемекс", що дозволяла користувачеві зв'язувати тексти і їхні фрагменти по різних типах зв'язків, переважливо по асоціативним відносинам. Відсутність комп'ютерної техніки зробило проект важкорезалізуемим, оскільки механічна система виявилася надмірно складною для практичного втілення.

Ідея Буша в 1960-і роки одержала другу народження в системі "Ксанаду" Т.Нельсона, що вже припускала використання комп'ютерної техніки. "Ксанаду" дозволяв користувачеві прочитувати сукупність уведених у систему текстів різними способами, у різній послідовності, програмне забезпечення давав можливість як запам'ятовувати послідовність переглянутих текстів, так і вибрати з їх практично кожної в довільний момент години. Безліч текстів з єднальними їхніми відносинами (системою переходів) було названо Т.Нельсоном гіпертекстом. Багато дослідників розглядають створення гіпертексту як початок нової інформаційної епохи, протипоставленій ері друкарства. Лінійність аркуша, що зовні відбиває лінійність мови, виявляється фундаментальною категорією, що обмежує мислення людини й розуміння тексту. Мир змісту нелінійний, тому стискання значенєвої інформації в лінійному мовному відрізку вимагає використання спеціальних "комунікативних упакувань" - членування на тему й рему, поділ плану змісту висловлення на експліцитні (твердження, речення, фокус) і імпліцитні (пресуппозиція, наслідок, імплікатура дискурса) шари. Відмова від лінійності тексту й у процесі його подання читачеві (тобто при

читанні й розумінні) і в процесі синтезу, на думку теоретиків, сприяв би "звільненню" мислення й навіть виникненню його нових форм.

У комп'ютерній системі гіпертекст подань у вигляді графові, у вузлах котрого перебувають традиційні тексти або їхні фрагменти, зображення, таблиці, відеоролики тощо. Вузли зв'язані різноманітними відносинами, типи яких задаються розроблювачами програмного забезпечення гіпертексту або самим читачем. Відносини задають потенційні можливості пересування, або навігації по гіпертексту. Відносини можуть бути односпрямованими або двунправленими. Відповідно, двунправлені стрілки дозволяють рухатися користувачеві в обидва боки, а односпрямовані - тільки в одну. Ланцюжок вузлів, через які проходить читач при перегляді компонентів тексту, утворить шлях, або маршрут.

Комп'ютерні реалізації гіпертексту бувають ієрархічними або мережними. Ієрархічне - деревоподібне - будова гіпертексту істотно обмежує можливості переходу між його компонентами. У такому гіпертексті відносини між компонентами нагадують структуру тезауруса, заснованого на родо-видових зв'язках. Мережний гіпертекст дозволяє використати різні типи відносин між компонентами, не обмежуючись відносинами "рід - вид". По способі існування гіпертексту виділяються статичні й динамічні гіпертексти. Статичний гіпертекст не міняється в процесі експлуатації; у ньому користувач може фіксувати свої коментарі, однак смороду не міняють суть справи. Для динамічного гіпертексту зміна є нормальною формою існування. Звичайно динамічні гіпертексти функціонують там, де необхідно постійно аналізувати потік інформації, тобто в інформаційних службах різного роду. Гіпертекстової є, наприклад, Арізонська інформаційна система (AAI), що щомісяця поповнюється на 300-500 рефератів на місяць.

Відносини між елементами гіпертексту можуть споконвічно фіксуватися творцями, а можуть породжуватися щораз, коли відбувається обіг користувача до гіпертексту. У першому випадку мова йде про гіпертексти твердої структури, а в іншому - про гіпертексти м'якої структури.

Тверда структура технологічно цілком зрозуміла. Технологія організації м'якої структури винна ґрунтуватися на семантичному аналізі близькості документів (або інших джерел інформації) друг до друга. Це нетривіальне завдання комп'ютерної лінгвістики. У цей година широке поширене використання технологій м'якої структури на ключових словах. Перехід від одного вузла до іншого в мережі гіпертексту здійснюється в результаті пошуку ключових слів. Оскільки набір ключових слів щораз може розрізнятися, щораз міняється й структура гіпертексту.

Технологія побудови гіпертекстових систем не робить розходжень між текстовою й нетекстовою інформацією. У той же час включення візуальної й звукової інформації (відеороликів, картин, фотографій, звукозаписів і т.п.) вимагає істотної зміни інтерфейсу з користувачем і могутнішої програмної й комп'ютерної підтримки. Такі системи одержали назву гіпермедіа, або мультимедіа. Наочність мультимедійних систем визначила їхнє широке використання в навчанні, у створенні комп'ютерних варіантів енциклопедій. Існують, наприклад, прекрасно виконані CD-роми з мультимедійними системами по дитячих енциклопедіях видавництва "Дорлін Киндерсли".

*Автоматичні словники.* У рамках комп'ютерної лексикографії розробляються комп'ютерні технології складання й експлуатації словників. Спеціальні програми - бази даних, комп'ютерні картотеки, програми обробки тексту - дозволяють в автоматичному режимі формувати словникові статті, зберігати словникову інформацію й обробляти її. Безліч різних комп'ютерних лексикографічних програм розділяються на дві великих групи: програми підтримки лексикографічних робіт й автоматичні словники різних типів, що включають лексикографічні бази даних. Автоматичний словник - це словник у спеціальному машинному форматі, призначений для використання на ЕОМ користувачем або комп'ютерною програмою обробки тексту. Іншими словами, розрізняються автоматичні словники кінцевої користувача-людини й автоматичні словники для програм обробки тексту. Автоматичні словники, призначені для кінцевого користувача, по інтерфейсі й структурі словникової

статті істотно відрізняються від автоматичних словників, включених у системи машинного перекладу, системи автоматичного реферування, інформаційного пошуку й т.д. Найчастіше смороду є комп'ютерними версіями добрі відомих звичайних словників. На ринку програмного забезпечення є комп'ютерні аналоги тлумачних словників англійської мови (автоматичний Вебстер, автоматичний тлумачний словник англійської мови видавництва Коллінз, автоматичний варіант Нового великого англо-російського словника під ред. Ю.Д.Апресяна й Э.М.Меднікової), існує й комп'ютерна версія словника Ожегова. Автоматичні словники для програм обробки тексту можна назвати автоматичними словниками в точному змісті. Вони, як правило, не призначені для звичайного користувача. Особливості їхньої структури, сфера охоплення словникового матеріалу задаються тими програмами, які з ними взаємодіють.

*Комп'ютерне моделювання структури сюжету.* Комп'ютерне моделювання структури сюжету - ще один перспективний напрямок комп'ютерної лінгвістики. Вивчення структури сюжету ставитися до проблематики структурного літературознавства (у широкому змісті), семіотики й культурології. Наявні комп'ютерні програми моделювання сюжету ґрунтуються на трьох базових формалізмах подання сюжету - морфологічному й синтаксичному напрямках подання сюжету, а також на когнітивному підході. Ідеї про морфологічний пристрій структури сюжету сходять до відомих робіт В.Я.Проппа (дів.) про російську чарівну казку. Пропп помітив, що при статку персонажів і подій чарівної казки кількість функцій персонажів обмежено, і запропонував апарат для опису цих функцій. Ідеї Проппа лягли в основу комп'ютерної програми TALE, що моделює породження сюжету казки. В основу алгоритму програми TALE покладена послідовність функцій персонажів казки. Фактично функції Проппа задавали безліч типізованих ситуацій, упорядкованих на основі аналізу емпіричного матеріалу. Можливості зчеплення різних ситуацій у правилах породження визначалися типовою послідовністю функцій - у тім виді, у якому це вдається

встановити з текстів казок. У програмі типові послідовності функцій описувалися як типові сценарії зустрічей персонажів.

Теоретичну основу синтаксичного підходу до сюжету тексту склали "сюжетні граматики", або "граматики оповідання" (story grammars). Вони з'явилися в середині 1970-х років у результаті переносу ідей граматики, що породжує Н.Хомського на опис макроструктури тексту. Якщо найважливішими складовими синтаксичної структури в граматиці, що породжує, були дієслівні й іменні групи, те в більшості сюжетних граматик у якості базових виділялися експозиція (setting), подія й епізод. У теорії сюжетних граматик широко обговорювалися умови мінімальності, тобто обмеження, що визначали статус послідовності з елементів сюжету як нормальний сюжет. Виявилось, однак, що чисто лінгвістичними методами це зробити неможливо. Багато обмежень носять соціокультурний характер. Сюжетні граматики, істотно розрізняючись набором категорій у дереві породження, допускали досить обмежений набір правил модифікації оповідальної (наративної) структури.

На початку 1980-х років однієї з учениць Р.Шенка - В.Ленерт у рамках робіт зі створення комп'ютерного генератора сюжетів був запропонований оригінальний формалізм емоційних сюжетних одиниць (Affective Plot Units), який виявився потужним засобом подання структури сюжету. При тім, що він був споконвічно розроблений для системи штучного інтелекту, цей формалізм використався в чисто теоретичних дослідженнях. Сутність підходу Ленерт полягала в тім, що сюжет описувався як послідовна зміна когнітивно-емоційних станів персонажів. Тим самим у центрі уваги формалізму Ленерт знаходяться не зовнішні компоненти сюжету - експозиція, подія, епізод, мораль, - а його змістовні характеристики. Щодо цього формалізм Ленерт почасти виявляється поверненням до ідей Проппа.

## 1.2. Частини мови і морфологія

Наскрізним питанням морфології є теорія частин мови (класів слів).

Класифікація частин мови В. Виноградова одна з найпоширеніших. На його думку, не всі слова є частинами мови, а до слів, які перебувають за межами частин мов, належить модальні слова, частки мови і вигуки. Отже, за функціями В. Виноградов усі слова поділяв на чотири групи: частини мови, модальні слова, частки мови, вигуки,

*Частини мови.* До них належать слова, які виконують номінативну функцію, є або можуть бути членами речення. На цій основі він виокремив сім частин мови: іменник, прикметник, числівник як іменні частини мови, залишки займенників, дієслово, прислівник, слова категорії стану.

*Модальні слова.* Такими він вважав слова, які виражають відношення висловлюваної думки до дійсності, яке встановлює мовець. Синтаксичні функції і семантична будова модальних слів відрізняється від функції і семантики частин мови та часток мови.

*Частки мови.* До цієї групи він відносив власне частки зв'язки, прийменники і сполучники. Службові слова В. Виноградов, як і французький мовознавець Жозеф Вандрієс (1375-1960), називав морфемами, тобто виразниками суто граматичних відношень.

*Вигуки.* За В. Виноградовим, вони перебувають за межами частин мови і часток мови й не мають нічого спільного з морфологією. На його думку, ці слова пов'язані з емоційним мовленням і не належать до структури інтелектуального мовлення.

Порядкові числівники В. Виноградов називав прикметниками із значенням ознаки (відчислівникові прикметники). До займенників він відносив як залишкові явища лише особові займенники (*я, ти, він, ми, ви, вони*), зворотний займенник (*себе*) та їх похідні. На його думку, займенник перебуває в стані занепаду.

Як і Л. Щерби, В. Виноградов виокремив і слова категорії стану, яка утворилася на основі деяких іменників, прикметників, прислівників і дієслів.

Усі частини мови, як зазначав В. Виноградов, між собою генетично пов'язані і взаємозумовлені: прислівник сформувався на основі іменників, прикметників, прислівників, займенників, дієприслівників, дієслів, дієприкметників; слова категорії стану — на основі іменників, прикметників, дієприкметників, прислівників, дієслів; дієприкметник — на основі прикметників і дієслів; дієприслівник — на основі дієслів, прислівників і дієприкметників.

Проблему поділу слів на класи намагалися розв'язати вчені різних епох і народів. У IV ст. до н. е. давньогрецький філософ Арістотель слова давньогрецької мови поділяв на чотири групи слів: ім'я, дієслово, член (вигук), сполучник (зв'язка). У V ст. до н. е. давньоіндійські граматисти Яска і Паніні в санскриті виділили п'ять груп слів: ім'я, дієслово, прийменник, частка, сполучник.

Грецький учений, представник олександрійської школи Арістарх Олександрійський і його учень Діонісій Фракійський у II ст. до н. е. вперше у грецькій мові виділили вісім груп слів і напровадили термін «частина мови (*partes orationis*): ім'я, дієслово, дієприкметник, член (вигук), займенник, прийменник, прислівник, сполучник. Цю класифікацію прийняли, замінивши назву “член” вигуком, римська граматики Варрона (I ст. до н. е.) та граматики слов'янської мови Мелетія Смотрицького (XVII ст.).

Російський учений Михайло Ломоносов (1711—1765) у «Российской грамматике» стосовно російської мови назвав десять частин мови: ім'я, прикметник, числівник, займенник, дієслово, дієприкметник, прислівник, прийменник, сполучник, вигук.

Наприкінці XIX ст. О. Потебня і П. Фортунатов розробили принципи класифікації частин мови (О. Потебня — семантичний, а П. Фортунатов — морфологічний). У XX ст. свою теорію частин мови сформулював В. Виноградов у книзі «Русский язык. Грамматическое учение о слове» (М., 1947).



Сучасна традиційна граматики виокремлює десять частин мови: іменник, прикметник, числівник, займенник, дієслово, прийменник, сполучник, частка, вигук.

Частини мови — граматичні класи (групи) слів, що розрізняються категоріальним значенням, притаманними їм морфологічними ознаками, синтаксичними властивостями в складі словосполучень та речень і за суфіксами.

Ці граматичні класи слів характеризуються:

- спільністю категоріального значення, абстрагованого від лексичних значень (напр., предметність в іменників, ознака предмета у прикметників, число у числівників, процесуальна дія у дієслів, ознака ознаки у прислівників);
- спільністю морфологічних категорій і правил словозміни (напр., рід, число, відмінок у іменників, вид, спосіб, час у дієслів);
- тотожністю синтаксичних властивостей слів (напр., підмет і додаток у іменників, присудок у дієслів, означення у прикметників, обставина у прислівників);
- спільністю суфіксів, з допомогою яких утворено відповідні слова (напр., суфікси іменників *-тель, -ець, -апин, -ик, -ник, -ист, -ізм, -ач, -ук*; суфікси прикметників *-н-ий, -ський, -аст-ий, -ап-ий, -ов-ий*; дієслівні суфікси *-нути, -ича-ти, -ува-ти*).

Класифікація за цими чотирма ознаками зафіксована в «Сучасній українській літературній мові. Морфології» (1969), «Русской грамматике» (1980) та шкільній граматиці.

Одним із принципів класифікації частин мови є здатність або нездатність слова (соловоформи) бути членом речення. Очні словоформи можуть бути самостійними Членами речення (напр., *Україна, будувати, радісна*), виконувати номінативну функцію і вживатися в реченні як його компоненти. Це *самостійні слова* (словоформи). Інші слова самостійними членами речення не бувають: пор. слова *і, але, якщо, щоб, бо, не, ні, би*,

*адже, під, через, біля* і под. Вони входять до складу речення разом із самостійними членами і вказують або по зв'язок між словоформами, або на відношення мовця до висловлюваної думки. Це *службові слова*. У зв'язку з цим частини мови поділяють на самостійні та службові. Але лише цієї ознаки для класифікації частин мови недостатньо, бо межа між самостійними і службовими словами умовна, оскільки самостійні слова, втрачаючи певною мірою лексичне значення, переходять у службові (*думати, вважати, почати, стати* і под.).

Українська морфологія (1669) класифікувала частини мови за “наявністю чи відсутністю у слові лексичного (співвідносного з поняттям) і граматичного значень”. Відповідно до цього слова, а отже, й частини мови, поділила на *повнозначні* й *неповнозначні*. Службові частини мови (прийменники, сполучники) відносила до тих, що не мають лексичного значення.

У сучасній морфології службові слова розглядають як такі, що мають лексичне значення, оскільки кожне слово становить єдність лексичного і граматичного значення.

Традиційно основним вважали морфологічний принцип поділу слів на частини мови. Прихильниками цього принципу були російський мовознавець Пилип Фортунатов (1848—1914) і його послідовники Олександр Пешковсікий (1878—1933) та Дмитро Ушаков (1873—1942). Морфологічні категорії є однією з суттєвих ознак частини мови: морфологічні категорії іменника (рід, число, відмінок) чітко протиставляються морфологічним категоріям дієслова (вид, спосіб, час). Але цей принцип придатний для класифікації лише тих слів, які мають систему словозміни, яким властиві флексії (іменник, прикметник, дієслово, займенник, числівник). Слова невідмінювані (*таксі, бюро, депо, вчора, сьогодні, череп, при. над. щоб, бо та ін.*) залишаються за межами цієї класифікації. На однобічність морфологічної класифікації вказували російські вчений Лев Щерба (1880—1944) і В. Виноградов.

Отже, морфологічний принцип суттєвий лише для певної групи слів, її тому є не визначальним, а допоміжним.

Існує класифікація частин мови і за лексичним значенням словоформи, носієм якого є корінь слова, наприклад, назви кольорів — прикметники (*білий, синій, голубий*), назви дій — дієслова (*читати, писати, боротися*), назви осіб — іменники (*сопілкар, читач, вихователя*) і ін. Але лексичне значення кореня є надто низьким рівнем узагальнення і надто дрібною семантичною одиницею для частин мови, бо практично неможливо всі слова мови поділити на групи за лексичним значенням із більш-менш чіткими межами між ними: колір *цукру, шоколаду, мокрого асфальту*. Тому для класифікації частин мови застосовується значно абстрактніше семантична одиниця — категоріальне значення, яке виражається суфіксами і закінченнями. Російський учений Михайло Панов (нар. 1920) довів, що належність слів до частин мови формально визначається не значенням коренів, а значенням суфіксів і закінчень.

В українському мовознавстві на базі закінчень і суфіксів розрізняють такі типи категоріального значення: категоріальне значення предметності, категоріальне значення ознаки, категоріальне значення дії. Всі слова, яким притаманне категоріальне значення предметності, є Іменниками (*українець, стілець, завдання*), слова з категоріальним значенням ознаки предмета — прикметниками (*блакитний, ранковий, капроновий*), слова з категоріальним значенням числа і кількості (*три, сім*) належать до числівників, слова з категоріальним значенням ознаки ознак — до прислівників (*читав швидко, ставився серйозно, вивчив учора*).

Отже, провідними принципами класифікації частин мови є категоріальне значення, морфологічні категорії та синтаксичні властивості слів.

### **1.3. Автоматичний морфологічний аналіз**

Обов'язковою складовою частиною будь-якої системи автоматичного опрацювання текстів є підсистема морфологічного аналізу.

Підсистема морфологічного аналізу застосовується до вирішення таких задач, як визначення для одиниць тексту інформації про місце їх у морфологічній системі відповідної мови та ідентифікація словоформ однієї лексеми.

Морфологічний аналізатор кожній словоформі тексту приписує коди частин мови і значення граматичних категорій (рід, число, відмінок, вид, час, особа тощо). Характер цієї інформації, обсяг її й методи, за допомогою яких встановлюється морфологічна інформація, залежать від мети дослідження, від орієнтації на характер текстів, що аналізуються.

Морфологічний аналіз присутній на всіх етапах аналізу тексту, тому що ані морфемний, ані синтаксичний, ані семантичний аналіз не можуть обійтися без визначення частин мови.

Наприклад, при морфемному сегментуванні тексту виділення префіксів можливе без знання частин мови, а суфіксів - ні: потрібні різні їх набори та процедури відсікання суфіксів для іменників, дієслів, прикметників, прислівників. При автоматичному синтаксичному аналізі лише за наявності лексико-граматичної та граматичної інформації до кожної словоформи можна синтаксично прив'язати словоформи у реченні. На рівні формального аналізу тексту саме морфологічна інформація забезпечує доступ комп'ютера до змісту, опосередкованого через співвіднесеність одиниць змісту з одиницями виразу.

Морфологічні ознаки одиниць тексту мають стати інструментом дослідження зв'язку між лексикою і граматиною, між використанням його у мовленні, між парадигматикою (в аспекті розгляду відмінкових форм відмінюваних слів) і синтагматикою (в аспекті лінійних зв'язків слів, сполучуваності у тексті). Саме таку функцію виконують частини мови.

Морфологія - це одна із частин граматичної будови мови, що охоплює граматичні класи слів (частини мови), граматичні (морфологічні) категорії цих частин мови та їхні форми.

Морфологія як наука передбачає розв'язання таких завдань: 1) вивчення граматичних класів слів - частин мови і принципів їхнього класифікаційного виділення; 2) виокремлення частини семантики слова як морфологічної; 3) обґрунтування набору морфологічних категорій та їхньої природи; 4) опис сукупності формальних засобів, закріплених за відповідними частинами мови та їхніми морфологічними категоріями.

Принципову відмінність стратегій традиційного і комп'ютерного морфологічного аналізу визначив Ю. Марчук: "<...>у комп'ютерній лінгвістиці поняття морфологічного аналізу є поняттям операційним. Якщо у традиційній лінгвістиці до морфологічного аналізу належить те, що характеризує форму і відповідає на питання "що" класифікують, то в обчислювальній (прикладній) лінгвістиці важливо не "що", а "як" одержують ту чи іншу інформацію". Справді, як буде далі показано, до етапу морфологічного аналізу входить велика кількість операцій, за допомогою яких можна одержати необхідну морфологічну інформацію.

Практично автоматичний морфологічний аналіз присутній в усіх видах аналізу тексту, оскільки жоден з них не може обійтися без аналізу форм слів, визначення належності слова до граматичного класу. Лінгвістичним поясненням цього може бути об'єктивно існуючий тісний зв'язок між лексичними і граматичними значеннями одиниць мови, а також: між системами парадигматичних і синтагматичних відношень. Саме морфологічна інформація забезпечує доступ комп'ютеру до змісту тексту, оскільки досі єдиним реальним шляхом автоматичного аналізу плану змісту залишається опосередкований шлях через співвіднесення його з одиницями плану вираження. Тому погляд на частини мови як на лексико-граматичні класи слів слугує теоретичною підставою обов'язковості розв'язання задачі

визначення граматичних класів при автоматичній переробці текстової інформації.

#### **1.4. Приклади систем автоматичного морфологічного аналізу**

Описуваний алгоритм морфологічного аналізу є словарним. Він використовує електронний морфологічний словник, що є розробкою лабораторії комп'ютерної лінгвістики ІППІ РАН. Об'єм словника – близько 90 тис. лексем (120 тис. словооснов). Зовнішній формат статей електронного морфологічного словника – формат системи ЕТАП. Програмна оболонка електронного морфологічного словника – Yandex-dict – розроблена підрозділом "Аркадія" фірми Comptek International. Алгоритм є частиною цієї програмної оболонки.

Основною особливістю алгоритму є те, що для словоформи, не описаної в словнику, алгоритм генерує її гадану модель словозміни – одну або декілька статей у форматі системи ЕТАП. Оскільки Yandex-dict дозволяє поповнювати словник новими лексемами, гіпотетичні статті, що згенерували, можна зберегти в цьому словнику (або іншому словнику такого ж типу) для подальшого використання.

Основні компоненти алгоритму:

1. Список унікальних ключів вигляду {основа, ідентифікатор лексеми, номер основи в лексемі} по всіх основах всіх лексем словника. Список впорядкований в інверсійному алфавітному порядку, тому основи слів з однаковим або схожим типом словозміни в цьому списку, як правило, знаходяться поруч (аналогічно для нормальних форм слів). Лексеми можуть мати одну або декілька основ. Так, лексеми з чергуванням в основі мають, як правило, дві основи (замок – замок, замк), суплетивні, – декілька основ (идти – ид, ш, ше, шел).
2. Уніфікований список закінчень всіх лексем словника в інверсійному алфавітному порядку.

3. Алгоритм морфологічного аналізу словоформи по окремій лексемі (текст лексеми – у внутрішньому форматі Yandex-dict). Його параметри: лексема, основа і закінчення аналізованої форми. Він повертає кількість варіантів розбору і набори граматичних характеристик по кожному варіанту розбору. Якщо словоформа не відповідає лексемі, то повертане значення кількості варіантів розбору дорівнює нулю.

Опис алгоритму.

- 1) Знаходяться всі варіанти основ – від аналізованої словоформи відрізуються всі варіанти закінчень.
- 2) Для кожного варіанту основи, починаючи з щонайдовшого, здійснюється бінарний пошук в інверсійному списку основ. Якщо варіант основи в цьому списку відсутній, то таким чином знаходяться "найбільш близькі" словарні основи – що мають максимальний по довжині загальний "хвіст". Позиція першої "найбільш близької" основи і міра її схожості – число символів, що збіглися, в основі і довжина закінчення – запам'ятовуються.
- 3) По всіх варіантах основ виробляється наступне:

Для всіх лексем, що мають однакову міру схожості (однакову довжину загального "хвоста" основи) здійснюється морфологічний. аналіз по лексемі.

Якщо варіант основи не збігається ні з однією з "найближчих" словарних основ, то це означає, що аналізоване слово з даним варіантом основи в словнику відсутнє. В цьому випадку по варіанту основи, закінченню і лексемі, відповідній "найближчій" словарній основі, генерується гіпотетична лексема – модель словозміни для цього невідомого слова. В разі успішної генерації ця гіпотеза подається на вхід морфологічному аналізатору по лексемі.

Успішні варіанти розбору запам'ятовуються у вигляді: {Лексема(текст статті), Варіанти розбору}.

Якщо результат є гіпотезою і при цьому така ж гіпотеза вже є, то вона не запам'ятовується повторно; замість цього збільшується лічильник "продуктивності" цієї гіпотези.

Якщо серед лексем з однаковою поточною мірою схожості є хоч би один варіант розбору, то перехід пункту 5 з успішним результатом. Якщо варіантів розбору немає, то довжина необхідного загального "хвоста" основи зменшується. Якщо після цього довжина необхідного загального "хвоста" основи стала менше двох, то перехід пункту 5 з відмовою; інакше – перехід до пункту 3.

4) Виробляється уніфікація гіпотез по парадигмах (оскільки формат допускає неоднозначний опис парадигми) і їх фільтрація по "продуктивності" – якщо "продуктивність" гіпотези менше максимальної "продуктивності" в 5 разів, то гіпотеза відсівається.

5) Кінець.

Описаний алгоритм використовується при побудові повнотекстових пошукових систем з врахуванням морфології російської мови в продуктах серії електронних наукових видань (ЕНВ), НТЦ "Інформрегістр", що розробляються, ІМЛІ ім. М.Горького РАН і фірмою Comptek International. Відомі продукти цієї серії – "Інформ-норматив" та "Грібоедов".

"Інформ-норматив" містить тексти ГОСТы, ANSI, ISO, IEC, ССІТТ, ЕСМА), правових документів, класифікаторів (ОКС,ГРНТИ,УДК,ОКП).

До складу "Грібоедова" входять тексти всіх творів А.С.Грібоедова по всіх найбільш авторитетних виданнях, література про життя і творчість А.С.Грібоедова (більше 200 текстів), бібліографічні описи видань творів А.С.Грібоедова і літератури про нього (більше 6000), анотовані покажчики імен, географічних назв, періодичних видань, словник мови А.С.Грібоедова.

У ЕНВ є конкорданс корпусу текстів, який є частиною пошукової системи. Зокрема, при лексичному пошуку за допомогою конкордансу для всіх слів із запиту синтезуються всі словоформи; після цього "розширений" запит передається пошуковому механізму для здійснення пошуку. Очевидно,



що чим повніше і точніше буде представлена лексика корпусу текстів в конкордансе, тим якісний буде лексичний пошук.

При підготовці ЕНВ конкорданс будується в три етапи:

Етап 1. За списком всіх словоформ корпусу текстів будується список всіх його лексем. У цьому списку для кожної "словарної" словоформи запам'ятовуються всі статті – варіанти її морфологічного розбору, а для кожної "несловарної" словоформи – всі гіпотетичні статті, що генеруються при її морфологічному аналізі. При морфологічному аналізі використовуються два словники – стандартний (об'ємом близько 90 тис. лексем) і словник імен власних (об'ємом більше 1000 лексем).

"Несловарних" словоформ в даному корпусі текстів – близько 50% (тут і далі наводяться статистичні дані по корпусу текстів ЕНІ "Інформ-норматив"). Для 90% з них згенерували гіпотези – в середньому по 3 на кожну словоформу.

Етап 2. Виробляється фільтрація гіпотетичних статей. Основним принципом фільтрації є аналіз парадигм лексем по корпусу текстів. Під парадигмою лексеми по корпусу текстів (ПКТ лексеми) тут розуміється список словоформ лексеми, які зустрічаються в цьому корпусі текстів.

Використовуються наступні евристичні правила аналізу ПКТ гіпотез:

- 1) "Включення". Якщо ПКТ однієї гіпотези є власною підмножиною ПКТ який-небудь іншої гіпотези, то дана гіпотеза відсівається. Вірогідність "спрацьовування" цієї евристики ~ 5%. Вірогідність правильної роботи цієї евристики – практично 100%.
- 2) "Відсутність нормальної форми". Якщо ПКТ гіпотези не має її нормальної форми, і існує гіпотеза з такою ж ПКТ, що має нормальну форму, то дана гіпотеза відсівається. Вірогідність "спрацьовування" цієї евристики ~ 15%. Вірогідність правильної роботи цієї евристики ~ 70–75%.

Якщо конкорданс системи використовується лише для пошуку, то третій етап побудови конкордансу можна вважати факультативним, оскільки

для пошукових цілей конкорданс виявляється, як правило, цілком прийнятним. По-перше, алгоритм генерації гіпотез дає відмова в основному за незмінними словами (аббревіатури, скорочення, слова типа "шоу", "мультимедіа") і для таких слів досить пошуку по точній формі. По-друге, за відсутності вірних гіпотез помилкові гіпотези незрідка виявляються сповна придатними для пошуку, оскільки вони можуть більш менш точно і повно описати список словоформ невідомої лексеми. Приклад неповного опису парадигми: формою "куздра" при використанні основного словника генерується гіпотеза за зразком "мездра", що має лише форми однини. Якщо ж гіпотези істотно помилкові – в основному це відноситься до помилкових дієслівних гіпотез для іменників типа "Байкал" – "байкати" – то словоформи з парадигм цих гіпотез є "неможливими", тобто в реальних текстах практично ніколи не зустрічаються, тому додаткового пошукового "шуму" при використанні цих гіпотез не виникає.

Програма, що демонструє можливості описаного алгоритму у складі Yandex-dict, знаходиться за адресою [www.comptek.ru/alta.html](http://www.comptek.ru/alta.html) і [www.comptek.ru/ramb.html](http://www.comptek.ru/ramb.html). Це оболонка над найбільш популярними пошуковими механізмами з тих, які не мають власних засобів роботи з російською морфологією Alta Vista і Rambler-ом.

## РОЗДІЛ 2. ТЕОРЕТИЧНІ ЗАСАДИ МОРФОЛОГІЧНОГО АНАЛІЗУ ПРИКМЕТНИКІВ

### 2.1. Загальна характеристика прикметника як частини мови

Прикметник протиставляється іменникові та дієслову як самостійним частинам мови.

Прикметник (лат. *adjectivum*) — частина мови з атрибутивною та предикативною функціями, яка має категоріальне значення непроцесуальної ознаки предмета, вираженої синтаксичними категоріями роду, числа, відмінка, категорією членності/нечленності й суфіксами.

Під ознакою розуміють властивості, якість, відношення, явища, якими характеризується предмет. Ознаками можуть бути:

- колір: *біла хустка, голубі очі, жовто-блакитний прапор;*
- смак: *солодкі: яблуко, гіркий перець;*
- розмір: *великий міст, малий палець;*
- матеріал: *цегляний будинок, шовкова блузка;*
- внутрішні (психологічні) риси людини: *добра людина, щедрий дідусь, надійний друг, сміливий воїн, мудрий наставник;*
- зовнішні риси людини чи тварини: *бадьорий хлопець, красива дівчина, міцний козак, худий кінь;*
- властивості предметів: *м'яке ліжко, зручне крісло, цікава книжка;*
- відношення до простору: *польові квіти, гірські потоки, лісова пісня, космічна ракета;*
- відношення до часу: *нинішні звичаї, вчорашній день, сьогоднішня газета, ранкова зарядка;*
- належність: *батьків портфель, материна сукня;*
- характеристик я за дією: *лякливий хлопець, сварлива жінка, вертляве дитя;*

— відношення до числа: *третій, сьомий*;

— якість: *золоті слова, умілі руки, професійний висновок, ґрунтовне пояснення, повна відповідь* тощо.

Ці ознаки формують відповідні лексико-семантичні групи слів, які становлять семантичну основу прикметника (як і іменника, дієслова, прислівника) як частини мови. Прикметникове значення ознаки має синтаксичний характер: до прикметників належать лише ті лексеми, які мають значення ознаки предмети, а тому здатні означати іменники та іменникові займенники і узгоджуватися з ними (*добрий день* — *доброго дня, я веселий* — *мене веселого*). Значення ознаки прикметника має свої засоби вираження — закінчення: в називному відмінку *-ий (-ій)* для чоловічого роду (*добрий, синій*), *-а, -я (-ая, -яя)* для жіночого (*добра, добрая, синя, синяя*), *-є, -е (-еє, -еє)* для середнього (*добрі; сине, добре. синєє*), *-і (-ії)* для множини (*добрі, добрії*).

Закінчення морфологічно оформляє прикметник як “ознакові” слова і тому є знаками якості (В. Виноградов). На думку О. Пешкопського, якщо в прикметнику *туманний* проста основа означає предмет (*туман*), а суфікс — відношення до нього (*-н-*), то, очевидно, значення ознаки може бути виражене тільки флексією *-ий*, тобто тим афіксом, який, власне, й робить його прикметником.

Прикметникові закінчення багатозначні. Вони означають частину мови, рід, число, відмінок, членність/нечленність.

Значення ознаки мають також іменники (*тиша, зелень, юнь. білизна, рань*), дієслівні форми (*зеленіти, веселитися, біліти, експлуатований, поранено*), прислівники (*навипередки, навперейми, тихо, навпростець, зелено*). Але вони не мають прикметникового плану вираження (флексій *-ий, -ій; -а, -я; -є, -е*), не спрямовані на предмет і не узгоджуються в іменником. Крім того, називаючи ознаку, такі іменники не передбачають її носія (пор. *Навколо тиша* і *Тихий вечір*), дієслово відносять до іменника процесуальну (пов'язану з дією, яка вказує на час, зв'язок із суб'єктом та об'єктом) ознаку

(земля зеленіє, море хвилюється), а прислівники відносять її не до іменника, а до дієслів і прислівників (*бігти навипередки, кидатися навперейми, навкруги тихо, скрізь зелено*).

Так можна довести, що: словоформи *теплий, хакі, електрик, комі, Шевченка* (вулиця) — прикметники; словоформи *навипередки, навперейми* (змагання *навипередки і змагатися навипередки, біг навперейми — бігти навперейми*), *по-варшавськи* (кава *по-варшавськи і каву зварити по-варшавськи*) і словоформи *перевірений, крокуючий* (*перевірений вчителем диктант, крокуючий полем навпростець бригадир*) не є прикметниками.

Отже, слово належить до категорії прикметника лише тоді, коли йому притаманні і семантичні (план змісту), і морфологічні (граматичні) ознаки (план вираження) прикметника. Визначення прикметників лише за семантичним або лише за морфологічним принципом дає різні результати. Це стосується лексем, що мають лише одну словоформу: *колір беж, пальто хакі, плаття електрик, література комі, колір бордо, машина кольору цукор, вулиця Франка, колір мокрого асфальту* та ін. З точки зору семантики — це прикметники, бо вони вказують на не процесуальну ознаку предмета і виконують атрибутивну функцію. З морфологічної точки зору вони не є прикметниками, бо не мають залежних від означуваного іменника категорій роду, числа, відмінка. Такі слова деякі мовознавці (М. Понов) виділяють її окрему групу і позивають *аналітичними прикметниками*.

Вживаючись у мові при іменниках, прикметники конкретизують їх зміст: *стіл — кухонний стіл, журнальний стіл, канцелярський стіл, однотумбовий стіл, паспортний стіл, солодкий стіл, святковий стіл, дерев'яний стіл, широкий стіл, круглий стіл* та ін. А інколи формують у них нове значення: *стіл* (предмет) — *паспортний стіл* (відділення, установа), *солодкий стіл* (їжа).

Поєднуючись з іменниками, прикметники конкретизують не лише іменникові значення, але й власні: *золота обручка* (матеріал), *золоте листя на деревах* (жовте, помаранчеве, червоне), *золотий характер* (гарний), *золоті*

*руки* (умілі), *золота осінь* (багата), *золоті слова* (мудрі), *золотий одяг*(дорогий) і под.

Деякі прикметники самі по собі не мають денотатів, а отже, й самостійного, як у іменника, лексичного значення. Воно формується на основі співвідношення прикметників з денотатами означуваних іменників: пор. *Залізні двері* (матеріал) - *залізна воля* (сильна), *стальні двері* (матеріал) — *стальні нерви* (міцні), *золоті сережки* (матеріал) — *золотий чоловік* (ідеальний).

Категоріальне значення прикметників виражається двома способами:

1) безпосередньо, поза відношенням ознаки до інших предметів, подій чи ознак: *солодкий, гіркий, дужий, зелений*;

2) опосередковано, шляхом відношення його до інших предметів, ознак, подій: *капроновий* (віднесений до предмета *капрон*), *білявий* (відношення до ознаки *білий*), *революційний* (відношення до події *революція*).

Словоформи, що належать до прикметника, за засобами вираження поділяють на морфологізовані і неморфологізовані. До морфологізованих належать ті, категоріальне значення яких виражене флексіями (субстанціальними або нульовими): *суверенний, вільного, гарною, веселе, годен, здоров, рад*. До неморфологізованих словоформ відносять ті, які не мають флексійного вираження, тобто невідмінювані: *бордо, теле, авіа, люкс, нетто, брутто, хакі, комі, хінд, урду*. Їх в українській мові дуже мало.

Морфологізовані прикметники значення ознаки виражають категоріями роду, числа, відмінка і членності/нечленності.

За здатністю чи нездатністю членуватися на основу і флексію прикметники поділяють на членні і нечленні. Членними (повними) бувають і якісні, і відносні прикметники: *розумний, добрий; вчорашній, теперішній, мідний*. Не членними (короткими) можуть бути лише якісні прикметники чоловічого роду в називному і знахідному відмінках з нульовим закінченням: *зелен сад*.

Для прикметників чоловічого роду звичайною є повна форма, яка виражається флексією -ий, -ій: *чарівний, сухий, далекий, хатний, могутній*. У жіночому і середньому роді в однині та множині повна форма, що виражається відповідно флексіями -ая, -ля, -єє, -ії, звичайному усному і писемному мовленню не властива і вживається лише у мові фольклору, поезії: *А в мого коня золотая грива, золотая грива, срібнії копита* (Нар. творчість); *Помандрував козаченько у чистеє поле* (Нар. творчість); *Земле моя, всеплодющая мати* (І. Франко); *Перед мене щасливе море, море голів* (П. Тичина); *На схід летять, пливають хмарки, як білії листки* (І. Вирган).

Коротка форма в чоловічому роді нечленною є історично, після занепаду кінцевих редукованих *ъ* та *ь* вона в називному відмінку стала вживатися з нульовою флексією: *новъ* → *нов*, *добръ* → *добр*. У сучасній українській мові прикметників чоловічого роду в короткій формі дуже мало: *винен, повинен, варт, винуват, годен, готов, жив, здоров, рад, ладен, повен* та ін. Вживаються вони переважно в усному мовленні, фольклорі, поезії: *Будь здоров! Будь готов! Я рад за тебе!*; *І шумить, і гуде, дрібен дощик іде* (Нар. творчість); *Як ішли ми зелен, яром к зелен гаю, солов'ї нам ткали пісню дивовижну* (Б. Олійник).

Отже, прикметники чоловічого роду існують у повній і короткій формах: *рад* — *радий*, *певен* — *певний*, *варт* — *вартий*. Але такі паралельні форми є нерегулярними, а тому несистемними. Більшість повних якісних прикметників (*розумний, веселий, стрункий, жвавий*) не має коротких відповідників.

Прикметники жіночого і середнього роду в однині та множині (*нова, нове, нові*) історично є членими, а в сучасній українській мові — стягненими. Стягнена форма розвинулася внаслідок стягнення голосних у повній формі:

У словоформах типу *новая* [новаја] після занепаду між голосними інтервокального [j] утворилася форма *новаа*, а після стягнення кінцевих [аа] в [а] — форма *нова*;

у словоформах типу новое [новоје], яка виникла з ново је новоје, внаслідок уподібнення [о] до [є] з'явилася форма новоје, що далі розвинулась у форму *новее*, з якої після сгягненіш [ее] утворилася форма *нове*;

членна множинна форми новыїѣ перетворилася на українському ґрунті и *новыїі*, потім - у *новиїі*, а після занепаду інтервакального [j] і стягнення кінцевих голосних з'явилася стягнена форми *нові*.

У процесі творення членних (повних), нечленних (коротких) і стягнених форм їх лексичне значення залишалось незмінним. Тому і категорія членності/нечленності як морфологічний засіб вираження категоріального значення ознаки є категорією словозмінною, а не класифікаційною. У прикметника вона має три грами (чоловічого, жіночого і середнього роду), в іменників — чотири (ще й граму парного роду). Ця категорія в іменників є самостійною, незалежною, несинтаксичною одиницею, а в прикметників вона синтаксична, залежна від означуваного іменника. Тому, на відміну від іменників, рід прикметників не поділяється на граматичний і семантико-граматичний, не виявляє варіантів, не регулюється власною семантикою. Прикметники не належать до одного з родів (як іменники), а змінюються за родами.

Категорія числа прикметників також відрізняється від категорії числа іменників. В іменників вона словотвірна, несинтаксична, граматично незалежна, а в прикметників — синтаксична, словозмінна, бо залежить від числа означуваного іменника: *навий будинок — нові будинки*.

Категорія відмінка у прикметника, як і в іменника, синтаксична, граматично залежна, з тією різницею, що форма відмінків прикметника залежить від іменника і формуються на основі узгодження (*фіолетова паста — фіолетовою пастою*), а форма відмінків іменника залежить від дієслів або віддієслівних іменників чи іншої частини мови і формується на основі керування (*читати книгу, читання книги, задоволений книгою*).

Прикметники послідовно маркують істотність чи неістотність іменників, з якими сполучаються. Ця їх властивість виявляється у формах



знахідного відмінка однини чоловічого роду і знахідного відмінка множини. Форма знахідного відмінка однина чоловічого роду і множини у прикметниках збігається з формою родового відмінка: *Зустрічаю давнього друга; Зустрічаю давніх друзів; Зловив великого коропа; Купив кількох коропів; Побачив сірих горобців.*

Форма знахідного відмінка однини чоловічого роду і множини у прикметниках при неістотних іменниках збігається з формою називного відмінка: *Побудували новий будинок; Підремонтували шкільні парти; Склали останній екзамен.*

Вибір прикметникових форм у цих позиціях зумовлюється значенням категорії істоти/неістоти відповідного іменника. У цих випадках категорія істоти/неістоти прикметника є синтаксичною, тобто позбавленою самостійної номінативності, залежною від відповідного іменника.

Інший характер має ця категорія (як і категорія числа відмінка) при сполучуваності з невідмінюваними іменниками та іменниками парного роду. В цих позиціях вона набуває самодостатнього, незалежного, несинтаксичного характеру і самостійно з допомогою відмінкових закінчень прикметника маркує неістотне або істотне значення іменника: а) *Зупинили вільне таксі; Прибули на профспілкове бюро;* б) *Показала дорослого шимпанзе і кількох великих какаду; Полюбили цих симпатичних двійнят; Нагородили колишнього старосту; Привітали нашу старосту; Відпочивали в зелених Карпатах.*

Отже, морфологічні ознаки прикметника є синтаксичними і словозмінними.

## 2.2 Суфікси прикметників

Прикметники відрізняються від інших частин мови й суфіксами. Найпоширенішими в українській мові є такі прикметникові суфікси:

Таблиця 2.1

### Суфікси прикметників

-ав-ий	кривавий, бідавий, ласкавий, жилавий, жовтавий
-ан (-ян) -ий	гречаний, копійчаний, дерев'яний, шерстяний
-ач-ий	лошачий, блошачий, мишачий, дівчачий, сорочачий
-н-ий	лісний, земний, мідний, лімітний, рігтий, лижний
-ов-ий	березовий, вербовий, липовий, дубовий, шовковий
-ськ-ий	сербський, мисливський, складський, сусідський
-к-ий	ламкий, грімкий, в'юнкий, хибкий, плавкий, давкий
-л-ий	задубілий, стиглий, смуглий, круглий, минулий
-ат-ий	вухатий, язикатий, лапатий, пихатий
-ів	батьків, братів, дядьків, бісів, Петрів
-ин	бабин, мамин, сестрин, тещин, Миколин
-уват-ий	темнуватий, сумнуватий, очеретуватий
-ист-ий	барвистий, йодистий, срібlistий, породистий
-аст-ий	вугластий, зубастий, довгастий, зеленастий
-ев-ий	бежевий, баржевий, металевий, сталевий
-ив-ий	злостивий, червивий, фальшивий, плаксивий, лінівий

Прикметникові суфікси відрізняються від іменникових і дієслівних будовою, функцією і значенням.

За будовою всі прикметникові суфікси консонантні (закінчуються на приголосний): *-ав-ий, -ан-ий, -ськ-ий, -ат-ий, -іє, -ин, -уват-ий, -аст-ий* та ін. Цим вони відрізняються від дієслівних, які всі вокативні (закінчуються на голосний): *-и- (бр-и-ти), -а- (чит-а-ти), -ти {нести), -ку-(стук-ну-ти), -ува- (бесід-ува-ти), -ка- (о-ка-ти)* тощо. У структурному відношенні

прикметникові суфікси подібні до іменникових, які переважно також є консонантними: *-ець (стілець), -ист (бандурист), -ап-и. (кияни), -ок (свисток), -ик-а (свитина), -ик (лісник), -ар (аптекарь), -енк-о (Шевченко), -ук (Шевчук)* та ін. Це пояснюється тим, що сучасні іменники і прикметники вийшли зі спільної давньої категорії імені, яка ще не була диференційована за предметністю і ознакою предмета.

За функцією прикметникові суфікси відрізняються від іменникових і дієслівних: функцією прикметникових суфіксів є вираження в слові непроцесуальної ознаки (*очеретуватий, зубастий*). Іменникові суфікси виражають предметність (*сміливість, сварка*), дієслівні — процесуальну ознаку (*зеленіти, обездолити*).

За значенням прикметникові суфікси відрізняються від іменникових і дієслівних. Усі прикметникові суфікси мають значення відношення (*йодистий, лижний, братів, березовий*), іменникові — значення особи, конкретного предмета, абстрактного поняття (*тракторист, свинарник, рівність*), дієслівні — значення дії, стану, вияву ознаки (*механізувати, спати, жовтіти*).

Моофи *-к-, -ськ-, -ист-* виявляються і в прикметниках, і в іменниках, але вони однакові лише за будовою, а за функцією і значенням різні: пор. прикметники *солодкий, гірський, складський, свійський, барвистий, вовнистий* і іменники *донька, виплавка, Придніпровськ, україніст*. Такі суфікси омонімічні.

Отже, суфікси є тією диференціальною ознакою, яка відрізняє прикметник від інших частин мови.

### 2.3. Номенклатура прикметникових афіксів української мови

Після детального опису диференціальної функції суфіксів українських прикметників розглянемо питання про номенклатуру афіксів – префіксів, суфіксів, флексій та біморфем у відповідності до зіставного на матеріалі української та російської мов словника службових морфем [12].

Серед префіксів в цьому виданні зустрічаються такі: *а, анти, архі, без, поза, внутрі, внутрішньо, до, за, між, межі, на, над, най, щонай, якнай, не, небез, по, під, пре, перед, при, проти, роз, понад, ультра, через.*

Номенклатура суфіксів представлена такими одиницями: *ав, яв, лав, альн, ільн, ан, ян, анн, янн, анськ, янськ, іанськ, аст, яст, ат, ач, яч, ащ, ящ, ш, іш, езн, елезн, ен, енн, енськ, еньк, есн, есеньк, ив, івськ, им, ин0, ін0, ин, ін, инськ, ист, іст, ит, ичн, ічн, инн, к, л, льн, н, ів0, їв0, ов, ев, єв, уват, юват, овит, овн, евн, івськ, ївськ, овськ, онн, ськ, зьк, цьк, тн, уч, юч, ущ, ющ, ч, чат, ш, іш, ин, ісіньк, юсіньк.*

Всі суфікси окрім позначених цифрою 0 (*ин0, ін0, ів0, їв0*) передбачають ненульову флексію.

Флексії (закінчення) прикметників для шести відмінків такі [14]:

1. тверда група, чоловічий рід – *ий, ого, ому, ий/ого, им, ому/ім;*
2. тверда група, жіночий рід – *а, ої, ій, у, ою, ій;*
3. тверда група, середній рід – *е, ого, ому, е/ого, им, ому/ім;*
4. тверда група, множина – *і, их, им, і/их, ими, их;*
5. м'яка група, чоловічий рід – *ій, ого, ому, ій/ого, ім, ому/ім;*
6. м'яка група, жіночий рід – *я, ої, ій, ю, ою, ій;*
7. м'яка група, середній рід – *є, ого, ому, є/ого, ім, ому/ім;*
8. м'яка група, множина – *і, їх, ім, і/їх, іми, їх.*

З прикметникових біморфем були враховані такі: *небез-н, про-ськ, про-зьк, про-анськ, про-н, спів-н, су-н, транс-н, транс-ов, транс-ськ, транс-ійськ.*

Названі афікси можуть бути використані для з'ясування можливої належності певного слова до класу прикметників.

## РОЗДІЛ 3. АВТОМАТИЧНИЙ МОРФОЛОГІЧНИЙ АНАЛІЗ ПРИКМЕТНИКІВ

### 3.1. Представлення афіксів

Для реалізації підсистеми автоматичного морфологічного аналізу було вибрано мову програмування SWI-Prolog – мову логічного програмування, орієнтовану на обробку символічних даних.

Флексії прийменників, описані в п. 2.3 представлені в програмі у формі фактів Прологу. Предикат `inflexion/5` має компоненти Флексія, ЧастинаМови, Відмінок, Число та Рід-Група, див. наведений далі фрагмент:

```
% inflexion(Флексія,ЧастинаМови,Відмінок,Число,Рід-Група)
```

```
% тверда група
```

```
% чоловічий рід
```

```
inflexion(ий,прикметник,нв,s,ч-а).
```

```
inflexion(ого,прикметник,рв,s,ч-а).
```

```
inflexion(ому,прикметник,дв,s,ч-а).
```

```
inflexion(ий,прикметник,зв,s,ч-а).
```

```
inflexion(ого,прикметник,зв,s,ч-а).
```

```
inflexion(им,прикметник,ов,s,ч-а).
```

```
inflexion(ому,прикметник,мв,s,ч-а).
```

```
inflexion(ім,прикметник,мв,s,ч-а).
```

```
% жіночий рід
```

```
inflexion(а,прикметник,нв,s,ж-а).
```

```
inflexion(ої,прикметник,рв,s,ж-а).
```

```
inflexion(ій,прикметник,дв,s,ж-а).
```

```
inflexion(у,прикметник,зв,s,ж-а).
```

```
inflexion(ою,прикметник,ов,s,ж-а).
```

```
inflexion(ій,прикметник,мв,s,ж-а).
```

```
% середній рід
```

inflexion(е,прикметник,нв,s,c-a).

inflexion(ого,прикметник,рв,s,c-a).

inflexion(ому,прикметник,дв,s,c-a).

inflexion(е,прикметник,зв,s,c-a).

inflexion(ого,прикметник,зв,s,c-a).

inflexion(им,прикметник,ов,s,c-a).

inflexion(ому,прикметник,мв,s,c-a).

inflexion(ім,прикметник,мв,s,c-a).

%множина

inflexion(і,прикметник,нв,p,0-a).

inflexion(их,прикметник,рв,p,0-a).

inflexion(им,прикметник,дв,p,0-a).

inflexion(і,прикметник,зв,p,0-a).

inflexion(их,прикметник,зв,p,0-a).

inflexion(ими,прикметник,ов,p,0-a).

inflexion(их,прикметник,мв,p,0-a).

Префікси, суфікси та біморфеми представлені у формі предикатів префікс\_прикм/1, суфікс\_прикм/1 та біморф\_прикм/1 відповідно. Єдиний компонент кожного з них є списком відповідних флексій, впорядкованих в зворотному лексикографічному порядку, див. Додаток А.

### 3.2. База прикметників

Для апробації застосовуваних методів використовувалась база (реєстр) прикметників Міжвідомчого центру моделювання природної мови та інтелекту. Реєстр містить понад 400 тис. словоформ, записаних та впорядкованих в зворотному порядку з кодами граєм від слова *слабі* – *прикметник*, *знахідний відмінок*, *множина* до слова *довговія* – *прикметник*, *називний відмінок*, *однина*, див приклад далі:

ібалс = 7 54 41

ібалсаз = 7 54 41

ібардолог = 7 54 41

іболодревт = 7 54 41

іболокориш = 7 54 41

.....

*юошьладйан* = 7 92 55 40 31

*юошьладоп* = 7 55 40 31

*юцилзеб* = 7 54 40 31

*яівоввод* = 7 51 40 31

Для представлення графем прикметників використовуються такі коди:

7 прикметник

30 чоловічий рід

31 жіночий рід

32 середній рід

33 спільний рід

40 однина

41 множина

51 називний відмінок

52 родовий відмінок

53 давальний відмінок

54 знахідний відмінок

55 орудний відмінок

56 місцевий відмінок

57 кличний відмінок

91 вищий ступінь порівняння

92 найвищий ступінь порівняння.

### 3.3. Автоматизація морфологічного аналізу

Для автоматизації морфологічного аналізу прикметників за префіксами та суфіксами було розроблено такі предикати.

прикметник(Словоформа,L):-

```

inflexion(Афікс,прикметник,Відмінок,Число,Рід-Відміна),
(atom_concat(Основа, Афікс, Словоформа),
(біморф(Основа,Рез), L=[Рез+Афікс,Відмінок,Число,Рід-Відміна],!;
прикмет(Основа,Рез), L=[Рез+Афікс,Відмінок,Число,Рід-Відміна],!
)
;
Основа=Словоформа,
прикмет0(Основа,Рез),
L=[Рез,нв,s,ч-a],!
).
```

Смисл його наступний: якщо входом є Словоформа, та факт *inflexion* відповідає наявному Афіксу з відповідними графемами, а конкатенація Основи з Афіксом становлять Словоформу, тоді успішність однієї з цілей *біморф(Основа, Рез)*, *прикмет(Основа, Рез)* чи *прикмет0(Словоформа, Рез)*, означає досяжність цілі *прикметник(Словоформа, L)*, де L – містить представлення структури прикметника та його графеми.

Предикат *прикмет/2*, конкретизує змінну *Рез* розкладенням Основи на Префікс, Суфікс та Залишок, або на Залишок та Суфікс, де Залишок – один чи кілька коренів, інтерфіксів і т.і

прикмет(Основа,Рез):-

```

(префікс_прикм(Префікси),
member(Префікс,Префікси),
atom_concat(Префікс,Залишок1,Основа)
;

```



```

Залишок1=Основа
),
суфікс_прикм(Суфікси),
member(Суфікс,Суфікси),
atom_concat(Залишок2,Суфікс,Залишок1),
(var(Префікс), Рез=Залишок2+Суфікс
;
Рез=Префікс+Залишок2+Суфікс
).
```

Наступні запити дають уявлення про роботу предиката прикмт / 2

?- прикмт(прекрасн,Рез).

Рез = пре+крас+н

?- прикмт(заправськ,Рез).

Рез = за+прав+ськ

?- прикмт(кепськ,Рез).

Рез = кеп+ськ

?- прикмт('трамвайно-тролейбусн',Рез).

Рез = 'трамвайно-тролейбус'+н

?- прикмт(великотонажн,Рез).

Рез = великотонаж+н

Аналогічно влаштований предикат біморф / 2:

біморф(Основа,Морф1+Залишок2+Морф2):-

```

біморф_прикм(VMorph),
member([Морф1,Морф2],VMorph),
atom_concat(Морф1,Залишок1,Основа),
```

atom\_concat(Залишок2,Морф2,Залишок1).

Приклади запитів:

?- біморф(співзвучн,Рез).

Рез = спів+звуч+н

?- біморф(транссибірськ,Рез).

Рез = транс+сибір+ськ

Предикат прикмт0/2, застосовується у випадку, коли флексія прикметника нульова. Відповідні суфікси представлені у формі факту суфікс\_прикм0 ([ин, ін, ів, їв]).

Запити:

?- прикмт0(сестрин,Рез).

Рез = сестр+ин

?- прикмт0('Іванів',Рез).

Рез = 'Іван'+ів

?- прикмт0(левів,Рез).

Рез = лев+ів

Тепер розглянемо приклади запитів до предикату прикметник/2:

?- прикметник(прекрасному,L).

L = [пре+крас+н+ому, дв, s, ч-а].

?- прикметник(заправських,L).

L = [за+прав+ськ+их, рв, р, 0-а].

?- прикметник(ліфтобудівними,L).

L = [ліфтобудів+н+ими, ов, р, 0-а].

?- прикметник(пересопницьке,L).

L = [пересопни+цьк+е, нв, s, с-а].

?- прикметник(сестриними,L).

L = [сестри+н+ими, ов, р, 0-а].

?- прикметник(сестрин,L).

L = [сестр+ин, нв, s, ч-а].

З метою перевірки повноти такого способу віднесення слова до класу прикметників, виконаємо фільтрацію реєстру за допомогою предиката `byLineFromFile`:-

```

open('adjective_not_dig.txt',read,Strm,
                                     [encoding(text),eof_action(eof_code)]),
open('Out_adjective.txt',write,StrmOut,
                                     [encoding(text),eof_action(eof_code)]),
repeat,
read_line_to_codes(Strm,LineCodes),
reverse(LineCodes,LCodes),
name(Словоформа,LCodes),
string_to_list(ССловоформа,LCodes),
(not(прикметник(Словоформа,_))),
    write_term(StrmOut,ССловоформа,[quoted(true)]),nl(StrmOut)
),
at_end_of_stream(Strm),
close(Strm),
close(StrmOut).
```

Він забезпечує порядкове читання вхідного файлу (реєстр) і, якщо слово не розпізнається як прикметник – виведення його до вихідного файлу. За запитом

?- byLineFromFile.

отримуємо у вихідному файлі близько 11 тис. словоформ, які не були розпізнані як прикметники. Ось деякі з них:

"слабі"

"заслабі"

"голодрабі"

"твердолобі"

"широколобі"

.....

"тисячоусте"

"пусте"

"напівпусте"

"непусте"

"напівзабуте"

"незабуте"

"нездобуте"

.....

"нищих"

"кущих"

"віщих"

"товщих"

"вищих"

"найвищих"

"якнайвищих"

"щонайвищих"

"нищих"

"довговію"

"слабою"

"безлицю"

"довговія"

Як видно з наведених фрагментів, це випадки, коли прикметник не має суфікса, але має флексію, причому основа закінчується на приголосний. Внівши відповідні приголосні в список суфіксів, очікуємо більш повного виявлення прикметників. Повторивши обробку, отримуємо 125 слів, не визначених як прикметники. Далі наводиться фрагмент списку:

"голоший"

"товстоший"

"безкраїм"

"неокраїм"

"цілокраїм"

"Дар'ін"

"Ліїн"

"Євсевіїн"

"Парасковіїн"

"Сільвіїн"

"Пелагіїн"

"Лідіїн"

"Надіїн"

З фрагменту видно, що це випадки, коли основа закінчується символом "й", який разом з голосним вихідної флексії дає флексії йотовані. Повний список таких словоформ (125 з 400 тис.) наведено у Додатку Б.

### **3.4. Лематизація прикметників**

Морфологічний аналіз, в т.ч. й автоматичний, передбачає лематизацію. Оскільки наша система виконує класифікацію, використовуючи також і флексії, то можна, скориставшись грамемами флексії, знайти флексію леми.

Вона буде мати ті ж грами, що і вихідна, але зі значенням називного відмінка. Реалізуємо цю ідею за допомогою предиката лема\_прикм/2:

лема\_прикм(Прикметник,Лема):-прикметник(Прикметник,L),  
 L=[Основа+\_\_,Число,Рід-\_,  
 inflexion(ФлексіяЛеми,\_нв,Число,Рід-\_),  
 Лема=Основа+ФлексіяЛеми.

?- лема\_прикм(перцептивним,Лема).

Лема = перцептив+н+ий

?- лема\_прикм(зеленесенькім,Лема).

Лема = зеленесень+к+ий

?- лема\_прикм(зеленесенькою,Лема).

Лема = зеленесень+к+а

?- лема\_прикм(зеленесеньких,Лема).

Лема = зеленесень+к+і

?- лема\_прикм(суглинистим,Лема).

Лема = суглин+ист+ий

?- лема\_прикм(пропорційними,Лема).

Лема = про+порцій+н+і

?- лема\_прикм(щонайвищих,Лема).

Лема = щонай+ви+щ+і

?- лема\_прикм(чорнобрової,Лема).

Лема = чорнобр+ов+а

?- лема\_прикм(швидконогих,Лема).

Лема = швидконо+г+і

Отже лема обчислюється коректно для прикметників з суфіксами та з твердими приголосними у безсуфіксних випадках.

?- лема\_прикм(білолицької,Лема).

Лема = білоли+ць+а

?- лема\_прикм(літньому,Лема).

Лема = літ+нь+ий

Наведені приклади ілюструють проблеми при обробці м'яких звуків та доводять, що лематизація потребує додаткового аналізу.

Проте, понад 97% прикметників з реєстру в 400 тис. словоформ мають суфікси, отже можна очікувати, що і лематизація цих прикметників буде коректною.

## ВИСНОВКИ

З аналізу публікацій в наукових виданнях було зроблено висновок, що на ринку програмних продуктів не існує відкритої системи автоматичного морфологічного аналізу української мови.

Засобами мови SWI-Prolog було розроблено систему автоматичного морфологічного аналізу прикметників української мови на основі аналізу афіксальної структури словоформи. Розроблено простий метод лематизації прикметників.

Перевірена на реєстрі обсягом понад 400 тис. словоформ повнота системи складає понад 99%. Швидкість обробки даних перевищує 600 К/с.

Отримані результати можуть бути застосовані при викладанні дисципліни "Логічне програмування" та в системах автоматичної обробки текстів.



## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Зализняк А. А. Грамматический словарь русского языка (словоизменение). 2-е изд., М., "Русский язык", 1980.
2. Скобликова Е.С. Очерки по теории словосочетания и предложения. Изд-во Саратовского университета: Куйбышевский филиал. 1990.
3. Лингвистическое обеспечение системы ЭТАП-2. М., 1989.
4. И. Сегалович, М. Маслов, Русский морфологический анализ и синтез с генерацией моделей словоизменения для не описанных в словаре слов [<http://company.yandex.ru/articles/article1.html>]
5. Г.Г. Белоногов, Ю.П. Калинин, А.А. Хорошилов, Компьютерная лингвистика и перспективные информационные технологии - М.: Русский мир, 2004.
6. Андреев А.М., Березкин Д.В., Симаков К. В. Особенности проектирования модели и онтологии предметной области для поиска противоречий в правовых электронных библиотеках. 6-ая Всероссийская научная конференция RCDL'2004.
7. Брик А.В. Исследование и разработка вероятностных методов синтаксического анализа текста на естественном языке. Диссертация на соискание ученой степени к.т.н. - М.: МГТУ им. Н.Э. Баумана, 2002.
8. Дарчук Н.П. Комп'ютерна лінгвістика. – К.: Видавничо-поліграфічний центр «Київський університет», 2008. – 351 с.
9. Белоногов Г.Г. Компьютерная лингвистика и перспективные информационные технологии. - М.: Русский мир, 2004. - 248 с.
10. Грязнухина Т.А., Дарчук Н.П., Клименко Н.Ф. Использование ЭВМ в лингвистических исследованиях. - К., 1990. - 223 с.
11. «Использование ЭВМ в лингвистических исследованиях», отв. ред. В.И. Перебийнос, Киев, «Наукова Думка», 1990.
12. Камынина А.А. Современный русский язык. Морфология. – М.: Изд-во Московского ун-та, 1999. – 240 с.

13. Цыганенко Г.П. Словарь служебных морфем русского языка. – К.: Рад. школа, 1982. – 242 с.
14. Горпинич В.О. Морфологія української мови: Підручник для студентів вищих навчальних закладів. – К.: ВЦ “Академія”, 2004. – 336 с. (Альма-матер)

## ДОДАТКИ

## Додаток А

## Представлення афіксів українських прикметників

% тверда група

% чоловічий рід

inflexion(ий,прикметник,нв,s,ч-а).

inflexion(ого,прикметник,рв,s,ч-а).

inflexion(ому,прикметник,дв,s,ч-а).

inflexion(ий,прикметник,зв,s,ч-а).

inflexion(ого,прикметник,зв,s,ч-а).

inflexion(им,прикметник,ов,s,ч-а).

inflexion(ому,прикметник,мв,s,ч-а).

inflexion(ім,прикметник,мв,s,ч-а).

% жіночий рід

inflexion(а,прикметник,нв,s,ж-а).

inflexion(ої,прикметник,рв,s,ж-а).

inflexion(ій,прикметник,дв,s,ж-а).

inflexion(у,прикметник,зв,s,ж-а).

inflexion(ою,прикметник,ов,s,ж-а).

inflexion(ій,прикметник,мв,s,ж-а).

% середній рід

inflexion(е,прикметник,нв,s,с-а).

inflexion(ого,прикметник,рв,s,с-а).

inflexion(ому,прикметник,дв,s,с-а).

inflexion(е,прикметник,зв,s,с-а).

inflexion(ого,прикметник,зв,s,с-а).

inflexion(им,прикметник,ов,s,с-а).

inflexion(ому,прикметник,мв,s,с-а).

inflexion(ім,прикметник,мв,s,с-а).

%множина

inflexion(і,прикметник,нв,р,0-а).

inflexion(их,прикметник,рв,р,0-а).

inflexion(им,прикметник,дв,р,0-а).

inflexion(і,прикметник,зв,р,0-а).

inflexion(их,прикметник,зв,р,0-а).

inflexion(ими,прикметник,ов,р,0-а).

inflexion(их,прикметник,мв,р,0-а).

% м'яка група

inflexion(ій,прикметник,нв,с,ч-б).

inflexion(ого,прикметник,рв,с,ч-б).

inflexion(ому,прикметник,дв,с,ч-б).

inflexion(ій,прикметник,зв,с,ч-б).

inflexion(ого,прикметник,зв,с,ч-б).

inflexion(ім,прикметник,ов,с,ч-б).

inflexion(ому,прикметник,мв,с,ч-б).

inflexion(ім,прикметник,мв,с,ч-б).

inflexion(я,прикметник,нв,с,ж-б).

inflexion(ої,прикметник,рв,с,ж-б).

inflexion(ій,прикметник,дв,с,ж-б).

inflexion(ю,прикметник,зв,с,ж-б).

inflexion(ою,прикметник,ов,с,ж-б).

inflexion(ій,прикметник,мв,с,ж-б).

inflexion(є,прикметник,нв,с,с-б).

inflexion(ого,прикметник,рв,с,с-б).

inflexion(ому,прикметник,дв,с,с-б).

inflexion(є,прикметник,зв,с,с-б).

inflexion(ого,прикметник,зв,с,с-б).

inflexion(ім,прикметник,ов,с,с-б).

inflexion(ому,прикметник,мв,с,с-б).

inflexion(ім,прикметник,мв,s,c-б).

% множина

inflexion(і,прикметник,нв,p,0-б).

inflexion(їх,прикметник,рв,p,0-б).

inflexion(ім,прикметник,дв,p,0-б).

inflexion(і,прикметник,зв,p,0-б).

inflexion(їх,прикметник,зв,p,0-б).

inflexion(іми,прикметник,ов,p,0-б).

inflexion(їх,прикметник,мв,p,0-б).

префікс\_прикм([якнай,щонай,через,ультра,роз,під,проти,при,пре,понад,поза, по,перед,небез,не,най,над,на,між,межи,за,до,внутрішньо,внутрі,без,архі,анти, а]).

біморф\_прикм([[небез,н],[про,ськ],[про,зьк],[про,анськ],[про,н],[спів,н],[су,н] ,[транс,н],[транс,ов],[транс,ськ],[транс,ійськ]]).

суфікс\_прикм([ївськ,іш,ічн,ісіньк,іст,ін,ільн,івськ,іанськ,єв,ящ,яч,яст,янськ,я нн,ян,яв,ющ,юч,юсіньк,юват,шн,ш,чат,ч,цьк,ущ,уч,уват,тн,ськ,онн,овськ,овн, овит,ов,н,ляв,льн,л,к,ичн,ит,ист,инськ,инн,ин,им,ив,зьк,есн,есеньк,еньк,енськ ,енн,ен,елезн,езн,евн,єв,аш,ач,ат,аст,анськ,анн,ан,альн,ав]).

суфікс\_прикм0([ин0,ін0,ів0,їв0]).

## Додаток Б

## Результати фільтрації реєстру прикметників

"безкрає"	"неокрає"	"цілокрає"
"безкраї"	"неокраї"	"цілокраї"
"довговійй"	"безкрайй"	"неокрайй"
"цілокрайй"	"кривошийй"	"довгошийй"
"тонкошийй"	"широкошийй"	"короткошийй"
"білошийй"	"голошийй"	"товстошийй"
"безкраїм"	"неокраїм"	"цілокраїм"
"Дар'їн"	"Ліїн"	"Євсевійн"
"Парасковійн"	"Сільвійн"	"Пелагійн"
"Лідійн"	"Надійн"	"Леокадійн"
"Аркадійн"	"Терезійн"	"Теодозійн"
"Домнікіїн"	"Докійн"	"Євдокійн"
"Лукіїн"	"Ліліїн"	"Еміліїн"
"Теофіліїн"	"Цеціліїн"	"Юліїн"
"Азаліїн"	"Віталіїн"	"Наталіїн"
"Анеліїн"	"Корнеліїн"	"Авреліїн"
"Анатоліїн"	"Артеміїн"	"Єфиміїн"
"Юхиміїн"	"Соломіїн"	"Віргініїн"
"Юніїн"	"Меланіїн"	"паніїн"
"Степаніїн"	"Стефаніїн"	"Феофаніїн"
"Євгеніїн"	"Ксеніїн"	"Поліксеніїн"
"Аксеніїн"	"Оксеніїн"	"Арсеніїн"
"Парфеніїн"	"Єфросиніїн"	"Фотиніїн"
"Піоніїн"	"Леоніїн"	"Феоніїн"
"Аполлоніїн"	"Февроніїн"	"Вероніїн"
"Агапіїн"	"Євлампіїн"	"Даріїн"

"Маріїн"	"Аполлінаріїн"	"Гликеріїн"
"Лукеріїн"	"Валеріїн"	"Флоріїн"
"Гаїсіїн"	"Афанасіїн"	"Аспасіїн"
"Євфрасіїн"	"Настасіїн"	"Онисіїн"
"Денисіїн"	"Хрисіїн"	"Євпраксіїн"
"Олексіїн"	"Федосіїн"	"Феодосіїн"
"Февросіїн"	"Мелетіїн"	"Леонтіїн"
"Феодотіїн"	"фійн"	"Гафійн"
"Агафійн"	"Євстафійн"	"Софійн"
"Євстахіїн"	"Євтихіїн"	"Лукреціїн"
"Констанціїн"	"Люціїн"	"уабаїн"
"тебаїн"	"Аглаїн"	"Леїн"
"Пелагеїн"	"Саломеїн"	"Віринейн"
"Доротеїн"	"Дорофеїн"	"антофеїн"
"Майїн"	"орлього"	"безкраїх"
"неокраїх"	"цілокраїх"	