

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДЕРЖАВНИЙ ВИЩИЙ НАВЧАЛЬНИЙ ЗАКЛАД
«ДОНЕЦЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ»
КАФЕДРА КОМП'ЮТЕРНИХ НАУК

КОНСПЕКТ ЛЕКЦІЙ

з дисципліни «Сучасні методи обробки природно-мовних текстів»
для студентів спеціальності 122 Комп'ютерні науки
денної та заочної форми навчання

Конспект лекцій з дисципліни «Сучасні методи обробки природно-мовних текстів» для студентів спеціальності 122 Комп'ютерні науки денної та заочної форми навчання / [укл. Є.Є. Федоров, І.В. Ярош, Т.О. Черняк]. – Покровськ: ДВНЗ «ДонНТУ», 2018. – 60 с.

Конспект лекцій містить основні поняття граматики, що включає в себе словотвір, морфологію і синтаксис

Конспект лекцій складений відповідно до вимог державних освітніх стандартів і рекомендований для студентів, які здобувають освітній ступень «магістр», освітньо-професійної програми підготовки зі спеціальності 122 Комп'ютерні науки.

Укладачі: Є.Є. Федоров, д.т.н., доцент, завідувач кафедри комп'ютерних наук
І.В. Ярош, старший викладач кафедри комп'ютерних наук
Т.О. Черняк, асистент кафедри комп'ютерних наук

Рецензент: С.О. Ковальов, к.т.н., доцент, декан ФКНТ

Відповідальний за випуск: Є.Є. Федоров, зав. каф. КН

Затверджено навчально-методичним відділом ДВНЗ «ДонНТУ»,
протокол № 12 від 24.04.2018 р.

Розглянуто на засіданні кафедри комп'ютерних наук,
протокол №13 від 12.04.2018 р.

ЗМІСТ

ВСТУП.....	4
ЗМІСТОВИЙ МОДУЛЬ 1 Вступ в граматику мови.....	6
Тема 1. Словотвір.....	6
Тема 2. Морфологія.....	10
Тема 3. Синтаксис.....	12
ЗМІСТОВИЙ МОДУЛЬ 2 Дослідження алгоритмів і методів аналізу мовних конструкцій.....	23
Тема 4. Морфологічний аналіз.....	24
Тема 5. Синтаксичний аналіз.....	26
Тема 6. Семантичний аналіз.....	46
Тема 7. Об'єднання морфологічного, синтаксичного і семантичного аналізів.....	56
СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ.....	58

ВСТУП

На даний час в дослідженнях з інтелектуальних систем виділяють наступні проблеми [1]:

1. Представлення знань.
2. Маніпулювання знаннями.
3. Спілкування.
4. Сприйняття.
5. Навчання.
6. Поведінка.

В роботі [2] виділені такі проблеми як:

1. Представлення знань.
2. Комп'ютерна лінгвістика.
3. Комп'ютерна логіка.

Стосовно до інтелектуальних систем управління ці проблеми можна сформулювати більш вузько:

1. Формалізація і уявлення знань про об'єкти управління різної природи.
2. Розпізнавання команд і повідомлень, які супроводжують управління процесом на візькоспеціалізованій технічній мові.
3. Комунікація взаємодії особи, яка приймає рішення, і інтелектуальної системи.
4. Навчання розпізнаванню командуправління об'єктом і повідомлень про стан об'єкта.

Для фахівця в області теорії управління природна мова виступає як засіб опису тих об'єктів, якими він управляє; ситуацій, що складаються на цих об'єктах; процедур управління.

В процесі функціонування системи управління можуть коригуватися мови опису ситуацій, змінюватися знання про об'єкт і методи управління ним. Тому в роботі [3] замість традиційної формальної логічної моделі з постійними елементами використовується семіотична модель, яка дозволяє модифікувати

лексику (базові поняття, відносини, дії та ін.), синтаксис, семантику і прагматику.

В роботі [3] висунуто дві гіпотези:

1. Всю інформацію про об'єкт управління і засобах управління ним можна висловити засобами звичайної природної мови.

2. Всякий текст на природній мові, що відноситься до того, про що йдеться в гіпотезі 1, можна перевести на формальну мову семіотичної моделі.

Таким чином, виникає проблема знаходження способу природно-мовного опису необхідної інформації про об'єкт управління і засобах його управління, який дозволив би реалізувати процедури аналізу і класифікації ситуацій, кореляції (визначення правила переходу від поточної ситуації до нової) і екстраполяції (вибір найкращого правила переходу).

Спочатку розглянемо основні поняття граматики, що включає в себе словотвір, морфологію і синтаксис.

ЗМІСТОВИЙ МОДУЛЬ 1

ВСТУП В ГРАМАТИКУ МОВИ

Тема 1. Словотвір

Тема 2. Морфологія

Тема 3. Синтаксис

Тема 1. Словотвір

Словотвір [4, 5] вивчає будову слів (з яких частин вони складаються) і способи їх утворення.

1.1. Основні поняття словотвору

Одне зі слів, пов'язаних відносинами похідності (мотивації), є виробляючим (базовим), а інше – похідним. Похідне слово складається з виробляючої частини і форманта.

Непохідне слово – слово, що не утворене від інших слів.

Словотворчий формант – сукупність всіх засобів, що беруть участь в утворенні слова.

Словотвірний ланцюжок – ряд однокореневих слів, що знаходяться у відносинах послідовної похідності.

Словотвірна парадигма – сукупність всіх похідних слів від одного виробляючого.

Словотворче гніздо (див. рис.1.1) – сукупність слів із загальним коренем, упорядкована відповідно до відносин похідності. Вершиною (вихідним словом) гнізда є непохідне слово. Гніздо може бути визначено як сукупність словотворчих ланцюжків (наприклад, откупорить→откупориться→откупориваться), що мають одне й те ж вихідне слово, і як сукупність словотворчих парадигм (откупорить→откупориться, откупоривать, откупорка).

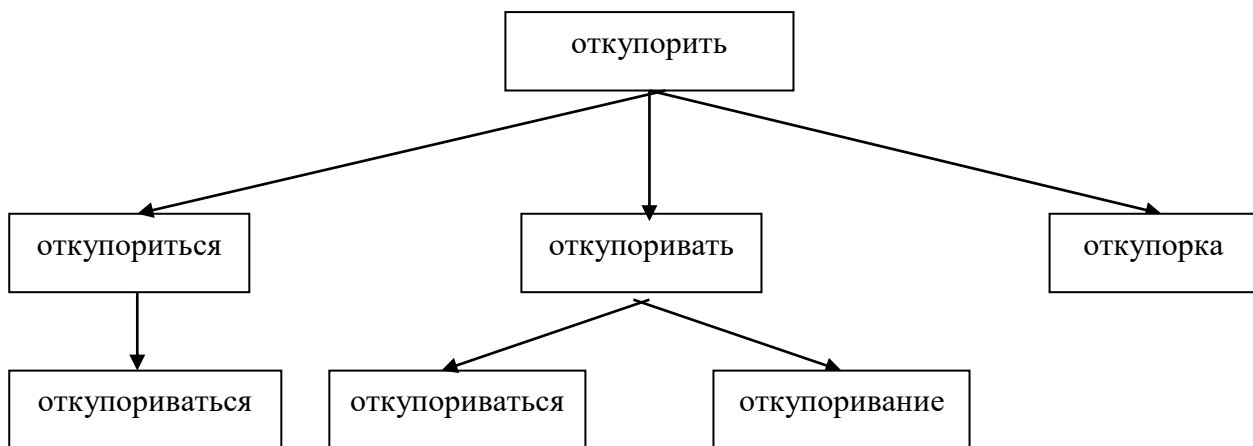


Рисунок 1.1 – Гніздо, утворене вихідним словом «откупорить»

Крім словотворчих гнізд виділяють кореневі гнізда. *Кореневе гніздо* включає два чи більше словотворчих гнізда, вихідні слова яких характеризуються спільністю кореня (див. рис.1.2).

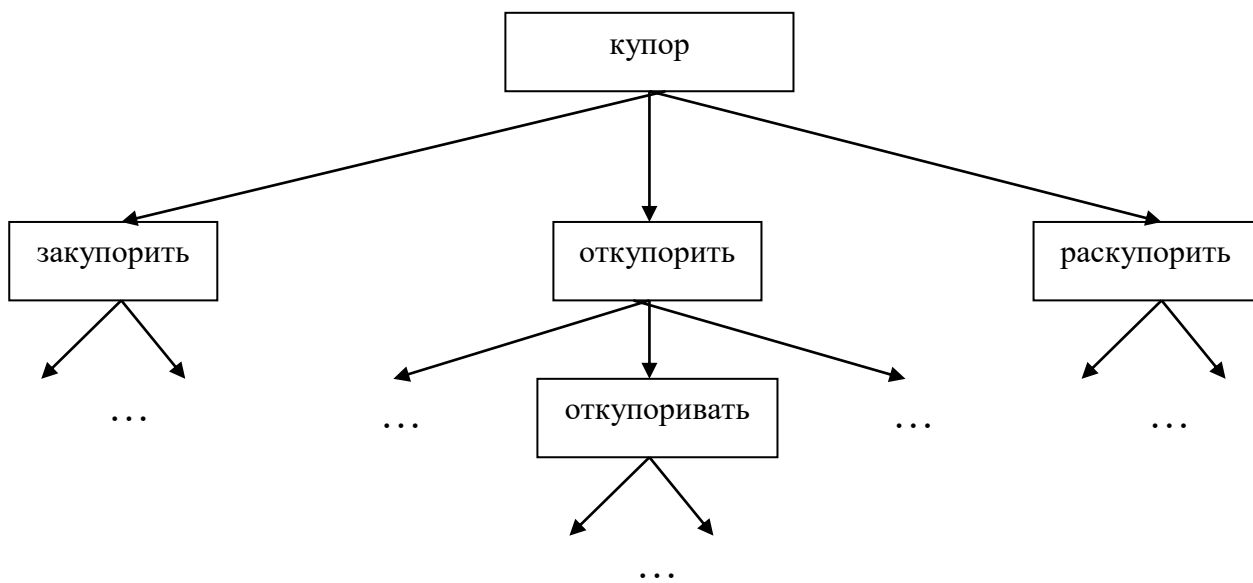


Рисунок 1.2 – Гніздо, утворене коренем «купор»

1.2. Частини слова

Морфема – мінімальна значуща частина слова.

Морф – мінімальна значима частина словоформи (конкретного представника слова).

Типи морфем і морфів:

1. Кореневі;

2. Афіксальні:

– префікс (приставка);

– інтерфікс (з'єднувальний афікс);

– суфікс;

– флексія (закінчення);

– постфікс.

Сукупність однойменних (що належать до одного класу, наприклад до кореня) морфів, що володіють тотожністю значення і фонетичною близькістю (близькі за складом фонеми), складають морфему (наприклад, морфи «-ств-» і «-еств-» об'єднуються в одну морфему). Таким чином, морфема – це узагальнена одиниця, а морф – конкретні представники (репрезентанти) морфеми.

1.3. Способи словотворення

Виділяють наступні способи словотворення, що використовують словотвірний формант:

– суфіксація (до складу форманта входить суфікс і флексія);

– префіксація (формантом є префікс);

– постфіксація (формантом є постфікс);

– префіксально-суфіксальний (до складу форманта входить префікс, суфікс і флексія);

– префіксально-постфіксальний (до складу форманта входить префікс і постфікс);

– суфіксально-постфіксальний (до складу форманта входить суфікс, флексія і постфікс);

- префіксально-суфіксально-постфіксальний (до складу форманта входить префікс, суфікс, флексія і постфікс);
- субстантивація прикметників і дієприкметників (формантом є флексія) (наприклад, операционный→ операционная);
- додавання (до слова зліва додається основа з інтерфіксом, який входить в формант);
- префіксально-складний (додавання з префіксацією, до складу форманта входить префікс і інтерфікс);
- суфіксально-складний (додавання з суфіксацією, до складу форманта входить інтерфікс, суфікс і флексія);
- префіксально-суфіксально-складний (додавання з префіксацією і суфіксацією, до складу форманта входить префікс, інтерфікс, суфікс і флексія);
- зрощення (до слова зліва додається основа без інтерфіксу);
- зрощення з суфіксацією (до складу форманта входить суфікс);
- аббревіація (формантом є флексія).

1.4. Морфологічні явища в словотворенні

Морфологічні перетворення – це перетворення фонемного складу, обумовлені сполучуваністю з певними морфемами, що входять до складу форманта.

Це перетворення мають місце:

- перед суфіксальними морфами при суфіксальному, префіксально-суфіксальному, суфіксально-постфіксальному, суфіксально-складному способах словотворення;
- перед інтерфіксальними морфами в складдних і суфіксально-складних словах.

Морфонологічні перетворення в словотворенні діляться на:

- лінійні;
- нелінійні.

До *лінійних перетворень* відноситься усічення або нарощення основи за рахунок фонем або поєднання фонем на початку або кінці.

До *нелінійних перетворень* відноситься чергування фонем в межах морфеми (в тому числі чергування гласної з нулем, тобто побіжність гласної), поєднання морфів (основи і суфікса, основи і інтерфікса), імена наголоси.

Тема 2. Морфологія

Морфологія [4, 5] об'єднує граматичні класи слів (частини мови), що належать цим класам граматичні (морфологічні) категорії і форми слів. Термін «словозміна» часто вживають як синонім терміну «морфологія», тобто морфологія вичає правила словозміни.

2.1. Частини мови і морфологічні категорії

Словозмінна парадигма – сукупність всіх форм слів.

До частин мови відноситься:

- іменник;
- прикметник;
- числівник;
- займенник;
- дієслово;
- прислівник;
- прийменник;
- сполучник;
- частка;
- вигук.

До морфологічних категорій іменників відноситься: рід, число, відмінок, одухотворений-неодухотворений, власне-загальне.

До морфологічних категорій прикметників відносяться: рід, число, відмінок, розряд, ступінь порівняння.

До морфологічних категорій числівників відносяться: рід, число, відмінок, розряд, кількісне-порядкове, просте-складне.

До морфологічних категорій займенників відноситься: рід, число, відмінок, розряд, особа.

До морфологічних категорій дієслів, дієприкметників і дієприслівників відносяться: рід (для дієслів, дієприкметників), число (для дієприкметників), відмінок (для дієприкметників), форма (для дієприкметників), особа (для дієслів), вид (для дієслів, дієприкметників, дієприслівників), перехідність (для дієслів), нахил (для дієслів), час (для дієслів, дієприкметників), заставу (для дієприкметників).

До морфологічних категорій прислівників відносяться: ступінь порівняння, розряд.

При словозміні також враховується схилення, словозмінний клас дієслів, тип фінальної фонемі основи і ударність флексії.

2.2. Морфологічні явища в словозміні

Морфологічні перетворення в словозміні діляться на:

- лінійні;
- нелінійні.

До *лінійних перетворень* відноситься усічення або нарощення основи за рахунок фонемі або поєднання фонем на початку або кінці.

До *нелінійних перетворень* відноситься чергування фонем в межах морфеми (в тому числі чергування голосної с нулем, тобто швидкість голосної), поєднання морфів (основи і суфікса, основи і інтерфікса), імена наголоси.

Тема 3. Синтаксис

Синтаксис вивчає правила зв'язування слів і форм слів і синтаксичні одиниці, в складі яких ці правила реалізуються.

3.1. Синтаксичні одиниці та синтаксичні зв'язки

Основними синтаксичними одиницями є:

- словосполучення;
- прості речення (монопредикативні);
- складні (поліпредикативні) речення.

Словосполучення – це непередикативна синтаксична одиниця, компонентами якої є слово і форма слова або декілька форм слів, з'єднаних між собою синтаксичним зв'язком.

Просте речення – це передикативна синтаксична одиниця, що складається з декількох з'єднаних між собою синтаксичним зв'язком форм слів або з однієї форми слова.

Складне речення – це синтаксична одиниця, компонентами якої є передикативні одиниці, з'єднані між собою синтаксичним зв'язком.

Під *передикативністю* розуміється співвідношення змісту пропозиції з дійсністю – пропозиція завжди містить модальну (повідомлення, питання, спонукання, побажання) і часову (минуле, сучасне, майбутнє) характеристику повідомлюваного.

Синтаксичні зв'язки поділяються на:

- сурядні;
- підрядні.

Сурядний зв'язок встановлюється між рівноправними однорідними членами речення.

Підрядні зв'язки встановлюються як для речень, так і для словосполучень.

Синтаксичні підрядні зв'язки:

– узгодження (це вид підрядного зв'язку, при якому залежний компонент словосполучення ставиться в тих самих формах, що і головний) (наприклад, швидкий вентилятор);

– управління (це вид підрядного зв'язку, при якому залежний компонент словосполучення ставиться при головному в певному відмінку без прийменника або з прийменником, при цьому форма головного компоненту не впливає на форму залежного) (наприклад, кращий з вентиляторів);

– примикання (це вид підрядного зв'язку, при якому залежний компонент словосполучення не змінює своєї форми і зв'язується з головним тільки за змістом, при цьому в якості залежних компонентів виступають незмінні слова і такі форми змінюваних слів, які не виражають їх синтаксичної залежності) (наприклад, повернути праворуч, люблю працювати).

3.2. Формальна організація пропозиції

Традиційно виділяють наступні члени речення:

1. Головні:

- підмет;
- присудок.

2. Другорядні:

- довповнення;
- визначення;
- обставина.

3.3. Смилова організація речення

Ч.Філлмор встановив кілька відмінкових функцій, які, на його погляд, виступають достатніми для опису семантичних відносин, в яких бере участь дієслово, в англійській мові: Спочатку він виділяв тільки п'ять функцій

- агентів – одухотворений збудник дії;
- ергатив – об'єкт (в агнлійській є підметом в реченні);
- датив – одухотворений предмет, на користь чи на шкоду якому відбувається дія;
- комітатив – висловлює сумісність (наприклад, «с ...», «вместе с ...»);
- інструменталіс – неживий предмет, за допомогою якого виконується дія (наприклад, вдарити палкою).

Потім Філлмор виділив наступні функції:

- агентив (A) – одухотворений збудник дії;
- інструменталіс (I) – нежива сила або предмет, залучений в дію;
- датив (D) – особа, що зачеплена станом або дією;
- фактитив (F) – предмет або істота, що виникає в результаті дії;
- локатив (L) – місце дії або його просторова орієнтація;
- об'єктив (O) – предмет, що зачеплений дією (колишній ергатив).

У.Л.Чейф вніс зміну в певну Ч.Філлмором відмінкову систему. Так, датив у нього замінюється на:

- експрієнцер – той, хто відповідно налаштований, для кого щось є бажаним: (наприклад, Том хотел выпить);
- бенефіціант – той, хто має користь з того, про що повідомляється в іншій частині речення (наприклад, у Тома есть билеты, мать варит картошку для Тома).

В подальшому Філлмор запропонував виділяти дев'ять функцій:

- агент – одухотворений збудник дії;
- контрагент – сила, яка заважає дії (наприклад, сопротивляться кому-либо);
- об'єкт – предмет, що зачеплений дією;
- місце – фізичне тіло, яке відчуває безпосередній вплив з боку діяча (наприклад, задеть чей-то нос) і відрізняється від об'єкта тим, що допускає переифразу типу «задеть чей-то нос – задеть кого-то по носу»; в разі об'єкта така переифраза є неприпустимою (наприклад, «разбьет чей-то нос – разбьет кого-то по носу»);

– результат – предмет або істота, що виникає в результаті дії (колишній фактив);

– інструмент – нежива сила або предмет, залучений в дію;

– джерело – місце, з якого щось направлене (наприклад, он продает агрегат);

– адресат (мета) – особа, що зачеплена станом або дією (колишній датив);

– пацієнс – предмет, який відчуває ефект дії (наприклад, сломать агрегат).

Для опису семантичних відносин між дієсловом і його іменним оточенням Ю.Д.Апресян пропонує двадцять п'ять семантичних функцій (ролей), хоча вважає список неповним:

– суб'єкт (наприклад, поезд движется);

– контрагент (наприклад, защищаются от врага);

– голова (наприклад, ходатайствовать перед коллективом);

– об'єкт (наприклад, стрелять в мишень);

– зміст (наприклад, знать об отъезде);

– адресат (наприклад, сообщать собравшимся);

– одержувач (наприклад, давать людям);

– посередник (наприклад, передавать через секретаря);

– джерело (наприклад, брать в кассе);

– локатив (місце) (наприклад, находится на заводе);

– початкова точка (наприклад, ехать из города);

– кінцева точка (наприклад, ехать в село);

– маршрут (наприклад, ехать через реку);

– засіб (наприклад, прибить гвоздями);

– інструмент (наприклад, резать ножом);

– спосіб (наприклад, относиться с почетом);

– умова (компромисс в этих условиях);

– причина (радоваться подарку);

- мотивування (награждать за храбрость);
- результат (красить в красный цвет);
- мета (стремится к добру);
- аспект (превосходит по качеству);
- кількість (увеличивать в три раза);
- термін (отпуск на два месяца);
- час (начать в полночь).

Класифікація Ю.Д. Апресяна видається дещо дробовою за рахунок того, що деякі валентності, близькі за змістом, наприклад, як місце, початкова точка, кінцева точка, маршрут; адресат і одержувач; причина і мотивування, представлені роздільно, хоча вони можуть бути заведені в одну: валентність локатива; адресата, каузатора. Валентність локатива об'єднує початкову точку, кінцеву точку і маршрут; валентність адресата включає адресата і одержувача; каузатор поєднує в собі значення мотивування і причини.

Як видно з наведеного огляду літератури, не існує єдиної думки про семантичні відношення.

У сучасних дослідженнях для опису семантично-синтаксичної структури речення використовується поняття синтаксеми.

Синтаксема – мінімальна семантико-семантична одиниця речення.

Пропозиція представляє собою набір синтаксем.

У простому реченні можуть перебувати:

- базові предикатні синтаксеми (предикати);
- базові субстанціональні синтаксеми (актанти);
- вторинні предикатні синтаксеми (сирконстанти і атрібути);
- вторинні субстанціональні синтаксеми (актанти).

Семантико-синтаксичну структуру простого речення визначає базова предикатна синтаксема (предикат).

Базові предикатні синтаксеми володіють валентністю, яку можна розглядати як максимальну кількість аргументів для даного предиката. У ролі аргументів виступають актанти, пов'язані з даним предикатом. Мінімальна

валентність предиката – 1. Максимальна валентність предиката (для української і російської мови) – 7.

3.3.1. Базові субстанціональні синтаксеми

Українські лінгвісти, виділяють наступні базові субстанціональні синтаксеми (актанти):

- суб'єктна (діяч або носій процесу, стану, якості, кількісної ознаки);
- об'єктна (предмет, на який спрямована дія, процес або стан);
- адресатна (живий предмет, на користь чи на шкоду якому відбувається дія);
- інструментальна (знаряддя або спосіб, за допомогою якого виконується дія);
- локативна (місце події; вихідний, транзитивний ікінцевий пункт).

3.3.2. Вторинні субстанціональні синтаксеми

Вторинні (неелементарні, складні) субстанціональні синтаксеми (актанти) об'єднують декілька елементарних.

3.3.3. Базові предикатні синтаксеми

Базові предикатні підсистеми мають наступну структуру:

Одновалентні. Мають суб'єктну синтаксему (наприклад, дерева ростуть), яка може бути не виражена (наприклад, холодає).

Двовалентні. Мають суб'єктну і об'єктну синтаксеми (наприклад, професор веде лекції).

Тривалентні. Мають суб'єктну, об'єктну і адресатну синтаксеми (наприклад, Петро розповів історію людям).

Чотирьохвалентні. Мають суб'єктну, інструментальну, об'єктну (з прийменником) і об'єктну (без прийменника) синтаксеми (наприклад, Петр приклеил клеєм на агрегаті табличку).

П'ятивалентні. Мають суб'єктну, об'єктну, адресатну, локативну (вихідний пункт) і локативну (кінцевий пункт) синтаксеми (наприклад, Петр принес посылку людям из города в село).

Шестивалентні. Мають суб'єктну, об'єктну, адресатну, інструментальну, локативну (вихідний пункт) і локативну (кінцевий пункт) синтаксеми (наприклад, Петр отправил посылку людям поездом из города в село).

Семивалентні. Мають суб'єктну, об'єктну, адресатну, інструментальну, локативну (транзитний пункт), локативну (вихідний пункт) і локативну (кінцевий пункт) синтаксему (наприклад, Петр перевез багаж людям машиной через річку из города в село).

Виділяють наступні предикати:

- дії;
- процесу;
- стану;
- кількості;
- якості;
- локативні.

1. Предикати дії

Предикати діх позначають діяльність, яка породжується діячем і їм активно і безпосередньо стимулюється. Виражені дієсловом.

Предикати дії висловлюють наступні лексичні значення:

- зі значенням створення, видів діяльності (наприклад, строить, печь, ткать, шить, точить, сеять);
- із зазначенням на більшу чи меншу ступінь інтенсивності діяльності, спрямованої на об'єкт (наприклад, валить, бить, вязать, тереть, сечь);
- зі значенням переміщення в просторі (наприклад, везти, вести, нести);
- зі значенням руху (наприклад, идти, ехать, прибывать, бежать).

Максимальна валентність предиката дії – 6. Можливі наступні субстанціональні синтаксеми:

- суб'єктні;
- об'єктні;
- адресатні;
- інструментальні;
- локативні.

2. Предикати процесу

Предикати процесу характеризують динамічну ситуацію, яка не передбачає активного виконавця, а стосується суб'єктів процесу, які їм не породжуються і активно і безпосередньо не стимулюються (або характеризують динамічну ситуацію, яка пов'язана зі зміною стану або зміною ознак предмета). Виражені дієсловом.

Максимальна валентність предиката процесу – 1. Мають тільки об'єктну синтаксему (наприклад, дерева ростуть).

3. Предикати стану

Предикати стану безпосередньо пов'язані з тимчасовим планом (або вказують на непостійну ознаку предмета). Виражені прислівником і дієсловом.

Вони вказують на стан, пов'язаний на певним відрізком часу, тобто такі стани вказують на тимчасову характеристику об'єкта.

Їх специфіку визначає їх сполучуваність з суб'єктною синтаксеми, що виражає значення пасивності.

Предикати стану висловлюють наступні лексичні значення:

- фізичний стан живого предмета (наприклад, больно, жарко, тепло, холодно);
- емоційний або інтелектуальний стан, психічні переживання людини (наприклад, жалко, погано, весело, тривожно, любить, ненавидить);
- стан навколишнього середовища (наприклад, тепло, холодає, замерзла).

Максимальна валентність предиката стану – 2. Одновалентність – має суб'єктну синтаксему (наприклад, мене хорошо), яка може бути не виражена (наприклад, холодаєт), двовалентні – мають суб'єктну і об'єктну синтаксеми (наприклад, студенти люблять лектора).

4 Предикати кількості

Предикати кількості передають кількісні характеристики об'єкта. Найбільш периферійні. Виражені числівниками (наприклад, було два, було багато).

Максимальна валентність предиката кількості – 1.

5 Предикати якості

Предикати якості позначають постійною, внутрішньою, невід'ємною ознакою предмета. Виражені прикметником.

Ознаки предметів, що використовуються найчастіше:

- колір і його відтінки;
- масть тваринного;
- звукові властивості;
- смакові властивості;
- нюхові властивості;
- дотикові властивості;
- температурні властивості;
- розмір, об'єм, відстань;
- вікові ознаки;
- риси характеру, інтелектуальні можливості;
- фізичні якості живих предметів;
- зовнішня форма;
- загальна оцінка та ін.

Максимальна валентність предиката якості – 2. Одновалентні – мають суб'єктну синтаксему (густой лес), двовалентні – суб'єктну і об'єктну синтаксеми (наприклад, сын похож на отца).

6. Локативні предикати

Локативні предикати вказують на просторові стани, процеси або ознаки предмета. Виражені дієсловом.

Виділяються наступні локативні предикати:

– властиво-локативні предикати (наприклад, *бути, знаходитися*).

Максимальна валентність цього предиката – 2. Двовалентні – суб'єктна і локативна синтаксеми (наприклад, *Петр находится в городе*);

– процесуально-локативні (наприклад, *сидеть, лежать, падать, капать*).

Максимальна валентність цього предиката – 3. Двовалентні – суб'єктна і локативна синтаксеми (наприклад, *Петр сидит на стуле*), тривалентні – суб'єктна, локативна (начальный пункт движения), локативна (конечный пункт движения) синтаксеми (наприклад, *книга падает со стола на пол*);

– акціонально-локативні (зі значенням діяльності та локативності) (наприклад, *везти*). Максимальна валентність цього предиката – 6. Трьох, чотирьох, п'яти, шестивалентні.

3.3.4. Вторинні предикатні синтаксеми

Виділяються наступні вторинні предикатні синтаксеми:

– адverbальні (сирконстанти);

– модальні (сирконстанти);

– атрибутивні (атрибути).

1. Адverbальні синтаксеми

Адverbальні синтаксеми залежать не від окремого слова, від граматичної основи (предиката або головного члена в односкладових реченнях і субстанціональної синтаксеми)

– тимчасові (висловлюють тимчасові відносини) (наприклад, *после дождя, на протяжении дня*);

– причинні (відображують причинно-наслідкові зв'язки) (наприклад, *из-за недостатка времени, вследствие изучения*);

- цільові (наприклад, ради слави, для потехи);
- припустимі (вказують на характер і явища) (наприклад, незважаючи на);
- умовні (висловлюють реальні або нереальні умови з якими може відбутися або не відбутися дія) (наприклад, в разі смерті).

2. Модальні синтаксеми

Модальні синтаксеми залежать не від окремого слова, а від граматичної основи (предиката або головного члена в односкладових реченнях і субстанціональної синтаксеми). Висловлюють ставлення мовця до висловленої ним думки.

Виділяють наступні модальні синтаксеми:

- зі значенням достовірності повідомлення, впевненості (наприклад, безумовно, дійсно);
- зі значенням непевненості, можливості, ймовірності (наприклад, ймовірно, можливо);
- зі значенням джерела повідомлення (наприклад, на мій погляд);
- з логічним значенням, зі значенням виділення основного в повідомленні, підкреслення виводу (наприклад, крім того, наприклад, в першу чергу);
- зі значенням емоційної оцінки що повідомляється (наприклад, з жалем);
- з інтимно-контактуючим значенням (наприклад, побачте, прошу вас).

3. Атрибутивні синтаксеми

Атрибутивні синтаксеми розширюють субстантивні синтаксеми (наприклад, потужний вентилятор)

Далі розглядається лінгвістична модель технічної мови команд оператора, яка зазвичай включає в себе граматичну та семантичну моделі.

Література: [1-11]

ЗМІСТОВИЙ МОДУЛЬ 2

ДОСЛІДЖЕННЯ АЛГОРИТМІВ І МЕТОДІВ АНАЛІЗУ МОВНИХ КОНСТРУКЦІЙ

Тема 4. Морфологічний аналіз

Тема 5. Синтаксичний аналіз

Тема 6. Семантичний аналіз

Тема 7. Об'єднання морфологічного, синтаксичного і семантичного аналізів

Спроби формалізувати інтелектуальну діяльність людини призвели до постановки фундаментальної лінгвістичної задачі, що складається в моделюванні його мовної поведінки, тобто в побудові функціональної моделі природної мови.

Мовна діяльність людини зводиться до розуміння текстів (аналіз) і їх виробництва (синтез).

Лінгвістичний процесор [12], що входить до складу інтелектуальної системи, реалізує функціональну модель мови, яка встановлює відповідність між змістом і текстом, і містить дві основні функції:

- аналіз;
- синтез.

Методи аналізу мовних конструкцій зазвичай включають в себе:

- морфологічний аналіз;
- синтаксичний аналіз;
- семантичний аналіз.

Структура лінгвістичного процесора представлена на рис.2.1.

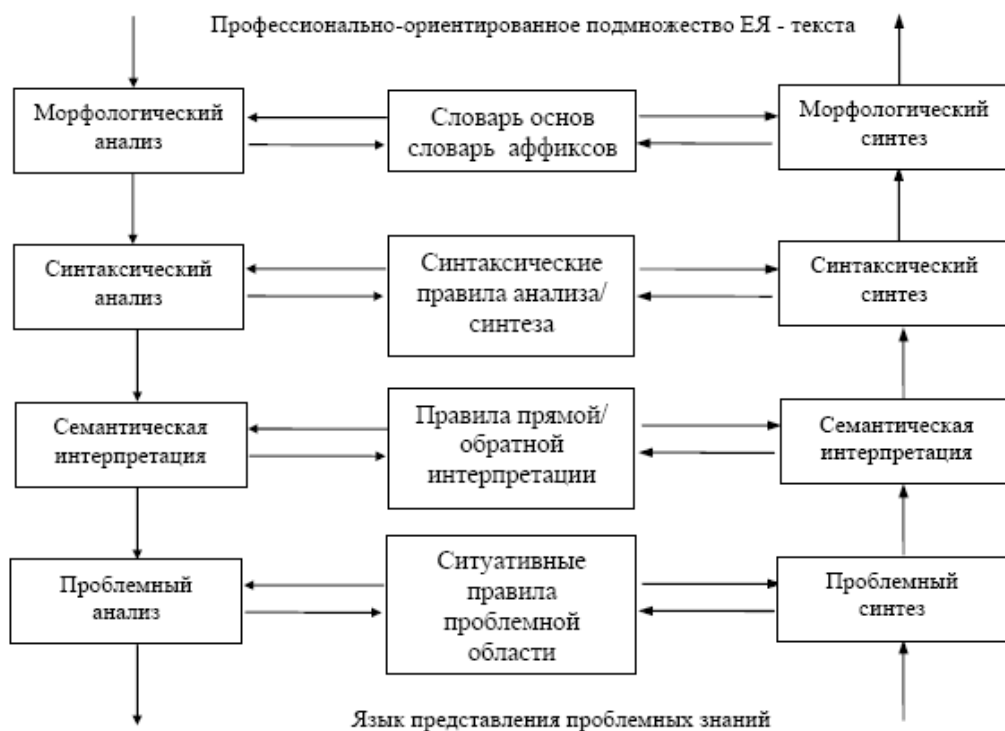


Рисунок 2.1 – Структура лінгвістичного процесора

Тема 4. Морфологічний аналіз

Морфологічний аналіз застосовується до форм слів речень. Результатом морфологічного аналізу є отримання вихідної форми слова (леми), його лексичного значення, частини мови та морфологічних ознак (наприклад, рід, число, відмінок), які подаються на вхід синтаксичного аналізатора.

Існують наступні методи морфологічного аналізу [12]:

- декоаративний;
- процедурний;
- комбінований (поєднання декларативного і процедурного);
- заснований на універсальних морфологічних моделях.

При декларативному методі в словнику зберігаються всі можливі форми слів з відповідною морфологічною інформацією. У цьому випадку задача морфологічного аналізу полягає в пошуку форми слова в словнику і переписування із словника морфологічної інформації. Перевагою методу є швидкість аналізу, тому цей метод є найбільш поширеним. Недоліком методу є

висока трудомісткість створення і підтримки словника, а також великий обсяг інформації, що зберігається, оскільки форм у кожного слова в російській і українській мові багато.

Процедурний метод виділяє основу форми слова; ідентифікує її та приписує формі слова морфологічну інформацію. Недоліком методу є висока трудомісткість складання словників «приставка-корінь», «корінь-суфікс-флексивний клас», «флексивний клас-флексія морфологічна інформація», оскільки в російській і українській мові велике число слів-винятків, які дозволяють автоматизувати цей процес.

Комбінований метод використовує як словник форм слів, так і словник основ. Якщо пошук за словником форм не успішний (тобто декларативний метод не дав результату), то задіють словник основ (тобто процедурний метод).

В рамках *методу, заснованого на універсальних морфологічних моделях*, використовуються:

- моделі у вигляді системи рівнянь алгебри скінченних предикатів;
- моделі морфології системи «Сенс-Текст».

Системи рівнянь алгебри скінченних предикатів [13] запропоновані Ю.П. Шабановим-Кушнарєнко і дозволяють проводити аналіз і синтез форм слів.

Морфологічна функція представляється у вигляді рівняння

$$Y = F(X),$$

морфологічне відношення представляється у вигляді рівняння

$$L(X, Y) = 1,$$

де Y – частина слова (основа і афікси), X – морфологічна інформація.

Наприклад, граматичне правило, яке описує те, що для іменників форма слова закінчується на «а», якщо родовий відмінок, чоловічий рід, однина, одухотворений, основа закінчується на «п», представлено у вигляді

$$y_1^a = X_1^p X_2^m X_3^e X_4^o X_5^n,$$

або у вигляді

$$y_1^a X_1^p X_2^m X_3^e X_4^o X_5^n = 1,$$

де y_1^a – перша буква закінчення «а», X_1^p – родовий відмінок, X_2^m – чоловічий рід, X_3^e – однина, X_4^o – одухотворений, X_5^n – остання буква основи «п».

Недоліком цього методу є складність практичної реалізації, оскільки в російській і українській мовах є велика кількість слів-винятків, до яких цей підхід непридатний.

Система «Сенс-Текст» [14] запропонована І.А. Мельчук і використовує синтез форм слів. У загальному вигляді правила морфологічного синтезу представлені як

$$(\lambda, \chi) \rightarrow v,$$

де λ – лексема, χ – морфологічна характеристика, v – форма слова.

У цьому підході враховувалися всі особливості морфології російської мови. Тому за допомогою відповідних правил можна синтезувати будь-яку форму слова.

Тема 5. Синтаксичний аналіз

Для синтаксичного аналізу використовуються:

- детерміністський підхід;
- стохастичний підхід.

5.1. Детерміністський підхід

Детерміністський підхід заснований на формальних граматиках.

Формальна граматики [15...19] представлена у вигляді

$$G = (V1, V2, P, S)$$

де $V1$ – кінцева множина нетермінальних символів (букв) або нетермінальний словник (алфавіт),

V_2 – кінцева множина термінальних символів (букв) або термінальний словник (алфавіт),

P – кінцева множина продукційних правил виду $\alpha \rightarrow \beta$, $\alpha, \beta \in V_1 \cup V_2$,

$S \in V_2$ – початковий нетермінальний символ (буква).

Довільна кінцева послідовність символів з словника $V = V_1 \cup V_2$ називається *ланцюжком*. Порожній ланцюжок позначається символом λ . Вхідження символів в ланцюжок називається її *точками*. Для будь-яких двох точок α_1, α_2 ланцюжки w таких, що $\alpha_1 \leq \alpha_2$, множина точок β , що задовольняють нерівності $\alpha_1 \leq \beta \leq \alpha_2$, називається *відрізком* ланцюжка w . Над ланцюжками визначена операція конкатенації.

Хомський визначив чотири типи формальних граматик [15]:

- автоматні;
- контекстно-вільні (безконтекстні);
- контекстно-залежні (безпосередньо складових);
- без обмежень.

Для лексичного аналізу природної мови і мови програмування використовуються автоматні граматики. Для синтаксичного аналізу мови програмування використовуються контекстно-вільні граматики. Для синтаксичного аналізу природної мови зазвичай використовуються граматики безпосередньо-складових і без обмежень або еквівалентних їм по потужності.

Далі розглядаються НС-граматика, трансформаційна граmatика, мережева граmatика. З детерміністським підходом також пов'язана нейросеть SRN, яка навчається граматиці.

5.1.1. Граmatика безпосередньо складових

Граmatика G називається *граматикою безпосередньо складових* (НС-граматикою) або *контекстно-залежною* [20], якщо кожне її правило має вигляд

$$\alpha_1 A \alpha_2 \rightarrow \alpha_1 \beta \alpha_2,$$

де α_1, α_2 – довільні ланцюжки в словнику $V_1 \cup V_2$, $A \in V_2$, β – довільна непорожній ланцюжок в словнику $V_1 \cup V_2$.

Така граматика дозволяє виконати заміщення не термінального символу A рядком β , якщо A з'являється в контексті рядків α_1, α_2 .

Мова $L(G)$, породжуваний НС-граматикою G , називається мовою безпосередніх складових.

Завдання формальної граматика полягає в породженні пропозиції мови і приписування йому структурного опису. У сучасній лінгвістиці використовують такі способи опису синтаксичної структури речення [21]:

- опис за допомогою дерева безпосередньо складових;
- опис за допомогою дерева синтаксичних залежностей (синтаксичного підпорядкування).

Нехай x – довільний непорожній ланцюжок в словнику (алфавіті) V . Множина C відрізків ланцюжка x називається системою складових цього ланцюжка, якщо воно задовольняє двом умовам:

- множина C містить відрізок, що складається з усіх точок ланцюжка x , і все одно точкові відрізки x ;
- будь-які два відрізки з C або не перетинаються, або один з них міститься в іншому.

Елементи C називають складовими. Одні точкові відрізки називають точковими складовими. Відрізок, що складається з усіх точок ланцюжка, називають повною складовою. Повну і точкові складові називають тривіальними.

На рис.5.1 наведено приклад дерева безпосередньо складових.

При описі речень природної мови за допомогою системи складових часто використовують спосіб введення в цю систему додаткової інформації шляхом відображення у безліч всіх підмножин деякої кінцевої множини, елементи якого називаються мітками і інтерпретуються як символи синтаксичних класів слів і словосполучень.

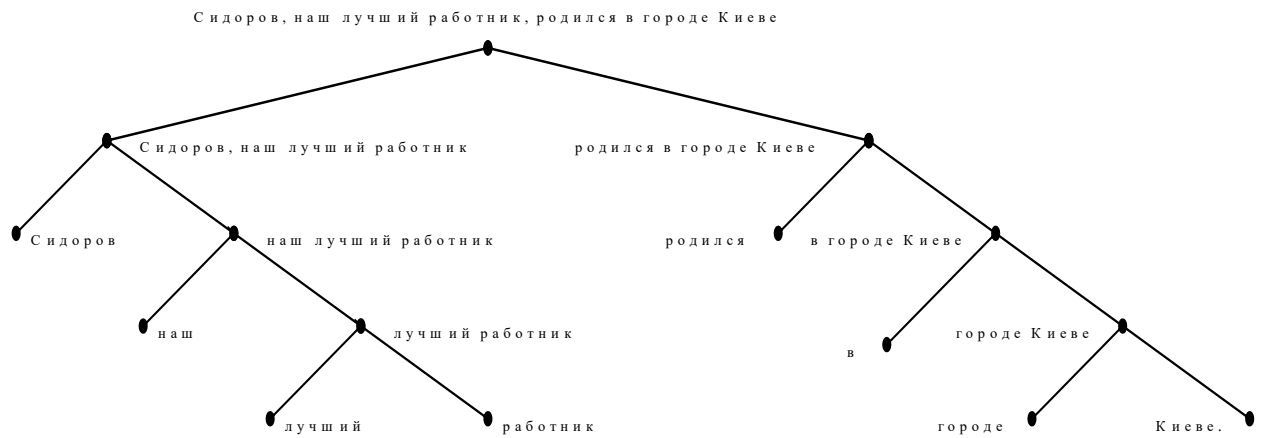


Рисунок 5.1 – Приклад розміченого дерева складових

Упорядковану трійку

$$\langle C, W, \varphi \rangle,$$

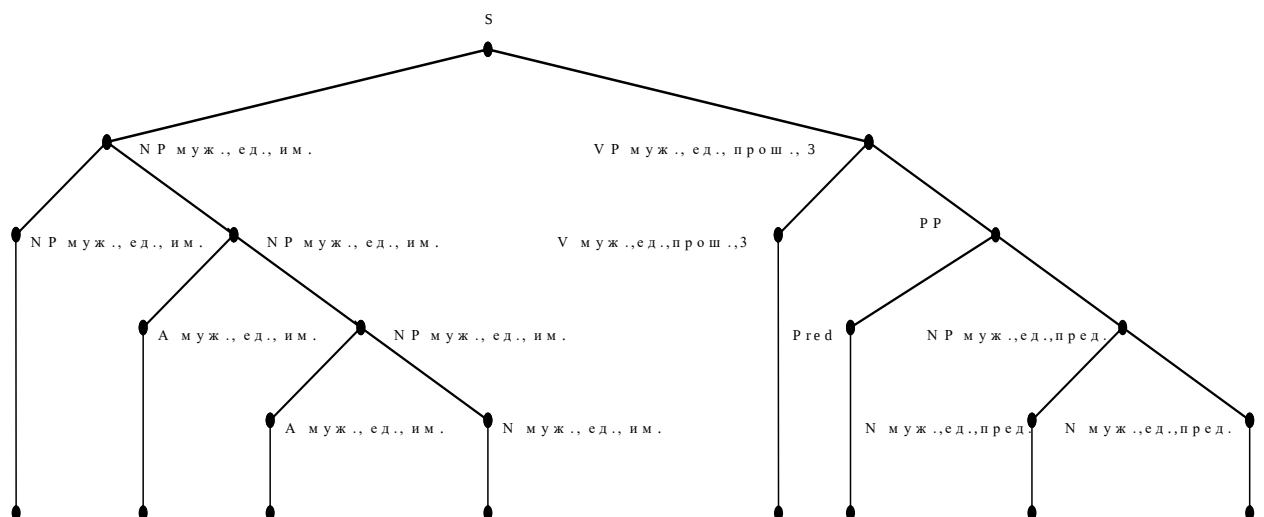
де C – система складових,

W – множина міток,

φ – відображення C в 2^W ,

називають *розміченою системою складових*.

На рис.5.2 наведено приклад розміченого дерева безпосередньо складових.



(Сидоров, (наш (лучший работник))), (родился (в (городе Киеве)))

Рисунок 5.2 – Приклад розміченого дерева складових

Множина міток W наведено в табл.5.1

Визначимо дерево підпорядкування.

$$G = (V1, V2, P, S)$$

Таблиця 5.1 Множина міток W

Елементи множини W	Коментарі
S	Речення
VP_{xyvw}	Група дієслова в роді x , числі y , часу v , особі w
V_{xyvw}	Дієслово в роді x , числі y , часу v , особі w
NP_{xyz}	Група іменника в роді x , числі y , відмінку z
N_{xyz}	Іменник в роді x , числі y , відмінку z
A_{xyz}	Прикметник в роді x , числі y , відмінку z
PP	Прийменникова група
$Pred$	Прийменник

Нехай x – довільний непорожній ланцюжок в словнику (алфавіті) V , X – множина всіх точок в x . Довільне бінарне відношення $\rightarrow$ на X , при якому граф $\langle X; \rightarrow \rangle$ є деревом, називається відношенням синтаксичного підпорядкування. Саме дерево $\langle X; \rightarrow \rangle$ називається деревом синтаксичного підпорядкування.

При графічному зображенні дерева підпорядкування точки ланцюжка x розміщують на горизонтальній прямій, і для будь-якої пари точок α, β , для якої $\alpha \rightarrow \beta$, проводиться стрілка з α в β , таким чином, щоб всі стрілки були по одну сторону від прямої.

На рис.5.3 наведено приклад дерева синтаксичного підпорядкування.

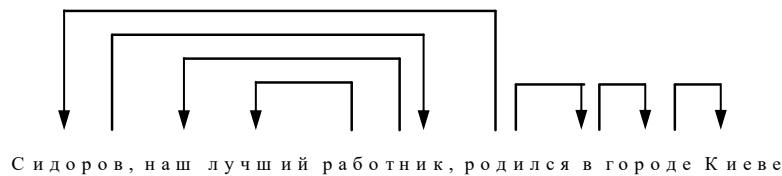


Рисунок 5.3 – Приклад дерева синтаксичного підпорядкування

Недоліком НС-граматик є те, що вони описують тільки синтаксичний рівень і не охоплює семантичний рівень.

5.1.2. Трансформаційна породжуюча граматика

Трансформаційні породжуючі граматика (ТПГ) [20] охоплюють синтаксичний і семантичний рівень.

Основні положення ТПГ викладені в роботах Хомського і полягають в наступному.

ТПГ складаються з трьох компонентів:

- синтаксичного;
- фонологічного;
- семантичного.

Фонологічний компонент інтерпретує поверхневі структури (пропозиція на природній мові). Семантичний компонент інтерпретує глибинні структури (контекст речення). Синтаксичний компонент породжує поверхневі і глибинні структури.

Центральна роль відводиться синтаксичному компоненту, який складається з двох субкомпонентів:

- базового;
- трансформаційного.

Базовий субкомпонент служить для породження глибинних структур. Трансформаційний субкомпонент забезпечує перехід від глибинної структури до поверхневої, використовуючи граматичні (нелексіческіе) трансформації.

На рис.5.4 приведена глибинна структура простого речення «Досвідчений конструктор створив надійний агрегат».

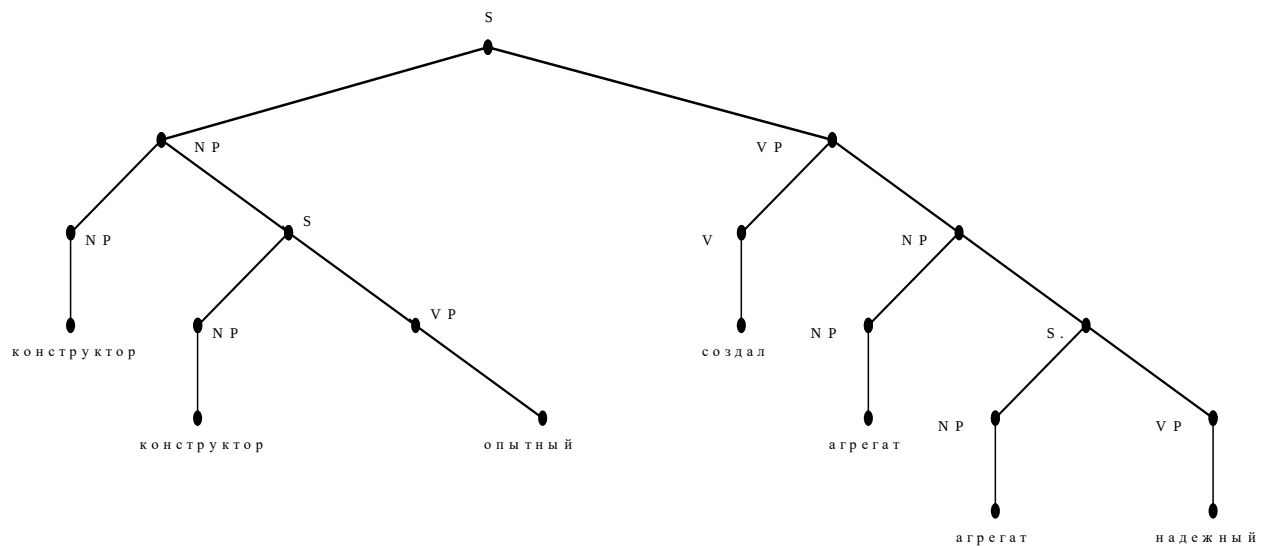


Рисунок 5.4 – Глибинна структура простого речення

Формально глибинні структури являють собою дерева (С-маркери), що породжуються:

- категоріальним правилами;
- лексиконом.

Категоріальні правила подаються у вигляді правил НС-граматики, що не містять термінальні символи, тобто породжувані ними ланцюжки складаються тільки з категоріальних символів (наприклад, N – іменник, V – дієслово).

Підстановка термінальних символів в породжуванні речення здійснюється за допомогою трьох видів правил:

- субкатегоризації;
- селекції;
- лексичного включення.

Субкатегоріальні правила і правила селекції комбінують категоріальні символи разом з субкатегоріальними в комплексні символи, здійснюючи заміни типу

$$N \rightarrow [+N, +\text{Нарицательность}, +\text{Исчисляемость}, +\text{Одушевленность}, -\text{Абстрактность}]$$

Правила лексичного включення підставляють в породжене дерево замість правої частини приведенного правила (комплексної одиниці) лексичну одиницю з відповідним набором ознак (наприклад, «оператор»).

В табл. 5.2 наведено приклад лексикона ТПГ.

Таблиця 5.2 Лексикон ТПГ

Комплексна одиниця	Лексична одиниця
[+N,+Нарицатель ность,+Исчисляемо сть, + Одушевлени ость,+ Человечнос ть]	оператор
[+V]	створити
[+N,+Нарицатель ность,+Исчисляемо сть, – Одушевлени ость,– Человечнос ть]	агрегат

Згідно табл.5.2, лексикон складається з записів, які вводять лексичні одиниці (лексеми) в ланцюжки, породжені категоріальним компонентом, утворюючи узагальнений С-маркер. Усі контекстуальні обмеження в базі забезпечуються цими лексичними трансформаціями.

Після породження глибинної структури, запускається трансформаційний субкомпонент. Кожна нелексична трансформація представляється у вигляді правила, умова застосовності якого задається у вигляді С-маркера.

Маючи узагальнений С-маркер, будується трансформаційний висновок шляхом застосування нелексичних трансформаційних правил від низу до верху, тобто застосовується послідовність правил до даної конфігурації тільки в тому випадку, якщо ця послідовність застосована до всіх базових С-маркерами цієї конфігурації.

Отримана поверхнева структура відображається за допомогою фонологічних правил фонологічного компонента в фонетичні уявлення.

З іншого боку, глибинні структури, отримані за допомогою базового субкомпоненту, відображаються за допомогою семантичного компонента в семантичні уявлення.

Семантичний компонент складається з:

– системи проєкційних правил;

– словника.

Правила проєкції йдуть знизу по дереву і поєднують значення лексичних одиниць, враховуючи обмеження на допустимість поєднання значень. Отриманий сумарний зміст являє собою семантичну інтерпретацію.

Згодом Хомський визнав, що семантичне уявлення залежить не тільки від глибинної, але і від поверхневої структури. Концепцію семантичного компонента, розвинену в результаті цього рішення, називають *інтерпретуючою семантикою*.

В рамках інтерпретуючої семантики була створена *відмінкова граматики Філлмора*.

Окрім інтерпретуючої семантики, була створена породжує семантика, в якій головна роль відводиться семантичному компоненту. Цей компонент перетворює семантичні уявлення в поверхневі структури, минаючи глибинні. У цьому полягає принципова різниця між інтерпретуючою і породжуючою семантикою.

ТПГ забезпечує потужність граматики без обмежень.

Недоліком ТПГ є те, що вона не призначена для аналізу пропозицій.

5.1.3. Мережева граматики

Мережева граматики [20], запропонована Вудсом, являє собою множину спрямованих підграфів (мереж переходів) з кінцевим числом станів і міченими дугами (рис.2.6).

Серед станів виділяються:

- початкове (наприклад, S_0);
- множина проміжних (наприклад, $S_1, S_2, S_3, S_4, S_8, S_9$);
- множина кінцевих станів (наприклад, $S_{5/1}, S_{6/1}, S_{7/1}, S_{10/1}$).

Мітки на дугах можуть бути:

- лексичними категоріями (наприклад, V – дієслово, MV – модальне дієслово, N – іменник, Par – частка, A – прикметник, Pred – прийменник);

– найменуваннями станів (наприклад, РР – прийменникова група іменників, VP – група дієслова, NP – група іменника).

Дуга, позначена найменуванням стану, інтерпретується наступним чином:

- найменування стану на кінці дуги запам'ятовується в стеку;
- управління передається до стану, яким позначена дуга.

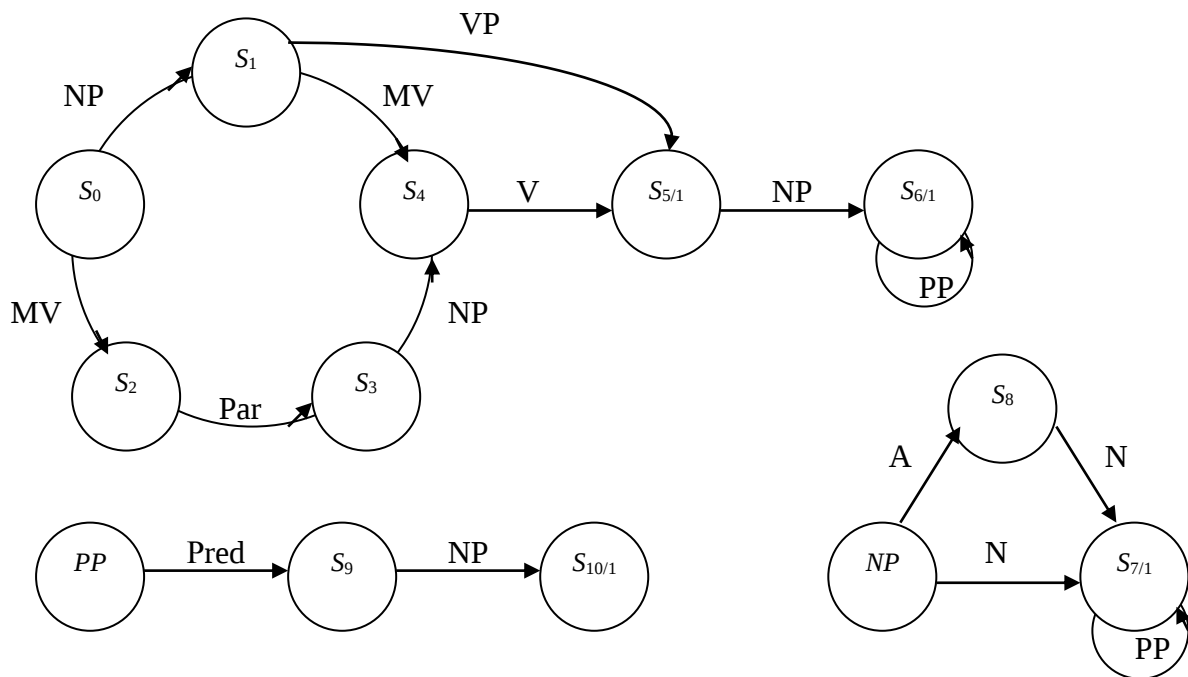


Рисунок 2.6 – Подання граматики у вигляді мережі переходів

Операції запам'ятовування в стеку і передачі управління називаються *зануренням*.

Якщо поточний стан є кінцевим, то відбувається виштовхування останнього запам'ятованого в стеці стану і передача управління в цей стан. Критерієм правильності вхідного рядка є спроба виштовхнути порожній стек, коли останній символ вхідного речення оброблений.

Щоб мережа могла виконувати трансформаційні перетворення (змінювати фрагменти синтаксичної структури), до кожної дуги додається:

- довільна умова (перехід в стан, що вказується дугою, дозволяється тільки при виконанні цих умов);

– довільну дію (дії виконуються, якщо здійснюється перехід в стан, на яке вказує дуга).

Отримана в результаті такого додавання мережа називається розширеною мережею переходів (РАМП).

РАМП будує структурний опис речення при переході з одного стану мережі в інше. Частини цього опису зберігаються в регістрах (мають стекову структуру) і їх вміст автоматично прошивується (занурюється), коли рекурсивне застосування викликає перехід до більш нижнього рівня. З дугами пов'язані дії, які змінюють вміст цих регістрів, і умови. Опис мережевий граматики задається за допомогою нормальної форми Бекуса (БНФ).

$$\langle \text{грамматика} \rangle ::= (\langle \text{словники} \rangle [\langle \text{нестандартні_процедури} \rangle] \langle \text{РАМП} \rangle)$$
$$\langle \text{словники} \rangle ::= (\text{СЛОВНИК} \langle \text{словник} \rangle^*)$$
$$\langle \text{словник} \rangle ::= \langle i' \text{ мя_словника} \rangle (\langle \text{функція_вводу} \rangle) |$$
$$\langle i' \text{ мя_словника} \rangle (\langle \text{словникова_стаття} \rangle^*)$$
$$\langle \text{РАМП} \rangle ::= (\langle \text{множена_дуг} \rangle \langle \text{множена_дуг} \rangle^*)$$
$$\langle \text{множена_дуг} \rangle ::= (\langle \text{стан} \rangle \langle \text{дуга} \rangle^*)$$
$$\langle \text{дуга} \rangle ::= (\text{КАТЕГОРІЯ} \langle i' \text{ мя_категорії} \rangle \langle \text{умова} \rangle \langle \text{дія} \rangle^* \langle \text{перехід} \rangle) |$$
$$(\text{ЗАНУРЕННЯ} \langle i' \text{ мя_категорії} \rangle \langle \text{умова} \rangle \langle \text{дія} \rangle^* \langle \text{перехід} \rangle) |$$
$$(\text{УМОВА} \langle \text{мітка} \rangle \langle \text{умова} \rangle \langle \text{дія} \rangle^* \langle \text{перехід} \rangle)$$
$$(\text{ВИШТОВХУВАННЯ} \langle \text{форма} \rangle \langle \text{умова} \rangle)$$
$$\langle \text{дія} \rangle ::= (\text{ПРТЕК} \langle \text{регістр} \rangle \langle \text{форма} \rangle) |$$
$$(\text{ПРПОГР} \langle \text{регістр} \rangle \langle \text{форма} \rangle) |$$
$$(\text{ПРВИТ} \langle \text{регістр} \rangle \langle \text{форма} \rangle)$$
$$\langle \text{перехід} \rangle ::= (\text{ПУ} \langle \text{стан} \rangle) |$$
$$(\text{ПП} \langle \text{стан} \rangle)$$
$$\langle \text{умова} \rangle ::= (\text{І} \langle \text{аргумент} \rangle \langle \text{аргумент} \rangle \langle \text{аргумент} \rangle^*) |$$
$$(\text{ЧИ} \langle \text{аргумент} \rangle \langle \text{аргумент} \rangle \langle \text{аргумент} \rangle^*) |$$

$(НЕ \langle аргумент \rangle)|$
 $(РАВНО \langle аргумент \rangle \langle аргумент \rangle)$
 $\langle форма \rangle ::= (ЗНАЧЕННЯ \langle реєстр \rangle)|$
 $\Delta |$
 $(ВЛАСТИВІСТЬ \langle властивість \rangle)$
 $(ДЕРЕВО \langle фрагмент_деревя \rangle \langle реєстр \rangle^*)|$
 $(СПИСОК \langle форма \rangle^*)|$
 $(ОБ'ЄДНАТИ \langle форма \rangle \langle форма \rangle)|$
 $(ГМЯ \langle структура \rangle)$

Опис словникової статті опускається, оскільки її змістом є значення різних синтаксичних і семантичних ознак слів природної мови.

Опис набору нестандартних функцій також опускається, так як їх конкретизація залежить від реалізації.

Типи дуг:

1. КАТЕГОРІЯ – задає лексичну категорію, якою позначена дуга.
2. ЗАНУРЕННЯ – вказує, що даний тип дуги викликає операцію занурення. Використовується для обробки дуг, позначених найменуванням стану.
3. УМОВА – задає умову переходу в стан.
4. ВИШТОВХУВАННЯ – вказує на необхідність виконання операції виштовхування стека. Пов'язує з кінцевим станом умови (їх виконання необхідно для підйому на більш високий рівень) і дії. Ця дуга є непомічені (порожньою).

Дія присвоює реєстру значення форми:

1. ПРТЕК – присвоювання на поточному рівні.
2. ПРПОГР – присвоювання на рівні занурення (більш глибокий, ніж поточний рівень).
2. ПРПОГР – присвоювання на рівні виштовхування.

Типи переходів:

1. ПУ (передача управління) – з вибором в якості поточного оброблюваного символу чергового символу вхідного рядка (речення).

2. ПП (перехід до підпрограми) – без зміни поточного оброблюваного символу.

Типи форм:

1. ЗНАЧЕННЯ – значенням цієї функції є вміст регістра.

2. Δ – значенням цієї функції є поточне слово вхідного речення.

3. ВЛАСТИВИСТЬ – значенням цієї функції є значення властивості.

4. ДЕРЕВО – значенням цієї функції є побудоване дерево або його фрагмент. Фрагмент, який виступає другим аргументом, задається за допомогою дужок, символів і знака «+», що позначає параметр.

5. СПИСОК – значенням цієї функції є список (форма) з значень аргументів.

6. ОБ'ЄДНАТИ – значенням цієї функції є об'єднання двох списків (форм) в один.

7. ІМ'Я – значенням цієї функції є аргументи форми.

Опишемо дуги, пов'язані з станами S0-S6 (рис.3.5)

((S0 (ЗАНУРЕННЯ *NP*

ІСТИНА

(ПРТЕК *S Δ*)

(ПРТЕК *ТИП* (І'МЯ ПОВЕСТВ))

(ПУ *S1*))

(КАТЕГОРІЯ *MV*

ІСТИНА

(ПРТЕК *MV Δ*)

(ПРТЕК *ТИП* (І'МЯ ВОПР))

(ПУ *S2*)))

(S1 (ЗАНУРЕННЯ *VP*

ІСТИНА

(ПРТЕК $MV \emptyset$)
 (ПРТЕК $VP \Delta$)
 (ПУ $S5$))
 (КАТЕГОРІЯ MV
 ІСТИНА
 (ПРТЕК $MV \Delta$)
 (ПУ $S4$)))
 (S2 (КАТЕГОРІЯ Par
 ІСТИНА
 (ПУ $S3$)))
 (S3 (ЗАНУРЕННЯ NP
 ІСТИНА
 (ПРТЕК $S \Delta$)
 (ПУ $S4$)))
 (S4 (КАТЕГОРІЯ V
 ІСТИНА
 (ПРТЕК $V \Delta$)
 (ПУ $S5$)))
 (S5 (ВИШТОВХУВАННЯ (ДЕРЕВО ($S+++ (VP+)$)) *ТИП* $S MV \Delta$)
 ІСТИНА)
 (ЗАНУРЕННЯ NP
 ІСТИНА
 (ПРТЕК VP (ДЕРЕВО ($VP(V+) \Delta$) V))
 (ПУ $S6$)))
 (S6 (ВИЩТОВХУВАННЯ (ДЕРЕВО ($S++++$)) *ТИП* $S MV VP$)
 ІСТИНА)
 (ЗАНУРЕННЯ PP
 ІСТИНА
 (ПРТЕК VP (ОБ'ЄДНАННЯ (ЗНАЧЕННЯ VP
 (СПИСОК Δ)))) (ПУ $S6$)))

Символи ПОВЕСТВ і ВОПР позначають типи пропозицій (розповідний і питальне).

Наприклад, для речення «Чи буде оператор управляти агрегатом?» Будуть виконані наступні дії.

S0: Δ =будет, MV :=будет;

$ТИП$:=ВОПР;

перехід до S2;

S2: перехід до S3;

S3: Δ =оператор, S :=(NP оператор);

перехід до S4;

S4: Δ =управлять, V :=управлять;

перехід до S5;

S5: Δ =агрегатом, VP :=($VP(V$ управлять)(NP агрегатом));

перехід до S6;

S6: (S ВОПР(NP оператор) буде ($VP(V$ управлять)(NP агрегатом))).

Для S0 ЗАНУРЕННЯ не виконується

Для S5 ВИШТОВХУВАННЯ не виконується.

Для S6 ЗАНУРЕННЯ не виконується.

Для простоти замість умови підставляється ІСТИНА.

Вирази типу

(ДЕРЕВО ($S+++ (V+)\Delta$) $ТИП S MV V$)

означає, що замість перших трьох + підставлятиметься значення $ТИП$, S , MV , а замість четвертого - значення V .

Глибинна структура речення представлена на рис.5.6.

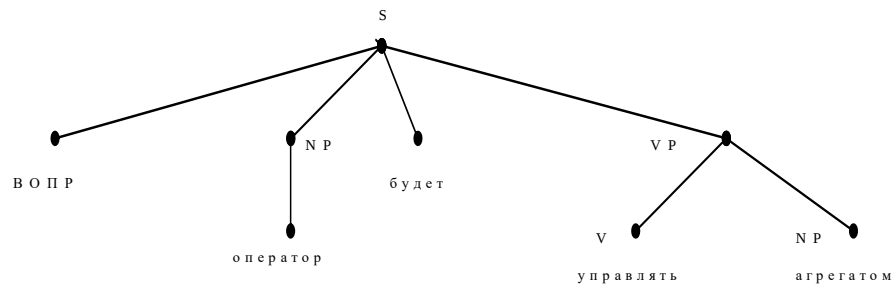


Рисунок 5.6 – Глибинна структура речення

У ряді випадків виділяють два види реєстрів:

- змісту (містять структури);
- флагової (містять тільки умови).

До переваг мережевої граматики відноситься те, що вона

- володіє наочністю (на відміну від ТПГ);
- забезпечує потужність граматики без обмежень (на відміну від НС-граматики);
- здатна об'єднувати загальні частини правил (на відміну від НС-граматики), що дає компактне уявлення, ефективність аналізу і усуває зайву обробку при розборі речення;
- може використовуватися для синтаксичного аналізу (на відміну від НС-граматики і ТПГ, які використовуються тільки для генерації);
- здатна змінювати побудову фрагмента дерева, тобто забезпечує ефективність аналізу;
- адаптивна (множина операцій може поповнюватися в разі необхідності).

Недоліком цієї граматики є відсутність ефективних алгоритмів синтаксичного розбору (наприклад, утруднене застосування алгоритму Ерлі, використовуваного для КС-граматик).

5.1.4. Нейромережевий підхід

У цьому підході синтаксичний аналізатор являє собою нейромережу Елмана (ENN) або просту рекуррентну мережу (SRN), яка є рекуррентною

двошаровою мережею (рис.2.8). Нейрони, виділені чорним кольором, називаються контекстними. Кількість контекстних нейронів збігається з кількістю нейронів прихованого шару. Кількість нейронів прихованого шару не повинно бути менше кількості нейронів вхідного шару.

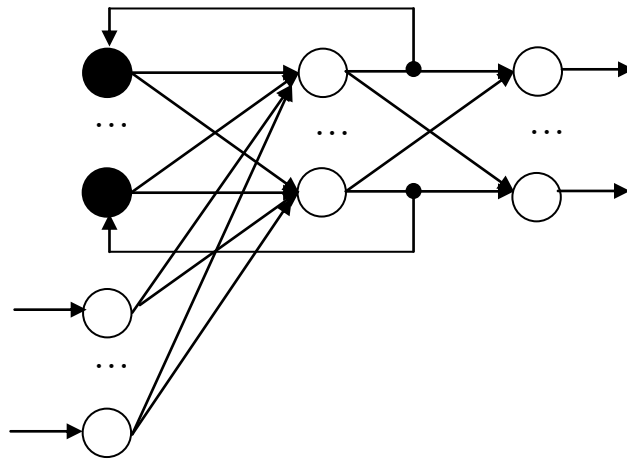


Рисунок 2.8 – Нейромережа Елмана (ENN)

Для синтаксичного аналізатора кількість нейронів вхідного шару збігається з кількістю нейронів вихідного шару, а кожне слово кодується бінарним вектором (містить тільки одну 1, що відповідає номеру слова). Для синтаксичного аналізатора за поточним словом, представленому бінарним вектором, прогнозується наступне слово, представлене бінарним вектором, на основі контекстно-вільної грамматики, приклад якої наведено нижче

$S \rightarrow NP VP \text{ '.'}$

$NP \rightarrow PropN \mid N \mid N RC$

$VP \rightarrow V (NP)$

$RC \rightarrow \text{who } NP VP \mid \text{who } VP (NP)$

$N \rightarrow \text{boy} \mid \text{girl} \mid \text{cat} \mid \text{dog}$

$PropN \rightarrow \text{John} \mid \text{Mary}$

$V \rightarrow \text{chase} \mid \text{feed} \mid \text{see} \mid \text{hear} \mid \text{walk} \mid \text{live}$

Слід зазначити, що дієслова chase і feed вимагають об'єкта безпосередньо, see і hear опціонально дозволяють об'єкт безпосередньо, walk і live забороняють об'єкт безпосередньо.

Для даної граматики кількість нейронів вхідного, вихідного і прихованого шару дорівнює 14 (знак '.' і 13 слів).

Для навчання SRN найчастіше використовується алгоритм зворотного поширення.

Процедура навчання (алгоритм зворотного поширення)

1. Номер ітерації навчання $n = 1$, $y_j^{(1)}(0) = 0$, $j \in \overline{1, N^{(1)}}$, ініціалізація всіх зсувів (порогів) $b_j^{(k)}(n)$ і ваг мережі у всіх шарах $w_{ij}^{(k)}(n)$, $i \in \overline{1, N^{(k-1)}}$, $j \in \overline{1, N^{(k)}}$, $k \in \overline{1, 2}$, де $N^{(k)}$ – кількість нейронів в k -м шарі.

2. Задається навчальна множина $\{(\mathbf{x}_\mu, \mathbf{d}_\mu) \mid \mathbf{x}_\mu \in \{0,1\}^{N^{(0)}}, \mathbf{d}_\mu \in \{0,1\}^{N^{(2)}}\}$, $\mu \in \overline{1, P}$, де $\mathbf{x}_\mu$ – μ -й бінарний вектор слова, за яким робиться прогноз, $\mathbf{d}_\mu$ – μ -й бінарний вектор прогнозованого слова, $N^{(0)} = N^{(2)}$, $N^{(0)}$ – кількість нейронів вхідного шару (або кількість слів, за якими робиться прогноз), $N^{(2)}$ – кількість нейронів вихідного шару (або кількість прогнозованих слів), P – потужність навчальної множини. Номер поточної пари з навчальної множини $\mu = 1$.

3. Обчислення вихідного сигналу для кожного шару (прямий хід)

$$y_j^{(1)}(n) = f^{(1)}(s_j^{(1)}(n)),$$

$$s_j^{(1)}(n) = \sum_{i=0}^{N^{(0)}} w_{ij}^{(1)}(n) x_{\mu i} + \sum_{i=N^{(0)}+1}^{N^{(0)}+N^{(1)}} w_{ij}^{(1)}(n) y_{i-N^{(0)}}^{(1)}(n-1), \quad j \in \overline{1, N^{(1)}},$$

$$y_j^{(2)}(n) = f^{(2)}(s_j^{(2)}(n)), \quad s_j^{(2)}(n) = \sum_{i=0}^{N^{(1)}} w_{ij}^{(2)}(n) y_i^{(1)}(n), \quad j \in \overline{1, N^{(2)}},$$

де $N^{(k)}$ – число нейронів в k -м шарі, k – номер шару, $w_{ij}^{(k)}(n)$ – вага зв'язку від i -го нейрона до j -го нейрону на k -м шарі в момент часу n , $y_j^{(k)}(n)$ – вихід j -го нейрона на k -му шарі, $f^{(k)}$ – функція активація нейронів k -го шару.

Вважається що $w_{0j}^{(k)}(n) = b_j^{(k)}(n)$, $x_{\mu 0} = 1$.

4. Обчислення середньоквадратичної помилки нейромережі

$$E(n) = \frac{1}{2} \sum_{j=1}^{N^{(2)}} e_j^2(n), \quad e_j(n) = y_j^{(2)}(n) - d_{\mu j},$$

5. Налаштування синаптичних ваг (зворотний хід)

Для налаштування вагових коефіцієнтів використовується рекурсивний алгоритм, який спочатку застосовується до вихідних нейронів мережі, а потім проходить мережу у зворотному напрямку до першого шару. Синаптичні ваги налаштовуються відповідно до формули

$$w_{ij}^{(k)}(n+1) = w_{ij}^{(k)}(n) - \eta \frac{\partial E(n)}{\partial w_{ij}^{(k)}(n)},$$

де η – параметр, що визначає швидкість навчання (при великому η навчання відбувається швидше, але збільшується небезпека одержати невірне рішення), $0 < \eta < 1$.

$$\frac{\partial E(n)}{\partial w_{ij}^{(2)}(n)} = y_i^{(1)}(n) g_j^{(2)}(n), \quad i \in \overline{0, N^{(1)}}, \quad j \in \overline{1, N^{(2)}},$$

$$\frac{\partial E(n)}{\partial w_{ij}^{(1)}(n)} = \begin{cases} y_{i-N^{(0)}}^{(1)}(n-1) g_j^{(1)}(n), & i > N^{(0)} \\ x_{\mu i} g_j^{(1)}(n), & i \leq N^{(0)} \end{cases},$$

$$i \in \overline{0, N^{(0)} + N^{(1)}}, \quad j \in \overline{1, N^{(1)}},$$

$$g_j^{(k)}(n) = \begin{cases} f'^{(2)}(s_j^{(2)}(n))(y_j^{(2)}(n) - d_{\mu j}), & k = 2 \\ f'^{(1)}(s_j^{(1)}(n)) \sum_{l=1}^{N^{(2)}} w_{jl}^{(2)}(n) g_l^{(2)}(n), & k = 1 \end{cases}$$

6. Перевірка умови завершення

Якщо $n \bmod P > 0$, то $\mu = \mu + 1$, $n = n + 1$, перехід до 3.

Якщо $n \bmod P = 0$ і $\frac{1}{P} \sum_{s=1}^P E(n - P + s) > \varepsilon$, то $n = n + 1$, перехід до 2.

Якщо $n \bmod P = 0$ і $\frac{1}{P} \sum_{s=1}^P E(n - P + s) < \varepsilon$, то завершитися.

Процедура прогнозу

$$y_i^{(0)}(1) = x_i, i \in \overline{1, N^{(0)}}, y_j^{(1)}(0) = 0, j \in \overline{1, N^{(1)}},$$

$$y_j^{(1)}(n) = f^{(1)}(s_j^{(1)}(n)),$$

$$s_j^{(1)}(n) = b_j^{(1)} + \sum_{i=1}^{N^{(0)}} w_{ij}^{(1)} y_i^{(0)}(n) + \sum_{i=N^{(0)}+1}^{N^{(0)}+N^{(1)}} w_{ij}^{(1)} y_{i-N^{(0)}}^{(1)}(n-1),$$

$$j \in \overline{1, N^{(1)}},$$

$$y_j^{(2)}(n) = f^{(2)}(s_j^{(2)}(n)), s_j^{(2)}(n) = b_j^{(2)} + \sum_{i=1}^{N^{(1)}} w_{ij}^{(2)} y_i^{(1)}(n), j \in \overline{1, N^{(2)}}.$$

5.2 Стохастичний підхід

Стохастичний підхід базується на N-грам (зазвичай біграмах або тріграмах) [16] і передбачає уявлення речення у вигляді спільної ймовірності появи послідовності слів $w_1 \dots w_n$, тобто

$$P(w_1 \dots w_n) = \prod_{i=1}^n P(w_i | w_1, \dots, w_{i-1}).$$

Наприклад, спільна ймовірність послідовності слів «отправить сляб в печь» визначається за формулою

$P(\text{отправить сляб в печь}) = P(\text{отправить})P(\text{сляб}|\text{отправить}) P(\text{в}|\text{отправить сляб})P(\text{печь}|\text{отправить сляб в})$

Для біграм речення подається у вигляді

$$P(w_1...w_n) = \prod_{i=1}^n P(w_i | w_{i-1}).$$

Для тріграм речення подається у вигляді

$$P(w_1...w_n) = \prod_{i=1}^n P(w_i | w_{i-1}, w_{i-2}).$$

N-грам використовуються в разі об'єднання систем розпізнавання мови і аналізу мовних конструкцій, коли можливо невірне розпізнавання вимовлених слів.

При аналізі речення за допомогою метод біграм і тріграм результатом аналізу є ймовірність такої структури.

До недоліків цього методу відноситься ігнорування синтаксичної структури речення, а також слабка адаптація до нової предметної області. Цей підхід не формує деревоподібну синтаксичну структуру, і семантичний аналіз після нього не проводиться.

Тема 6. Семантичний аналіз

Семантичний аналіз здійснює перетворення сформованої деревовидної синтаксичної структури, сформованої за допомогою детерміністского підходу, в семантичне представлення.

Для частин пропозицій (синтаксем) відповідно до предиката (присудком) визначаються їх семантичні ролі.

Цей аналіз використовує:

- продукційні правила;
- семантичні мережі;

- фрейми;
- сценарії;
- числення предикатів;
- алгебру скінченних предикатів;
- нейромережі.

6.1. Продукційні правила

Підхід, заснований на продукційних правилах [1, 25, 26], дозволяє уявити знання про предметну область у вигляді пропозицій типу «Якщо (умова), то (дія)», тобто

$$X \rightarrow Y$$

Цей підхід забезпечує високу модульність, легкість внесення доповнень і змін, простоту реалізації. Крім того, продукційні правила зручно зберігати у реляційній базі даних, що дозволяє здійснити їх сортування і підвищити ефективність пошуку. Однак при великій кількості правил можуть виникнути складності.

6.2. Семантична мережа

Семантична мережа [1, 25, 26] має вигляд орієнтованого графа, вершини якого - поняття, події або властивості предметної області, а дуги - відносини між ними (наприклад, мережа може описувати команду або технологічний ланцюг). Дана модель подання знань була запропонована американським психологом М.Р.Куїлліаном. Її основна перевага - у тому, що вона більше за інших відповідає сучасним уявленням про організацію довготривалої пам'яті людини. Недоліком цієї моделі є відсутність модульності, складність реалізації, ускладнена процедура управління знаннями.

На рис.6.1 зображена семантична мережа. В якості вершин виступають поняття “Черный”, “Цвет” та ін. Дугами є “значение”, ”свойство” та ін.

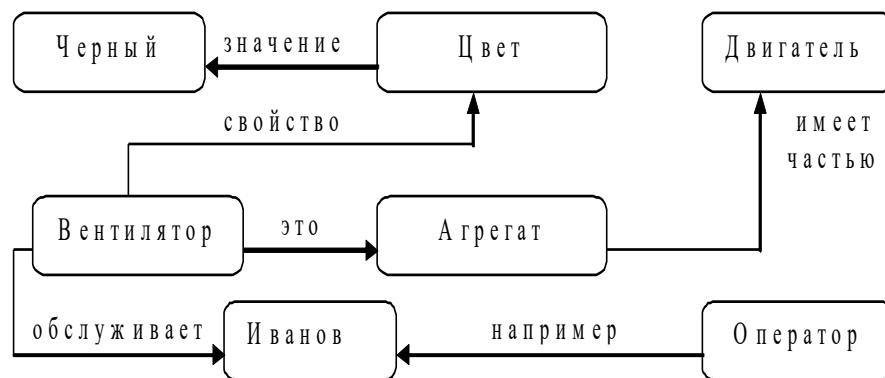


Рисунок 6.2 – Семантическая сеть

6.3. Фрейм

Фрейм [1, 19, 20] має вигляд локальної семантичної мережі і призначений для опису деякого поняття або події предметної області з усією сукупністю властивих їм властивостей (наприклад, фрейм може описувати нагрівальну піч з її режимами). Але на відміну від звичайних семантичних мереж він забезпечує перетворення інформації всередині себе і зв'язок з іншими фреймами. Якщо в мережі фреймів не використовуються процедури перетворення інформації, то така мережа буде відповідати семантичній мережі. Дана модель подання знань була запропонована американським вченим М.Минским. Цей підхід забезпечує високу модульність. До недоліків відноситься ускладнена процедура управління знаннями.

Розрізняють фрейми-зразки (прототипи), що зберігаються в базі знань, і фрейми-екземпляри, які створюються для відображення реальних ситуацій на основі даних, що надходять.

Структура фрейму може бути представлена як список властивостей об'єкта:

(ІМ'Я ФРЕЙМА,

(ім'я 1-го слота, значення 1-го слота, ім'я 1-й приєднаної процедури),

...

(ім'я N-го слота, значення N-го слота, ім'я N-й приєднаної процедури)),

Слот відповідає властивості об'єкта, яке може змінювати своє значення (наприклад, кількість вікон, колір стін, висота стелі у фреймі кімната). В якості значення слота може виступати ім'я іншої фрейма. Так утворюються мережі фреймів.

Існує кілька способів отримання слотом значень у фреймі-екземплярі:

- за замовчуванням від фрейму-зразка (Default-значення);
- через успадкування властивостей від фрейму, зазначеного у слоті АКО (A-Kind-Of = це);
- за формулою, зазначеною в слоті;
- через приєднану процедуру (виробляє обчислення всередині цього фрейму);
- явно з діалогу з користувачем;
- з бази даних.

Найважливішою властивістю теорії фреймів є запозичення з теорії семантичних мереж - так зване успадкування властивостей. І у фреймах, і в семантичних мережах спадкування відбувається за АКО-зв'язками. Слот АКО вказує на фрейм більш високого рівня ієрархії, звідки неявно успадковуються, тобто переносяться, значення аналогічних слотів.

На рис.6.2 представлена мережа фреймів.

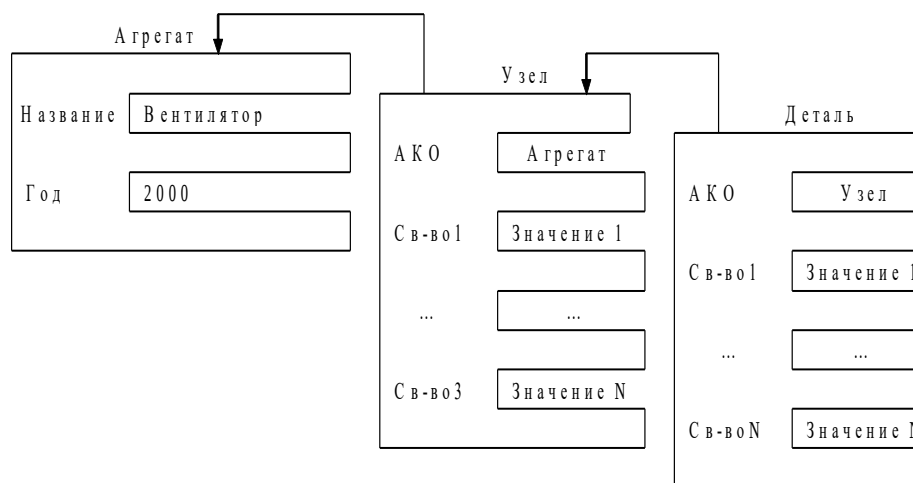


Рисунок 6.2 – Мережа фреймів

Наприклад, в мережі фреймів на рис.6.2 поняття «деталь» успадковує властивості фреймів «вузел» і «агрегат», які знаходяться на більш високому рівні ієрархії. Так, на питання «год изготовления узла 2000» слід відповісти «да», так як цією властивістю володіють всі вузли, що зазначено у розіграші «вузел». Спадкування властивостей може бути частковим, так як не всі властивості для деталей успадковуються з фрейма «вузел».

Фрейми зручно реалізовувати за допомогою мов об'єктно-орієнтованого програмування.

6.4. Сценарій

Сценарій [1] являє знання про типову ситуації предметної області у вигляді формалізованого опису стандартної послідовності взаємопов'язаних фактів. Він має фреймоподібну структуру і може бути представлений у вигляді мережі, вершини якої - факти, а дуги - причинно-наслідкові зв'язки (наприклад, відсутність слябів на складі призводить до зупинки печі, що в свою чергу призводить до зупинки кліті «трио»). Вперше поняття сценарію було введено Р.Шенком і Р.Абельсоном при розробці нових засобів розуміння історії. Сценарії в їх системі представлялися у вигляді фреймоподібних структур і використовувалися для зв'язування подій історії. Кожен сценарій складався з набору слотів і їх значень, що описують ролі, причини і послідовності сцен, які були послідовністю певних дій. Слот «роль» ставив виконавців сценарію, слот «цель» - мотивування дій, що робляться. Будь-яка попередня дія створює умови для здійснення подальшої дії. До недоліків цієї моделі відноситься складність практичної реалізації, ускладнена процедура управління знаннями.

Приклад

<сценарий: агрегат

роли: оператор

цель: управление агрегатом

сцена 1: включение агрегата

зайти в помещение

включить электричество

подойти к агрегату

осмотреть агрегат

включить агрегат

сцена 2: работа агрегата

установить небольшую нагрузку

проверить работоспособность агрегата

установить требуемую нагрузку

сцена 3: включение агрегата

выключить агрегат

выключить электричество

выход из помещения>

6.5. Обчислення предикатів

Підхід, заснований на обчисленні предикатів [1, 25, 26], дозволяє описати знання про предметну область у вигляді сукупності логічних формул. До недоліків цієї моделі відноситься відсутність модульності, складність практичної реалізації, ускладнена процедура управління знаннями.

Наприклад, команда «Каждая деталь должна сначала пройти фрезерование торцов, а затем токарную обработку» описується у вигляді $\forall x \in \text{Деталь} \exists y, z \in \text{Операция} (\text{кон}(y) \leq \text{нач}(z) \wedge \text{дет}(y) = x \wedge \text{дет}(z) = x \wedge \text{фрез_торц}(y) \wedge \text{ток_обр}(z))$.

6.6. Алгебра кінцевих предикатів

В рамках підходу, що базується на алгебрі скінченних предикатів [13], семантичну структуру речення можна представити у вигляді деякого рівняння.

До недоліків цієї моделі відноситься відсутність модульності, складність практичної реалізації, і ускладнена процедура управління знаннями.

Наприклад, всі базові субстанціональні синтаксеми можна описати як сукупність суб'єктів, об'єктів, адресатів, інструментів, локов у вигляді рівняння

$$y^c C(X) \vee y^o O(X) \vee y^a A(X) \vee y^i И(X) \vee y^l Л(X) = 1$$

6.7. Нейромережевий підхід

В якості нейромережі може використовуватися XRAAM [22...24], на вхід якої надходить фрагмент тексту у вигляді сценарію.

Мережа XRAAM – це мережа RAAM (аналогічна мережі Елмана і використовується для стиснення і шифрування інформації, кількість вхідних нейронів завжди збігається з кількістю вихідних) з глобальним словником, який зберігає пари «символ – распределенное семантическое представление». Розподілене семантичне уявлення відображає різні ролі слів (наприклад, слово «печь» може бути як ОБЪЕКТ, ИНСТРУМЕНТ або ЛОК). З кожним реченням зв'язується структура, що складається з ролей, відповідних синтаксемам речення (наприклад, ПРЕДИКАТ-СУБЪЕКТ-ОБЪЕКТ-ЛОК). Мається двосторонній зв'язок між словами-поняттями і реченнями - семантичний зміст речення залежить від кожного з його слів-понять, і кожне слово-поняття залежить від кожного з речень, в якому воно зустрічається. Щоб вивчити слова-поняття, використовується кодуюча поняття мережа, а щоб вивчити речення, використовується кодуюча речення мережа. Обидві ці мережі є RAAM і пов'язані з глобальним словником. Навчання обох мереж взаємопов'язане - на вхід мережі кодування понять подаються уявлення, створювані мережею кодування речень, і навпаки. В ході навчання розподіленим семантичним уявленням кодуються всі слова-поняття, потім кодуються всі речення, і процес повторюється до тих пір, поки не будуть знайдені стійкі зразки розподілених семантичних уявлень. Слова-поняття і речення подаються на вхід мережі у вигляді трійок, тобто використовується потрібна RAAM. Слово-поняття

структурується як «слово-понятіє, роль, метка_предложения», речення структурується як «метка_предложения, роль, слово-понятіє».

Наприклад, слова «печь» зустрічається в реченнях «Оператор подает сляб в печь» і «Печь расплавляет сляб». У цих реченнях роль слова «печь» відрізняється. У першому випадку вона - лок, у другому - інструмент.

Деревоподібна структура кодування слова «печь», яка використовується в двох реченнях, приведена на рис.6.3.

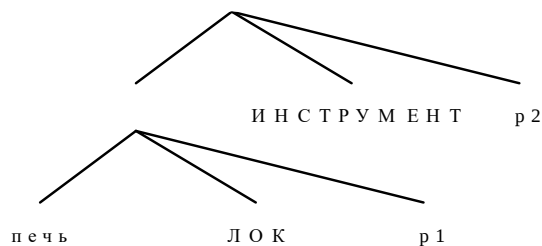


Рисунок 6.3 – Деревоподібна структура кодування слова «печь»

Деревоподібна структура кодування речення «Оператор подает сляб в печь» приведена на рис.6.4.

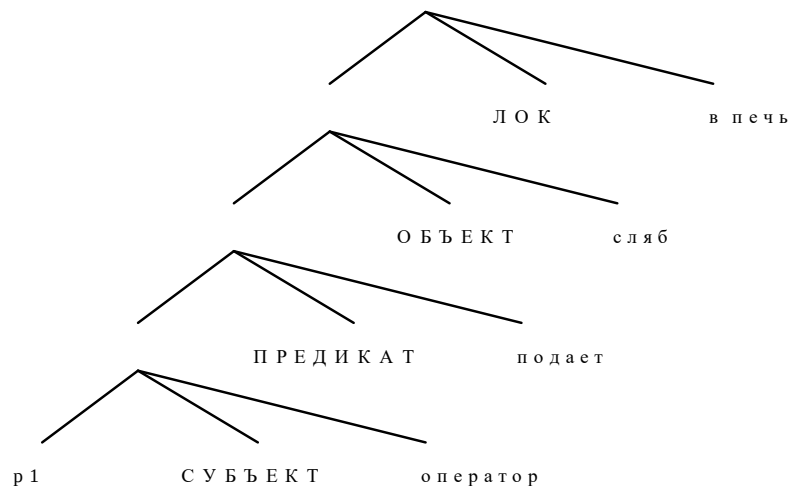


Рисунок 6.4 – Деревоподібна структура кодування речення «Оператор подает сляб в печь»

Деревоподібна структура кодування речення «Печь расплавляет сляб» приведена на рис.6.5.

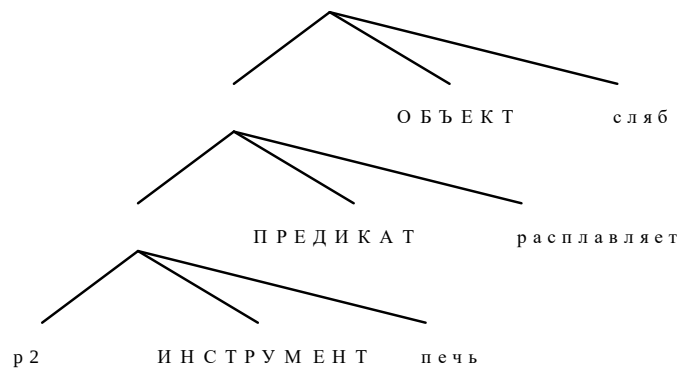


Рисунок 6.5 – Деревоподібна структура кодування речення «Печь расплавляет сляб»

6.5. Реляційні моделі

Реляційна модель дозволяє представити знання про предметну область у вигляді формул реляційної алгебри або реляційного числення.

В основі цього подання знань лежить уявлення інформаційних зв'язків між поняттями у вигляді відносин і таблиць, що складаються з рядків (кортежів) і стовпців (атрибутів).

Основні операції реляційної алгебри:

1. Об'єднання відносин R_1, R_2

$$R = R_1 \cup R_2$$

2. Різниця відносин R_1, R_2

$$R = R_1 - R_2$$

3. Декартовий добуток відносин R_1, R_2

$$R = R_1 \times R_2$$

4. Проекція відносини R_1 на компоненти (номера стовпців) $i_1, \dots, i_r$

$$R = \pi_{i_1, \dots, i_r}(R_1)$$

5. Селекція відносини R_1 за формулою F

$$R = \sigma_F(R_1), \text{ де формула } F \text{ утворена:}$$

– операндами, які є номерами стовпців;

– логічними операторами $\neg$ (не), $\wedge$ (и), $\vee$ (или);

– арифметичними операторами порівняння $<$, $=$, $>$, $\leq$, $\geq$, $\neq$

Інші операції реляційної алгебри (перетин відносин, частне відносин, з'єднання відносин, природне з'єднання) виходять за допомогою основних операцій.

Основні операції обчислення висловлювань:

1. Об'єднання відносин R_1, R_2

$$\{t \mid R_1(t) \vee R_2(t)\}$$

2. Різниця відносин R_1, R_2

$$\{t \mid R_1(t) \wedge \neg R_2(t)\}$$

3. Декартовий добуток відносин R_1, R_2

$$\{t^{(m+n)} \mid (\exists u)(\exists v)(R_1(u) \wedge R_2(v) \wedge t[1]=u[i_1] \wedge \dots \wedge t[m]=u[m] \wedge \\ \wedge t[m+1]=v[1] \wedge \dots \wedge t[m+n]=v[n])\}$$

4. Проекція відносини R_1 на компоненти (номера стовпців) $i_1, \dots, i_r$

$$\{t^{(r)} \mid (\exists u)(R_1(u) \wedge t[1]=u[i_1] \wedge \dots \wedge t[r]=u[i_r])\}$$

5. Селекція відносини R_1 за формулою F

$\{t \mid R_1(t) \wedge F'\}$, де вираз F' є вираз F , в якому кожен операнд, що позначає компонент I , замінений на $t[i]$

Приклад відносини R_1 наведено в табл.2.3.

Таблиця 6.1 Таблиця відносини R_1

Продавець	Покупець	Купівля	Властивість (вік) покупки	Час покупки
Том	Джон	автомобиль	новый	вчера
Билл	Гарри	мотоцикл	старый	сегодня
...	...	...	...	...

Запит до реляційної БЗ «какой продавец какому покупателю продал какую покупку» виконаний за допомогою:

– реляційної алгебри

$$R = \pi_{i_1, i_2, i_3} (R_1)$$

– реляційного числення

$$\{t^{(3)} \mid (\exists u)(R_1(u) \wedge t[1] = u[i_1] \wedge t[2] = u[i_2] \wedge t[3] = u[i_3])\}$$

Тама 7. Об'єднання морфологічного, синтаксичного і семантичного аналізів

В даний час стає популярним об'єднання різних методів аналізу мовних конструкцій в рамках однієї нейромережі. В роботі [27] було запропоновано підхід, заснований на семантичній нейромережі (рис.2.14), який об'єднує морфологічний, синтаксичний, семантичний аналіз, і синтез тексту.

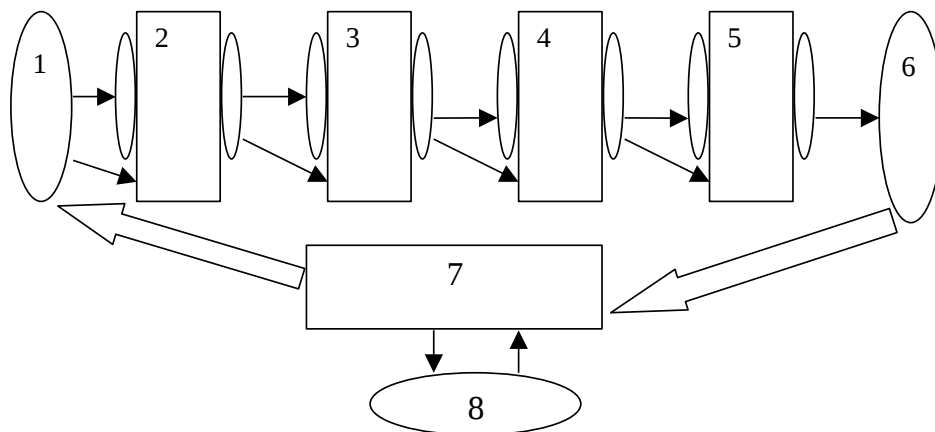


Рис. 7.1. Блок-схема системи, що обробляє природну мову

Вхідний текст із зовнішнього середовища 8 потрапляє в комутатор 7, і через нього на шар рецепторів 1. Шари 2 і 3 проводять морфологічний і синтаксичний аналіз відповідно. Результати семантичного аналізу з системи 4 потрапляють в шар синтезу символічних послідовностей 5 і далі на шар ефекторів 6. З шару ефекторів символічна послідовність потрапляє в комутатор 7. Система отримує можливість оцінити і скорегувати проміжні результати, сформувати остаточний результат і передати його в зовнішнє середовище.

До недоліків цього методу відноситься те, що він недостатньо враховує граматику мови.

Література: [1, 12-27]

СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

1. Искусственный интеллект: [в 3-х кн.] / под ред. Д.А. Поспелова. – М.: Радио и связь, 1990. – Кн.2: Модели и методы: Справочник. – М.: Радио и связь, 1990. – 304 с.
2. Поспелов Г.С. Искусственный интеллект – основа новой информационной технологии / Г.С. Поспелов. – М.: Наука, 1988. – 280 с.
3. Поспелов Д.А. Ситуационное управление: теория и практика / Д.А. Поспелов. – М.: Наука, 1986. – 288 с.
4. Русская грамматика: [в 2 т.] / под ред. Н.Ю. Шведовой. – М.: Наука, 1982. – Т. 1: Фонетика. Фонология. Ударение. Интонация. Введение в морфемику. Словообразование. Морфология. – 1982. – 784 с.
5. Современный русский язык / под ред. В.А. Белошапковой. – М.: Азбуковник, 1997. – 928 с.
6. Филлмор Ч. Дело о падеже / Ч. Филлмор // Новое в зарубежной лингвистике. – 1981. – Вып.10. – С. 369-495.
7. Апресян Ю.Д. Избранные труды: [в 2 т.] / Ю.Д. Апресян. – М.: «Языки русской культуры», 1995. – Т. 1: Лексическая семантика (синонимические средства языка). – М.: «Языки русской культуры», – 1995. – 230 с.
8. Золотова Г.А. Синтаксический словарь. Репертуар элементарных единиц русского синтаксиса / Г.А. Золотова. – М.: 1988. – 300 с.
9. Вихованець І.Р. Граматика української мови. Синтаксис: Підручник / І.Р. Вихованець.. – К., Либідь, 1993. – 368 с.
10. Загнітко А.П. Теоретична граматика української мови: Морфологія: Монографія / А.П. Загнітко. – Донецьк: ДонДУ, 1996. – 437 с.
11. Теньер Л. Основы структурного синтаксиса / Л. Теньер. – М.: Прогресс, 1988. – 654 с.
12. Евдокимова И.С. Естественно-языковые системы: курс лекций / И.С. Евдокимова. – Улан-Удэ: Изд-во ВСГТУ, 2006. – 92 с.

13. Шабанов-Кушнарченко Ю.П. Теория интеллекта. Технические средства / Ю.П. Шабанов-Кушнарченко. – Х.: Вища шк. Изд-во при Харьк. ун-те, 1986. – 136 с.
14. Мельчук И.А. Опыт лингвистических моделей «смысл-текст» / И.А. Мельчук. – М.: Наука, 1982. – 314 с.
15. Компаниец Р.И. Системное программирование. Основы построения трансляторов / Р.И. Компаниец, Е.В. Маньков, Н.Е. Филатов. – СПб.: Корона принт, 2000. – 256 с.
16. Бондарев В.Н. Искусственный интеллект / В.Н. Бондарев, Ф.Г. Аде. – Севастополь: Изд-во СевНТУ, 2002. – 615 с.
17. Luger G.F. Artificial intelligence: Structures and Strategies for complex problem solving / G.F. Luger, W.A. Stubblefield. – Redwood: The Benjamin Publishing Company Inc., 1993. – 738 p.
18. Ахо А. Теория синтаксического анализа, перевода и компиляции: [в 2 т.] / А. Ахо, Дж. Ульман. – М.: Мир, 1978. – Т.1: Синтаксический анализ. – М.: Мир, 1978. – 612 с.
19. Ахо А. Компиляторы: принципы, технологии и инструменты / А. Ахо, Р. Сети, Дж. Ульман. – М.: Изд. дом «Вильямс», 2001. – 768 с.
20. Попов Э.В. Алгоритмические основы интеллектуальных роботов и искусственного интеллекта / Э.В. Попов, Г.Р. Фирдман. – М.: Наука, 1976. – 455 с.
21. Гладкий А.В. Синтаксические структуры естественного языка в автоматизированных системах общения / А.В. Гладкий. – М.: Наука, 1985. – 144 с.
22. Каллан Р. Основные концепции нейронных сетей / Р. Каллан. – М.: Изд. дом «Вильямс», 2001. – 288 с.
23. Pollack J. Recursive distributed representations / J. Pollack // Artificial Intelligence. – № 46. – P.77-105.

24. Lee G. Learning distributed representations of conceptual knowledge and their application to script-based story processing / G. Lee, M. Flowers, M. Dyer // Connection Science. – Vol.2, № 4. – P.313-345.

25. Гаврилова Т.А. Базы знаний интеллектуальных систем / Т.А. Гаврилова, В.Ф. Хорошевский. – СПб.: Питер, 2001. – 384 с.

26. Кузин Л.Т. Основы кибернетики: [в 2 т.] / Л.Т. Кузин – М.: Энергия, 1979. – Т.2: Основы кибернетических моделей. – М.: Энергия, 1979. – 584 с.

27. Шуклин Д.Е. Модели семантических нейронных сетей и их применение в системах искусственного интеллекта: дис. ... кандидата техн. наук: 05.13.23 // Д.Е. Шуклин. – Х., 2003. – 192 с.