

УДК 004.82 (045)

АЛГОРИТМ ПОРІВНЯННЯ ЛОГІКО-ЛІНГВІСТИЧНИХ МОДЕЛЕЙ РЕЧЕНЬ ПРИРОДНОЇ МОВИ

А.І. Вавіленкова

Національний авіаційний університет

e-mail: a_vavilenkova@mail.ru

Детальне дослідження текстових документів показало, що всі тексти, крім авторських, мають певну структуру, характерну для всіх стилів: вступ, зав'язка, постановка проблеми, розв'язка, висновки і т.д. По аналогії з синтаксичною та лексичною структурами, можна виділити основні змістовні (семантичні) елементи, що притаманні кожному тексту.

Так, важливими для подальшого семантичного аналізу текстів є поняття *когезії* (зв'язок елементів тексту, при якому інтерпретація одних елементів залежить від інших) та *когерентності* (прагматична зв'язність, що забезпечується зовнішнім впливом на текст) [1]. Саме завдяки цим поняттям можна робити певні висновки та будувати зв'язки, відсутні в документі у явному вигляді.

Основним засобом когезії є повторення одних і тих самих елементів, що можна використати для здійснення порівняльного аналізу речень природної мови. Наприклад, *лексичний повтор* полягає у вживанні слів чи словосполучень певну кількість разів, з чого можна робити висновки про ключові слова документів. *Семантичний повтор* дає змогу простежити ряд речень з однаковим змістом (синтаксична структура таких речень буде однаковою, проте суб'єкт або об'єкт може бути замінений синонімами). В тексті також існують *анафоричні відношення*, що означає наявність в деяких реченнях слів, які містять посилання на інший фрагмент тексту. Найпоширенішим прикладом антецедента у таких зв'язках є займенник. *Еліпсис* – ще один засіб когезії, що означає пропуск у текстах певної мовної одиниці, яку можна відтворити за змістом попередніх речень. Не меншу роль у формуванні текстів грають *дискурсивні елементи* – сполучники та частки. Враховуючи всі вище описані елементи когезії, можна перейти до формалізації процесу порівняння текстової інформації за змістом.

Всім реченням природної мови притаманний певний порядок слів: прямий (суб'єкт, предикат, аргументи і т.д.) або інверсний (аргументи, предикат, суб'єкт, означення і т.д.). Тому для аналізу змісту речень природної мови необхідно використовувати такі формальні моделі, які б могли врахувати синтаксичну структуру речень. Такими засобами є логіко-лінгвістичні моделі, типу [2]:

$$P(x_1 \{ \bigwedge_{d_1 \in C_1(x_1)} c_{1d_1} \}, \bigwedge_{q \in J(S)} [\bigwedge_{q \in X(S)} [x_q \{ \bigwedge_{d_2 \in C_2(x_q)} c_{qd_2} \}]]), \quad (1)$$

де P - предикат, що відображає зміст речення;

x_1 - предикатна змінна (суб'єкт), знаходиться у предикативному відношенні з P ;

c_{1d_1} - предикатна константа, що вказує на ознаку суб'єкта;

d_1 - номер предикатної константи, що вказує на ознаку суб'єкта;

$C_1(x_1)$ - множина предикатних констант суб'єкта x_1 ;

x_q - предикатна змінна (аргумент);

q - номер предикатної змінної (аргументу), початкове значення якого $q = 2$;

$X(S)$ - множина предикатних змінних (аргументів);

c_{qd_2} - предикатна константа, що вказує на ознаку q -тої предикатної змінної (аргументу або об'єкта);

d_2 - номер предикатної константи, що вказує на ознаку предикатної змінної (аргументу);

$C_2(x_q)$ - множина предикатних констант предикатної змінної x_q ;

$J(S)$ - множина предикатних змінних, які виконують у реченні рівнозначну роль, $J(S) \in X(S)$;

q_1 - номер предикатної змінної із множини $J(S)$; якщо речення не має ієрархічної будови або в ньому не зустрічаються аргументи, рівносильні за своєю роллю, то $J(S) = \emptyset$.

Логічна формула (1) є інтерпретацією синтаксичної структури речень з урахуванням семантичних зв'язків, що є формальним засобом відображення змісту текстової інформації.

Значення речення представляє собою не елементарну, складну структуру. Для порівняння речень природної мови, представлених у вигляді логіко-лінгвістичних моделей, необхідно враховувати відмінкову граматику – опис семантики речення без модальної, комунікативної та інших рамок, за допомогою поняття глибинного відмінку. Принциповою для відмінкової граматики є різниця декількох рівнів глибинних та поверхневих структур, в її основі лежить поняття відмінкової рамки, функція якої полягає у простеженні зв'язків між описом ситуації та глибинними синтаксичними структурами. Відмінкова рамка приписує семантико-синтаксичні ролі конкретним учасникам ситуації. Аналіз логічної структури речення передбачає визначення пропозиції – відношень та зв'язків, що створюють певну ієрархію.

Ч.Філлмор ввів шість необхідних відмінків:

- *агентив* – відмінок ініціатора дії, що ідентифікується дієсловом;
- *інструменталіс* – відмінок певного предмета, що впливає на дію або стан, позиціонується як причина;
- *датив* – відмінок певного об'єкта, на який впливає дія або стан;
- *фактив* – відмінок предмета або істоти, що виникає внаслідок здійснення дії або розуміється як частина дієслова;
- *локатив* – відмінок, яким характеризується місце знаходження або просторова орієнтація дії;
- *об'єктив* – семантично найбільш нейтральний відмінок, що визначається семантичною інтерпретацією самого дієслова.

Якщо розглянути два речення ідентичної будови «Викладач відчинив двері» та «Ключ відчинив двері», то, не зважаючи на те, що структура речень однакова, слово «викладач» виступає в реченні агентивом, а слово «ключ» – інструменталісом. При побудові логіко-лінгвістичних моделей цих двох речень і «викладач» і «ключ» позиціонуються як суб'єкти:

$P(x_1, x_2)$ та $P(x_1', x_2)$, при чому $x_1 \neq x_1'$.

Відчинив(викладач, двері) та *Відчинив(ключ, двері)*.

Тобто побудова логіко-лінгвістичної моделі згідно формули (1) дає змогу встановити відношення між структурними елементами речення, проте не бере до уваги їх зміст (вважається, що речення істинне і несе в собі певні знання).

Дослідити істинність того чи іншого речення природної мови можливо лише у контексті порівняння його з іншими реченнями. Саме у цьому випадку велику роль відіграє відмінкова граматику.

Розглянемо декілька речень «Викладач відчинив двері», «Ключ відчинив двері», «Викладач відчинив двері ключем», «Двері відчинили», «Двері було відчинено ключем». Всі перераховані вище речення несуть знання про факт відкриття дверей, проте суб'єкти цих речень різні, що пояснюється вживанням різних відмінкових рамок. Флективна мова, крім вживання великої кількості синонімів, передбачає різноманітність структурних засобів побудови речень, так, вживання активної та пасивної форми дієслів, вживання

займенників, прямий та інверсний порядок слів у реченні. Відповідно, для порівняння за змістом необхідно простежити всі можливі варіації побудови речень, створити систему продукцій, побудовану на основі правил природної мови та формалізувати процес відшукування збігів і протиріч.

Для цього необхідно побудувати логіко-лінгвістичні моделі кожного з речень.

1. $P(x_1, x_2)$, *Відчинив(викладач, двері)*
2. $P(x_1, x_2)$, *Відчинив(ключ, двері)*
3. $P(x_1[x_3], x_2)$, *Відчинив(викладач[ключем], двері)*
4. $P'(x_2)$, *Відчинили(, двері)*
5. $P''(x_2, x_3)$, *Було _відчинено(, двері, ключем)*

Аналіз утворених логіко-лінгвістичних моделей речень природної мови дає змогу зробити наступні висновки:

1. Предикати P , P' та P'' є спільнокореновими словами, P і P' - дієслова в активній формі, проте у різному числі, а P'' є їх пасивною формою, про що свідчить перша складова предиката «було». Це дає можливість говорити про те, що зміст P , P' та P'' однаковий, тому у всіх логіко-лінгвістичних моделях можна записувати одну змінну P .

2. Четверта логіко-лінгвістична модель *Відчинили(, двері)* є істинною по відношенню до *Відчинив(викладач, двері)* і *Відчинив(ключ, двері)*, так як не має суб'єкту, але об'єкт для всіх трьох логіко-лінгвістичних моделей однаковий.

3. Аналогічно логіко-лінгвістична модель *Було _відчинено(, двері, ключем)* є істинною по відношенню до *Відчинив(викладач, двері)*, *Відчинив(ключ, двері)*, *Відчинив(викладач[ключем], двері)*, *Відчинили(, двері)*, так як також не містить суб'єкта, а об'єкти у цих реченнях співпадають.

4. Не можна говорити про ідентичність за змістом висловлювань *Відчинив(викладач, двері)*, *Відчинив(ключ, двері)* та *Відчинив(ключ, двері)*, *Відчинив(викладач[ключем], двері)*, тому що суб'єкт у наведених вище парах висловлювань різний, з чого можна припустити, що двері було відчинено не викладачем, а, наприклад, учнем.

5. Ідентифікація різних форм предиката, а також слова «ключ» у різних реченнях, дає змогу зробити заміни для спрощення процесу порівняння.

Після заміни змінних у логіко-лінгвістичних моделях будемо мати наступний набір формул:

1. $P(x_1, x_2)$, *Відчинив(викладач, двері)*
2. $P(x_3, x_2)$, *Відчинив(ключ, двері)*
3. $P(x_1[x_3], x_2)$, *Відчинив(викладач[ключем], двері)*
4. $P(x_2)$, *Відчинили(, двері)*
5. $P(x_2, x_3)$, *Було _відчинено(, двері, ключем)*

Таким чином, дослідження структури речень природної мови, а також врахування основних елементів когезії дає змогу сформулювати базові етапи алгоритму порівняння логіко-лінгвістичних моделей текстової інформації.

1. Побудова логіко-лінгвістичних моделей. На цьому етапі кожному реченню природної мови ставиться у відповідність логічна формула, що представляє собою одновимірний масив слів, з яких складаються речення, упорядкованих у відповідності до того, яку синтаксичну роль вони виконують. Таким чином вся текстова інформація буде формалізована і записана у вигляді моделей, створених за єдиним шаблоном (1).

2. Ідентифікація. Відбувається почерговий перегляд елементів всіх логіко-лінгвістичних моделей: предикатів, предикатних змінних (суб'єктів), предикатних

змінних (аргументів), предикатних констант. Серед складових логіко-лінгвістичних моделей шукаються спільнокореневі слова, синоніми, активні та пасивні форми спільнокоренових дієслів. Почерговий перегляд відбувається не лише між елементами, що виконують одну і ту ж синтаксичну роль, але й між усіма структурними складовими, що дає змогу повністю ідентифікувати всі слова речень, навіть якщо вони мають різні грами, проте відображають один і той самий зміст.

3. Заміна тотожних змінних. Якщо на етапі ідентифікації знайдено тотожні змінні, у всіх логіко-лінгвістичних моделях відбувається їх перепозначення, завдяки чому одні й ті самі слова (навіть якщо вони мають різні граматичні рамки) будуть позначатися однаково, і, відповідно, мати ідентичний зміст.

4. Логічний вивід. На останньому етапі алгоритму порівняння логіко-лінгвістичних моделей після ідентифікації та заміни тотожних змінних застосовується система продукцій, що містить правила порівняння логіко-лінгвістичних моделей.

Найбільшу цікавість представляє собою останній етап. Саме тут використовується база знань у вигляді продукційної моделі, сформована на основі правил формування речень флективних мов. Такі правила дозволяють через встановлені кореляції між словами переходити до представлення значень слів у вигляді комбінацій елементарних компонентів змісту. Тобто побудова таких правил передбачає компонентний аналіз лексичних значень.

Відношення, у яких знаходяться слова у флективних мовах, мають різний ступінь системності. Якщо розуміти під системністю організацію деякої множини об'єктів як можливість розміщення цих об'єктів за допомогою правил, кількість яких менша, ніж кількість об'єктів, то можна вважати, що системними є ті кореляції, які забезпечують системність організації множини слів [3]. Із цього випливає, що системні кореляції не повинні бути унікальними, повинні характеризувати не одну пару взаємопов'язаних слів, а цілий ряд таких пар. Саме на основі таких системних відношень між словами будуються пропорційні рівності, тобто семантичні пропорції.

Компонентний аналіз передбачає співставлення значень слів у декілька етапів [3]:

1. визначається та одиниця змісту, яка включає значення слова (початкова форма та словосполучення, що дає пояснення компоненти змісту);
2. знаходяться синоніми та гіпероніми;
3. досліджуються одиниці одного ієрархічного рівня і знаходяться у відношенні перетину із шуканим словом;
4. створення мінімального списку тих компонентів, що відрізняють шуканий елемент від всіх інших;
5. формулювання дефініції слова на основі діагностичних компонент.

Порівняння логіко-лінгвістичних моделей речень природної мови дає можливість аналізувати текстові документи за змістом, знаходити у них збіги та протиріччя. Алгоритм порівняльного аналізу, запропонований автором, включає в себе базові етапи, кожен з яких представляє собою об'ємну процедуру використання логічних, лексичних, синтаксичних та семантичних правил формування текстів.

Список використаної літератури

1. Кронгауз М.А. Семантика – М.: Издательский центр «Академия», 2005. – 352 с.
2. Вавіленкова А.І. Логіко-лінгвістична модель як засіб відображення синтаксичних особливостей текстової інформації /А.І. Вавіленкова // Математичні машини та системи. – 2010. – № 2. – С. 134–137.
3. Кобозева И.М. Лингвистическая семантика. – М.: Эдитореал УРСС, 2000. – 352.