The II[nd] International Conference «Advanced Information Systems and Technologies, AIST 2013»
21-24 May 2013, Sumy, Ukraine

Рівень образного аналізу та синтезу в традиційній тріаді морфологія-синтаксис-семантика

Бісікало О. В., Богач І.В.
ВНТУ, obisikalo@gmail.com, http://aivt.inaeksu.vntu.edu.ua/ksklad/1153.html

The level of figurative analysis and synthesis in the traditional triad of morphology, syntax and semantics is examined in the article. New possibilities of using figurative analysis and synthesis the sentences of text are suggested. The characteristics and requirements to the formal procedures of the obtained approach are discussed.

ВСТУП

Класичний підхід до аналізу речень тексту в лінгвістиці загалом та комп'ютерній лінгвістиці зокрема передбачає послідовність дій на 3-х рівнях [1]: морфологічному, синтаксичному та семантичному. Синтез природно-мовних конструкцій (ПМК) базується на зворотному порядку тих же самих рівнів. Проте значні складності у формалізації семантичних правил змушують дослідників вдаватися до введення штучних рівнів на зразок глибинного синтаксичного або поверхневого семантичного [2], що на практиці не дає бажаного наслідку для розуміння текстового контенту [3].

Нейропсихологічний аналіз мовленнєвої діяльності людини демонструє паралельний розвиток процесів різного рівня [4] та дозволяє висунути гіпотезу щодо існування базового образного рівня розуміння [5] природної мови.

Мета роботи полягає в обґрунтуванні основних вимог до формалізації рівня образного аналізу та синтезу природно-мовного контенту.

ОБРАЗНИЙ АНАЛІЗ ТА СИНТЕЗ РЕЧЕННЯ

Основним змістовним елементом тексту будемо вважати речення як природно-мовну форму відображення окремої події [3].

Формальні задачі виокремлення речення та морфологічного і синтаксичного аналізу його слів з прийнятною якістю вирішені для більшості природних мов.

Мовленнєва практика показує, що розуміння основного сенсу речення не вимагає його повного семантичного аналізу. Наприклад, ті ж самі міміка та жестикуляція значно покращують розуміння висловлювань співбесідника навіть в умовах невизначеності окремих почутих слів. Отже, існує образний рівень сприйняття світу, для якого вербальні ознаки є лише одним з можливих типів формальних ознак. У певному колі задач комп'ютерної лінгвістики достатньо досягти оцього загального розуміння ПМК, який у [5] формально введено у вигляді поняття образного сенсу. Тоді зображений на рисунку 1 рівень образного аналізу та синтезу забезпечить розв'язок таких задач, оминаючи труднощі семантичного рівня.

Рисунок 1 – Образний рівень у традиційній тріаді

Прикладами подібного роду задач можна вважати: пошук за змістом ПМК, побудова природно-мовних онтологій, кластеризація, анотування та реферування текстів, переклад, підтримка спрощених типів діалогу, коли відповіді надаються у вигляді множини слів, асоціативно пов'язаних з питанням («дельфійський оракул»), цитат з літературних джерел («магістр Йода»), з

недотриманням синтаксичних правил («Basic English»).

Останні приклади задач підтримки спрощених типів діалогу є обмеженими випадками відомого тесту Тьюринга, який відносять до так званих AI-повних задач. Запропонований підхід до формалізації рівня образного аналізу та синтезу ПМК має забезпечити можливість отримання таких обмежених поняттям образного сенсу випадків класу AI-повних задач, для розв'язку яких достатньо поліноміальної складності обчислювальних процедур. Ще одним перспективним, на думку авторів, прикладом дослідження AI-повної задачі на основі окресленого підходу є формалізація жарту та інших природно-мовних форм гумору.

Концептуально рівень образного аналізу та синтезу ПМК базується на використанні тезауруса мовних образів [5]. Останніми вважають множини однокореневих слів, які характеризують окремий образ з нескінченної множини $I = \{i_1, i_2, ..., i_n, ...\}$, що забезпечує морфемну класифікацію та гніздовий принцип об'єднання словоформ у тезаурусі. Запропонований підхід до аналізу та синтезу текстового контенту має забезпечувати такі операції:

- відокремлення речення або автономної частини речення з тексту;
- перетворення речення у список слів, що ідентифікуються модулем тезауруса;
- побудова та модифікація тезауруса мовних образів обраної природної мови;
- розробка парсеру обраної мови для отримання дерева підлеглості (залежностей) речення через синтаксичні зв'язки;
- образна індексація множини текстів за рахунок накопичення синтаксичних зв'язків між мовними образами тезауруса у вигляді семантичної мережі;
- перетворення базових словоформ з дерева мовних образів у речення або ПМК за допомогою модулів синтаксичного та морфологічного рівня.

ВИСНОВКИ

Доцільність введення рівня образного аналізу та синтезу ПМК у традиційну тріаду морфологія–синтаксис–семантика пов'язана з можливістю отримання прийнятних рішень актуального кола задач комп'ютерної лінгвістики. При цьому зникає необхідність долати відомі труднощі семантичного рівня. З формальної точки зору підхід базується на отриманні обчислювальних процедур поліноміальної складності для обмежених поняттям образного сенсу випадків класу AI-повних задач.

З метою подальшого розвитку запропонованого підходу необхідно реалізувати 6 формальних операцій, найважливішою з яких є розробка парсеру обраної природної мови для отримання дерева підлеглості (залежностей) речення через синтаксичні зв'язки.

ЛІТЕРАТУРА

[1] Apresyan J. ETAP-3 Linguistic Pricessor: a full-fledged NLP implementation of the MTT / J. Apresyan, I. Boguslavsky, L. Iomdin etc. // MTT 2003, First International Conference on Meaning – Text theory (Paris, June 16-18, 2003).– Paris, ENS, 2003. – P. 279–288.

[2] Кобозева И.М. Лингвистическая семантика: учебное пособие / Кобозева И.М. – М.: Эдиториал УРСС, 2000. — 352 с. – ISBN 5-8360-0165-0

[3] Бісікало О.В. Концептуальні основи моделювання образного мислення людини / Бісікало О.В. – Вінниця: ПП Балюк І.Б., ВДАУ, 2009. – 163 с. – ISBN 978-966-2959-62-3.

[4] Лурия А.Р. Язык и сознание / Лурия А.Р.; под ред. Е.Д. Хомской. – М.: Издательство Московского университета, 1979. – 320 с.

[5] Бісікало О.В. Формалізація понять мовного образу та образного сенсу природномовних конструкцій / О.В. Бісікало // Математичні машини і системи. – 2012. – № 2. – С. 70–73. – ISSN 1028-9763.