

Література

1. Tilk O., Alumäe T. Bidirectional recurrent neural network with attention mechanism for punctuation restoration // Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH. 2016. Т. 08-12-Sept. № September. С. 3047–3051.
2. Воркачев С.Г. Лингвокультурология, языковая личность, концепт: становление антропоцентрической парадигмы в языкознании // Филологические науки. 2001. Т. 1. № 1. С. 64–72.
3. Геодакян В.О. Организация систем - живых и неживых // Системные исследования 1970. Ежегодник. Москва: Наука, 1970. С. 49–62.
4. Загнітко А.П. Лінгвоіндивідуації та лінгвоіндивідуалізації в епістолярному дискурсі: закономірності реалізації мовної особистості // Лінгвокомп'ютерні дослідження. 2016. Т. 9. С. 121–143.
5. Іваницька М.Л. Мовна особистість перекладача як чинник впливу на продукт перекладу (на матеріалі роману Б. Леберта Крейзі) // Проблеми семантики, прагматики та когнітивної лінгвістики. 2012. № 21. С. 147–159.
6. Караулов Ю.Н. Русский язык и языковая личность. : УРСС, 1987. 262 с.
7. Красных В.В. Виртуальная реальность или реальная виртуальность?: человек, сознание, коммуникация. Москва: Dialog-Mgu, 1998. 352 с.
8. Мацько Л.І. Мовна особистість Андрія Малишка як лінгвокультурологічний феномен національного виховання у світлі лінгвофілософії // Науковий часопис НПУ імені М. П. Драгоманова. Серія 16 Творча особистість учителя проблеми теорії і практики. 2013. № 20. С. 141–146.
9. Редько Є.О. Українська мовна особистість і арго (на прикладі семантичних назв осіб за спорідненістю і свояцтвом) // Науковий часопис НПУ імені М. П. Драгоманова. Серія Філологічні науки (мовознавство і літературознавство). 2014. № 5. С. 212–218.

Наталія Дарчук
(*м.Київ*)

ДЕРЕВО ЗАЛЕЖНОСТЕЙ ЯК ПАРАМЕТР АВТОРСЬКОГО СТИЛЮ

У багатьох системах автоматичного опрацювання тексту і вітчизняних, і зарубіжних використовується спосіб формального синтаксичного аналізу, який базується на граматиці залежностей (ГЗ), а структура речення має вигляд дерева залежностей (ДЗ). В основі ГЗ лежить синтаксис Теньєра [3]. Будова речення уподібнюється молекулі, складається з атомів – слів, пов'язаних між собою. Принципи такі:

1) при аналізі увагу скеровано на зовнішню мовну форму, лінгвістичні одиниці ототожнюються і класифікуються не на основі значення, як це мало місце у традиційній граматиці, а на основі їх розподілу (дистрибуції) у мовленні [1]; кожне явище вивчається в оточенні інших явищ й у взаємодії з іншими явищами.

2) лінгвістичний синтаксичний опис являє собою набір процедур опрацювання тексту, виконання яких у певному порядку має привести до створення моделі мовленнєвої діяльності.

3) у реченні розгортаються синтагматичні зв'язки або синтагматичні залежності;

4) керування – синтаксична особливість слова.

Теоретичні засади, практичне значення і перспектививи формального синтаксису у вигляді ДЗ для мовознавства були розроблені в І.П.Севбо у ґрунтовній праці «Графічне представлення синтаксичних структур і стилістична діагностика» [2].

І.П. Севбо пропонує такі основні домовленості ГЗ [2, с.8 – 9].

1. Речення будується з одиниць одного таксономічного (англ. taxonometry – систематизація) рівня – зі словоформ. Службові слова, як і повнозначні, вважаються мінімальними елементами синтаксичної структури.

2. Словоформи у реченні зв'язані відношенням підпорядкування, яке є узагальненням традиційних синтаксичних відношень керування, узгодження, прилягання (сурядність й однорідність вписуються в загальну систему підпорядкування).

3. У реченні є одна абсолютно незалежна словоформа – присудок. Усі інші словоформи підпорядковуються якомусь «хазяїну», причому тільки одному (у словоформи не може бути одночасно відразу двох керувальних). Однак одна словоформа може підпорядковувати більше двох словоформ. У системі підпорядкування виділяються два типи зв'язку в групі словоформ:

- гілкування, коли від однієї словоформи залежить кілька словоформ;
- ланцюжок підпорядкувань, коли кожна наступна словоформа залежить від попередньої.

Структура залежностей закінчується словоформами, яким не підпорядковані жодні інші словоформи.

4. Для кожної словоформи можна задати набір підпорядкувань, який відображає її поведінку в мовленні при побудові конкретних речень. Такі моделі описують валентності словоформ, тобто потенційні синтаксичні сполучуваності певної словоформи із залежними від неї словоформами.

5. Ієрархічна структура залежностей між словоформами речення зображується графічно за допомогою дужок, стрілок або у вигляді дерева залежностей.

Слід зазначити, що увагу розробників синтаксичного модуля автоматичної обробки українськомовного тексту АГАТ було сконцентровано

саме на останніх двох положеннях (4-5), які використовувалися в більшості систем: числення моделей валентностей; графічна побудова дерева залежностей.

Мета даного проекту полягає у комплексній та аспектній розробці Корпусу текстів української мови стилів і дискурсів з можливістю «портретувати» речення, визначаючи морфологічну характеристику кожного члена речення із набором підпорядкувань і синтаксичну у вигляді ДЗ. Цей шлях забезпечує можливість, з одного боку, одержання однотипної, точно у лінгвістичному смислі інформації, з іншого – створення за результатами розмітки набору параметрів стилістичного характеру, що передбачає наявність комп'ютерних інструментів – пакету програм для різнобічної лінгвістичної обробки текстового матеріалу, в тому числі стилістичної.

ДЗ дає важливу інформацію для стилістичних висновків, оскільки слугує параметричною інформацією. Напр.:

- 1) вузлові параметри, або середня кількість вузлів у дереві;
- 2) параметр глибини дерева, або середня кількість рівнів у дереві;
- 3) параметр ширини, або середня кількість вузлів на одному рівні дерева;
- 4) асиметричний параметр, який являє собою співвідношення у кількості вузлів між піддеревами, отриманими поділом навпіл другого рівня;
- 5) гілковий параметр, який вимірюється числом кінцевих вузлів до числа дерев;
- 6) параметр кратності – відношення числа кратних вузлів до числа дерев (кратним називаються вузли, у якого кілька підлеглих);
- 7) наскрізний параметр, або середня довжина наскрізного ланцюжка (наскрізний ланцюжок – це шлях у дереві, який веде від кореня до кінцевого вузла).

Вся перерахована інформація, яка стосується ДЗ, будується для кожного речення конкретного тексту. Накопичення її в кінцевому підсумку дасть чимало корисних даних про словосполучення, їх синтаксичну модель із кількісними характеристиками, структуру речення і стилістичні особливості синтаксису (Рис. 1).

Контекст

— Громадяни, дотримуйтесь тиші, — вимагають чотири щити навколо високої вежі, де великими літерами зазначено " Всесоюзний науково-дослідний центр по акліматизації картоплі на Марсі " .

Джерело

>>

Показати дерево залежностей

Ламане дерево: Ні
 непроективна структура: Ні
 кількість вузлів у графі (словоформ) у реченні: 21
 кількість простих речень у складному: 2
 кількість рівнів у графі: 8
 ширина гілкування у кореня: 3
 максимальна кількість змін у шляху гілки графа: 1
 максимальна протяжність дуги графа: 8
 загальна кількість вузлів у графі: 0
 асиметрія графа: 1/19
 середня кількість рівнів: 9/21
 середня кількість вузлів у рівні графа: 2,625
 співвідношення всіх вузлів речення, які не є терміальними (не є листями), до всіх вузлів цього речення: 12/21
 середня глибина гілки речення: 5

Рис. 1 Параметризація дерева залежностей (на прикладі речення В.Стуса)

Дослідження здійснювалося на матеріалі чотирьох вибірок з поетичних творів М.Вінграновського, І.Драча, В.Стуса й А. Мойсієнка (по 100 речень). Автоматично створювалися ДЗ і параметризація кожного речення, встановлювалися особливості авторської манери та неоднорідність як метод стилізації, різномантність ДЗ в однорідному тексті (перелік типів структур).

Література

1. Атаян Э. Р. Предмет и основные понятия структурального. Ереван : Митк, 1968. 279 с.
2. Севбо И. П. Графическое представление синтаксических структур и стилистическая диагностика. Київ : Наукова думка, 1981. 192 с.
3. Теньер Л. Основы структурного синтаксиса. Москва : Прогресс, 1988. 656 с.