

переводе с английского языка на русский // Проблемы кибернетики. – Вып. 1. – М., 1958. – С. 213–221; ⁵Малаховский Л.В. Теория лексической и грамматической омонимии. – Л., 1990. – 240 С; ⁶Шевченко І. Алгоритмічна словозмінна класифікація української лексики // Мовознавство. – 1996. – С. 40–44.

О.В. Бугаков, асп.

АВТОМАТИЧНА ІДЕНТИФІКАЦІЯ СКЛАДЕНИХ ПРИЙМЕННИКІВ В УКРАЇНСЬКОМУ ТЕКСТІ

Подано формальні правила розмежування складених прийменників від синтаксичних конструкцій типу “однослівний прийменник + повнозначне слово” в тексті (модуль автоматичного контекстного аналізу).

Formal rules that distinguish complex prepositions from syntactic structures “one-word preposition + full meaning word” in text (module of automatic context analysis) are given.

Розробка систем автоматичної обробки текстової інформації (АОТ), зокрема, автоматичного синтаксичного аналізу (АСА), вимагає розв’язання проблеми ідентифікації аналітичних конструкцій – сполучень повнозначного і службового слів, в яких службове слово виражає граматичне значення повнозначного і, відповідно, усієї конструкції в цілому¹. Кількість компонентів може коливатися від 2 до 5: згідно з; з огляду на те що. До таких аналітичних конструкцій (“складені об’єкти” – Л.В.Ричкова², “еквіваленти слова” – Р.П.Рогожнікова³, А.А.Лучик⁴) відносимо: аналітичні форми дієслова майбутнього часу (*буду писати*), минулого часу (*зробив був*), умовного способу (*пішов би*), описові форми ступенів порівняння прикметників і прислівників (*найбільш ерудований*), аналітичні сполучники (*у разі коли*), еквіваленти прислівників (*не в міру*), еквіваленти часток (*лише б*), складені прийменники (*відповідно до*) тощо. Останні і є об’єктом нашого дослідження.

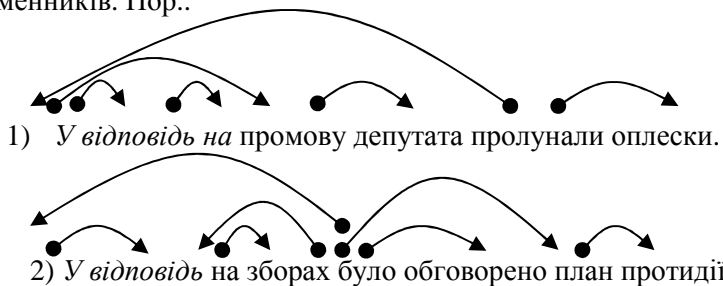
За даними попереднього дослідження, проведеного на текстах обсягом 100 тис. слововживань, частота вживання складених прийменників (СП) серед усіх прийменникових вживань становить 3,2%.

У прийменникових сполученнях, що одержують статус прийменника, повнозначні слова в поєднанні з конкретними

прийменниками повністю або значною мірою втрачають свої лексичні, парадигматичні, синтаксичні, зв'язки з відповідними їм лексико-граматичними класами, перебираючи на себе категоріальне значення прийменника і уточнюючи його.

При АСА функціональна єдність (синтаксична і лексична) складених прийменників знаходить своє відображення на етапі формального представлення синтаксичної структури речення в особливому маркуванні СП як єдиної синтаксичної одиниці, що бере участь у керуванні залежною іменною формою в певному відмінку. Відмінок залежного слова визначається не валентностями прийменникового компонента аналітичної конструкції, а усім сполученням в цілому.

В алгоритмі АСА інформація про конкретний тип керування, в якому бере участь той чи інший СП, належить до вихідної і може задаватися різними способами. У нашому випадку вона задається через номер списку, до якого належить конкретний СП (в один список попередньо об'єднуються складені прийменники згідно з конкретним відмінком залежних від них форм). Пошук залежної форми в тексті здійснюється за правилами обробки однослівних прийменників. Пор.:



При переході до глибокого семантико-синтаксичного представлення речення складені прийменники одержуватимуть статус однієї лексики з приписуванням їй особливого функціонального значення – семантичного відношення, виразником якого є даний СП.

Метою дослідження є розробка формальних правил ідентифікації складених прийменників у тексті, тобто розмежування синтаксичних конструкцій типу *однослівний прийменник + повнозначне слово* від тих, де відповідна конструкція виступає у функції прийменника і вимагає вживання залежної іменної форми в певному відмінку. Необхідність у таких правилах зумовлена тим, що

в результаті автоматичного морфологічного аналізу (АМА) тексту, що здійснюється в межах кожної текстової словоформи (тобто графемного ланцюжка між пробілами), компоненти складеного прийменника і компоненти сполучення однослівного прийменника з залежною іменною формою мають однакові морфологічні коди, і отже, на етапі АМА не можуть розрізняватися. Пор.:

3) Сергій<ЙИ> на<ПВ> знак<ЙВ> (СП) згоди<КР> хитнув<ГЕ> головою<КТ>.

4) Ібрагім<ЙИ> підвів<ГЕ> очі<ЛА> на<ПВ> знак<ЙВ> (синтаксична конструкція), що<СП> знаходився<ГЭ> ліворуч<Н0>. (Приклади подано з кодами, уточненими за допомогою контекстного аналізу (КА), призначеного для усунення граматичної омонімії).

Нерозмежування цих ситуацій на подальших етапах автоматичної обробки текстової інформації, зокрема, при машинному перекладі, призводитиме до некоректних варіантів перекладу, тому що пошук перекладних еквівалентів для СП здійснюється за іншими принципами, ніж для сполучень прийменника з залежною відмінковою формою:

5) Регулювання напруги в мережі **за допомогою** (СП) статичного компенсатора. – Tension regulation in electricity supply network **by means of** static jack (Неправильним був би переклад: ... ***for help** static jack).

6) Учора Арафат звернувся **за допомогою** (синтаксична конструкція) до Європейського Союзу. – Yesterday Arafat asked European Union **for help**.

Без розрізнення СП і конструкції *прийменник + залежна форма* переклад на російську мову прикладу 3) мав би такий вигляд: Сергей ***на знак** согласия кивнул головой (Правильно: ... **в знак** согласия).

Крім того, ідентифікація СП уточнює результат КА щодо однозначного зняття омонімії в ланцюжку іменних форм, розташованому в постпозиції до СП. Продемонструємо це на прикладі аналізу фрагмента речення: здійснено<Г@> за<ПТ> допомогою<КТ> моделі<КРКДКПКАКУКШ> (результат роботи АМА). Після АКА одержуємо: здійснено<Г@> за<ПТ> допомогою<КТ> моделі<КРКДКУ>. Надання сполученню “за допомогою” статусу складеного прийменника автоматично залишає в омонімічному ланцюжку словоформи *моделі* один код <КР>

(родовий однини), оскільки саме родовий відмінок передбачає прийменник *за допомогою*.

Предметом дослідження був список зі 144 прийменникових сполучень, які в тексті можуть виконувати функцію прийменника. До нього ввійшли сполучення, визнані складеними прийменниками в сучасних українських граматиках та наукових дослідженнях, представлених у монографіях І.Р.Вихованця⁵, А.С.Колодяжного⁶, А.А.Лучик⁷. До всіх сполучень зі списку було застосовано 4 диференційні ознаки, за якими можна відрізнити складені прийменники від сполучень повнозначних слів з простими прийменниками: 1) контактність компонентів; 2) наявність зліва від прийменникового сполучення іменної форми у чітко визначеному відмінку; 3) трансформація лексичного значення повнозначного слова у складі СП; 4) можливість заміни СП простим прийменником без зміни значення: з метою (чого) ↔ для, у разі (чого) ↔ при, під час (чого) ↔ протягом.

У результаті було отримано список із 69 СП, які за структурою зводяться до чотирьох типів: 1) простий прийменник + іменник у непрямому відмінку (PrepN: *з метою, за допомогою*); 2) простий прийменник + іменник у непрямому відмінку + простий прийменник (Prep₁NPrep₂: *у відповідь на, у зв'язку з*); 3) прислівник + простий прийменник (AdvPrep: *разом з, залежно від*); 4) складені прийменники *незважаючи на, неважаючи на, зважаючи на*.

Складені прийменники, на відміну від простих, характеризуються однозначністю щодо вимоги відмінка залежної від них форми (родового, давального, знахідного або орудного). Відповідно до цього всі СП було поділено на 4 групи: 1) СПД – складені прийменники, які беруть участь у керуванні залежною іменною формою в давальному відмінку: *на зміну, на противагу* (2); 2) СПЗ – беруть участь у керуванні залежною формою в знахідному відмінку: *зважаючи на, неважаючи на, незважаючи на, з огляду на, у відповідь на* (5); 3) СПО – беруть участь у керуванні залежною формою в орудному відмінку: *в/у порівнянні з, вслід/услід за, згідно з, нарівні з, одночасно з, паралельно з, порівняно з, поруч з, поряд з, разом з, спільно з, слідом/слідком за, у зв'язку з* (13); і, нарешті, до кількісно найбільшої групи 4) – СПР відносяться складені прийменники, які беруть участь у керуванні залежною формою в родовому відмінку: *відповідно до, на адресу, у справі* тощо (49).

Як показав попередній аналіз, частота вживання прийменникових сполучень – претендентів на роль складених прийменників – у функції прийменника становить близько 93%. З 69 досліджуваних сполучень 20 актуалізують тільки функцію прийменника: *відповідно до, залежно від, згідно з, незважаючи на, паралельно з, під час, подібно до, порівняно з, стосовно до* тощо. Ці прийменникові сполучення було виділено в окремі списки згідно з відмінком залежних від них іменних форм. При знаходженні таких сполучень у тексті алгоритм передбачає об'єднання їх у СП і визначення відмінка залежної іменної форми, за яким і буде уточнено граматичний код у першій іменній словоформи праворуч від даного СП. Для решти 49 прийменникових сполучень, які в текстах нашої вибірки актуалізують функції як прийменника, так і сполучення, укладалися правила розмежування цих функцій за контекстними умовами. Розпізнавальними вважаються синтагматичні контексти, які постійно повторюються для одного елемента омонімічного ряду і ніколи не зустрічаються з іншими його елементами⁸.

Матеріалом дослідження слугував електронний корпус текстів обсягом 23 млн. слововживань. На ньому для кожного прийменникового сполучення – претендента на роль складеного прийменника – зі списку за допомогою спеціальної програми, розробленої в Українському мовно-інформаційному фонді НАН України, було сформовано конкорданси, в яких межі контекстів визначалися початком і кінцем речень, що містять досліджувані мовні одиниці. Обсяг одержаної бази становить 41 789 контекстів.

Правила ідентифікації аналітичних конструкцій, в тому числі складених прийменників, пропонується об'єднувати в окремий модуль КА, за допомогою якого конструкціям надаватиметься код однієї текстової одиниці. Потреба у виділенні окремого модуля пояснюється тим, що в деяких дослідженнях, зокрема, в морфемному сегментаторі текстових словоформ (МСТС), зайвим є зведення аналітичних конструкцій до однієї одиниці, тому що вихідною одиницею для МСТС є саме окрема словоформа з однозначними граматичними ознаками.

При визначенні розпізнавальних контекстів, за якими здійснюється уточнення в реченні функцій прийменникових сполучень – претендентів на роль складених прийменників (ПСП) в окрему групу виділено ті, що у функції СП вимагають вживання залежної від них форми в давальному відмінку. Це зумовлено тим,

що іменна форма цих СП може вживатися в тексті як в постпозиції до них, так і в препозиції: “Головацький став йому **на зміну**”. Тоді як для решти ПСП пошук залежної форми здійснюється лише в правому контексті.

Необхідними і достатніми контекстуальними ознаками, які не дозволяють переводити ПСП в СП, є:

1. Безпосереднє розташування праворуч від ПСП розділових знаків { ; ! ? ... [] () – } : “*І не кожному це **на користь**.*”;
2. Розташування справа від ПСП безпосередньо або через вставні слова, прислівник, частку, кому і двокрапку граматичних класів: дієслова, прийменника і сполучника: “*Студент, радий з **нагоди** агітувати когось, охоче пояснював.*”;
3. Правим контекстом ПСП є кома з наступним сполучником підрядності: “*Руйнації він піддається тільки у **разі**, коли цьому сприяють погодні умови.*”;
4. Перший справа від ПСП іменник, ад’єктив, займенник або числівник має відмінкову форму, що не збігається з тією, яку вимагає відповідний СП: “*Лише так можна пояснити азарт, з яким чеченці розправлялись із спійманими **на базі** бандитами.*”

Для першої групи ПСП наведені вище контекстуальні ознаки можуть також розглядатися як достатні для ідентифікації прийменникових синтаксичних конструкцій за умови, що в постпозиції до них не вжито іменну форму в давальному відмінку.

До числа необхідних контекстуальних ознак, що діагностують СП для всіх ПСП, є контактне (постпозиційне для другої групи, і постпозиційне або препозиційне для першої) розташування іменної форми у відмінку, що закріплений за відповідним СП. Як показав аналіз матеріалу, дана ознака для ряду ПСП (*в/у зв’язку з, в/у залежності від, в/у світлі, в/у справі, на базі, на адресу, на зміну, на знак*) не може вважатися достатньою при визначенні функціонального навантаження досліджуваних ПСП у тексті. Для них є необхідним також аналіз лексичного оточення.

Лексична інформація в алгоритмі КА може бути представлена двома способами: 1) виділенням семантичних класів текстових одиниць (дієслова, іменника, прикметника), що виконують функцію головного або залежного слова в прийменникових конструкціях. Семантичні класи задаються списками квазіоснов або словоформ. Наприклад, для визначення функції синтаксичної конструкції у сполучень *в/у зв’язку з* і *в/у залежності від* перевіряється головне

слово-дієслово у препозиції на приналежність його до семантичного класу “існування”: *перебувати, стояти, бути, опинятися, опинитися, залишатися, залишитися*. Пор.: “Республіка опинилася в залежності від (синтаксична конструкція) князя.” і “Уточнено класифікацію податків в залежності від (СП) їх функціональної спрямованості.” 2) використання певних лексико-граматичних ознак головного і залежного слів, поданих у словнику, на якому базується АМА (категорія “істота/неістота” для іменників, категорія перехідності для дієслів тощо). Наприклад, для ПСП в/у *справі* має значення категорія “істота/неістота”. Лише у випадку, коли залежний іменник належить до категорії неістоти досліджуване сполучення переводиться в СП. “У *справі* (синтаксична конструкція) Андрія Бабицького виявлено нові обставини, які стосуються його участі в незаконних озброєних формуваннях.” – сполучення залишається без змін, “Процедурні питання дедалі виразніше виступають на перший план у *справі* (СП) імплементації.”

Як показав аналіз матеріалу, на перевірку лексичного контексту – як лівого, так і правого – орієнтовано близько 30% правил.

Роботу алгоритму КА було перевірено на текстах публіцистичного стилю обсягом 100 тис. слововживань. Аналіз результатів (94%) свідчить про високу розпізнавальну силу запропонованого методу розмежування складених прийменників від прийменникових синтаксичних конструкцій в тексті.

¹Жирмунский В.М. Об аналитических конструкциях // Аналитические конструкции в языках различных типов. – М.–Л., 1965. – С. 5 – 57; ²Рычкова Л.В. Проблема саставных аб'ектаў у корпусах славянских моў і лінгвістичных базах дадзеных // Мовознаўства. Літаратура. Культуралогія. Фалькларыстыка. – Мінск, 2003. – С. 184 – 195; ³Рогожникова Р.П. Словарь эквивалентов слова: наречные, служебные, модальные единства. – М., 1991; ⁴Лучик А.А. Російсько-український та українсько-російський словник еквівалентів слова. – К., 2003; ⁵Вихованець І.Р. Прийменникова система української мови. – К., 1980; ⁶Колодяжний А.С. Прийменник: Матеріали до лекцій з курсу сучасної української літературної мови. – Х., 1960; ⁷Лучик А.А. Значч. праця; ⁸Грязнухіна Т.О., Братищенко Л.Г., Дарчук Н.П., Критська В.І., Пузирева Т.К., Орлова Л.В. Шляхи усунення граматичної омонімії в системі автоматичного морфологічного аналізу наукового тексту // Мовознавство. – 1989. – №5. – С. 3 – 12.