

УДК 81'322

Ольга Шипнієвська
Військовий інститут
Київського національного університету імені Тараса Шевченка

ВИЗНАЧЕННЯ ТИПІВ СИНТАКСИЧНОЇ НЕОДНОЗНАЧНОСТІ У ЗНАННООРІЄНТОВАНІЙ СИСТЕМІ МАШИННОГО ПЕРЕКЛАДУ

Розглянуто теоретико-практичні засади визначення синтаксично неоднозначних конструкцій в автоматичному синтаксичному аналізі. На матеріалі текстів військової тематики проаналізовано основні типи синтаксичної омонімії.

Ключові слова: автоматичний синтаксичний аналіз, знаннеорієнтована система машинного перекладу, синтаксична неоднозначність, типи синтаксичних омонімів, синтаксична структура.

Синтаксична неоднозначність як один із типів мовної неоднозначності вивчалася багатьма лінгвістами. У традиційному мовознавстві ця проблема розглядалась найчастіше в стилістичному аспекті, теоретичне лінгвістичне обґрунтування неоднозначного тлумачення синтаксичних одиниць висвітлено у працях, присвячених вивченню синтаксичної структури речення [Гладкий 1985 : 110-127; Колесников 1976; Севбо 1981 : 30–65], з автоматичного опрацювання текстів природної мови [Дрейзин 1966; Дрейзин 1966; Иорданская 1967].

У поданому дослідженні, зважаючи на специфіку розроблюваного модуля автоматичного синтаксичного аналізу в знаннеорієнтованій системі машинного перекладу [Замаруєва 1999; Толубко 2010], розглядаються теоретико-практичні передумови визначення синтаксично неоднозначних одиниць тексту в автоматичному обробленні та на основі створеної лінгвістичної бази даних таких одиниць проведено аналіз синтаксично неоднозначних одиниць в українськомовних текстах військової тематики.

Вивчення синтаксичної омонімії як одного із типів мовної неоднозначності має як теоретичне, так і практичне значення [Колесников 1976; Иорданская 1967]. Особливо актуальним це

питання постає у контексті проблем автоматичного опрацювання текстів [Гладкий 1985 : 110–127; Дрейзин 1966; Дрейзин 1966]. Актуалізація досліджень із автоматичного синтаксичного аналізу ставить завдання формального представлення синтаксичної структури тексту та визначення семантично адекватних синтаксичних структур з-поміж безлічі отриманих у ході автоматичного синтаксичного аналізу правильних синтаксичних структур [Иорданская 1967]. Проте, зважаючи на специфіку явища, принципове вирішення синтаксичної омонімії, на думку дослідників, можливе тільки на рівні семантики. На рівні синтаксису можна говорити лише про зменшення кількості синтаксично неоднозначних одиниць [Иорданская 1967 : 10].

У дослідженні синтаксичними омонімами вважаються омонімічні синтагми речення, відмінність значень яких зумовлена неоднозначністю їхньої синтаксичної структури [Гладкий 1985 : 110]. Неоднозначність синтаксичних структур, на відміну від морфологічних, лексичних омонімів, зумовлена власне синтаксичними засобами: порядком членування речення на синтагми, типом зв'язку слів у реченні [Иорданская 1967; Колесников 1976]. Процедурно ж синтаксичну омонімію визначаємо як можливість приписати фразі більше аніж одну правильну синтаксичну структуру [Иорданская 1967 : 9]. За теоретичним обґрунтуванням Л. Йорданської, правильною синтаксичною структурою називається така синтаксична структура, яку приписує фразі носій мови, базуючись на своєму інтуїтивному уявленні про правильність синтаксичної структури [Иорданская 1966 : 215]. Вона повинна відповідати обраному принципу представлення синтаксичної структури, певному списку встановлених заздалегідь синтагм.

У лінгвістиці запропоновано два підходи зменшення кількості синтаксично неоднозначних структур [Иорданская 1967 : 10–11]. Перший – розроблення правил переваг. В основу цього підходу покладено процедуру порівняння отриманих для даної фрази структур та виявлення тієї, що найточніше відповідає аналізованій фразі. Здійснюється це із залученням процедури синтезу: для кожного із значень, встановлених для окремої фрази, синтезу-

ються всі можливі фрази, які стилістично оцінюються. У наборі фраз для кожного із значень є і вихідна фраза. Із отриманих структур перевага надається тій, за моделлю якої синтезується вихідна фраза з найвищою оцінкою. Другий підхід базується на виборі найбільш достовірного варіанта. При цьому підході алгоритм аналізу повинен мати список неоднозначних ситуацій, для кожної із яких визначено найбільш достовірну інтерпретацію.

У модулі автоматичного синтаксичного аналізу система будує тільки правильні синтаксичні структури, враховуючи можливі неоднозначні інтерпретації [Толубко 2010]. Синтаксична структура представлена за допомогою певного набору бінарних синтаксичних відношень. Список синтагм, за яким визначаються компоненти структури речення, на сьогодні є відкритим. При цьому під синтагмою розуміється сполучення двох елементів речення, для якого визначено головний та залежний елементи і тип синтаксичних відношень між ними [Иорданская 1966 : 216].

Описані у лінгвістиці класифікації типів синтаксичних омонімів демонструють виділення синтаксично неоднозначних одиниць на рівні словосполучень, синтагм, синтаксичних груп, речень [Колесников 1967]. Безперечно, для синтаксичних омонімів, як і для інших лінгвістичних одиниць, характерна знакова двоплановість. В одних випадках план вираження залишається тотожним, а план змісту зумовлює омонімічність, в інших – неоднозначність конструкції зумовлює не тільки план вираження, а й план змісту. У дослідженні за планом вираження визначається декілька типів синтаксичних омонімів: синтаксичні омоніми на рівні словосполучення (в тому числі прийменникових словосполучень), синтаксичні омоніми на рівні синтагми, синтаксичні омоніми на рівні речення (між частинами складного речення), синтаксичні омоніми на рівні синтаксичних конструкцій (як-от, конструкцій із однорідними членами речення, поєднаними сурядним зв'язком). У плані вираження синтаксична омонімія визначається за типом синтаксичних відношень, що існують між головним та залежним компонентами синтаксичної структури (об'єктні, атрибутивні, обставинні) [Мельчук 1974 : 221–236].

Для зменшення кількості встановлених правильних синтаксичних одиниць у нашому дослідженні пропонується аналіз структурно-семантичних типів синтаксичних омонімів, встановлених на основі розробленої бази даних, із метою їх подальшого опрацювання на рівні семантичного аналізу¹.

Класичним прикладом синтаксичної омонімії є іменникові словосполучення із родовим відмінком, які у реченні виступають як неузгоджене означення. Наприклад, у словосполученні *фотографія брата* можна виокремити родовий приналежності (кому належить), родовий автора (хто зробив), родовий тотожності (хто зображений). Значення такого родового відмінка часто залежить від значення головного слова, як, скажімо, у словосполученнях *хатина дядька Івана* та *начальник дядька Івана* – правильне розуміння таких словосполучень не залежить від функцій родового відмінка (*хата* належить дядькові Івану, а *начальник* не належить).

І. Чередниченко наголошує на синтаксичній омонімії родового відмінка при віддієслівних іменниках: *рішення зборів* (збори вирішили), *рішення справи* (справу вирішили) [Чередниченко 1955 : 32]. Важливо, що дослідник встановив понад двадцять семантичних моделей, в яких родовий відмінок виражає означальну функцію, та близько п'ятнадцяти моделей із родовим відмінком у значенні додатка. Звісно, такі моделі потребують докладного вивчення на значному масиві текстів з урахуванням практичних потреб та можливостей лінгвістичних ресурсів. Наприклад, у межах нашої системи автоматичного синтаксичного аналізу може бути розпізнана модель: головне слово – назва абстрактного поняття, властивості, залежне – особа, якій приписано цю властивість. З урахуванням розробленого нами модуля словотвірного аналізу у словосполученнях *спеціальність студе-*

¹ Докладно про базу даних див. Zamarueva I., Shypnivska O. The instrumental environment for the automatic syntactical analysis of Ukrainian // NLP, Multilinguality / Proceedings of the Sixth International Conference: Corpus Linguistics, Corpus Based Grammar Research. 20–21 October 2011, Bratislava, Smolenice. – 2011. – 163–173.

нта, професія інженера для кожної із складових можна буде визначити необхідну семантичну інформацію і в результаті написати словосполученню атрибутивне значення.

Одним із найпоширеніших типів синтаксичної неоднозначності у проаналізованому матеріалі і є неоднозначність на рівні словосполучень, зокрема іменникових (63,3 %)¹. Альтернативним для аналізу є здебільшого неоднозначне визначення об'єктних та атрибутивних відношень, причому варто зауважити, що саме така інтерпретація зумовлена їх подальшим опрацюванням у перекладі іншою мовою: *застосуванню сили, країнами-членами Альянсу, із виконання Договору, керівництво повсякденним життям, військовослужбовці жіночої статі*. Хоча незначну частку серед них становлять саме ті одиниці, для яких у результаті кінцевого аналізу за допомогою семантичного аналізатора можна визначити тільки атрибутивне відношення (5 %): *літаками подвійного призначення, дельтовидне крило досить великої площі, засобів довготривалого зберігання, куртку камуфльованого кольору*. Для решти типів пріоритетним є об'єктне відношення (досить часто із значенням “частина/ціле”): *потенціалу Збройних Сил України, війська берегової охорони, війська протиповітряної оборони, полків армійської авіації, придатності громадян, пенсіонерів Збройних Сил*.

Синтаксичне відношення із об'єктним значенням уже на поверхнево-синтаксичному етапі аналізу тексту можуть отримати словосполучення, в яких головним словом є віддієслівний іменник: *виконання договору, здійснення повітряної розвідки, охорони повітряного простору, вирішення питань, супроводження служби*. Щоправда, не завжди це обмеження може бути визнане пріоритетним, як, скажімо, у словосполученні *звання офіцерського складу*, в якому за розробленою нами системою словотвірного аналізу словоформі *звання* приписується словотвірне значення “віддієслівний іменник зі значенням «звати»”.

¹ Загальна кількість встановлених на сьогодні синтаксично неоднозначних конструкцій становить близько 1300 одиниць.

Окрему групу становлять словосполучення, в яких головним словом є віддієслівний іменник, а залежне – іменник на позначення назви істоти: *рішення командира, передачі кореспондента, втручання противника, ураження противника, допит полонених, показання полонених*. Остаточне рішення щодо визначення синтаксичного відношення як об'єктне чи як атрибутивне здійснюється із урахуванням семантичної інформації. Об'єктні відношення можуть бути встановлені на рівні семантичного аналізу, подібно до випадку зі словосполученнями семантичної моделі “частина/ціле”: *комплект предметів, вихованці військових ліцеїв*.

Синтаксично неоднозначні прийменникові конструкції ставали об'єктом багатьох лінгвістичних досліджень. Л. Йорданська в російській мові виділяє вісім типів потенційно можливих синтаксично неоднозначних структур речення з прийменником [Йорданская 1967 : 14–15]. Ф. Дрейзин визначає у тексті специфіку розподілу таких конструкцій щодо омонімічного вузла [Дрейзин 1966 : 56–57]. Традиційно прийменникові конструкції розглядаються у контексті розроблення системи автоматичного синтаксичного аналізу. Зокрема, дослідники визначають зони прийменникових зв'язків у тексті [Бугаков 2006; Грязнухина 1985].

Синтаксична неоднозначність структур із прийменником (30 % у базі) визначається типом синтаксичних відношень між прийменником та залежним від нього компонентом при визначенні головного слова, якому підпорядкована прийменникова конструкція. Неоднозначність зумовлюється типом синтаксичних відношень між варіантами додаток/означення за умови, що головним словом конструкції є іменник з дієслівною семантикою: *конфлікт із застосуванням сили, переговорного процесу між сторонами, злочинів проти миру і людства*. Звісно, варіант кінцевого аналізу залежить і від синтаксичного типу речення та від функції сегмента у речення (наприклад, визначена група підмета). Так, у заголовку така конструкція може отримати позначку “атрибутивне”: *контроль за ядерними озброєннями*. Визначення неоднозначності прийменникових словосполучень залежить здебільшого від лексичного значення головного слова

прийменникової конструкції та від лексичного значення прийменника. Обмеження, що можна накласти вже на рівні поверхнево-синтаксичного аналізу, зумовлені семантичною характеристикою одного із компонентів словосполучення, яку можна отримати завдяки модулю словотвірного аналізу. Наприклад, якщо головне слово прийменникової конструкції є віддієслівним іменником, для неї приписуємо значення – “об’єктні відношення”: *сприяння в наданні, нагляду за дотриманням, зв’язків з громадськістю, зарахування до списків, підготовка до роботи, розгортання у бойові порядки, зіткненні з перешкодою.*

Значну частку аналізованих конструкцій становлять усталені словосполучення із характерним для текстів військової тематики значенням. Такі текстові структури із визначеним типом синтаксичних відношень у межах змодельованого тематичного корпусу можна подавати заздалегідь укладеним списком: *право на носіння, доповідей про хід бойових дій, інформацію про повітряну обстановку, дані про температуру.*

Більшість випадків неоднозначних прийменникових конструкцій отримують правильну синтаксичну ідентифікацію на семантичному рівні лінгвістичного аналізу за семантичною моделлю “частина/ціле”: *затворна рама з газовим поршнем, газова трубка зі ствольною накладкою.*

Поширеними серед синтагм із прийменниковими конструкціями є обставинні відношення, зокрема способу дії та місця дії, в яких головним словом є віддієслівний іменник: *пересування на лижах, бій на великих відстанях, досягнення успіху у наступі, забезпечення безпеки в Україні.* У групі присудка в інших прийменникових структурах обставинні відношення способу дії можна визначити у поєднанні з об’єктними відношеннями: *за масштабами виділяють, поряд з традиційними джерелами безпеки виділяють, утворені відповідно до законів України.*

У досліджуваному матеріалі головним словом прийменникового словосполучення вважається перше ліворуч. Проте у лінійній структурі речення часто залишається не з’ясованим питання вибору головного слова прийменникової конструкції з-поміж

декількох, наприклад у реченнях: *(придатності) (громадян) України до професійної військової служби; (резерву) (кандидатів) для проходження служби; (сорочку) кольору (полину) з погонами; (придатні) до проходження військової (служби) за станом здоров'я та віком; (приймання) військовослужбовців (командуванням) з особистих питань.*

Для синтаксично неоднозначних синтагм характерною є здебільшого варіантна визначеність головного слова: *(застосовується) для (ураження) (противника) в ближньому бою; що (працює) в ультрафіолетовому (діапазоні) з підвищеною стійкістю до теплових пасток; Гармата Т-84 (оснащена) ефективною (стабілізацією) у двох площинах; (використовувати) (показання) (полонених) в інтересах рішення бойових задач.*

Отримані дані дозволяють зробити висновок, що навіть незначний опрацьований матеріал свідчить про можливість застосування у визначенні синтаксичної неоднозначності ймовірного критерію (фільтру). Так, у синтагмах із двома претендентами на роль хазяїна найбільш достовірним є перший ліворуч від омонімічного гнізда.

Особливої уваги в автоматичному синтаксичному аналізі потребують конструкції із однорідними членами речення. Пояснюється це як невизначеністю єдиного підходу формалізації цих зв'язків між членами конструкцій, так і лінгвістичною природою цих конструкцій [Дрейзин 1966; Дрейзин 1966; Иорданская 1967; Санников 1989 : 32–79]. Дослідники визначають декілька типів синтаксичної неоднозначності однорідних членів: 1) слово залежить від одного або від декількох однорідних; 2) омонімія за протиставленням однорідність/немає однорідності; 3) омонімія типу “з чим однорідне”.

У проаналізованому нами матеріалі незначну частку становлять синтаксично неоднозначні синтагми із сурядним зв'язком (19 випадків). Неоднозначність таких конструкцій полягає у визначенні головного та залежних слів, наприклад:

- *організація (взаємодії) та (управління) ЗСУ у мирний та воєнний час* (однорідними можуть бути *організація* та *управління*).

ня або взаємодії та управління; неоднозначним є визначення головного слова для прийменникового словосполучення – ЗСУ, організація та управління, взаємодії та управління);

- складається з (**парів**) (розпечених речовин) ядерного боєприпасу, **повітря і часток ґрунту** (визначення однорідності: складається з парів, повітря, і часток...чи складається із з парів ... боєприпасу, повітря, і часток ...);

- **зв'язок механізованих і танкових підрозділів з артилерією** (неоднозначність однорідних членів: що є однорідним механізованих та танкових залежать від підрозділів чи механізованих підрозділів та танкових підрозділів з артилерією; словосполучення з артилерією залежить від танкових підрозділів чи від зв'язок);

- **рухомого стержня і муфти** (чи є слово рухомого залежним і від муфти). Остаточне визначення такого типу омонімії вимагає більш ґрунтовного дослідження.

Отже, аналіз речень на рівні поверхнево-синтаксичного аналізу демонструє таку властивість тексту, як синтаксично неоднозначні конструкції на рівні словосполучення, синтагми, речення. Простішими щодо однозначного визначення синтаксичних відношень виявились синтаксично неоднозначні словосполучення, для яких вже на етапі поверхнево-синтаксичного аналізу можна застосувати певні фільтри розмежування. Інші типи потребують додаткового опрацювання, зокрема на рівні семантичного аналізу. Одними із пріоритетних завдань є визначення максимально повного списку властивих українській мові типів синтаксичної неоднозначності та розроблення семантичного аналізатора речень.

1. Бугаков О. В. Функціонування прийменників в українському тексті: морфологічний та семантико-синтаксичний аспекти / Автореф. дис. канд. філол. наук / О. В. Бугаков. – К., 2006. 2. Гладкий А. В. Синтаксические структуры естественного языка в автоматизированных системах общения / А. В. Гладкий. – М., 1985. 3. Грязнухина Т. А. Анализ предложных связей в научном тексте / Т. А. Грязнухина. – К., 1985. 4. Дрейзин Ф. А. Частота появления основных видов синтаксической омонимии в русских текстах / Ф. А. Дрейзин // НТИ. – № 12. – 1966. – С. 55–59. 5. Дрейзин Ф. А. Зависимые слова и группы

слов при однородных существительных с точки зрения автоматического синтаксического анализа (на материале русского языка) / Ф. А. Дрейзин // НТИ. – № 7. – 1966. – С. 43–45. 6. Замаруєва І. В. Комп'ютерна модель розуміння природно-мовної текстової інформації / І. В. Замаруєва // Проблеми програмування. – 1999. – № 2. – С. 96–102. 7. Иорданская Л. Н. Свойства правильной синтаксической структуры и алгоритм ее обнаружения (на материале русского языка) / Л. Н. Иорданская // Проблемы кибернетики. – Вып. 11. – М., 1964. – С. 215–243. 8. Иорданская Л. Н. Синтаксическая омонимия в русском языке (с точки зрения автоматического анализа и синтеза) / Л. Н. Иорданская // НТИ. – № 5. – 1967. – С. 9–17. 9. Колесников Н. П. Омонимия в предложении и вопросы ее устранения / Н. П. Колесников. – М., 1976. 10. Мельчук И. А. Опыт теории лингвистических моделей “Смысл-Текст”. Семантика, синтаксис / И. А. Мельчук. – М., 1974. 11. Санников В. З. Русские сочинительные конструкции. Семантика. Прагматика. Синтаксис / В. З. Санников. – М., 1989. 12. Толубко В. В. Задачі автоматичної обробки синтаксичної структури в знання-орієнтованій системі машинного перекладу / В. В. Толубко, О. О. Шипнівська, А. В. Ляшенко // Вісник Київського нац. ун-ту ім. Т. Шевченка. Серія: Військово-спеціальні науки. – Вип. 27. – 2010. – С. 136–140. 13. Севбо И. И. Графическое представление синтаксических структур и стилистическая диагностика / И. И. Севбо. – К., 1981. 14. Чередниченко І. Г. Принципи розрізнення неузгоджених означень і присубстантивних додатків, виражених формою родового відмінка без прийменників / І. Г. Чередниченко // Українська мова. – 1955. – № 2. – С. 33–38.

Рассматриваются теоретико-практические основания определения синтаксически неоднозначных конструкций в автоматическом синтаксическом анализе. На материале текстов военной тематики проанализированы основные типы синтаксической омонимии.

Ключевые слова: автоматический синтаксический анализ, знаниеориентированная система машинного перевода, синтаксическая неоднозначность, типы синтаксических омонимов, синтаксическая структура.

In the paper the theoretical and practical basis of the syntactical ambiguity for the automatic syntactical analysis is described. The description of main types of syntactical homonyms for the automatic syntactical analysis in military texts is presented.

Key words: automatic syntactical analysis, knowledge-based machine translation system, syntactical ambiguity, types of syntactical homonyms, syntactical structure.

Стаття надійшла до редакції 25.09.2012