

22 червня 2020 року доктор філологічних наук, професор кафедри української мови та прикладної лінгвістики Київського національного університету імені Тараса Шевченка Наталія Петрівна Дарчук відзначила 75-річчя від дня народження. Упродовж багатьох років життєва й наукова доля вельмишановної Ювілярки пов'язана з відділом структурно-математичної лінгвістики. У цьому колективі вона почала свій шлях у науці, сформувалася як високоавторитетний фахівець, знаний і шанований в Україні та за кордоном. Пропонована стаття — свідчення її невтомної багаторічної праці над розбудовою першого в Україні Корпусу текстів української мови (www.mova.info), який вона створила.

Редколегія журналу «Українська мова», колеги, друзі та учні щиро вітають Наталію Петрівну Дарчук і зичать міцного здоров'я і творчої наснаги для подальшої плідної праці на ниві українського мовознавства.

<https://doi.org/10.15407/ukrmova2020.04.089>

УДК 811.161.2'42:004

Н.П. ДАРЧУК, доктор філологічних наук,
професор кафедри української мови та прикладної лінгвістики,
Київський національний університет імені Тараса Шевченка
бул. Тараса Шевченка, 14, Київ, 01601
E-mail: nataliadarchuk@gmail.com
<https://orcid.org/0000-0001-8932-9301>

АГАТ-СИНТАКСИС ЯК КОМПОНЕНТ КОМП'ЮТЕРНОЇ ГРАМАТИКИ УКРАЇНСЬКОЇ МОВИ

Статтю присвячено описові синтаксичного модуля комп'ютерної граматики української мови АГАТ. Простежено, як унаслідок роботи правил АГАТ-граматики лінійна морфологічна структура речення перетворюється на двомірну деревоподібну синтаксичну структуру із приписуванням типу синтаксичного зв'язку і типу синтаксичного відношення. Це дало змогу автоматично побудувати алфавітно-частотні словники сполучуваності всіх частин мови.

Ключові слова: автоматичний синтаксичний аналіз, дерево залежностей, синтаксичні зв'язки, синтаксичні відношення.

Комп'ютерна граMATика АГАТ (Автоматичний граматичний аналіз тексту) — це інтегральна комп'ютерна модель української мови, у якій онтологічна система працює у взаємодоповнювальному режимі до гносеологічного аспекту. Її модель — чинний автомат аналізу тексту — ієрархічно розв'язує всі необхідні завдання аналогічно до людини-лінгвіста, але робить це за правилами комп'ютерної граматики, яка складається за об'єктами опису з двох розділів — морфології і синтаксису.

Першою причиною до розширення АГАТ синтаксичним модулем було завершення морфологічного аналізу — зняття лексико-граматичної і граматичної омонімії — через процедури синтаксичного аналізу, коли встановлюються зв'язки слів. Друга причина — теоретична необхідність

Цитування: Дарчук Н.П. АГАТ-синтаксис як компонент комп'ютерної граматики української мови. *Українська мова*. 2020. № 4 (76). С. 89–102. <https://doi.org/10.15407/ukrmova2020.04.089>

у вивченні сполучуваності лексичних одиниць, синтаксичних моделей речень тощо. Автоматичний синтаксис надає змогу каталогізувати синтаксичні одиниці, що є базою для розв'язання багатьох теоретичних проблем. Третя причина — це прикладні проблеми: автоматизація лінгвістичних досліджень, автоматичне визначення меж словосполучень, установлення критеріїв членування фрази на синтагми, автоматичний синтаксичний аналіз речення, автоматичне реферування й анотування тексту на основі сполучувальних критеріїв, машинний переклад, автоматичне редагування тощо [2].

Є кілька різновидів синтаксичних моделей, використовуваних під час автоматичного опрацювання текстів (АОТ): граматики безпосередніх складників (БС) (ланцюжкова граматика); граматики залежностей (граматика Л. Теньєра, моделі традиційної граматики за О.М. Пешковським; розмічені граматики залежностей «Смисл — Текст»); теорії синтаксичних груп О.В. Гладкого. Під час побудови робочої системи АОТ доводиться їх сполучати, оскільки в кожній з них є свої переваги та недоліки [6: 82]. Наприклад, традиційна граматика членів речення визначає глобальну ієрархію одиниць (синтаксичних груп) у складі висловлення — підмет—присудок—додаток; граматику БС зручно використовувати на початкових стадіях синтаксичного аналізу, оскільки за її правилами формуються складові одиниці — синтаксичні групи. Граматика залежностей унаочнює ієрархію одиниць, які є основою подальшого підрахунку інформаційної ваги одиниць семантичного рівня (семантичних вузлів), а за граматику Л. Теньєра збираються групи як заготовки до побудови «середніх» одиниць семантичних структур — багатомісних предикатів, фреймів [6: 83].

Створюючи синтаксичний модуль АОТ, як і інші модулі, наразі застосовують два основних підходи: перший заснований на правилах (rule-based), другий — інженерний, який названо «машинним навчанням» (machine learning). Перший є лінгвістичним, оскільки представляє лінгвістичну інформацію як формальні правила, іноді вбудовані у програмний код, або спеціально створену формальну мову. Правила формулюють самі лінгвісти. У межах підходу, який базується на машинному навчанні, джерелом лінгвістичної інформації слугують не правила, а відібрані тексти проблемної галузі. Серед методів, застосовуваних за цього підходу, є навчання з учителем (supervised), часткове навчання без учителя (unsupervised) та часткове навчання з учителем (bootstrapping). Здебільшого застосовують навчання з учителем, коли будують математичну та програмну модель машинного класифікатора, що вміє розпізнавати різні класи одиниць тексту (слів, словосполучень тощо) або самі тексти. Побудову класифікатора здійснюємо на спеціально розміченому текстовому корпусі (навчальна вибірка), у якому аналізованим одиницям тексту надано мітки — коди певної граматичної чи семантичної інформації. Навчання базовано на загальних закономірностях, властивих текстам природної мови на основі даних навчальної вибірки, отже, знання

декларативні (правила) сполучені зі знаннями процедурними (машинне навчання).

Обидва методи мають переваги і недоліки. Створення правил — процес трудомісткий, але глибоко лінгвістичний, що враховує навіть часткові складні випадки, яких досить багато в різностильових текстах. Правила, звісно, декларативні, легко зрозумілі, їх неважко модифікувати залежно від результатів роботи модуля. Машинне навчання не потребує ручної праці з укладання правил, скорочує час на розроблення систем, але класифікатори непрозорі для розуміння, оскільки не мають лінгвістичної інтерпретації. Машинне навчання передбачає наявність розміченого корпусу текстів, що потребує значних обсягів ручної праці. І що більше буде розмічених текстів у Корпусі, то кращими будуть результати його синтаксичного аналізу [7].

У межах АГАТ створювався глибоко лінгвістичний тип автоматичного синтаксичного аналізу (АСА), за допомогою якого можна одержати різноманітну інформацію про функціонування синтаксичних одиниць та їхніх категорій, зокрема формально-синтаксичних: предикативність, підрядність, сурядність, а також категорій підмета, присудка, другорядного члена речення тощо. Розроблено лінгвістичне та програмне забезпечення, завдяки якому на *першому етапі* здійснювано синтаксичний аналіз цілого речення з його результатом у формі дерева залежностей із подальшою можливістю його редагування. Утім, створення аналізатора, який безпомилково здійснював би розпізнавання українського тексту, узагалі неможливе, тому якісне анотування тексту завжди пов'язане з його ручним доопрацюванням. На *другому етапі* АСА накопичення правильно розмічених дерев залежностей становитиме вихідний матеріал для машинного навчання за допомогою векторного аналізу, що лежить у його основі. Одержана на першому етапі узагальнена репрезентація сполучуваності кожного слова може бути екстрапольована на нові контексти різних дискурсів Корпусу текстів української мови, після чого може бути обчислена ймовірність появи аналізованого слова або словосполучки.

Зупинимось коротко на особливостях першого етапу роботи системи АГАТ-синтаксису.

1) АГАТ-синтаксис — це сукупність операцій, виконуваних комп'ютером над послідовностями інформації морфологічного характеру (результатом роботи модуля автоматичного морфологічного аналізу (АМА)), що представляють вхідний текст для встановлення синтаксичних зв'язків між його одиницями;

2) завданням АГАТ-синтаксису на цьому етапі є виявлення всіх різновидів сполучуваності — **предикативної, підрядної і сурядної** — кожного слова в тексті;

3) граматичні характеристики словосполучення безпосередньо залежать від того, до якої частини мови належить його головне слово. Як відомо, лексико-граматична природа слова визначає його здатність сполучатися

з іншими словами. Відповідно до цього словосполучення поділяють на **іменникові, прикметникові, займенникові, числівникові, дієслівні та прислівникові**. За нашою концепцією АСА, виокремлення словосполучень базовано на граматичній валентності, а саме: на підграматичній для дієслова (31206 правил) ([2] Додаток Б.1), підграматичній для іменника (40023 правил) (Додаток Б.4), підграматичній для ад'єктива (6205 правил) (Додаток Б.6), а також на словнику фразеологізмів (близько 27200 одиниць) (Додаток Б.8). Комп'ютерні підграматичні валентностей зазначених частин мови ми побудували за єдиним принципом: указано лексему, прийменник, який бере участь у керуванні, та відмінок керованої, залежної іменникової словоформи у вигляді двобуквенного коду, що позначає її частиномовну належність (граматичний клас) та належність до категорії в межах такої частини мови (граматичний підклас);

4) у граматичній теорії за складом словосполучення поділяють на **прості, складні та комбіновані** [5]. Ми розглядаємо тільки прості бінарні словосполучення, що можуть бути поширені в ускладнені, бо для визначення їхнього складу потрібен аналіз змістової структури. Автоматичний режим аналізу на цьому етапі усуває звернення до семантики слів. Оскільки в процесі аналізу зберігається номер кожної словоформи в базі даних (БД), можна подавати на екран ускладнене словосполучення, «склеєне» з елементарних, простих;

5) для кожного слова визначено такі зв'язки: **підрядні, предикативні і сурядні**, оскільки вони відповідають відтворенню загальної системи відношень між компонентами ситуації, описуваної в реченні. Ми відмовилися від традиційних видів підрядного зв'язку — узгодження, керування та прилягання, ґрунтуючись на широкому розумінні поняття синтаксичного зв'язку: це будь-які поєднання з головним словом прислівника або відмінкової форми іменника, прикметника;

6) у комп'ютерній граматичній підрядні зв'язки поділено на **ядрові і неядрові**. **Ядровим** називаємо такий зв'язок, за якого аналізоване слово є керувальним, головним. **Неядровим** вважаємо зв'язок, коли аналізоване слово є залежним, керованим. **Предикативний, або координаційний**, зв'язок устанавлюємо між основними компонентами речення «підмет — присудок». Його ґрунтовано на двобічній залежності головних членів речення, тобто на їх взаємозалежності. **Сурядний** зв'язок наявний між словами, що не є ані домінувальними, ані домінованими одне щодо одного. Уважаємо, що два слова перебувають у сурядному зв'язку, якщо кожне з них підпорядковано тому ж третьому слову або якщо вони пов'язані між собою сполучником чи відокремлені один від одного комою. Формальною підставою для виявлення сурядного ряду є перевірка кодів аналізованої пари слів за спеціальною таблицею сурядності;

7) під час автоматичного укладання словника словосполучень на будь-якому тексті КУМ (Корпусу текстів української мови) одержуємо чотири типи моделей: **ядрові, неядрові** (ад'юнктивні, що відображають підрядні зв'язки), **сурядні, предикативні** (докладніше про це в [3]).

<<	3	>>
Так (Н0) одвіку (Н0) було (ГР) й (СС) так (Н0) воно (МС) буде (ГД) довіку (Н0) , () поки (Н0) житимуть (ГЗ) люди (ІА) і (СС) поки (СП) ростимуть (ГЗ) терни (ІА) . ()		
було	▼ Так	дієслівна безприменникова сполука
було	▼ одвіку	дієслівна безприменникова сполука
було	▼ й	Сурядний зв'язок у складному реченні
й	▼ буде	Сурядний зв'язок у складному реченні
буде	▼ воно	координаційний зв'язок, сполука підмета і присудка
буде	▼ довіку	дієслівна безприменникова сполука
буде	▼ так	дієслівна безприменникова сполука
буде	▼ житимуть	зв. пред. част. через сполучне слово
житимуть	▼ люди	координаційний зв'язок, сполука підмета і присудка
житимуть	▼ поки	дієслівна безприменникова сполука
житимуть	▼ і	Сурядний зв'язок у складному реченні
і	▼ ростимуть	Сурядний зв'язок у складному реченні
ростимуть	▼ терни	координаційний зв'язок, сполука підмета і присудка
ростимуть	▼ поки	дієслівна безприменникова сполука
Зберегти		

Рис. 1. Бінарні сполуки і типи синтаксичних зв'язків у поезії Лесі Українки «Завжди терновий вінець...» із циклу «Невольницькі пісні» (реч. 3)

На рис. 1 представлено речення з поезії Лесі Українки «Завжди терновий вінець...» із циклу «Невольничі пісні» і таблицю підпорядкування, яку програма будує автоматично. Над робочою таблицею вміщено речення з морфологічною анотацією, інформацією про частиномовні й категорійні характеристики слів. У таблиці три стовпчики: у першому — головний член бінарної сполуки («хазяїн»), у другому — підпорядкований член сполуки («слуга»), а в третьому — синтаксична інформація про тип сполуки. Це вможливило після завершення роботи програми утворення алфавітно-частотних словників словосполучень вірша, циклу або творів автора загалом (рис. 2).

АЛФАВІТНО-ЧАСТОТНИЙ СЛОВНИК ЛЕКСИКИ ПРИЖИТТЄВИХ ВИДАНЬ ЛЕСІ УКРАЇНКИ ("НА КРИЛАХ ПІСЕНЬ", "ДУМИ І МРІЇ", "ВІДГУКИ", ПОЕМИ "ДАВНЯ КАЗКА", "РОБЕРТ БРЮС"). - СЛОВОСПОЛУЧЕННЯ

ініціал пошуку

мова

словосполучення

Залежне слово

Частина мови:

Тип зв'язку:

пошук словосполучень

любити

(початкова форма)

Всі

Всі

Шукати

Слово	Типи зв'язків	Словосполучення
любити	2	3

зв'язок	Слово	Частота
дієслівна безприменникова сполука	любити	ГР 4
координаційний зв'язок	любити	ГР 2
координаційний зв'язок	любити	ГЯ 3
дієслівна безприменникова сполука	любити	ГФ 2
координаційний зв'язок	любити	ГЙ 1
дієслівна безприменникова сполука	любити	ГН 2
дієслівна безприменникова сполука	любити	ГЯ 3
дієслівна безприменникова сполука	любити	ГЙ 2
координаційний зв'язок	любити	ГН 1

Рис. 2. Інтерфейс програми зі створення словника словосполучень із зазначенням типу синтаксичного зв'язку між їхніми складниками

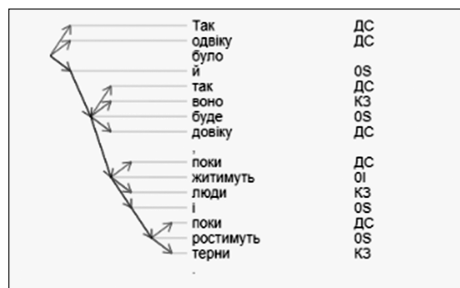


Рис. 3. Дерево залежностей у поезії Лесі Українки «Завжди терновий вінець...» із циклу «Невольницькі пісні» (реч. 3)

За правилами цієї програми побудовано алфавітно-частотні словники для конкретних лексем, класів слів з показником частоти їхнього вживання, необхідного ілюстративного контексту.

На рис. 3 розміщено дерево залежностей, що є наслідком автоматичного інвертування таблиці сполучуваності, це його графічний вигляд. Останній стовпчик справа на рисунку — двійковий код усіх синтаксичних зв'язків.

Дерева залежностей складаються з двох елементів — вузлів і дуг. Вузли — словоформи, а дуги — відношення між вершиною («хазяїном») і залежним словом («служогою») сполуки. Це дає змогу описати конфігурацію, форму, зовнішні параметри речення, але цього недостатньо для опису структури речення. У термінах граматики залежностей сполуки *її принесли*, *їй принесли* ідентичні з погляду зображення в дереві (стрілка від *принесли*) і лексично, але розбіжність є в будові конструкції (що принесли? і кому принесли?). Множині дуг дерева автоматично надано інформацію про тип зв'язку між складниками сполуки, тобто дерево залежностей становить розмічене дерево підпорядкування, але цієї інформації недостатньо. Варті на увагу семантико-синтаксичні відношення (ССВ) між словами, які можуть бути відображені в ієрархічно впорядкованій структурі — дереві залежностей.

Для побудови системи автоматичного визначення ССВ треба спочатку уточнити їхні типи. У сучасній теоретичній граматиці їх кваліфікують як:

1) семантичні відношення — граматичні значення словосполучення, які встановлюють між окремими членами речення та які репрезентують синтаксичне **значення** членів речення;

2) семантико-синтаксичні відношення, які, відображаючи реальні відношення між предметами та явищами позамовної дійсності, є об'єктивними (тому **семантико-**), а вказуючи на функціональні відношення між компонентами структури речення, є власне мовними, структурними (синтаксичними) і становлять синтаксичне значення членів речення (тому **синтаксичні**) [1: 59–60].

ССВ пов'язані з поверхнево-синтаксичним аналізом речення; уводимо їх як спеціально сконструйовані одиниці проміжної природи для спрощення опису складних відповідностей між змістовими відношення-

ми та формальним способом їх вираження, оскільки розмічування тексту здійснювано автоматично з оперттям на формальні ознаки його одиниць. Отже, автоматичний синтаксичний аналіз речення має два рівні: **перший** — приписування кожній парі слів типу синтаксичного зв'язку на рівні морфологічного способу вираження «хазяїна»; **другий** — кожній дузі тип синтаксичного відношення.

Надамо спеціальні пояснення до кожного типу ССВ.

Суб'єктне — відношення між підметом і присудком, які утворюють змістове ядро речення.

Об'єктне — відношення предмета до ознаки, які поділено на **прямо-об'єктні** (залежне слово — «слуга» — у знахідному відмінку) і **непрямо-об'єктні** («слуга» в інших відмінках).

Атрибутивне — відношення між компонентами словосполучення, коли один із них виражає ознаку іншого.

Обставинне — відношення обставин до дії, яку ці обставини характеризують.

Комплетивне — відношення між компонентами цілісно вираженого складеного члена речення, а не між членами речення.

Апозитивне — відношення між прикладкою і означуваним нею головним компонентом словосполучення. Перераховані типи ССВ характеризують семантико-синтаксичну організацію словосполучення на найвищому рівні узагальнення. Їх уточнювали за правилами алгоритму, який потрібно читати так: (див. нижче) у фігурних дужках код типу синтаксичного зв'язку, напр., АС — код ад'єктивної безприменникової сполуки, поруч ff7800 — її програмний код. Шляхом перебору всіх типів зв'язку й збігу з пропонуваними умовами сполуці автоматично приписується певний тип синтаксичного відношення між її складниками.

Якщо {"АС", "ff7800", "ад'єктивна безприменникова сполука"}, то

1) перевірка на наявність «слуги» — прислівника (*величний тоді*), тоді **обставинне**, якщо ні, то

2) **об'єктне** (напр., *пов'язаний ниткою, сильний духом*).

Якщо {"АП", "ff7800", "ад'єктивна применникова сполука"}, то **об'єктно-применникове** (напр., *пов'язаний з друзями*).

Якщо {"ДС", "0090ff", "дієслівна безприменникова сполука", "говорити мовою"}:

1) причому залежне у знахідному відмінку (коди: ЙВЙУКВКУЛВ-ЛУИУМВМЛМХЧУМУ), то **прямо-об'єктне** (напр., *бачу хлопчика*);

2) якщо інші відмінки, то **непрямо-об'єктне** (*сказав мені*);

3) якщо дієслово-зв'язка (*бути, стати / ставати, робитися / зводитися, вважатися, називатися / назватися, іменуватися, почати / почина-ти, продовжувати / продовжити, кінчати / кінчити, припиняти / припинити, могли, вміти, мати, мусити, хотіти, бажати, погодитися*), то

а) перевірка залежного слова на наявність орудного відмінка (ЙТКТЛТЙЮКЮЛЮИЮАТАМАЦАЮ), **призв'язкове** (напр., *стане керівником, був хворим*);

б) якщо залежне слово прислівник (код Н0) (напр., *буде завжди*), то **обставинне**;

в) якщо залежне слово інфінітив (коди ГФГП) (напр., *вміти говорити*), то **об'єктне**;

4) решта **прямо-об'єктне** (напр., *насіпати борщу, налити чаю*).

Якщо {"ДП", "0090ff", "дієслівна прийменникова сполука", "говорити з друзями"}, **об'єктно-прийменникове**.

Якщо {"ІС", "ba00ff", "іменникова безприйменникова сполука", "заняття мовою"},

1) а залежне слово прикметник (код А), то атрибутивне;

2) якщо залежне слово має код родового відмінка (ЙРКРЛРЙЕКЕЛЕ),

а) при цьому головне слово позначає предмет (код r0concr), а залежне — особу (код t0hum), то **посесивне** (напр., *зошит учня*);

б) або якщо головне слово позначає частину (код pt0part), то **атрибутивне** (напр., *дах будинку, сторінка зошита*);

в) або якщо головне слово має код особи (код t0hum), то **атрибутивне** (*викладач вишу*);

г) або якщо головне слово позначає сукупність (код pt0aggr), то **атрибутивне** (напр., *стадо корів, натовп людей*);

г) або якщо головне слово має значенням опредметненої дії (код der0v), а залежне позначає діяча (код t0hum), то **посесивне** (напр., *наказ директора*);

д) або якщо головне слово називає опредметнену ознаку (код der0a), а залежне — особу (код t0hum) чи предмет (код r0concr), то **атрибутивне** (*білизна снігу, цікавість учнів*);

3) якщо залежне слово іменник (перша літера код: Й/К/Л/И в будь-якому відмінку) **об'єктне** (напр., *прагнення перемоги*);

4) якщо залежне слово інфінітив (коди ГФГП) (напр., *необхідність пояснити*), то **атрибутивне**;

5) якщо залежне слово прислівник (код Н0) (напр., *прогулянка восени*), то **обставинне**.

Якщо {"ІП", "ba00ff", "іменникова прийменникова сполука", "заняття з друзями"}, то **об'єктно-прийменникове**.

Якщо {"РС", "a60800", "прислівникова безприйменникова сполука", "дуже сильно"}, то **обставинне**.

Якщо {"РР", "a60800", "прислівникова прийменникова сполука", "далеко від неї"}, то **обставинне**.

Якщо {"ЗС", "ffcc00", "займенникова прийменникова сполука", "він з нею"}, то **комплетивне**.

Якщо {"ЗБ", "ffcc00", "займенникова безприйменникова сполука", "він нею"}, **комплетивне**.

Якщо {"ЧС", "22d6d4", "числівникова безприйменникова сполука", "двісті них"}, то **комплетивне**.

Якщо {"ЧП", "22d6d4", "числівникова прийменникова сполука", "двісті з них"}, то **комплетивне**.

Якщо {"КЗ", "ff0000", "координаційний зв'язок, сполука підмета і присудка", ""}, то **суб'єктне**.

Якщо {"ПМ", "b75c03", "сполука з цифрою", "200 осіб"}, то **комплетивне**.

Якщо {"СУ", "22d626", "сурядна сполука", "яблук і груш"}, то **однорідно-сурядне**:

- 1) єднально-сурядне (код s0cnj0sur0jed);
- 2) протиставно-сурядне (код s0cnj0sur0prot);
- 3) зіставно-сурядне (код s0cnj0sur0zist);
- 4) пояснювально-сурядне (код s0cnj0sur0poj);
- 5) приєднувально-сурядне (код s0cnj0sur0pryj);
- 6) розділово-сурядне (коds0cnj0sur0rozd).

Якщо {"ЧК", "029d90", "сполука з часткою", "не читаю"}, то

- 1) при цьому, частки *тільки не, ні*, то **обмежувальне**;
- 2) решта — **дискурсивні**.

Якщо {"СО", "029d90", "однорідна сполука", "яблук, груш"}, то **однорідно-сурядне**.

Якщо {"ПП", "0012ff", "прийменникова сполука", "на столі"}, то **відприйменникове**.

{"17", "f9d7ab", "прикладка", ""}, **апозитивне** (напр., *трикутник ABC*).

{"0V", "E2EE2E", "приєднання вставних слів", ""}, **модальне**.

{"0D", "E2EE2E", "дієприслівниковий зворот", ""}, **обставинне**.

{"0P", "E2EE2E", "дієприкметниковий зворот", ""}, **означальне**.

Є словосполучення, які поєднують кілька значень, напр., атрибутивні й обставинні (*мандрівка восени*), атрибутивні й об'єктні (*думи про тебе, написання книжки*). Їм приписується одне відношення: *мандрівка восени* — обставинне; *думи про тебе* — об'єктно-прийменникове, *написання книжки* — об'єктне [5], тобто за формальною ознакою — морфологічним кодом класу залежного слова — «слуги».

ССВ встановлювано не тільки в підрядних словосполученнях, а й у сурядних. Основними в сурядних словосполученнях є такі типи семантико-синтаксичних відношень:

- 1) єднальні відношення — значення переліку;
- 2) протиставні відношення, що уявляють значення протиставлення (*суворий, але добрий*);
- 3) розділові відношення, що вказують на несумісність, чергування дій, явищ, ознак тощо (напр., *сніг або дощ, то сніг, то дощ*);
- 4) градаційні відношення, які увиразнюють змістову вагу одного з однорідних членів речення, напр., *не тільки батьки, але й діти*.

Формальним виразником сурядності є сполучник, хоч ми розуміємо, що цей виразник може мати розгалужений функціонально-семантичний діапазон (напр., можуть бути *і...і* з часово-єднальним або причиново-наслідковим відношенням) і под. [5: 48].

Майже всі правила базовано на типах синтаксичного зв'язку, але є блок правил іменникового керування, до якого залучено інформацію семантичного характеру, тому що семантика компонента словосполучення

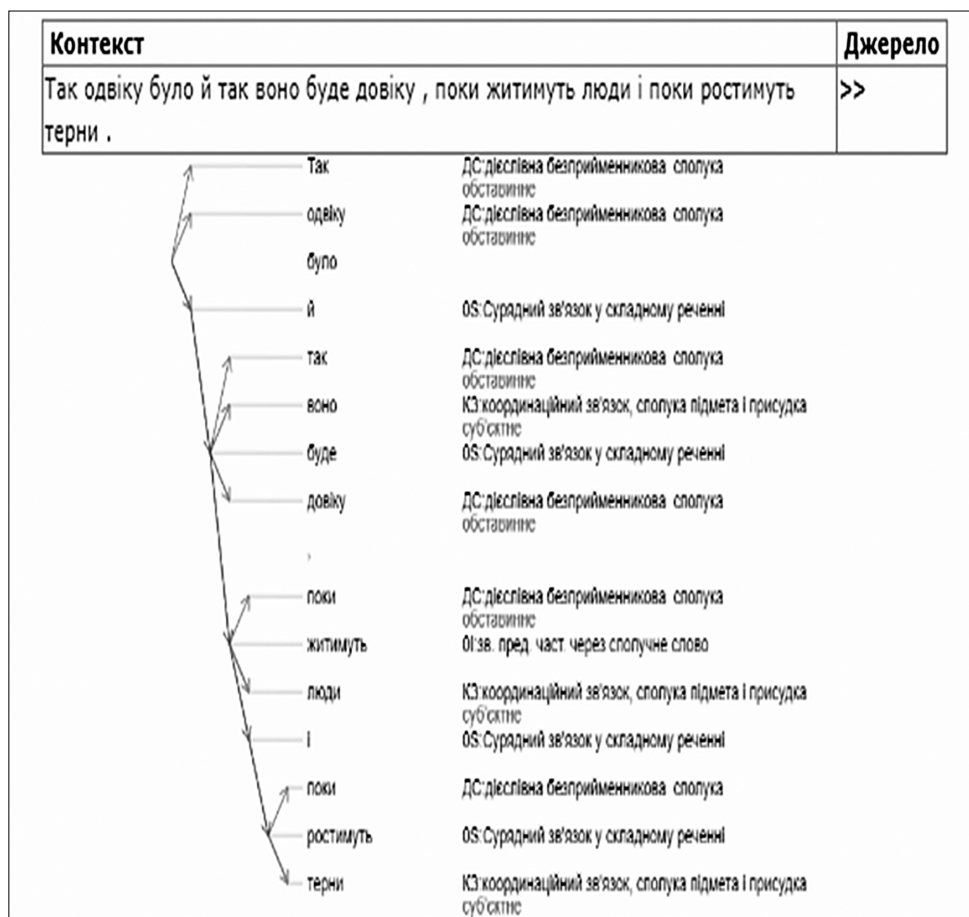


Рис. 4. Дерево залежностей речення поезії Лесі Українки «Завжди терновий вінець...» із циклу «Невольницькі пісні» (реч. 3) із синтаксичними зв'язками та синтаксичними відношеннями

взаємодіє із семантикою відношень. Ми протегували лексико-семантичні варіанти всіх лексем, які входять до лексику української мови — їм приписано коди доменів семантичної класифікації. Звертання до цієї інформації допомагає уточнити ССВ. Деякі відношення (модальне, дискурсивне) уведено для подальшого дослідження текстових зв'язків і розроблення відповідних лінгвістичних і програмних засобів аналізу цілого тексту.

На рис. 4 представлено результати автоматичного встановлення типу семантичного відношення, яке записується під типом синтаксичного зв'язку. За єдиним алгоритмом опрацьовано всі речення тексту.

Отже, повний цикл автоматичного синтаксичного аналізу українськомовного тексту завершено: встановлено синтаксичні словосполучення, їм надано автоматично тип синтаксичного зв'язку і тип синтаксичного відношення. Можливість каталогізування їх створює умови для системного вивчення семантико-синтаксичних проблем: граматична і

<< 7 >>

Зложили (ГН) всі (МА) обітницю (КВ) врочисту (АЛ) ховати (ГФ) таємницю (КВ) до (ПР) загину (ЙР) . ()

Зложили	всі	координаційний зв'язок, сполука підмета і присудка
Зложили	обітницю	дієслівна безприйменникова сполука
обітницю	врочисту	іменникова безприйменникова сполука
обітницю	ховати	іменникова безприйменникова сполука
ховати	до	дієслівна прийменникова сполука
ховати	таємницю	дієслівна безприйменникова сполука
до	загину	прийменникова сполука

Зберегти

Editor: darchuk
08.07.2019 10:39:16
Status: OK+
stats статистичні
параметри речення

Словосполучення на базі синтаксису

Виключення 0B|0F|0S|0I|0G|SI|RS|0L|C Зберегти

Зложили [КЗ] всі

Зложили [ДС] обітницю [ІС] врочисту

Зложили [ДС] обітницю [ІС] ховати [ДП] до [ПП] загину

Зложили [ДС] обітницю [ІС] ховати [ДС] таємницю

Рис. 5. Дерево залежностей речення вірша Лесі Українки «Віче» із циклу «Невольницькі пісні» (реч. 7)

лексична валентність слів, типова частиномовна сполучуваність, лексична і граматична валентність як критерій синонімічності, закони комбінаторики словосполучень різних типів і розрядів тощо. У прикладному аспекті система АГАТ-синтаксису буде вбудована в масштабний проєкт «Аналітико-консультативна автоматизована онтологічна система предметно-галузевого спрямування», виконуваний двома колективами лінгвістів і математиків.

Крім відношень між словами, у реченні спостерігаємо ще один важливіший вид упорядкованих відношень — відношення між групами слів, або словосполученнями. Для їх відображення потрібна формальна структура іншого типу — структура складників. Аналізуючи речення на рис. 5, інтуїтивно можемо поділити його на відрізки, що мають ієрархічну структуру, причому в деяких є спільна частина, тобто одна частина входить в іншу. Речення *Зложили всі обітницю врочисту ховати таємницю до загину* автоматично поділене на відрізки, які утворюють ієрархічну структуру: *зложили всі* / *зложили обітницю врочисту* / *зложили обітницю ховати до загину* / *Зложили обітницю ховати таємницю*. Якщо в них є спільна частина — один відрізок цілком входить в інший — формальною моделлю речення слугує система складників. У граматиці важливе значення мають два види відношень між складниками — **домінація** і **командування** [8: 105]. Складник *обітницю врочисту* поєднується з домінувальним *зложили* й утворює більший складник. *Загину* поєднується підпорядковувальним зв'язком з прийменником *до*, який визначає іменник *обітницю*. Складник *обітницю ховати до загину* — це теж складник, який включає не тільки окремі слова, а й вкладені один в одний «складні» складники, тобто

такі, які включають більше одного слова. Нарешті, складник *зложили обітницю ховати до загину*, який вбирає в себе всі перераховані одиниці, може бути визнаний цілим складним.

Отже, список безпосередніх складників, укладений за деревом залежностей, засвідчує, що синтаксичними одиницями в реченні є не тільки окремі слова, а й цілі сполуки слів, або **групи**: *[зложили всі]/ [обітницю врочисту]/ [ховати до загину]/ [обітницю [ховати таємницю]]*.

Звернімо увагу, що складники не можуть перетинатися, але можуть «вкладатися» один в одний. Це означає, якщо одне слово або множина слів охоплює одночасно два складники, то один з них цілком уходить у другий. Таким чином, можна, спираючись на інтуїтивні міркування, поділити речення на структурні синтаксичні одиниці — **складники**. Складниками є не тільки окремі слова, а й більші синтаксичні одиниці — **групи**. Далі передбачаємо більш або менш однозначно встановити, які множини таких структурних одиниць (складників) належать до того самого граматичного класу. Усі ці міркування підсумовано в структурі складників. Вона може бути зображена як розмічене дерево. Я.Г. Тестелець дає таке визначення системи складників: **системою складників** у S є така множина C відрізків S , яка містить в статусі елементів саме S і кожне слово, яке входить в S , причому будь-які два відрізки, що входять в C , не перетинаються, або один з них міститься в другому. Елементи C і є **складниками**. Якщо ланцюжок S є реченням, то C називається **системою складників речення S** [8: 117].

Ми не будемо заглиблюватися в типи відношень між складниками, структуру складників, способи вираження синтаксичних відношень тощо — це попереду. Звернімо увагу на інше, а саме: на змістову частину складників. І.Р. Вихованець зауважує, що дослідники семантики речення нерідко пов'язують змістову організацію речення з його формальною організацією, використовуючи поняття семантичної структури речення, кваліфікуючи її як зміст речення, поданий в узагальненому вигляді з урахуванням тих елементів значення, які накреслює форма речення [2: 121]. Проте об'єктивний зміст речення найкраще відбиває поняття **пропозиції**. Це стабільне ядро, константа речення, яка відображає структуру описуваної ситуації. Лінгвіст зазначає, що структуру пропозиції визначає предикат, який указує на характер ситуації, у нашому випадку — це корінь дерева, і на відповідні місця для предметів — учасників ситуації — це актанти, аргументи, репрезентовані групами присудка і підмета, якість і функції яких зумовлені предикатом. І тільки семантична природа предиката зумовлює кількість, ролі актантів. А відтак актуальними є два аспекти вивчення: семантика предикатних слів і семантичні ролі актантів. Якщо перший аспект — семантику ми вже опрацювали — це семантичні домени класів дієслів, предикативних прислівників, які мають номер семантичного класу, то семантичні ролі актантів ще належить вивчати.

Автоматично отримані складники можна розглядати як вихідний матеріал для формування n -місних предикатів, як первинний спосіб вираження їхнього змісту. Належність предиката до певного семантичного

класу визначає його семантику. Отже, ми можемо визначати способи вираження пропозицій (пропозиції руху, мовлення, звуку, ментальної сфери, емоцій тощо), кількість актантів та ін. На великому корпусному матеріалі формовано підкорпуси безпосередніх складників, закріплені за певними пропозиціями, з яких вирізняється такий, який найчастіше і найвиразніше оформляє саму пропозицію.

Працюючи в галузі комп'ютерного оброблення мовного матеріалу, розробляючи синтаксичний модуль автоматичного аналізу тексту, ми усвідомили величезне значення теорії та зважали на українські лінгвістичні традиції, закладаючи їх у правила комп'ютерного синтаксису для чіткого розуміння українськими граматистами. Водночас уважаємо, що йти шляхом накопичення окремих фактів безперспективно. Перед сучасною лінгвістикою стоїть завдання створити повний опис мови жорсткими науковими методами. Як зазначав Я.Г. Тестелець, треба шукати рівновагу між регулярністю і багатством синтаксичних структур [8: 15]. А шлях формального моделювання — перспективний шлях розв'язання багатьох лінгвістичних завдань.

ЛІТЕРАТУРА

1. Арібжанова І.М. Структура простого речення (формально-граматичний аспект). Навчальний посібник. Київ, 2001.
2. Вихованець І.Р. Граматика української мови. Київ, 1993.
3. Дарчук Н. Дерево залежностей (ДЗ): роль і місце в синтаксисі. *Матеріали IV міжнародної наукової конференції «Людина. Комп'ютер. Комунікація»*. 25–27 вересня 2019 р. Львів, 2019. С. 20–28.
4. Дарчук Н.П. Комп'ютерне анотування українського тексту: результати і перспективи. Київ, 2013.
5. Загнітко А.П. Основи українського теоретичного синтаксису. Ч. 1. Горлівка, 2004.
6. Леонтьева Н.Н. Автоматическое понимание текстов: системы, модели, ресурсы. Москва, 2006.
7. Маннинг К., Рагхаван П., Шютце Ч. Введение в информационный поиск. Москва, 2011.
8. Тестелец Я.Г. Введение в общий синтаксис. Москва, 2001.

REFERENCES

1. Aribzhanova, I.M. (2001). Structure of the simple sentence (formal-grammatical aspect). Kyiv (in Ukr.).
2. Vykhovanets, I.R. (1993). Grammar of the Ukrainian language. Kyiv (in Ukr.).
3. Darchuk, N.P. (2019). Dependences tree (DT): the role and the place in the syntax. *Papers of IV international scientific conference "Human being. Computer. Communication"*. P. 20–28. Lviv (in Ukr.).
4. Darchuk, N.P. (2013). Computer annotation of the Ukrainian text: results and perspectives. Kyiv (in Ukr.).
5. Zahnitko, A.P. (2004). Foundations of the Ukrainian theoretical syntax. Horlivka (in Ukr.).
6. Leontjeva, N.N. (2006). Automatical understanding of texts: systems, models, resources. Moscow (in Rus.).
7. Manning, K., Ragkhavan, P., Shutce, Ch. (2011). Introduction to the information retrieval. Moscow (in Rus.).
8. Testelec, Ya.G. (2001). Introduction to the general syntax. Moscow (in Rus.).

Статтю отримано 11.08.2020

Nataliia Darchuk, Ph.D, professor
at the Department of the Ukrainian language and applied linguistics,
Taras Shevchenko National University of Kyiv
14 Taras Shevchenko Blvd., Kyiv 01601, Ukraine
E-mail: nataliadarchuk@gmail.com
<https://orcid.org/0000-0001-8932-9301>

AGAT-SYNTAX AS COMPONENT OF COMPUTER-BASED UKRAINIAN GRAMMAR

The article describes functional features of the syntactic module of computer-based Ukrainian grammar AGAT. This is a linguistic type of computer-aided syntactic analysis, which provides full information about syntactic units and categories, in particular, predicativity, coordinate and subordinate clauses, the categories of subject and predicate etc. The developed linguistic software provides syntactic analysis of a whole sentence in the form of a dependency tree and indicates the types of syntactic relations and links.

The AGAT-syntax task is to identify all varieties of compatibility — **predicative, subordinate, and coordinate** — of each word in the text. The grammatical characteristics of the phrase directly depend on which part of the language its keyword belongs to. The lexical and grammatical nature of the word determines its compatibility to the other words. Accordingly, phrases can be divided into **substantive, adjective, pronouns, numeral, verbal and adverbial**. Computer subgrammars of valencies of the said parts of the language are built by us on a single principle: a lexema is indicated, preposition that participates in government and a case of a substantive word form in the shape of a two-letter code. In theory, according to their composition words combinations (phrases) are divided into **simple, complex and combined**.

Dependency tree is built from two elements — nodes and connections. Nodes are wordforms and connections are relationships between the main element (“master”) and dependent element (“slave”). It enables to describe a configuration, a form, external parameters of a sentence but this is not sufficient to describe a sentence structure. Thus, the syntactic analysis has two levels: the first one attributes to each binary pair a type of syntactic relationships on the level of morphological way of expression of a “master”; the second level attributes to the connection a type of syntactic relationships, which include: subjective, objective, attributive, adverbial, completive and appositive modifying.

In such a way, the cycle of automated syntactic analysis of Ukrainian texts is completed by determining the syntactic word-combination, identifying a type of syntactic link and a type of relationship. It provides full range of characteristics that can be used for systemic study of semantic and syntactic problems.

Keywords: *automated syntactic analysis, dependency tree, syntactic relations, syntactic links.*