

(Editorial URSS), 2009. – 264 с. **4.** Fader, A. Identifying relations for open information extraction. [Text] / Fader, S. Soderland, O. Etzioni. // Conference on Empirical Methods in Natural Language Processing. – Edinburgh, Scotland, 2011. – P. 1535 – 1545. **5.** Барахнин, В. Б. Проблемы разработки технологии фактографического поиска [Текст] / В. Б. Барахнин, А. М. Федотов. – М.: Институт вычислительных технологий СО РАН, 1980. – 150 с. **6.** Baeza-Yates, R. Modern Information Retrieval. [Text] / R. Baeza-Yates, B. Ribeiro-Neto // Addison-Wesley, 1999. — 340 p. **7.** Ritter, A. Named entity recognition in tweets: an experimental study. [Text] / A. Ritter, S. Clark, K. Mausam, O. Etzioni // Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics. – Edinburgh/Scotland, 2011. – P. 1524–1534. **8.** Филлмор, Ч. Дело о падеже открывается вновь // Новое в зарубежной лингвистике [Текст] / Ч. Филлмор. – М.: Изд. иностр. лит., 1981, вып. 10. – С. 496-530. **9.** Хайрова, Н. Ф. Використання логіко-алгебраїчної моделі семантичних відмінків для семантичного аналізу речення [Текст] / Н. Ф. Хайрова // Збірник наукових праць Військового інституту Київського національного університету імені Тараса Шевченка. – К.: ВІКНУ, 2012.- Вип. № 38. – С. 239 – 245. **10.** Бондаренко, М. Ф. Теория интеллекта [Текст] / Бондаренко М. Ф., Шабанов-Кушнаренко Ю. П. // Харьков: Комп. СМІТ, 2007. – 576 с.

**Bibliography (transliterated):** **1.** Kiselev, S. (2015). Model of information system business dissolve. Otkritie sistemi. #05-06/2005: <http://www.osp.ru/os/2005/05-06/185595/> **2.** Andersen, P. M., Huettner, A. K. (1994). Knowledge engineering for the JASPER fact extraction system. Intrgated Computer-Aided Enginrereng. – 1 (6), 473–493. **3.** Lande, D. V., Snarskiy, A. A., Bezsudnov, I. V. (2009). Internetika: Navigation in complex networks: models and algorithms. Moskow: Libkom (Editorial URSS). **4.** Fader, S. Soderland, O. Etzioni, A. (2011). Identifying relations for open information extraction. Conference on Empirical Methods in Natural Language Processings. Edinburgh, Scotland, 1535–1545. **5.** Barahnin, V. B., Fedotov, A. M. (1980). Problems of development of technology factual search. Moscow: Institute of Computational Technologies. **6.** Baeza-Yates, R., Ribeiro-Neto, B. (1999). Modern Information Retrieval. Addison-Wesley. **7.** Ritter, A., Clark, S., Mausam, K., Etzioni, O. (2011). Named entity recognition in tweets: an experimental study. Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics. – Edinburgh/Scotland, 1524–1534. **8.** Charles, J. Fillmore (1968). The Case for Case. In Bach and Harms (Ed.): Universals in Linguistic Theory. New York: Holt, Rinehart, and Winston, 1-88. **9.** Khairova, N. (2012). Using logic-algebraic model of semantic rules for semantic analysis of the sentence. Kiev: Zbirnik naukovich prach Viyskovogo institutu, Vol. 38, 239 – 245. **10.** Bondarenko, M. F., Shabanov-Kushnarenko, U. P. (2007). The theory of intelligence. Kharkov: SMIT Comp.

Надійшла (received) 21.02.2015

УДК 004.912

**О. В. ЛОЗИНСЬКА**, асистент, НУ «Львівська політехніка»;

**М. В. ДАВИДОВ**, канд. техн. наук, доц., НУ «Львівська політехніка»

## МАТЕМАТИЧНА МОДЕЛЬ ГРАМАТИЧНО-ДОПОВНЕНОЇ ОНТОЛОГІЇ

Дослідження відомих методів вирішення проблеми багатозначності слів з використанням онтологій показало, що відомі методи обмежені лише контекстом слова і не надають додаткових переваг для граматичного і семантичного розбору речення. Для вирішення цієї проблеми розроблено математичну модель граматично-доповненої онтології. Ця модель використана для граматичного розбору речень української мови. Отримані результати показали адекватність розробленої моделі, але її використання вимагає наповнення словників нового типу.

**Ключові слова:** українська жестова мова, онтологія, граматично-доповнена онтологія, синсет.

© О. В. ЛОЗИНСЬКА, М. В. ДАВИДОВ, 2015

**Вступ.** Проблема розроблення комп'ютеризованої системи перекладу жестової мови вже довгий час є об'єктом дослідження науковців [1]. Вирішення цієї проблеми надасть нові можливості для спілкування людей із вадами слуху. Задача перекладу української жестової мови (УЖМ) на українську словесну мову (УСМ) належить до задач комп'ютерного перекладу.

Однією з проблем, яка виникає при перекладі з однієї мови на іншу, є проблема усунення багатозначності слів, тобто вибору одного з можливих значень слова, перерахованих в лінгвістичному ресурсі. Не зважаючи на численні евристичні методи вирішення цієї проблеми, серед яких методи навчання з вчителем, що використовують розмічені корпуси, та без вчителя [2], методи на основі знань [3], проблема усунення багатозначності досі не вирішена повністю, оскільки вирішення даної проблеми вимагає структурування знань людини з різних предметних областей для кожної з мов перекладу.

Довгий час джерелом знань для систем перекладу були тлумачні та двомовні словники [4]. Через те, що тлумачні словники не володіють необхідною повнотою і структурою для вираження семантичних зв'язків між словами і поняттями, замість них почали використовувати онтологічні словники [3]. Сучасні онтологічні словники містять не лише конкретні значення слів, а й лексичні (антонімія, слова-відношення, номіналізація та ін.) та семантичні (гіперонімія/ гіпонімія, меронімія/голонімія та ін.) зв'язки між ними, що дає змогу використовувати їх для усунення багатозначності слів на основі цих зв'язків.

Проте, онтологічні словники не описують в повній мірі мовні конструкції для вираження семантичних зв'язків між поняттями, що зумовлює необхідність розроблення більш деталізованих моделей.

У статті описано нову модель граматично-доповненої онтології (ГДО) а також подано приклади її використання для граматичного та семантичного розбору речень української мови. Робота є частиною проекту з розроблення системи двостороннього перекладу української жестової мови.

**Аналіз відомих досліджень та постановка проблеми.** Для більш повного використання онтологічних словників в системах машинного перекладу необхідно розробити математичну модель для створення засобів вираження семантичних зв'язків у формі конструкцій речення.

Для вирішення цієї проблеми необхідно, щоб онтологія була доповнена інформацією про граматичні зв'язки. Для використання у системах перекладу граматично-доповнена онтологія повинна містити додаткові засоби для:

–вираження граматичних атрибутів слів (граматичної категорії, граматичного значення);

–опису основних граматичних конструкцій та їх зв'язку з поняттями.

Для практичного втілення граматично-доповненої онтології треба розробити математичну модель даної онтології, що є предметом цієї статті.

Сьогодні для структурування знань із різних предметних областей використовуються онтології. Онтології використовуються для формальної специфікації понять і відносин, які характеризують певну галузь знань. Перевагою онтологій як способу подання знань є їх формальна структура, яка спрощує комп'ютерне опрацювання [5].

Відомі роботи з розроблення онтологій як спеціалізованих, для опису окремих

предметних галузей, так і загальних, для опису загальноновживаних понять мови. Наприклад, розроблені онтології у галузі медицини (Gene Ontology для анотації біомедичних даних [6], система уніфікованої медичної мови (UMLS, the Unified Medical Language System) [7], онтологія у галузі товарів та послуг UNSPSC [8] та ін.

Загальні онтології широко використовуються для досліджень в області комп'ютерної лінгвістики. До таких онтологій належать WordNet для англійської мови [9], MultiWordNet project для італійської мови [10], BalkaNet project для шести європейських мов (болгарської, чеської, грецької, румунської, турецької та сербської) [11] та інші.

Для української мови розробленням WordNet-подібного словника для частини іменників займаються вчені Кульчицький І.М. та ін. [12]. У результаті дослідження вченими розроблено фрагмент WordNet-подібного словника української мови, в якому реалізовано 194 синсети, пов'язані між собою зв'язками гіпо-/гіперонімії, антонімії, а також додатково зв'язками меронімії/голонімії.

Створенням української онтологічної лексико-семантичної бази знань UkrWordNet (UWN) займаються науковці Київського національного університету імені Тараса Шевченка [13]. На даний момент онтологія містить близько 80000 концептів.

Онтології широко використовуються для вирішення проблеми багатозначності слів. У роботі [3] вчені А. Романюк та ін. дослідили проблему вирішення лексичної багатозначності та проаналізували основні методи її розв'язання. Вченими обрано методи на основі знань, досліджено сучасні інформаційні ресурси та встановлено, що мережа WordNet надає найбільше можливостей для автоматичного розрізнення значень слів.

При використанні відомих онтологій для усунення багатозначності слів використовуються статистичні моделі у яких враховується частота появи слів разом у певному контексті [2]. Такий підхід може забезпечити якісний переклад лише за умови використання великих навчальних корпусів. Використання онтологій для перекладу між мовами, для яких немає великих двомовних корпусів, вимагає розроблення альтернативного підходу, який зорієнтований на граматичні зв'язки та їх семантичне значення. Авторам не відомі роботи у яких моделі загальних онтологій доповнені такими граматичними зв'язками.

**Математична модель граматично доповненої онтології.** У літературі відомі декілька альтернативних моделей онтологій [14].

За основу обрано математичну модель онтології з [15], визначену як множину  $O = \langle L, C, F, R_c \rangle$ , де  $L = \{w_i\}$  – словник термінів предметної області,  $C = \{c_i\}$  – множина понять (концептів),  $F \subset L \times C$  – відношення відповідних термінів та понять,  $R_c$  – множина відношень на множині понять (гіпонімії, гіперонімії, меронімії, голонімії і т. д.).

Означимо математичну модель граматично-доповненої онтології як множину

$$O_G = \langle O, P, S, T, R_p \rangle,$$

де  $P = \{p_i\}$  – множина предикатів;  $S = \{s_i\}$  – множина висловлювань, які можуть бути задані переліком або граматикою, де  $s_i = ((w_1, g_1), (w_2, g_2), \dots, (w_n, g_n))$  – рядок граматично-доповнених термінів,  $n = \text{Len}(s_i)$  – довжина рядка,  $(w_i, g_i)$  – граматично-доповнений

термін;  $T = \{t_i\}$  – множина параметризованих висловлювань, де  $t_i = (s_k, f_i)$  – пара висловлювання-функція параметризації;  $f_i : \{1, 2, \dots, Len(s_i)\} \rightarrow \{0, 1, \dots, r\}$  – функція параметризації,  $r$  – кількість параметрів параметризованого висловлювання  $t_i$ ,  $f(k) > 0$  означає, що відповідний термін у позиції  $k$  є параметром з номером  $f(k)$ , і може бути замінений на інший термін із множини гіпонімів;  $I \subset T \times P$  – відношення інтерпретації між параметризованими висловлюваннями і предикатами;  $R_p$  – множина відношень на множині предикатів (гіпонімії, гіперонімії, меронімії і т. д.).

Таке означення граматично-доповненої онтології дає змогу виражати зв'язок між поняттями, предикатами і засобами їх вираження у вигляді мовних конструкцій. Наприклад, розглянемо предикат «вчити (кого, чому)». Застосуємо його у реченні «Я вчу школярів математиці». Це речення можна побудувати і таким чином «Я вчу математиці школярів». Таке речення не відповідає параметризованому висловлюванню згідно наведеної математичної моделі граматично-доповненої онтології, тому що ми можемо поміняти параметри місцями (у наведеному прикладі, параметри «кого, чому»). Під параметризованим розуміємо таке висловлювання, у якому параметри пронумеровані і відповідають номерам предметних змінних предикату. Наприклад, предикату «вчити (суб'єкт, кого, чому)» відповідає параметризоване висловлювання «1.хто вчити 2.кого 3.чому». Даному параметризованому висловлюванню, як приклад, відповідає речення «Я вчу школярів математиці». Враховуючи граматично-доповнені терміни у онтології, запишемо дане параметризоване висловлювання з використанням відмінків (позначено С, casus) для параметрів предикату «1.С1 вчити 2.С4 3.С3». Цифри зліва від літери С позначають позицію параметра у параметризованому висловлюванні, а цифри справа позначають номер відмінка в українській мові (називний – 1, родовий – 2, давальний – 3, знахідний – 4, орудний – 5, місцевий – 6, кличний – 7). У реченні деякі параметри можуть опускатися. Наприклад, у реченні «Я вчу школярів» відсутній третій параметр «чому» висловлювання, який відповідає предикату «вчити (суб'єкт, кого, чому)».

Крім того, предикат «вчити» вживається й у іншому значенні «вчити що» (вивчати). У даному випадку, згідно наведеної математичної моделі граматично-доповненої онтології запишемо його так: «вчити (суб'єкт, що)», якому відповідає параметризоване висловлювання «1.С1 вчити 2.С4». Наприклад, речення «Я вчу вірш» відповідає даному предикату.

Серед відношень на множині предикатів, розглянемо відношення гіпонімії/гіперонімії. Наприклад, у реченнях «Я вчу школярів математиці» та «Я вчу людей математиці» між поняттями «школяр» та «людина» встановлено відношення гіпонімії/гіперонімії (поняття «школяр» є гіпонімом поняттю «людина»). Дерево семантично-синтаксичного розбору для даних речень, коли розбір відповідає онтології (вага більша за 1), зображено на рис. 1, а. Отже, за допомогою відношень між поняттями у граматично-доповненій онтології на місця параметризованих змінних можна поставити лише правильні по змісту поняття. Таким чином, речення «Я вчу дошку математиці» згідно наведеної граматично-доповненої онтології є некоректним і для цього речення дерево синтаксичного розбору відповідає звичайному синтаксису речення (вага дорівнює 1) та зображено на рис. 1, б.

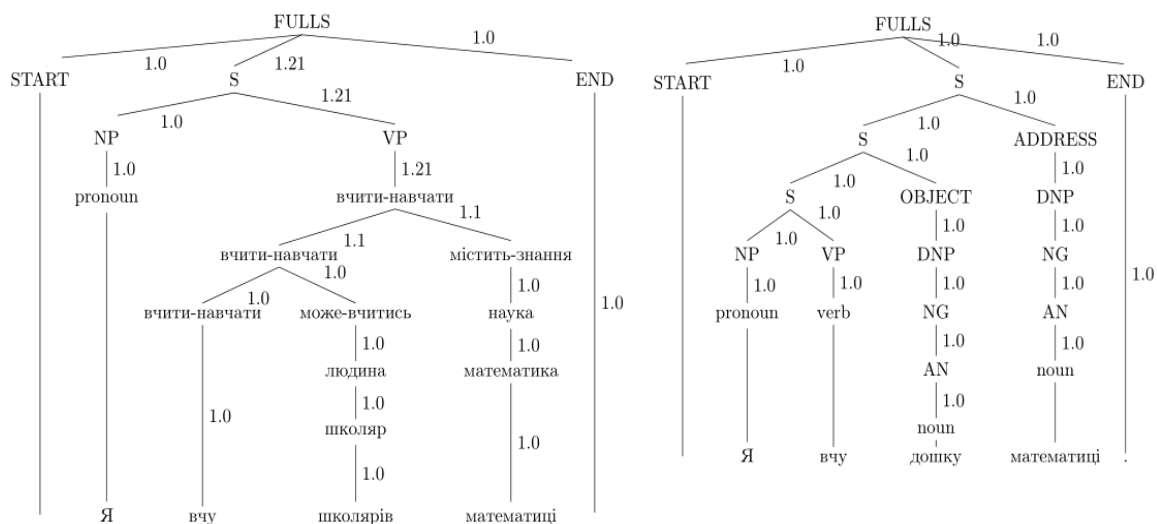


Рис. 1 – Дерева синтаксичного розбору для речень: а – «Я вчу школярів математиці» та «Я вчу людей математиці»; б – «Я вчу дошку математиці»

Відношення меронімії/голонімії у граматично-доповненій онтології відображає відношення «частина-ціле». Це відношення властиве іменникам. Наприклад, двері є частиною (меронімом) будинку. Отже, у предикаті «1.C1 заходити у 2.C4» у якості параметра на 2 позиції можна підставити термін «двері» або термін «будинок».

**Висновки.** У статті розроблено математичну модель граматично-доповненої онтології. Дана математична модель забезпечує можливість введення в онтологію граматичних категорій та граматичних значень, які виражені граматичними формами.

Застосування розробленої математичної моделі у парсері української мови на основі ймовірнісних афікських контекстно-вільних граматик дало змогу отримати семантичний розбір речення за умови відповідності речення правилам онтологічного словника.

Подальші дослідження будуть спрямовані на автоматизацію побудови граматично-доповненого онтологічного словника і розширення можливостей системи автоматизованого розбору речень.

**Список літератури:** 1. *Bungeroth, J.* Statistical Sign Language Translation / *J. Bungeroth, H. Ney* // Proc. LREC 2004, Workshop proceedings: Representation and Processing of Sign Languages. – Lisbon, Portugal. – 2004. – P. 105–108. 2. *Navigli, R.* Word Sense Disambiguation: A Survey / *R. Navigli* // ACM Computing Surveys, ACM Press, 2009. – № 41(2). – P. 1– 69. 3. *Романюк, А. Б.* Методи вирішення лексичної багатозначності. Використання WORDNET для вирішення проблем багатозначності / *А. Б. Романюк, І. А. Сундуютова, М. М. Романишин* // Вісник НУ "ЛПІ" "Комп'ютерні системи проектування. Теорія і практика" – 2011. – № 711. – С. 147–157. 4. *Кобрицов, Б. П.* Снятие семантической многозначности глаголов с использованием моделей управления, извлеченных из электронных толковых словарей. – Режим доступа : <http://download.yandex.ru/IMAT2007/kobricov.pdf>. – Дата звертання : 20 лютого 2015. 5. *Вороной, А. С.* Використання онтологій для підвищення якості пошуку інформації для поповнення баз знань інтелектуальних систем / *А. С. Вороной* // Матеріали Міжн. наук.-техн. конференції «Комп'ютерні науки і інженерія 2009». – Lviv: Львівська політехніка, 2009. – С. 364–366. 6. *Brochhausen, M.* Establishing and Harmonising Ontologies in an Interdisciplinary. Health Care and Clinical Research Environment. – Режим доступа : <http://ontology.buffalo.edu/medo/eHealth.pdf>. – Дата звертання : 15 лютого 2015. 7. *Kumar, A.* The Unified Medical Language System and the Gene

Ontology: Some Critical Reflections / A. Kumar, B. Smith // In: KI 2003: Advances in Artificial Intelligence. – Berlin: Springer, 2003. – P. 135–148. **8.** Проект UNSPSC. – [Електронний ресурс]. – Режим доступу: unspsc.org. – Дата звертання : 30 січня 2016. **9.** Al-Halimi, R. WordNet: An Electronic Lexical Database / R. Al-Halimi, R. C. Berwick, J. F. Burg // Edited by Christiane Fellbaum. – Cambridge, MA: MIT Press, 1998. – 422 pp. **10.** Pianta, E. MultiWordNet: Developing an aligned multilingual database / E. Pianta, L. Bentivogli, C. Girardi // In Proc. of the 1st International Conference on Global WordNet. – Mysore, India, 2002. – P. 21–25. **11.** Tufis, D. Balkanet: Aims, methods, results and perspectives. A general overview / D. Tufis, D. Cristea, S. Stamou // Romanian J. Sci. Tech. Inform. (Special Issue on Balkanet). – 2004. – № 7(1-2). – P. 9–43. **12.** Кульчицький, І. М. Розроблення wordnet - подібного словника української мови / І. М. Кульчицький, А. Б. Романюк, Х. Б. Харів // Вісник Національного університету "Львівська політехніка": Інформаційні системи та мережі. – 2011. – № 673. – С. 306–319. **13.** Анісімов, А. В. UWN: Універсальна онтологічна база знань української мови / А. В. Анісімов, О. О. Марченко, А. О. Никоненко // Проблеми програмування. – 2012. – № 2-3. – С. 348–355. **14.** Антонов, І. В. Модель онтології предметної області для систем семантично-орієнтованого доступу / І. В. Антонов // Труды Псковского политехнического института. Электротехника. Машиностроение. – 2011. – № 14.3. – С. 339–343. **15.** Захарова, І. В. Способи автоматического построения онтологий для задач анализа текстов / І. В. Захарова, М. С. Тимченко // Знання – Онтології – Теорії»: труды Всероссийск. конф. 30НТ-09. – Новосибирск, 2009. — Т. 2. – С. 164–167.

**Bibliography (transliterated):** **1.** Bungeroth, J., Ney, H. (2004). Statistical Sign Language Translation. Proc. LREC 2004, Workshop proceedings: *Representation and Processing of Sign Languages*. Lisbon, Portugal, 105–108. **2.** Navigli, R. (2009). Word Sense Disambiguation: A Survey. *ACM Computing Surveys*, ACM Press, 41(2), 1–69. **3.** Romanyuk, A. B., Sundutova, I. A., Romanishin, M. M. (2011). Metodi virishennya leksichnoї bagatoznachnosti. Viktoristannya WORDNET dlya virishennya problem bagatoznachnosti. *Visnik NU "LP" "Komp'yuterni sistemi proektuvannya. Teoriya i praktika"*, 711, 147–157. **4.** Kobricov, B. P. Snyatie semanticheskoy mnogoznachnosti glagolov s ispolzovaniem modelej upravleniya, izvlechennyx iz elektronnyx tolkovyx slovarej. – Rezhim dostupu : <http://download.yandex.ru/IMAT2007/kobricov.pdf>. – Data zvertannya : 20 lyutogo 2015. **5.** Voronov, A. S. (2009). Viktoristannya ontologij dlya pidvishhennya yakosti poshuku informacii dlya popovnennya baz znan intelektualnix system. *Materiali Mizhn. nauk.-texn. konferencii «Komp'yuterni nauki i inzheneriya 2009»*. L.: Lvivska politexnika, 364–366. **6.** Brochhausen, M. (2015). Establishing and Harmonising Ontologies in an Interdisciplinary. Health Care and Clinical Research Environment. Rezhim dostupu : <http://ontology.buffalo.edu/medo/eHealth.pdf>. **7.** Kumar, A., Smith, B. (2003). The Unified Medical Language System and the Gene Ontology: Some Critical Reflections. In: *KI 2003: Advances in Artificial Intelligence*. Berlin: Springer, 135–148. **8.** Проект UNSPSC.(2016). Rezhim dostupu: unspsc.org. **9.** Al-Halimi, R., Berwick, R. C., Burg, J. F. (1998). WordNet: An Electronic Lexical Database. Edited by Christiane Fellbaum. Cambridge, MA: MIT Press, 422. **10.** Pianta, E., Bentivogli, L., Girardi, C. (2002). MultiWordNet: Developing an aligned multilingual database. In *Proc. of the 1st International Conference on Global WordNet*. Mysore, India, 21–25. **11.** Tufis, D., Cristea, D., Stamou, S. (2004). Balkanet: Aims, methods, results and perspectives. A general overview. *Romanian J. Sci. Tech. Inform. (Special Issue on Balkanet)*, 7(1-2), 9–43. **12.** Kulchickij, I. M., Romanyuk, A. B., Hariv, X. B. (2011). Rozroblennya wordnet - podobnogo slovnika ukraїnskoї movi. *Visnik Nacionalnogo universitetu "Lvivska politexnika": Informacijni sistemi ta merezhi*, 673, 306–319. **13.** Anisimov, A. V. Marchenko, O. O., Nikonenko, A. O. (2012). UWN: Universalna ontologichna baza znan ukranskoї movi. *Problemi programuvannya*, 2-3, 348–355. **14.** Antonov, I. V. (2011). Model ontologii predmetnoj oblasti dlya sistem semanticheskii-orientirovannogo dostupa. *Trudy Pskovskogo politexnicheskogo instituta. Elektrotexnika. Mashinostroenie*, No14.3, 339–343. **15.** Zaxarova, I. V., Timchenko M. S. (2009). Sposoby avtomaticheskogo postroeniya ontologii dlya zadach analiza tekstov. *Znaniya – Ontologii – Teorii* : trudy Vserosijsk. konf. 30НТ-09. Novosibirsk, V2, 164–167.

Поступила (received) 27.02.2015