

М.П. Костіков (Національний університет харчових технологій, м. Київ, Україна)

### СТВОРЕННЯ БАЗИ ЗНАТЬ ГРАМАТИКИ ПОЛЬСЬКОЇ МОВИ\*

*У статті проаналізовано можливості використання баз знань для розробки інтелектуальної навчальної системи з граматики польської мови. Описано підхід до створення баз знань у морфології.*

*Ключові слова:* бази знань, електронні засоби навчання, електронні словники, продукційна модель, польська мова, граматика, морфологія.

*Рис. 1. Літ. 11.*

Н.П. Костиков (Национальный университет пищевых технологий, г. Киев, Украина)

### СОЗДАНИЕ БАЗЫ ЗНАНИЙ ГРАММАТИКИ ПОЛЬСКОГО ЯЗЫКА

*В статье проанализированы возможности использования баз знаний для разработки интеллектуальной обучающей системы по грамматике польского языка. Описан подход к созданию баз знаний в морфологии.*

*Ключевые слова:* базы знаний, продукционная модель, электронные словари, электронные средства обучения, польский язык, грамматика, морфология.

M.P. Kostikov (National University of Food Technologies, Kyiv, Ukraine)

### CREATING A POLISH GRAMMAR DATABASE

*The article analyzes the opportunities for using knowledge bases in developing an intelligent tutoring system on Polish grammar. An approach to creating a knowledge base in morphology is described.*

*Keywords:* knowledge base; product model; electronic dictionaries; electronic means of teaching; Polish language; grammar; morphology.

**Постановка проблеми.** В сучасних умовах зростає роль самостійного навчання у процесі вищої освіти, в тому числі у викладанні іноземних мов у ВНЗ України. Постає питання: як ефективніше організувати цей процес із допомогою сучасних інформаційних технологій? Для вивчення іноземних мов нині розробляються різноманітні електронні засоби навчання (ЕЗН). Однак, поперше, відчувається суттєвий брак програмного забезпечення для вивчення саме слов'янських мов, а по-друге, існуючі для них засоби не завжди є ефективними й складні для використання в освітньому процесі. Це підштовхує до пошуку шляхів створення якісно нових ЕЗН іноземних мов, зокрема до розробки інтелектуальних навчальних систем.

**Аналіз останніх досліджень і публікацій.** Проблему створення інтелектуальних і автоматизованих навчальних систем досліджували Ю. Машбиць [3], О. Меньяйленко [2] та інші. Питаннями штучного інтелекту, зокрема експертних і експертно-навчальних систем займались В. Петрушин [4], Д. Смолін [6] та інші.

**Невирішені раніше частини загальної проблеми.** Проведений нами аналіз сучасних ЕЗН польської мови [1] показав, що в них недостатньо розвинений зворотний зв'язок. У багатьох програмах бракує пояснень, коментарів і підказок. Лише деякі ЕЗН є адаптивними: система реагує на помилки користувача,

\* статтю підготовлено на основі доповіді на XI-му міжнародному науковому семінарі «Сучасні проблеми інформатики в управлінні, економіці, освіті та екології» (2–7 липня 2012 р., Київ – оз. Світязь).

пропонуючи ще раз проглянути попередній матеріал і/або пройти вправу повторно. Однак відсутні спроби проаналізувати причини помилок. Статистика навчання користувача часто збирається та візуалізується, проте ця індивідуальна історія рідко впливає на подальший хід курсу. Крім того, при викладанні споріднених мов важливим принципом є врахування особливостей рідної мови того, хто навчається [5, 54]. Однак у наявних засобах для польської не береться до уваги специфіка викладання цієї мови у слов'яномовних країнах. Зокрема, основну увагу часто приділяють вивченню слів, іноді зовсім нехтуючи граматику, в той час як для носіїв споріднених мов лексика викликає менше труднощів. Наприклад, польська й українська мови, за твердженням К. Тищенка, мають за словниковим складом лише 30% відмінної лексики [7, 266–267]. Таким чином, у сучасних ЕЗН польської мови є проблеми з реалізацією двох важливих принципів навчання – індивідуалізації та врахування рідної мови студента.

**Мета дослідження** – визначити шляхи використання баз знань як основи якісно нового ЕЗН польської мови, який би забезпечував індивідуалізоване навчання та враховував специфіку викладання мови в Україні.

**Основні результати дослідження.** Виходячи зі спорідненості української та польської мов, при створенні засобу навчання доцільно зробити змістом курсу граматику, акцентуючи увагу на особливостях побудови слівосполучень і взаємозв'язку слів. Це допоможе в реалізації принципу врахування рідної мови студентів. Індивідуалізація ж передбачає ґрунтовну розробку механізму зворотного зв'язку. Крім того, на думку Ю. Машбиця, індивідуалізованим можна вважати лише таке навчання, яке спирається на модель студента [3, 193].

Механізм зворотного зв'язку в загальному випадку означає наявність у схемі управління діагностуючого та керуючого органів. Перший із них визначає, чи є відхилення від необхідного стану об'єкта управління після подання на нього початкових керуючих впливів. Ця інформація подається на керуючий орган, який приймає рішення, чи треба і яким чином слід змінити керуючі впливи, аби привести об'єкт управління до необхідного стану. При застосуванні зворотного зв'язку у процесі навчання, по-перше, здійснюється контроль знань студента із виявленням і аналізом його помилок, а по-друге, фіксується реакція на них, тобто зміну поведінки системи залежно від отриманих результатів. Це може бути повторення матеріалу, додаткові запитання чи вправи, коментарі, пояснення, підказки, подальше коригування індивідуальної траєкторії навчання студента.

Цей механізм успішно реалізовується шляхом розробки автоматизованих навчальних систем. За словами О. Меньяйленка, їх використання дає змогу широко застосувати принципи індивідуалізації процесу навчання [2, 7]. Серед них важливе місце посідають експертно-навчальні системи (ЕНС). За В. Петрушиним, загальна архітектура експертно-навчальної системи включає в себе модель студента (в якій зберігається інформація про його поточні знання та інші характеристики) та три експертні системи (ЕС) – ЕС предметної області (ПО), ЕС діагностування помилок студента і ЕС управління навчанням [4, 45–47].

Якщо дві останні ЕС можна вважати відповідно діагностуючим і керуючим органами у схемі управління, то ЕС ПО безпосередньо розв'язує задачі.

Для процесу навчання важливою характерною рисою ЕС є наявність підсистеми пояснень. Вона слугує для обґрунтування процесу отримання системою розв'язку поточної задачі, що дозволяє користувачеві прослідкувати хід розв'язання та пересвідчитись в обґрунтованості кожного кроку [4, 24]. Отже, ЕС дають можливість не лише отримати відповідь, а й зрозуміти напрямок розв'язання. При застосуванні ЕС у створенні інтелектуальних навчальних систем ця можливість надає їм суттєві переваги над навчальними системами попередніх поколінь [3, 90]. Це здійснюється за рахунок того, що ЕС містять не просто набір даних, а знання експертів із ПО, формалізовані тим чи іншим чином і збережені у відповідній базі знань (БЗ) у вигляді фактів, правил, взаємозв'язків між об'єктами тощо. Наявність БЗ є головною відмінністю ЕС від інших програмних засобів [6, 151].

У нашому випадку при розробці ЕНС граматики польської мови БЗ ПО міститиме правила морфології (словозміни за різними граматичними категоріями) та синтаксису (зв'язку слів у словосполученнях і реченнях). Наразі проводиться робота зі створення БЗ із морфології польської мови. Завдання формалізації цієї ПО полягає в описі правил, за якими відбувається словозміна, тобто яким чином утворюються окремі граматичні форми (ГФ) слів.

До певної міри опис морфології можна побачити в електронних граматичних словниках. Наприклад, інтегрована лексикографічна система «Словники України», розроблена в Українському мовно-інформаційному фонді НАН України, містить модуль «Словозміна», що забезпечує візуалізацію всіх словоформ української мови.

Ще одним відомим прикладом реалізації такої опції є багатомовний словник "Lingvo" виробництва компанії АBBYY. У версії X5 (2011 р.) інформація щодо словозміни доступна для 11 мов: англійської, грецької, іспанської, італійської, казахської, латини, німецької, російської, турецької, української, французької. Для змінюваних слів відображаються повні таблиці парадигм, тобто всі їхні ГФ. Форми дієслів згруповані в таблиці за часами. Є можливість перемикання між частинами мови, копіювання форм слів, друку парадигми.

Існують також інші локальні та мережеві програмні засоби, що описують морфологію флективних мов: електронна версія граматичного словника російської мови, розроблена компанією "Solarix", граматичний тезаурус хорватської мови, онлайн-словники німецької, румунської та інших мов, низка онлайн-словників у рамках проекту "Wiktionary" тощо.

Першу версію електронного граматичного словника польської мови – "Słownik gramatyczny języka polskiego" (SGJP) – було випущено в 2007 р. у Варшаві видавництвом "Wiedza Powszechna". Як зазначають автори, вони прагнули до повного опису словозміни польської мови [10, 157]. Словник містить майже 245 тис. слів із усіма їхніми флективними формами (загалом понад 4 млн.) [9, 12]. Один із розробників SGJP М. Волінський стверджує, що польська словозміна може бути описана в уніфікованій та відносно компактній реляційній моделі [11, 96].

Реляційна модель дійсно є зручною формою представлення даних і компактна з точки зору використання дискового простору. Суть цієї моделі полягає в тому, що слова класифікуються та групуються за типами відмінювання, і

в окремих таблицях за частинами мови зберігаються набори закінчень ГФ для кожного з типів. Оскільки ж загальноприйнятих відмін і дієвідмін, що традиційно використовуються при викладанні іноземної мови, недостатньо для опису всіх особливостей словозміни, фахівцями з комп'ютерної лінгвістики додатково вводяться численні підтипи. Наприклад, при розробці системи «Словники України» всі слова української мови було поділено на так звані словозмінні парадигматичні класи, всередині яких діють єдині правила словозміни. Таких класів було виокремлено 2053, і близько половини з них припадає лише на дієслова [8, 78–79].

Таким чином, подібні реляційні моделі дають можливість достатньо повного опису словозміни флективних мов й успішно використовуються для побудови парадигматичних таблиць в електронних граматичних словниках. Однак усі перелічені програмні засоби можуть бути використані у навчальному процесі лише як інформаційно-довідкові джерела, тобто для перевірки правильності написання тієї чи іншої ГФ. Натомість для опанування студентами морфології при вивченні іноземної мови цих парадигматичних таблиць недостатньо, адже інформацію щодо самих правил утворення слівформ із них можна вивести хіба що інтуїтивно. Для набуття системних знань із граматики необхідно, аби ці правила були представлені в явному вигляді. Це може бути реалізовано саме з допомогою ЕС, яка пояснюватиме, яким чином будується кожна окрема слівформа і чому слова відмінюються так чи інакше.

Що стосується представлення знань у БЗ, існують різні моделі: логічні, продукційні, фреймові, семантичні мережі. Для формалізації граматики нами було обрано продукційну модель або модель, засновану на правилах, адже і сама по собі граMATика викладається через навчання студентів застосування відповідних правил.

Змоделюємо самий процес утворення ГФ слова. Оскільки для мов флективного типу правила словозміни визначаються закінченнями ГФ і характером чергування в основі [8, 78], саме ці ознаки будуть предметом вибору розв'язувача ЕС ПО при утворенні тої чи іншої ГФ. Тепер визначимо, що впливатиме на цей вибір. Утворення певної ГФ полягає у наданні слову відповідного граматичного значення (ГЗ), що є перетином граматичних категорій. Це категорії, змінювані для відповідної частини мови, такі як число, відмінок, рід прикметника, час дієслова тощо. Прикладом окремого ГЗ може бути орудний відмінок однини (для іменника), називний відмінок множини жіночого роду (для прикметника) тощо. У процесі надання ГЗ початковій формі слова будуть задіяні, по-перше, загальний шаблон словозміни для відповідної частини мови, по-друге, фонетичні характеристики слова, які впливатимуть на чергування звуків, а по-третє, різноманітні винятки для окремих слів, що стосуються як поодиноких ГФ, так і цілої парадигми (рис. 1).

Фонетичні характеристики можуть бути визначені розв'язувачем ЕС ПО шляхом аналізу слова, що подається на вхід системи. Натомість правила утворення ГФ залежно від цих характеристик, а також шаблони словозміни та винятки мають бути збережені безпосередньо в БЗ. Крім того, необхідна наявність словника, розміченого за частинами мови та класифікаційними (незмінними) граматичними категоріями (рід іменника, назви істот і неістот тощо),

оскільки ці ознаки неможливо однозначно вивести шляхом аналізу самого слова.

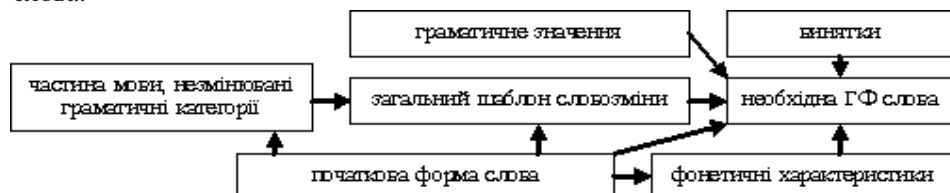


Рис. 1. Утворення граматичної форми слова від початкової, авторська розробка

Отже, на вхід ЕС ПО подаватиметься початкова форма слова та власне ГЗ, яке вона повинна виражати. Звірившись зі словником, розв'язувач звертатиметься до БЗ і застосовуватиме один із загальних шаблонів словозміни залежно від частини мови та незмінних граматичних категорій, що впливають на процес утворення конкретної ГФ. Також він враховуватиме фонетичні характеристики й можливі винятки. В результаті цієї обробки ЕС ПО видаватиме відповідь на поставлене завдання.

Для ілюстрації цього механізму роботи візьмемо звичайну вправу з граматики польської мови: наприклад, слово *zegar* («годинник») необхідно поставити в орудний відмінок. Маючи інформацію про початкову форму слова – *zegar*, його частину мови – іменник і рід – чоловічий, розв'язання завдання з допомогою БЗ відбуватиметься шляхом застосування відповідних правил: до основи слова додається закінчення *-em*, твердий *r* у кінці не впливає на утворення ГФ, винятків теж немає. У підсумку на виході маємо словоформу *zegarem*.

Створювана БЗ граматики польської мови як основа цілісної ЕНС стане джерелом для різноманітних тренувальних і контрольних завдань із морфології та синтаксису. В рамках ЕНС взаємодія студента з БЗ відбуватиметься через інтерфейс користувача. При поданні певної граматичної вправи в робочій пам'яті зберігатимуться дані відповідного завдання. Розв'язувач ЕС ПО застосовуватиме правила з БЗ і видаватиме відповідь. Завдяки ж підсистемі пояснень, що входить до складу ЕС ПО, можна буде не лише визначити необхідну ГФ слова, що дозволяють і електронні граматичні словники, а й обґрунтувати та прокоментувати через закладені в БЗ граматичні правила, як саме відбувається процес словозміни. Помилки студента аналізуватимуться ЕС діагностування, після чого ЕС управління прийматиме рішення про необхідні керуючі впливи. В подальшому інформація про знання користувача зберігатиметься в моделі студента, що дозволить забезпечити індивідуалізацію навчання на основі механізму зворотного зв'язку.

Інтерфейс навчальної оболонки для роботи з ЕНС, а відповідно й БЗ, доцільно розробити як веб-додаток, оскільки вони безпосередньо орієнтовані на кінцевого споживача та пропонують широкий спектр можливостей візуального представлення інформації. Крім того, веб-додатки зручні й прості у використанні, адже не потребують інсталяції, запускаються у більшості браузерів і зазвичай не залежать від операційної системи користувача. Деякі програмні

платформи також дозволяють створити локальну версію веб-додатка для використання за відсутності доступу до Інтернету.

**Висновки та перспективи подальших досліджень.** Розробка БЗ граматики польської мови є основою створення відповідної ЕНС. Така система в подальшому дасть змогу реалізувати принципи індивідуалізації навчання та врахування рідної мови при навчанні польської мови студентів ВНЗ України. Розробка інших складових ЕНС є предметом подальшого дослідження.

1. *Костіков М.П.* Можливості сучасних електронних засобів навчання польської мови для вивчення граматики студентами // Наукові записки Кіровоград. держ. пед. ун-те ім. В. Винниченка. – Вип. 108, Ч. 1. – Кіровоград, 2012. – С. 206–209.

2. *Меняйленко О.С.* Автоматизовані педагогічні навчальні системи: Монографія. – Луганськ: Альма-матер, 2003. – 272 с.

3. Основи нових інформаційних технологій навчання: Посібник для вчителів / Ю.І. Машбиць, О.О. Гокунь, М.І. Жалдак та ін.; Інститут психології ім. Г.С. Костюка АПН України, Інститут змісту і методів навчання. – К., 1997. – 260 с.

4. *Петрушин В.А.* Экспертно-обучающие системы / Отв. ред. А.М. Довгялло; АН УССР. Ин-т кибернетики. – К.: Наукова думка, 1992. – 196 с.

5. *Салистра И.Д.* Очерки методов обучения иностранным языкам. Система упражнений и система занятий. – М.: Высшая школа, 1966. – 252 с.

6. *Смолин Д.В.* Введение в искусственный интеллект: Конспект лекций. – М.: Физматлит, 2004. – 208 с.

7. *Тищенко К.М.* Метатеорія мовознавства. – К.: Основи, 2000. – 350 с.

8. *Широков В.А.* Елементи лексикографії: Монографія / Укр. мов.-інформ. фонд НАН України. – К.: Довіра, 2005. – 304 с.

9. *Saloni, Z., Gruszczynski, W., Wolinski, M., Wolosz, R.* (2007). Grammatical Dictionary of Polish. Presentation by the Authors. Studies in Polish Linguistics (Cracow : LEXIS), 4: 5–25.

10. *Saloni, Z., Gruszczynski, W., Wolinski, M., Wolosz, R.* (2008). Słownik gramatyczny języka polskiego. Warszawa : Wiedza Powszechna. 180 s.

11. *Wolinski, M.* (2009). A Relational Model of Polish Inflection in Grammatical Dictionary of Polish. In: Human Language Technology: Challenges of the Information Society. Berlin, Heidelberg: Springer-Verlag. P. 96–106.

Стаття надійшла до редакції 31.07.2012.