

# Розробка універсальної мови комп'ютерного перекладу

Дмитро Барсуков, Олексій Некрасов, Олександр Шапкін

Запорізький національний технічний університет, УКРАЇНА, м.Запоріжжя,  
E-mail: barsukov.dmytro@gmail.com, raacoon@rambler.ru.

*Abstract – The goal of the current work is development of universal computer translation language. This language is syntax-based and oriented on possibility of translation into any natural human language. The minimal element of language structure is word that defines single concept. Words are organized into sentences. The language can be used in translation programs, search systems, etc.*

Ключові слова – artificial language, universal language, computer translation

## I. Вступ

Метою даної роботи є розробка універсальної штучної мови, призначеної для комп'ютерного перекладу. Основною вимогою є можливість перекладу з цієї мови на будь-яку існуючу природну мову з найменшими витратами часу та найбільшою точністю. Мова призначена як проміжна при перекладі з однієї природної мови на іншу. Її суттєва перевага полягає в тому, що, по-перше, вона зберігає зміст виразу навіть при помилках перекладу, по-друге, вираз, перекладений на універсальну мову, може у подальшому бути швидко перекладеним на будь-яку кількість природних мов. Це дозволяє забезпечити високу якість перекладу порівняно з існуючими програмами-перекладачами, створювати багатомовні Інтернет-ресурси без копіювання інформації різними мовами, а також організувати комп'ютерний пошук незалежно від мови, якою написаний шуканий ресурс. Це досягається орієнтацією універсальної мови на відображення понять та простою організацією речень, що виключають синонімію, складні граматичні конструкції, що різняться в різних мовах, та ін.

## II. Основні положення

Загальна структура мови орієнтована на лексику та синтаксис, тому що морфологія та граматики різних природних мов суттєво відрізняється та реалізація всіх існуючих граматичних категорій практично неможлива. З іншого боку, лексичні поняття, що використовуються різними мовами, практично однакові. Синтаксис різних мов також відрізняється, але всі існуючі структури можуть бути зведені до схеми з суб'єктом, об'єктом, дією та їхніми характеристиками. Таким чином забезпечується асоціювання структури слів та речень універсальної мови з аналогічними структурами інших мов.

Оскільки морфологічні категорії в універсальній мові відсутні, слова несуть лише змістове (лексичне) навантаження та мінімальне граматичне значення, що потрібне лише для більш зручної класифікації та не відіграє ролі зв'язку слів. Зв'язок слів визначається

виключно синтаксичними конструкціями. Таким чином, універсальна мова може бути порівняна з ізолюючими мовами на зразок англійської чи китайської.

Слова складаються з літер, що є цифрами шістнадцяткової системи числення від 0 до E, цифра F зарезервована. Перша цифра означає частину мови, до якої належить слово, наступні цифри конкретизують значення слова. Словник універсальної мови має деревовидну структуру.

Речення складаються зі слів та спеціальних послідовностей. Спеціальні послідовності визначають синтаксичну роль слова та фактично відіграють роль розділових знаків та деяких службових слів.

## III. Структура слова та словника

Перша літера будь-якого слова визначає частину мови, до якої воно належить. Частини мови відображають синтаксичну роль слова та не обов'язково повинні співпадати з частинами мови в українській чи будь-якій іншій мові.

Перелік частин мови та цифр (літер), що їм відповідають, поданий нижче.

0-2 – об'єкти. В реченні відіграють роль підмета чи додатку. Відповідають іменникам та особовим займенникам української мови.

3-5 – дії. Відіграють роль присудку. Відповідають дієсловом.

6-8 – ознаки об'єктів. Відіграють роль означень. Відповідають прикметникам, дієприкметникам, присвійним займенникам, числівникам.

9-B – ознаки дій. Відіграють роль обставин. Відповідають прислівникам, дієприслівникам та прийменникам (наприклад, прийменник «в» фактично може бути замінений прислівником «всередині»).

C-E – знаки, що не перекладаються. Це можуть бути певні цифри, математичні оператори, позначення фізичних констант та ін.

Числівники відносяться до ознак об'єктів, але становлять окрему групу. Вони утворені зі службових літер, що повідомляють про те, що це числівник, та числа у десятковій системі числення.

Слова, аналогів яким немає в словнику універсальної мови (найчастіше власні іменники), можуть бути передані шляхом транслітерації.

Друга та наступні літери слова конкретизують його значення. Наприклад:

0 – істота (дехто)

00 – розумна істота

000 – людина

Довжина слова не обмежена. В реченнях можуть використовуватись поняття будь-якого рівня (як 00, так і 000 та його похідні).



Рис. 1. Приклад структури словника

Як видно з Рис. 1, словник організований у формі дерева. Його структура дозволяє швидко додавати чи вилучати цілі гілки, кожна з яких відповідає категорії чи групі понять.

#### IV. Структура речення

Речення є основною складовою одиницею тексту на універсальній мові. Кожне речення складається зі слів, розділених спеціальними послідовностями. Спеціальна послідовність – група символів, що починається з зарезервованої літери F. Кожна спеціальна послідовність визначає синтаксичну роль слова чи конструкції, що знаходиться безпосередньо за нею. Приклади основних спеціальних послідовностей:

- F0 – підмет
- F1 – присудок
- F2 – додаток
- F3 – означення
- F4 – обставина

Окрім членів речення, спеціальні послідовності визначають певні форми слів, наприклад, час дієслова.

Загальна структура речення також деревовидна та подана на Рис. 2.



Рис. 2. Приклад структури речення

В подальших прикладах слова будуть передані курсивом, а спеціальні послідовності – квадратними дужками.

Так, речення *Вранці він йшов вулицею* буде передано так:

[підмет]він[присудок][минулий час]йти[обставина]вранці[обставина]вздовж вулиця[кінець речення]

На місці членів речення можуть знаходитись як синтаксичні конструкції, наприклад, дієприкметникові чи дієприслівникові звороти, так і підрядні речення. Речення *Він йшов вулицею, що вела до його будинку* передається як

[підмет]він[присудок][минулий час]йти[обставина]вздовж вулиця [означення] [підмет] вона [присудок][минулий час] вести [обставина] до будинок [означення] його [кінець речення][кінець речення]

Сурядні речення та однорідні члени речення записуються поруч та розділяються спеціальними послідовностями, що замінюють коми чи сполучники «і».

Вставні слова, звернення та інші конструкції також визначаються певними спеціальними послідовностями, всередині яких розміщені слова чи речення. Ці ж послідовності використовуються для вставки різноманітних спеціальних символів. Така конструкція означає, що символ займає місце в реченні, не будучи при цьому його членом.

Загальні питальні речення будуються так, як і стверджувальні, але спеціальна послідовність «кінець речення» замінюється послідовністю «знак питання». Речення *Чи йшов він вранці вулицею?* передається як

[підмет]він[присудок][минулий час]йти[обставина]вранці[обставина]вздовж вулиця[знак питання]

Спеціальні питання, що використовують питальні займенники типу «хто», «що», «де» та т. і. також отримуються зі стверджувальних заміною відповідного слова таким займенником. Так, речення *Де він йшов вранці?* Передається як

[підмет]він[присудок][минулий час]йти[обставина]вранці[обставина]де[знак питання]

Умовний чи наказовий спосіб дії передається спеціальною послідовністю перед присудком таким же чином, як і час дієслова.

#### Висновок

Подана вище структура слів та речень забезпечує зручне та однозначне уявлення конструкцій будь-якої мови, в ній відсутня синонімія на інші неоднозначності. Загальна структура системи універсальна, проте програми-перекладачі для будь-якої мови повинні створюватися з урахуванням конкретних особливостей. Навіть з невеликою кількістю програм-перекладачів система може ефективно функціонувати для машинного перекладу чи пошуку інформації, будучи точніше та універсальніше існуючих систем.