

Проблематика морфологічного аналізу сучасної турецької мови під час машинного перекладу

Всю лексику мови можна поділити за різноманітними критеріями, такими як фонологічні особливості, соціальні фактори та історія мови. Проте всі ці класи слів відносяться до десяти традиційних частин мови (іменник, дієслово, прикметник, прислівник, займенник, прийменник або післяіменник, сполучник, числівник, артикль та вигук).

Термін «частиномовна класифікація» вперше був запропонований на початку XX ст. структуралістами. Існує дві найбільш актуальні проблеми частиномовної класифікації:

1. частину мови де-яких слів важко визначити, через можливість цих слів відноситись до різних категорій,

2. де-які класи слів мають різну кількісну характеристику (так є мови, в яких є безліч іменників та дієслів, а займенники та сполучники майже відсутні) [6: 16539].

Рішенням цих проблем є поділ частин мов на самостійні (іменник, дієслово, прикметник та прислівник) та службові (всі інші). Проте в прибічників структуралізму на початку XX ст. виникає потреба в більш чіткому та загальному описі кожної частини мови для окремо взятої мови, оскільки загальномовиний розподіл не був прийнятним для конкретної мови. Так було вирішено надати перевагу в цій класифікації морфологічним та синтаксичним критеріям. Але деякі науковці, нехтуючи синтаксичним критерієм, за морфологічними ознаками виділяли лише дієслово та іменник, об'єднуючи в одну групу іменник та прикметник.

Рішення цієї проблеми – це частиномовна класифікація для кожної окремо взятої мови, але за загальноприйнятими критеріями [8: 285-317].

Так, іменником називаються слова, які позначають предмет або особу, дієсловом – слова на позначення дії або процесу, прикметником – слова, що позначають ознаку.

Проте турецькі науковці виділяють наступні критерії, за якими відбувається частиномовна класифікація:

1. значення – досліджуючи частиномовну класифікацію турецької мови, зазвичай дослідники звертають увагу на семантичне значення слова. Самім Сінаоглу, керуючись цим критерієм, виділяє такі класи: слова, які позначають предмети, та слова, що їх не позначають [7: 200-201].

2. форма – цей критерій враховує всі додані афікси, граматичні зв'язки з іншими словами. На думку таких вчених, як Озан Башкан, Андре Мартінет, значення є змінним поняттям, а форма – більш сталим та точним [2: 100-101].

3. функція – сьогодні науковці визначають частини мови відповідно до функцій, які слово виконує в реченні.

4. позиція – класична морфологія визначає частини мови відповідно до місця, яке це слово займає в реченні.

Таким чином, в турецькій мові сполучник та прикметник завжди стоять в препозиції, тому Доан Аксан, Севгі Озель та Нуреттін Коч замість поняття «прикметник» використовують «передіменник».

Турецька мова займає особливе місце серед природніх мов, оскільки є аглютинативною, та всі притаманні їй аффікси виступають в ролі суфіксів. Слова цієї мови формуються за допомогою додавання суфіксів до кореня слова. Таким чином утворене слово може бути досить довгим, та відповідати навіть цілому реченню в англійській та українських мовах. Проте схема додавання аффіксів є чітко визначеною, це дозволяє створювати технічні (комп'ютерні) методи перекладу.

Ще одна особлива риса турецької мови – це те, що людина може аналізувати будову слова завдяки аффіксам, не знаючи значення самого кореню. Фонологічні особливості турецької мови так само впливають на її морфологічну будову.

Турецька, як аглютинативна мова, має дуже різноманітну морфологічну будову. Семантичне значення слова може змінюватись в залежності від аффіксів. Єдиним типом аффіксів в турецькій мові є суфікси, тому під час морфологічного аналізу слова потрібно віднімати аффікси з кінця слова, й таким чином рухатись до кореня. Цей метод може бути названим «аналіз справа наліво». Наприклад, слово "kalem" перший аффікс "-im" є присвійним аффіксом, за ним іде "-ler" аффікс множини, а вже далі сам корінь слова "kalem" [3: 2].

Аффікси турецької мови мають безліч варіантів, через закон гармонії голосних та приголосних. Хоча в загальному всі аффікси турецької мови можна класифікувати наступним чином:

№	клас	тип
1	іменні дієслівні суфікси	флексійні
2	дериваційні суфікси	дериваційні
3	номінативні суфікси	флексійні

4	дієслівні суфікси часу та особи	флексійні
5	дієслівні суфікси	флексійні

В цій статті ми зупинимось на дериваційних суфіксах, адже ці аффікси можуть змінювати часину мови слова, що є нажзвичайно важливим під час перекладу. Важливим в цьому питанні є те що ці дериваційні аффікси завжди додаються перед флексійними. Тому під час розробки комп'ютерних програм перекладу ми зтикаємось з проблемою цього морфологічного аналізу. Адже ми можемо визначити частину мови лише після аналізу флексійних аффіксів (число, особа, час і т.д.).

Проте завдяки частиномовній класифікації всі турецькі слова можуть бути поділені на два великих класи: іменні та дієслівні слова. Та будь-яке слово може змінити свій клас завдяки додаванню аффіксів, так, іменник може перейти в розряд дієслова та навпаки. Тому під час частиномовної класифікації варто керуватись сукупністю трьох принципів:

- морфологічна будова
- функція слова в реченні
- морфологічні зв'язки слова.

Отже, для визначення частин мови потрібно провести морфологічний аналіз основ слова та афіксів. Також встановити типи цих основ та афіксів та встановити, яке відношення вони мають до частин мови. Баскаков О.М. виділяє в слові наступні структурні елементи:

1. корінь слова, який у тюркських мовах може бути самостійною лексемою та може мати лексичне значення та граматичну форму;
2. перша основа, яка складається з кореня та словотворчого афіксу, що визначає лексичне значення;
3. друга основа, яка складається з кореня або першої основи та функціонально-граматичних афіксів;
4. афікси словозміни, які вказують на відношення даного слова до інших слів у складі словосполучення та речення [1: 65].

Кожен аффікс відображає певну морфологічну характеристику слова, включаючи частиномовний поділ. Морфологічний аналіз виявляє досить різноманітну структуру слова. Проте не всі ці риси є необхідними для комп'ютерного перекладу. Крім того, деякі особливості певних характеристик слова можуть залежати лише від функцій цього слова. Отже під час машинного перекладу морфологічний аналіз відбувається без семантичного аналізу.

Для аналізу іменних слів використовуються два різних метода, в залежності від того, чи це слово є залежним чи основним:

- якщо слово є залежним (а визначити, чи слово залежне, можна лише за допомогою останнього аффіксу), ми сприймаємо його як маркер відмінку, який в свою чергу, визначає синтаксичну функцію слова в реченні,
- якщо слово є основним, ми визначаємо лише його частину мови та аффікси присвійності (морфологічні зв'язки в ізафетах).

Прикметники вимагають особливої уваги, якщо вони використовуються у якості залежних слів, ми все рівно визначаємо їх частиномовну приналежність, якщо ж вони використовуються у якості головного слова, ми визначаємо і частиномовний аффікс і його морфологічні зв'язки [4: 92].

Zalal Güngözü та Kemal Oflazer розробили морфологічно граматичний аналіз турецької мови на базі машинного перекладу. Вони використали дворівневий аналіз морфологічної будови турецьких слів, взявши за основу 24 000 кореневих слів. Ця база використовувалась саме для морфологічного аналізу, тому дещо обмежила лексичний та синтаксичний аналіз. Важливим моментом цієї роботи є визначення прикметника як повнозначної частини мови. Проте вони зазначають, що всі прикметники турецької мови, можуть функціонувати як іменники [5: 5].

Отже, для створення комп'ютерних програм перекладу, перед усім потрібно розробити схему морфологічного аналізу кожного слова. Оскільки турецька мова є аглютинативною, будова всіх слів є подібною, вони складаються з основи та аффіксів, доданих до неї. Є певний порядок додавання цих аффіксів, спочатку до основи додаються словотворчі (дериваційні), а потім словозмінні (флексійні). Таким чином, морфологічний аналіз слова в турецькій мові можна провести без знання семантики кореня слова. Проте існує важлива проблема. Комп'ютерний аналіз проводиться справа наліво, поступово віднімаючи аффікси й просуваючись до кореня, але першим етапом морфологічного аналізу повинна бути частиномовна класифікація, а в нашому випадку це є неможливим, оскільки аффікси – показники частиномовної приналежності – стоять перед словозмінними. Вихід з цієї проблеми запропонували турецькі лінгвісти. Вони поділили всі слова на два великих класи: іменні слова та дієслівні. Так як для комп'ютерного перекладу це є достатнім, подальший аналіз проводиться, спираючись на цю класифікацію, пропускаючи такі важливі частини мови, як прикметник та прислівник, що в свою чергу призводить до неточностей у перекладі.

Література

1. Баскаков Н.А. Морфологическая структура слов и части речи в тюркских языках // Совет. востоковедение. – 1957. 2. Başkan Ö. Lengüistik Metodu. – Multilingual Yay. İstanbul, 2003. 3. Eryiğit G., Adalı E. An affix stripping morphological analyzer for Turkish. – Innsbruck, 2004. 4. Eryiğit G., Oflazer K. Statistical Dependency Parsing of Turkish. – İstanbul, 2004. 5. Güngözü Z., Oflazer K. Parsing Turkish using the Lexical Functional Grammar Formalism. – Ankara, 2003. 6. Haspelmath M. Word Classes and Parts of Speech. / International Encyclopedia of the Social & Behavior Sciences, 2001. 7. Sinanoğlu S.: Kelimelerin Bölüklere Ayrılması. – Türk Dili, 1959. 8. Wierzbicka A.: Lexical prototypes as a universal basis for cross-linguistic identification of "parts of speech". – Berlin, 2000.