

*Михайлова Є. В.,
кандидат філологічних наук,
доцент кафедри романо-германської філології
Рівненського державного гуманітарного університету*

ЛІНГВІСТИЧНІ ОСНОВИ МОДЕЛЮВАННЯ ТУРЕЦЬКО-УКРАЇНСЬКОГО ПЕРЕКЛАДУ

Анотація. У статті розглядаються лінгвістичні основи моделювання для систем автоматичного турецько-українського перекладу. Також аналізуються етапи моделювання, а саме принципи укладання словникової бази, способи граматичного аналізу та проектування граматик і алгоритмів, методи синтаксичного синтезу турецького та українського речень. Окремо визначаються морфологічні трансформації, які найчастіше реалізуються під час машинного перекладу в турецько-українській мовній парі.

Ключові слова: лінгвістичне моделювання, автоматичний переклад, морфологічні трансформації, синтаксична функція.

Постановка проблеми та обґрунтування актуальності її розгляду. До числа компонентів перекладацької діяльності, модельованої в теорії перекладу, входять протиставлені один одному мови, тексти, культури і ситуації. Таким чином, у процес перекладу залучаються як мовні, так і позамовні компоненти. Однак питання про їх співвідношення та роль кожного з них ще не повністю досліджена. Пропонована робота має на меті описати принципи лінгвістичного моделювання під час автоматизованого турецько-українського перекладу.

Аналіз останніх досліджень і публікацій. Нині моделювання мовних одиниць у системах машинного перекладу аналізуються у працях Ю.М. Марчука, Л.Л. Нелюбіна. Функціональні характеристики систем машинного перекладу описуються в наукових дослідженнях В.П. Бурдаєва, Г.А. Гайна, В.Г. Долі та ін. Морфологічний аналіз слів турецької мови, їх структурно-морфологічні моделі, що є основою для адекватного моделювання мовних одиниць у системах машинного перекладу, досліджуються у роботах Є.В. Михайлової та К. Офлазера.

Об'єктом дослідження є системи машинного перекладу Google Translate, Windows Live Translator Bing та Pragma 6.

Предметом дослідження є лінгвістичні принципи моделювання у турецько-українському машинному перекладі.

Мета цього дослідження полягає у визначенні граматичних моделей аналізу, опису принципів організації словника під час машинного перекладу, а також моделювання структури українського та турецького речень під час машинного перекладу.

Виклад основного матеріалу. Моделлю перекладу називається умовний опис розумових операцій, виконуючи які перекладач може здійснити переклад всього оригіналу або деякої його частини [6, с. 158]. У лінгвістичній теорії перекладу моделі перекладу являють процес перекладацької діяльності у вигляді низки розумових операцій над мовними одиницями, тобто у вигляді лінгвістичних операцій, вибір яких обумовлюється мовними особливостями оригіналу та відповідними явищами в мові перекладу. Модель перекладу носить умовний характер, оскільки вона необов'язково відображає реальні дії переклада-

ча в процесі створення тексту перекладу. Більшість таких моделей має обмежену пояснювальну силу і не претендує на те, що на їх основі може бути реально здійснений переклад будь-якого тексту з необхідним ступенем еквівалентності. Завдання моделі полягають лише в тому, щоб описати послідовність дій, за допомогою яких можна виконати перекладацьке завдання за умов процесу перекладу. Моделі перекладу розкривають окремі сторони функціонування лінгвістичного механізму перекладу. Необхідно звернутися до моделей перекладу, в яких відображаються найбільш істотні, з точки зору теоретиків перекладу, сторони цього процесу [6, с. 169].

Одну з таких моделей запропонував свого часу відомий фіхівець в області теорії перекладу О. Каде [5]. Відповідно до цієї моделі переклад розпадається на три фази: 1) комунікація між відправником і перекладачем, який виступає у функції одержувача вихідного тексту; 2) зміна коду, здійснювана перекладачем, який виступає у функції декодуючого; 3) комунікація між перекладачем, який виступає у функції відправника кінцевого тексту, і одержувачем цього тексту.

Під час письмового машинного турецько-українського перекладу можна виділити такі операції: основні – 1) читання та розуміння тексту-оригіналу; 2) переклад окремих слів; 3) буквальний переклад тексту; 4) еквівалентний переклад зв'язного тексту; 5) словесне оформлення зв'язного тексту адекватно; допоміжні – 1) письмове фіксування тексту-оригіналу і тексту-перекладу; 2) пошук необхідних слів у словнику; 3) перевірка орфографії тексту-перекладу: форматування, стилізація; 4) збір статистики про перекладений текст: кількість знаків, слів, рядків.

Практично всі операції за нинішнього розвитку інформаційних технологій досить легко автоматизуються. А саме всі допоміжні операції, переклад окремих слів та буквальний переклад тексту. Інші операції автоматизуються меншою мірою, проте більш-менш успішні спроби все одно здійснюються.

Граматичні моделі аналізу. Граматика і словник задають лінгвістичну модель, утворюють основну частину лінгвістичних даних. Алгоритми їх оброблення, тобто співвіднесення з текстовими одиницями, відносять до математично-алгебраїчного забезпечення системи. Поділ граматик і алгоритмів є важливим у практичному сенсі тим, що дозволяє змінювати правила граматики, не змінюючи алгоритмів (і відповідно програм), що працюють з граматиками. Найбільш чітко поділ граматик та алгоритмів спостерігається в системах, що працюють з контекстно-вільними граматиками, де модель мови – граматика з кінцевим числом станів, а алгоритм має забезпечити для довільно взятого речення його виведення за правилами граматики, і якщо таких висновків кілька, то перерахувати їх. Такий алгоритм називають аналізатором. Аналізатори будуються для

класів граматик, хоча врахування специфічних особливостей граматики може підвищити ефективність аналізатора.

Ланцюгова граMATика фіксує порядок проходження елементів, тобто лінійні структури речення, застосовуючи їх у термінах граматичних класів слів (іменник + прикметник) або в термінах функціональних елементів (підмет + присудок). Прикладом реалізації такої мовної моделі є певний синтаксичний аналіз: ідентифікуюча граматична категорія слова пророкує (з певною часткою ймовірності) появу граматичної категорії наступного за ним слова.

ГраMATика складових фіксує лексичну інформацію про групу граматичних елементів, наприклад, іменна група (складається з іменника, прикметника та інших модифікаторів), прикметникова група (складається з прикметника та іменної групи) та ін. до рівня речення.

ГраMATика залежностей задає ієрархію відносин елементів речення. Такий аналіз заснований на ідентифікації «головних» та їх «залежних». Головним у реченні є дієслово в особовій формі, оскільки воно визначає число і характер залежних іменників. Стратегія аналізу – зверху вниз: спочатку ідентифікуються «головні», потім «залежні», або знизу вгору: «головні» визначаються процесом підстановки.

Для *контекстно-вільних граматик* існує широкий вибір методики побудови аналізаторів. Проте чисті контекстно-вільні граматики недостатні для опису природної мови. Було запропоновано доповнити їх набором трансформаційних правил, що працюють зі складових.

Уніфікаційні граматики являють собою узагальнення моделі аналізу: вони здатні втілювати граматики різних видів. Уніфікаційні граматики містять чотири компоненти: пакет уніфікацій, інтерпретатор для правил і лексичних описів, програми обробки спрямованих графів, аналізатор, що працює за допомогою граф-схеми. Такі граматики об'єднують граматичні правила зі словниковими описами, синтаксичні валентності із семантичними [8, с. 56].

Незважаючи на поширену думку щодо вирішальної ролі граматики у створенні кращої системи машинного перекладу [1, с. 12–17], а також на те, що більшість існуючих систем базуються саме на підході використання граматичних правил, багато хто стверджує, що системи машинного перекладу можуть гуртуватися на інших принципах. Наприклад, автори роботи «Автоматичний переклад на інформаційній мові» доводять, що відсутність граматики у пропонованій ними системі «викликає п'яту частину інформаційного шуму», і відкидання синтагматичних відносин «цілком узгоджується з характером інформаційно-пошукової мови, яка є спрощенням природної мови» [9, с. 78], а творці Google Translate стверджують, що статистичний підхід зарекомендував себе краще, ніж інші [3].

Організація словника під час машинного перекладу. Електронні словники почали розроблятися значно раніше, ніж альтернативні їм системи машинного перекладу, тому натеper на ринку програмного забезпечення існує надзвичайно широкий вибір словників – від найпростіших (наприклад, DIC) до потужних систем, що поєднують в одній програмній оболонці кілька лексичних баз даних – спеціалізованих тематичних словників, послідовність підключення яких визначається користувачем (LINGVO).

Словники аналізу, як правило, одномовні. Вони мають містити всю інформацію, необхідну для включення даної лексичної одиниці в структурне подання. Часто поділяють слов-

ники основ (з морфолого-синтаксичної інформацією: частина мови, словозміна, характерна синтаксична функція лексичної одиниці), і словники словозначень, що містять семантичну і концептуальну інформацію: семантичний клас лексичної одиниці, семантичні відмінки (валентності), умови їх реалізації та ін. [2, с. 36].

Принципова відмінність словника для машинного перекладу полягає в тому, що в ньому містяться не граматично оформлені слова, а основи слів. При цьому використовується спеціальне поняття основи, що є відмінним від традиційного. Основою називають частину слова, яка залишається графічно незмінною у всіх його формах. Наприклад: *yaz-* (основа дієслова *yazmak /nucatum/*), що може у формі іменника мати закінчення *-ı* (*yazı /nupus/*) в однині не прислонує афікс, а у множині прислонує закінчення *-lar/ler*). Тобто, щоб перекладати речення, що містять слова *defteri /zouum/*, *deftere /zouumy/* і т.ін., добре було б мати спосіб співвіднесення словникової статті з автоматичного словника для слова *defter /zouit/* з відповідною словоформою з тексту. Тому для опису як вхідної, так і вихідної мови в системі має існувати певний формальний метод опису морфології, на якому ґрунтується вибір одиниці словника.

Деякі форми слів, які розглядаються як особливі, що не підкоряються загальним правилам формоутворення, були введені у словник як самостійні лексеми, кожне зі своєю граматичною інформацією. Наприклад, іменник *nehir /pıcka/* під час прислонування афіксів, втрачає свій голосний з останнього складу *-ı*: *nehre /do pıckı/*, *nehrin /pıckı/*. До словника включені також розділові знаки. Для кожної основи турецькою мовою у словнику міститься її перекладний еквівалент разом із додатковою спеціальною інформацією.

Для багатозначних слів використовують наступний метод: 1) декілька близьких значень підводять під одне загальне, наприклад, *gıtmek /ıtu, nımu/* та *yügmek /ıtu nıshı, ruhatısya/* отримало один переклад – йти; 2) значення, що є відмінними, наприклад: *çalmak /grati, krasıtı/* отримало один переклад – грати.

Тут розрізняють два випадки: 1) декілька перекладів турецького слова належить одній і тій самій частині мови; 2) різні переклади турецького слова належать різним частинам мови. У першому випадку ми маємо справу з полісемією, або лексичною омонімією. Вибір потрібного перекладу формально не визначений, він залежить від вмісту фрази, тобто у випадках лексичної омонімії машина буде видавати всі переклади даного слова. У другому випадку ми маємо справу з лексико-граматичною омонімією, усунення якої передбачене у даному алгоритмі. На основі формальних ознак за контекстом з'ясується, який з перекладів підходить у кожному конкретному випадку.

У словникових базах програм-перекладачів безліч закінчень для кожної мови зберігається у вигляді певних структур, що забезпечують не тільки ефективний спосіб зберігання, а й ефективний алгоритм морфологічного аналізу.

Проте розроблення опису морфології дозволяє вирішити лише проблему того, що є назвою словникової статті, за якою відбувається ідентифікація одиниці тексту й одиниці словника. Але ж ідентифікація слова з тексту зі словниковою статтею відбувається не заради ідентифікації, як це потрібно в електронних словниках, вона необхідна для виконання програмою власне процедури перекладу. Для того, щоб програма перекладала, кожна стаття має бути оснащена відповідною граMATикою.

Синтез структури українського та турецького речень.

Турецьке речення та його структура дуже своєрідні. Згідно з останніми дослідженнями, проведеними у Дослідницькій лабораторії мозку та мови у технічному університеті Близького Сходу (ÖDTÜ), через структуру турецької граматики, носії турецької мови здійснюють вдвічі більше мозкових операцій, ніж англо-мовне населення. Оскільки у турецькій мові дієслово стоїть у кінці речення, і необхідні граматичні показники також приєднуються до останнього складу, то для розуміння фрази людина має утримувати в свідомості її цілком від початку до кінця.

Структура турецького речення має такий порядок: означення до підмета, підмет, означення до додатку, додаток, означення до обставини, обставина, присудок. Тобто, означення завжди передують означальному: *Güzel kız büyük kitabı ferah odada okuyor. /Гарна дівчинка читає велику книгу у просторій кімнаті/. Alçak gönüllü kadın fakir insanlara çok para ve yemek bağışlamış. /Добра жінка дала на пожертву бідним людям багато грошей та їжі/. Benim eski arkadaşlarım evimin bahar bahçesinde geziyordu. /Мої давні друзі гуляли по весняному саду мого будинку/.*

В українській мові основою типового речення є підмет у його зв'язку з присудком, навколо яких згуртовуються у поширеному реченні його другорядні члени: означення й прикладки в групі підмета та додатки й обставини в групі присудка. Порядок слів є більш вільним [4, с. 369].

У турецьких формулах морфологічне оформлення відіграє головну роль, позаяк у турецькій мові багато граматичних форм і особливо таких, які є характерними для тієї чи іншої граматичної конфігурації. У процесі синтезу здійснюється послідовне перенесення граматичної інформації. При цьому переноситься не тільки конкретна граматична інформація, але і вказівки про узгодження тощо. За синтезу можуть зустрітися випадки протиріччя граматичної інформації, яка поступає з формул, та інформації, отриманої з різного роду узгоджень. У подібних випадках перевага надається інформації, що міститься у формулах. Вона враховується, а інформація, основана на вказівках про узгодження, відкидається.

Номери всіх турецьких формул, що застосовувались у процесі аналізу, зафіксовані, так само зафіксовані порядкові номери тих елементів речення, до яких ці формули застосовувались. У цьому випадку, оскільки вихідна мова українська, то її формули можуть містити багату морфологічну інформацію, зокрема інформацію про граматичні узгодження, що переноситься від формули до формули. У результаті синтезу отримується схема побудови українського речення, у якій зазначено порядок слів та їх граматичні форми, або безпосередньо, або у вигляді узгоджень одного слова з іншим. У випадку неоднозначної відповідності між українською та турецькою конфігураціями, тобто одній турецькій формулі відповідає декілька українських, додатково проводиться вибір однієї з них, що єдиним чином підходить у кожному конкретному випадку. При цьому іноді робиться не просто вибір однієї з декількох формул, але і запис додаткової інформації та перестановка слів.

Таким чином, у загальному вигляді задача синтезу полягає у підставленні на зазначені місця у реченні формул із вказаними номерами.

Морфологічні трансформації. Граматичні трансформації полягають у перетворенні структури речення в процесі перекладу відповідно до норм перекладної мови. Вони можуть відбуватися на синтаксичному та на морфологічному рівнях. Розглянемо заміни, що здійснюються на морфологічному рівні.

Морфологічні трансформації – такий спосіб перекладу, коли граматична одиниця в оригіналі перетворюється на одиницю

мови перекладу з іншим граматичним значенням. У перекладацькій практиці зустрічаються різні види граматичних заміни.

Заміни частин мови. Подібні заміни зазвичай викликані різним вживанням слів і різними нормами сполучуваності в турецькій і українській мовах, а в деяких випадках – відсутністю частини мови з відповідним значенням в українській мові.

Найпростішим видом такої заміни є перетворення іменника турецької мови в прикметник українською мовою, яке відбувається під час перекладу ізафетних конструкцій: **annenin elleri** /*маміні руки*/, **çocuğun arabası** /*дитяча машинка*/, **mermer merdiven** /*мармурові сходи*/, **pencerenin camı** /*віконне скло*/, **yün çorap** /*вовняні шкарпетки*/, **taş yürek** /*кам'яне серце*/.

Для турецької мови характерна тенденція до субстантивізації. Як і в українській, іменник, якому передують дієприкметник, може опускатися, таким чином, дієприкметник субстантивується, тобто потрапляє в синтаксично незалежну позицію. У цьому випадку дієприкметник отримує синтаксичні та морфологічні властивості іменника: змінюється за відмінками, приєднує афікс множини тощо: **artık satılanlarına benzeyeni de alabilirsiniz** /*ви можете придбати схоже на вже продане*/, **bunu bilenler de var** /*є ті, хто це знає*/, **Gelen geçer, konan göçer** /*все минає*/, **Minareyi çalan kılıfını hazır(lar)** /*той, хто вкрав мінарет, вже знайшов схованку для нього*/.

Використовуючи під час перекладу будь-які граматичні заміни частин мови, перекладачу часто доводиться супроводжувати їх також і заміною членів речення, тобто перебудувати синтаксичні структури речення. Таким чином, синтаксичні та морфологічні трансформації постійно перебувають у взаємодії.

Заміна числа. У турецькій та українській мовах існують форми однини і множини, і, як правило, відповідні іменники в оригіналі і в перекладі вжито в тому ж самому числі, за винятком випадків, коли формі однини в турецькій відповідає форма множини в українській (*para* /*гроші*/, *ayakkabı* /*взуття, туфлі*/ тощо).

Заміна граматичного часу. Як уже було зазначено, на відміну від української мови, у турецькій не завжди часова форма передає основне своє значення у реченні. У зв'язку з цим виникає необхідність заміни граматичного часу під час перекладу: **Yarın İstanbul'a gidiyorum.** /*Завтра я їду у Стамбул*/, **Şanslısın, tatile gidiyorsun.** /*Щасливчик, ти поїдеши відпочивати*/, **Bu işi bugün tammamlayacağım.** /*Я повинна завершити цю справу сьогодні*/, **Yarın hastaneye gideceksin.** /*Ти завтра маєш поїхати в лікарню*/, **Diyorlar ki, o hiç bir yere gitmemiş.** /*Кажуть, вона нікуди не їде*/, **Babam gelmiş.** /*Мій батько, напевно, вже прийшов*/, **Ablam yarın gelir.** /*Моя старша сестра прийде завтра*/.

Висновки та перспективи подальших досліджень у цьому напрямі. Проведений аналіз моделювання систем автоматичного перекладу дає змогу зробити висновки, що основними лінгвістичними етапами є створення словника, вибір граматики та алгоритмів відповідно до мовної пари та синтез синтаксичних структур. Принцип моделювання використовується у теорії перекладу не для побудови моделі як такої, а для усвідомлення та аналізу певних уявлень, що виникають у мозку та конструюють процес перекладу. Граматика і словник задають лінгвістичну модель, утворюють основну частину лінгвістичних даних. Поділ граматики і алгоритмів є важливим у практичному сенсі тим, що дозволяє змінювати правила граматики, не змінюючи алгоритмів (і відповідно програм), що працюють з граматиками. Принципова відмінність словника для систем машинного перекладу полягає в тому, що в ньому містяться не граматично

оформлені слова, а основи слів. Синтез структури речення українською мовою полягає у підставленні на місце кожної згорнутої турецької конфігурації відповідної української конфігурації.

Література:

1. Бархударов Л.С. Тетради переводчика. Вып. 15. – М.: Международные отношения, 1978. – 112 с.
2. Вейзе А.А., Мирончиков И.К. Перевод технического текста с английского языка на русский. – Мн.: МГЛУ, 1995. – 168 с.
3. Гензель Д.И. Создание системы машинного перевода (Электронный ресурс) – URL: http://google.ru/translate_tools?hl=ru 56.
4. Енциклопедія українознавства: Словникова частина / Ред. В. Кубійович; Наукове товариство ім. Т. Шевченка у Львові. – Л., 1993. – 400 с.
5. Каде О. Проблемы перевода в свете теории коммуникации: (пер. с нем. А. Батрака). Вопросы теории перевода в зарубежной лингвистике. – М.: Международные отношения 1978. – С. 54–73.
6. Комиссаров В.Н. Теория перевода. М.: Высшая школа, 1990. – 253 с.
7. Михайлова С.В. Помилки на трансформаційних рівнях турецько-українського перекладу під час роботи із СМП. Науковий вісник Дрогобицького державного педагогічного університету імені Івана Франка. Серія: «Філологічні науки» (мовознавство): Збірник наукових праць. – Дрогобич, 2017. – № 8. – Том 1. – С. 156–159.
8. Попов С.В. Искусственный интеллект. Системы общения и экспертные системы. – М.: Радио и связь, 1990. – 464 с.
9. Семчинський С. Л. Загальне мовознавство. – К.: АТ «ОКО», 1996. – 416 с.

Михайлова Е. В. Лингвистические основы моделирования турецко-украинского перевода

Аннотация. В статье рассматриваются лингвистические основы моделирования систем автоматического турецко-украинского перевода. Также анализируются этапы моделирования, а именно принципы создания словарной базы, способы грамматического анализа и системы грамматик и алгоритмов, методы синтаксического синтеза турецкого и украинского предложений. Отдельно определяются морфологические трансформации, которые чаще всего реализуются во время машинного перевода в турецко-украинской языковой паре.

Ключевые слова: лингвистическое моделирование, автоматический перевод, морфологические трансформации, синтаксическая функция.

Mykhaylova E. Linguistic basis for modeling the Turkish-Ukrainian translation

Summary. The article deals with the linguistic basis of the modeling of automatic Turkish-Ukrainian translation systems. Also the stages of modeling are analyzed, mostly the principles of the dictionary base, the methods of grammatical analysis and the system of grammars and algorithms, the methods of syntactic synthesis of the Turkish and Ukrainian sentences are viewed. Morphological transformations, which are most often realized at the machine translation in the Turkish-Ukrainian language pair, are defined.

Key words: linguistic modeling, automatic translation, morphological transformations, syntactic function.