



DOI: <https://doi.org/10.15688/jvolsu2.2023.3.4>

UDC 81'322.4
LBC 81.184

Submitted: 12.12.2022
Accepted: 06.03.2023

RECURSIVE COMPARISON AS A BASIS TO TRANSLATION EQUIVALENTS SEARCH IN AGGLUTINATIVE AND INFLECTIONAL LANGUAGES

Sergey N. Bredikhin

North-Caucasus Federal University, Stavropol, Russia

Stanislav G. Sidorenko

North-Caucasus Federal University, Stavropol, Russia

Abstract. The study offers a developed algorithm, aimed to observe cognitive-communicative syntactic specifiers based on recursive comparison in the process of computer-aided translation of statements in agglutinating languages. The key method that enables fast and adequate search for translation equivalent in the languages with different structures is the multilevel recursive comparison of interim forms with dictionary equivalents, as well with contextual matches within the frameworks of prospective modelling in the course of an invariant thesaurus development according to dynamic complicated codes. The combination of cognitive-semantic and semantemic-morphological parallel comparison with spiral references will make it possible to create not only an equivalent version of the target text that meets the requirements of lexico-morphological correctness, but also to ensure the transfer of cognitive and consituational elements of the original utterance. The inclusion of cognitive-communicative syntactic samples in the recursive comparison algorithm during automatic processing of the utterance, as the final stage of generating the target utterance, is designed to solve the problem of polysemic, synonymous and homonymous barriers that arise at the stage of generating the target text. The described automated analysis algorithm is demonstrated on the utterances in languages with different structures (Turkish, Russian and English). The article also provides examples of the interaction of the invariant thesaurus of commonly used constructions with variant correspondences while connecting recursive comparison with elements of cognitive-communicative syntax.

Key words: automatic translation system, translation equivalent, cognitive-communicative syntactic specifier, thesaurus of invariant bases, agglutinating languages, recursive comparison, conceptual-perceptual model.

Citation. Bredikhin S.N., Sidorenko S.G. Recursive Comparison as a Basis to Translation Equivalents Search in Agglutinative and Inflectional Languages. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 2. Yazykoznanie* [Science Journal of Volgograd State University. Linguistics], 2023, vol. 22, no. 3, pp. 43-56. (in Russian). DOI: <https://doi.org/10.15688/jvolsu2.2023.3.4>

УДК 81'322.4
ББК 81.184

Дата поступления статьи: 12.12.2022
Дата принятия статьи: 06.03.2023

РЕКУРСИВНОЕ СРАВНЕНИЕ КАК ОСНОВА ПОИСКА ТРАНСЛЯЦИОННЫХ ЭКВИВАЛЕНТОВ В АГГЛЮТИНАТИВНЫХ И ФЛЕКТИВНЫХ ЯЗЫКАХ

Сергей Николаевич Бредихин

Северо-Кавказский федеральный университет, г. Ставрополь, Россия

Станислав Геннадьевич Сидоренко

Северо-Кавказский федеральный университет, г. Ставрополь, Россия

Аннотация. В статье представлен авторский алгоритм учета когнитивно-коммуникативных синтаксических спецификаторов на основе рекурсивного сравнения в процессе автоматического перевода высказыва-

ваний на агглютинирующих языках. Ключевым методом обеспечения быстрого и адекватного поиска трансляционного соответствия в разноструктурных языках признается многоуровневое рекурсивное сравнение промежуточных решений как со словарными эквивалентами, так и с контекстуальными соответствиями в рамках проспективного моделирования при создании инвариантного тезауруса по динамическим усложненным кодам. Установлено, что сочетание когнитивно-семантического параллельного сравнения со спиральными отсылками позволяет создать эквивалентный вариант транслята, удовлетворяющий требованиям лексико-морфологической корректности, и обеспечить передачу когнитивных и конситуативных элементов исходного высказывания. Включение в алгоритм рекурсивного сравнения при автоматической обработке высказывания когнитивно-коммуникативных синтаксических образцов в качестве финального этапа генерации целевого высказывания позволяет устранить полисемический, синонимический и омонимический барьеры, возникающие на этапе формирования выходного целевого текста. Возможности предложенного алгоритма автоматизированного анализа демонстрируются на примере высказываний на разноструктурных языках (турецком, русском и английском). Показано взаимодействие инвариантного тезауруса общеупотребительных конструкций с вариантными соответствиями при подключении рекурсивного сравнения с элементами когнитивно-коммуникативного синтаксиса.

Ключевые слова: автоматическая система перевода, трансляционное соответствие, когнитивно-коммуникативный синтаксический спецификатор, тезаурус инвариантных основ, агглютинирующие языки, рекурсивное сравнение, концептуально-перцептивная модель.

Цитирование. Бредихин С. Н., Сидоренко С. Г. Рекурсивное сравнение как основа поиска трансляционных эквивалентов в агглютинативных и флективных языках // Вестник Волгоградского государственного университета. Серия 2, Языкознание. – 2023. – Т. 22, № 3. – С. 43–56. – DOI: <https://doi.org/10.15688/jvolsu2.2023.3.4>

Введение

В рамках перманентных изменений, связанных как с научно-техническими инновациями, так и с развитием интернациональных связей в социально-культурной сфере, резко возрастают объемы национальных информационных потоков, что требует их интеграции в общецивилизационное когнитивное пространство. Создание автоматизированных систем, обеспечивающих оперативную обработку, анализ и трансляцию таких потоков, является на настоящий момент доминантной проблемой современной лингвистической кибернетики и прикладного языкознания. Основанием для адекватной трансляции когниогенеративных элементов в разноструктурных языковых системах, например в некоррелирующих парах агглютинативных и флективных языков, может служить многоуровневая концептуально-перцептивная модель [Bredikhin, Babayants, Pelevina, 2021], предполагающая создание алгоритма с учетом когнитивно-коммуникативных синтаксических спецификаторов на основе рекурсивного сравнения.

Различные виды системного анализа дискретных единиц являются базовым компонентом, на котором строятся модули автоматизированных систем переработки текста. В публикациях, посвященных таким системам, боль-

шинство исследователей пишут о «лексемно-фразовой переработке» (см., например: [Большакова и др., 2011, с. 18]), однако данный подход следует признать ошибочным, поскольку он приводит к отказу от более прецизионного анализа морфологических структур, которые в некоторых языковых системах обеспечивают финальную смыслодеривацию во фразе. В частности, автоматизированная переработка текстов на наиболее представленных в современном информационном континууме тюркских языках (относящихся к агглютинативным) должна основываться на иных принципах, чем переработка текстов на последовательно флективных и флективно-аналитических языках.

В большинстве индоевропейских языковых систем словоизменение осуществляется посредством весьма ограниченного количества аффиксов (вне зависимости от их типа), а значит, не представляет особых трудностей стереотипное извлечение исходной словоформы и приписывание ей контекстуального актуализируемого лексического значения. При оперировании словоформами агглютинирующих языков программный код сталкивается с большим количеством формантов и приведение словоформ к исходной основе существенно затрудняется. Кроме того, исходные основы с включением потенциально «бесконечного» числа формантов в парадигме аффиксирую-

шего словоизменения могут образовывать «тысячи» форм (см. об этом: [Пинес, 1974]). Большинство компаративистских исследований показывает значительное различие статистических характеристик условного формантного покрытия (достигающего семидесяти процентов) в любых произвольно взятых прецедентных текстах, а именно – объем выборки текстов, равно как и объем самих контекстов в агглютинирующем языке, должен быть в пять раз больше, чем сопоставимый объем выборки во флективно-аналитических языках и более чем в два раза превосходить объем текстовой выборки в последовательно флективных (см., например: [Бектаев, 1978, с. 37]). Данная ситуация возникает по причине вариаций супертиповых характеристик различных естественных языков, поскольку, как отмечает П.В. Дурст-Андерсен, «...грамматические системы разных языков грамматикализованы, но на основе структур сознания, которые, каждая по-своему, отражают определенные структуры действительности» [Дурст-Андерсен, 1995, с. 31]. Именно поэтому создание инварианта (канонического образца) для алгоритмизированной экстраполяции текстовых словоформ представляется одним из наиболее перспективных путей решения задачи автоматической трансляции текстов агглютинативных языков.

Составление инвариантного компаративного тезауруса с алгоритмами вычленения и параллельного транскодирования морфологических компонентов внедренного в автоматизированную систему перевода наиболее важно для тюркских языков, в которых лексемно-фразовый контекстуально-семантический анализ, сводящий к минимуму морфологический анализ словоформы, является малоэффективным в силу специфической структуры генерализованного смысла, основанного на сочетании формантов в словоформе.

Целью настоящего исследования является создание алгоритмической модели рекурсивного спуска в рамках трехчленного формантного анализа для обеспечения наибольшей степени эквивалентности трансляционного соответствия в условиях формирования высказывания на флективном языке при сохранении исходных иллокутивных компонентов высказываний на агглютинирующих языках.

Материал и методы исследования

Материалом исследования послужили некоторые устные высказывания, обладающие определенными характеристиками отнесенности, модальности, эксплицируемыми различными формантами, осложняющими структуру транслатемы. При этом единицы анализа подбирались таким образом, чтобы в их составе наличествовали форманты, которые могут быть отнесены к каждой из трех подгрупп вербализаторов лексических и грамматических значений.

Для проверки функциональности предложенного алгоритма трехчленного формантного анализа использовалась комплексная методология, включающая приемы компаративного анализа переводческих соответствий, дистрибутивного анализа для определения элементов окружения формантов (для агглютинативных языков) и словоформ (для флективных языков), а также контекстуального и интерпретативного анализа для определения степени эквивалентности иллокутивно-перлокутивного соответствия высказывания на исходном и переводящем языках.

Еще в середине прошлого века формальный анализ лексико-морфологических экспликативов большей частью носил теоретический характер, что объяснялось невозможностью эмпирической верификации моделей по причине недостаточного объема памяти и быстродействия компьютеров. Кроме того, ключевые междисциплинарные связи структурной, математической и когнитивной лингвистики были неустойчивы, отсутствовал единый методологический и терминологический аппарат. Несмотря на разработку формализованных описаний грамматических подсистем, семантическое транслирование в языковые системы несходных морфотипов оставалось недоступным (см., например: [Семенов, 2008]).

Разработка практических моделей в 1970–80-е гг. XX в. велась преимущественно на основе последовательно флективных и флективно-аналитических языков, для которых модель сравнения и выявления тезаурусных соответствий достаточно проста и строится на базе двуязычного словаря без создания инвариантного тезауруса. В случае трансляции генера-

лизованного смысла высказывания на агглютинативном языке создание простого словаря соответствий усложняет сравнение в условиях отсутствующего форманта или при введении форманта в качестве дополнительного элемента основы. Кроме того, возникают многочисленные девиации по причине омонимии и синонимии основ и формантов, что снижает скорость и качество транслята. Переход от стандартного лексемно-фразового моделирования к усложненным кодам иерархического сравнения дал возможность постепенного пополнения инвариантного тезауруса и банка формантов. Таким образом, рубеж XX–XXI вв. ознаменовал слом принципов статистического автоматизированного перевода и появление когнитивно-семантических автоматизированных систем с интеллектуальными возможностями translation memoгу. Именно на основе данных систем возможно построить принципиально новую модель лексико-морфологического анализа, представляющего базу взаимотрансляции высказываний на разноструктурных языках.

В современных исследованиях применяется трехуровневое моделирование вариативных формантов с поэтапной проверкой контекстуальных эквивалентов посредством сравнения с имеющимися соответствиями. Однако рассмотрение вариантов в рамках трехуровневой модели не дает возможности исчислить все корреляты, поскольку недостроенными остаются любые формы, организованные более чем тремя формантами. Внедрение в трехуровневую модель вне осуществления дополнительных прямых операций сравнения может производиться на основе рекурсивной отсылки при учете мультидоменного представления различных видов значения и их классификации в поддиректориях инвариантного тезауруса (см. об этом: [Наккани-Тюр et al., 2016, p. 716]).

Как показали Б.В. Орехов и Е.А. Слободян, формальные критерии выделения компонентов лексического и морфологического значений в условиях трансляции позволяют лишь частично заполнить парсеры фразовых основ высказывания [Орехов, Слободян, 2010]. Для анализа инвариантных основ и последующего построения вариативных соответствий в рамках трансляции высказывания зачастую используют модули, построенные на основе метода рекурсивного спуска [Ксалов, Гошкова, Денисенко, 2013,

с. 39]. Данный метод представляется весьма перспективным в аспекте включения в цепочку переменных таких компонентов синтаксического анализа, которые выявлены на основе взаимно возвратных парсящих процедур контекстуально свободного синтаксиса [Proudian, Pollard, 1985]. Трансформационные возможности парсинга как линейной последовательности [Fodor, 1978] дополняются сходными рекурсивными отсылками к инвариантному тезаурусу и модельным схемам распределения генерализованного смысла, на основе чего и выстраивается финальная фраза как часть целевого текста.

Упомянутые работы фрагментарно решают задачи контаминированного анализа, включающего лексико-морфологические и коммуникативно-синтаксические аспекты и служащего базой для создания семантико-ориентированной автоматизированной системы перевода для языков агглютинативного типа посредством формирования инвариантного тезауруса с параллельным банком аффиксальных формантов. Комплексность и иерархизированность многоуровневого анализа в процессе трансляции высказываний агглютинативных языков подтверждается признанной корреляцией семантики и грамматики таких систем, поскольку, как отмечается лингвистами, строгих норм сочетаемости аффиксов, репрезентирующих грамматическое значение, и корневых морфем, реализующих основное лексическое значение, в типично агглютинативных языках не существует [Рыбаков, 2004, с. 67]. В связи с отсутствием однозначных грамматических и лексических соответствий рекурсивная отсылка при сохранении трехуровневой модели может быть признана наиболее эффективным способом создания алгоритма поиска релевантного эквивалента с ограниченным количеством шагов.

Результаты и обсуждение

Классический алгоритм тюркско-русско-английского лексическо-морфологического анализа

В рамках автоматической трансляции высказывания в условиях языковой пары разноструктурных систем (агглютинирующий – флективный, изолирующий – полисинтетичес-

кий и т. п.) процесс адекватного поиска формальных соответствий значительно усложняется. Рассмотрим некоторые случаи осложненного ввода / вывода фраз на примере разноструктурных языков: триады турецкого (агглютинативного) – русского (последовательно флективного) – английского (флективно-аналитического). При таком структурном несоответствии видовременная парадигма глагола, например, будет демонстрировать существенные различия в экспликационных формантах. Так, реализация абсолютного будущего времени в английском языке осуществляется аналитически (с помощью аналитических предпозиционных вспомогательных глаголов *shall / will*), в русском (в зависимости от вида) – аналитически (глагол *быть* в будущем времени) либо синтетически (в рамках префиксального формообразования посредством *про-, с-, по-* и т. п. в зависимости от выражения инкоации, дуративности, финализации), в агглютинирующем турецком языке – синтетически (в рамках постфиксального агглютинирующего формообразования *-(y)acak*). Например: *I'll read – Я буду читать / Я прочту – Ben okuyacağım*.

При работе с бесконтекстным синтаксисом в рамках построения рекурсивного спуска строится парсер, состоящий из двух фаз взаимно-возвратной отсылки к инвариантно-

му тезаурусу. На основе этого двухфазного алгоритма создается модель комплексного анализа поверхностных структур формантов в процессе понимания глубинного содержания единиц со сложной семантикой, интерпретация которого прогнозируется самими формами выражения на основе лексических и морфологических ожиданий в рецепирующей системе [Bredikhin, Serebriakov, 2019, p. 2549]. Следует указать, что прогностика лексических ожиданий в данном случае превалирует над элементами грамматических спецификаторов глубинного содержания, подобное осуществляется в рамках «Garden-path-Modells», предполагающих, однако, расширение количества прямых операций [Konieczny et al., 1994, p. 137]. Увеличение операций происходит, как видно из схемы, в прокурсивных скачках с пропуском операции (рис. 1).

Автоматический словарь формантных соответствий в бесконтекстной привязке к словоизменительной парадигме может представлять на выводе целый ряд возможных вариативных соответствий, в которых не снимается имманентная многозначность. При этом создается не вариант адекватного перевода, а осуществляется вывод аналитической или флективной формы, которая будет формально соответствовать аффиксальной цепочке исходного высказывания.



Рис. 1. Модель лексико-морфологического анализа / синтеза в автоматизированной системе обработки текста без учета когнитивного-коммуникативного синтаксиса

Fig. 1. A model of lexico-morphological analysis/synthesis within an automated text processing system without cognitive-communicative syntax consideration

Создание единственного адекватного исходной агглютинативной форме варианта может быть осуществлено только при введении подпрограммного кода рекурсии (сверки с инвариантным тезаурусом основ), в котором соответствующее лексическое значение повторно перепроверяется на основе имеющихся в инициальной памяти формантов. Причем данный «промежуточный» результат достигается посредством рекурсивной отсылки как к знаменательным, так и к вспомогательным формам, без экспликации дополнительных обертонов, реализуемых в горизонтальном контексте. В данном случае происходит учет всех возможных параметров актуализации фразовой основы, содержащихся в ядерном компоненте высказывания (подробно об этом см.: [Бредихин, 2017, с. 420]), то есть каждый из проверяемых вариантов трансляции получает четкое место в иерархии компонентов для инвариантного тезауруса.

При установлении однозначной связи формантов грамматического значения производится соотнесение полученной формы с элементами инвариантного тезауруса для корректировки лексического значения. Выявленные в результате компаративного анализа критериальные признаки агглютинативного (единичного) грамматического значения транслируются в ряд соответствующих флективных формантов (каждый признак получает единичный модельный эквивалент), затем при рекурсивном сопоставлении с основой (лексическим значением), содержащейся в инвариантном тезаурусе, производится отбор наиболее релевантных (частотных) формантов. Таким образом, выдаются образцы эквивалентных внеконтекстуальных соответствий, которые могут быть установлены в качестве базовых для дальнейшего распознавания и опредмечивания на материале систем другой структуры, то есть получения высказывания, адекватного исходному. Сравнение внеконтекстуальных соответствий с коммуникативной задачей происходит в процессе вторичной рекурсивной обработки отдельных вариативных элементов при объяснении эффектов работы языка в пограничных случаях, например таких, как предложения с «подвохом», и в случаях предпочтительных толкований смысла высказывания [Tanaka, Branigan, Pickering, 2011, p. 123].

Сходным образом формируется и система грамматической экспликации модального пространства в разноструктурных языках. В качестве примера проанализируем трансляцию аффиксирующей формы предельно выраженного внутреннего долженствования: *Sen gitmelisin*.

В подпрограмме инвариантного тезауруса основ при первом соотнесении выделяется лексическое значение слова *gitmek* (*git* основа в инвариантном тезаурусе), соответствующее инфинитивному значению в неличной форме, без учета грамматических категорий в ряду языков флективного строя (*to go* и *идти*).

При следующем шаге в имплементированном параллельном тезаурусе формантов производится поиск наиболее релевантных в сочетаемости плане аналитических компонентов, выражающих генерализованный компонент «долженствование». Таким образом, определяется подмножество зоны аффикса долженствования (*-meli-*), что условно констатируем как промежуточный результат, который соответствует форме *быть должным* и частично соответствует английским модальным глаголам *must, should* и обороту (*to have to*). Выбор того или иного соответствия из предложенных производится на следующем шаге когнитивно-коммуникативного компаративного анализа.

Формант, эксплицирующий подмножество вариативных соответствий, представляет собой аффикс сказуемости *-sin*, который манифестируется во флективных языках с помощью аналитического или энклитического (в меньшинстве флективных языков) личного местоимения *ты* (в русском), *you* (в английском).

На выходе в результате анализа исходной турецкой аффиксирующей формы *Sen gitmelisisin!* возможна единственная форма в последовательно флективном русском языке *идти + должен + ты = Ты должен идти!* и вариативное соответствие в качестве промежуточной формы для дальнейшего когнитивно-коммуникативного синтаксического анализа в флективно-аналитическом английском *go + must / should / have to + you = You must go!*

**Общая структура
тюркско-русско-английского
лексическо-морфологического анализа
с учетом когнитивного-коммуникативного
синтаксиса в трансляционном процессе**

При формировании автоматизированного поиска соответствий с использованием инвариантного тезауруса закономерно используется несколько иерархических компонентов, которые представляют собой следующую последовательность шагов рекурсивного спуска:

- инвариантный тезаурус основ;
- инвариантный тезаурус общеупотребительных оборотов;
- список фразовых основ, функционирующих как финитная основа предикации (2-й компонент оборотов);
- список лексем, функционирующих как дополнительный классификатор предикации (1-й компонент оборотов);
- алгоритмизированная модель иерархического компаративного анализа – лексико-морфологического анализа с учетом когнитивного-коммуникативного синтаксиса.

Общеизвестно, что адекватный исходной фразе вариант автоматизированного перевода может быть осуществлен только в рамках прецизионного лингвистического анализа, который не может строиться исключительно на лексико-морфологических основаниях, как указанные выше способы, он должен включать и когнитивно-коммуникативные синтаксические образцы. Созданная на их основе алгоритмизированная модель позволяет с легкостью решить задачи преодоления полисемического, синонимического и омонимического барьеров, возникающих на этапе формирования выходного целевого текста. Именно поэтому в предлагаемом алгоритме нами учтены подпрограммные шаги рекурсивного сравнения как в направлении сокращения формантов до минимума $0 < N^1 < 2$, так и в направлении перманентного их расширения за счет не учтенных ранее в инвариантном тезаурусе единиц. Кроме того, в новую модель включаются стандартные контекстные валентности, на их основе строится промежуточный результат выходной фразы, которая на следующем шаге сравнивается с образцами, отвечающими задачам когнитивно-коммуникативного синтаксиса.

Следует, однако, оговориться, что неконтролируемое добавление рекурсивных отсылок может происходить при отсутствии функциональных ограничений, определяющих для автоматизированного процесса обработки данных супертип языковой системы – ориентировка на облегчение производства высказывания или восприятия высказывания (подробно об этом см.: [Бредихин, 2013, с. 30–31]). Супертип языка детерминирует национальную специфику актуализации и интенсификации глубинного содержания высказывания и иерархическую организацию базовых функций в процессе речепорождения – от экспрессивной (соотносимой с говорящим), как это происходит в большинстве агглютинирующих языков, и репрезентативной (соотносящей язык с предметами и ситуациями реальности) функций до примарно-апеллятивной, то есть функции, соотносимой со слушающим, как это происходит в большинстве случаев в языковых системах флективно-аналитического типа. Именно соотнесение спецификаторов производителя и рецептора высказывания, в условиях пары разноструктурных языков осложненное субъектно-объектной асимметрией (о ней см.: [Бредихин, Серебрякова, 2016, с. 115]), должно снять транслатологический барьер супертиповых характеристик при автоматизированном переводе.

Порядок введения промежуточных результатов в инвариантный тезаурус и рекурсивный поиск контекстуально релевантного соответствия, позволяющего на основе снижения степени вариативности и конситуативной детерминации словоформы создать адекватное оригиналу высказывание на флективном языке, может быть представлен графически (см. рис. 2).

При поиске соответствий основ в инвариантном тезаурусе в качестве промежуточного результата фиксируются не более двух внеконтекстуальных эквивалентов, которые являются наиболее частотными в конситуации общения (сфера науки, бытового употребления, газетного стиля и т. д.). Возникающая в данном случае омонимия нивелируется на этапе компаративного анализа формантов-экспликаторов грамматического значения.

Например, омонимичные в турецком основы *yüz* / *yüzme* (*face* / *swim*, *лицо* /

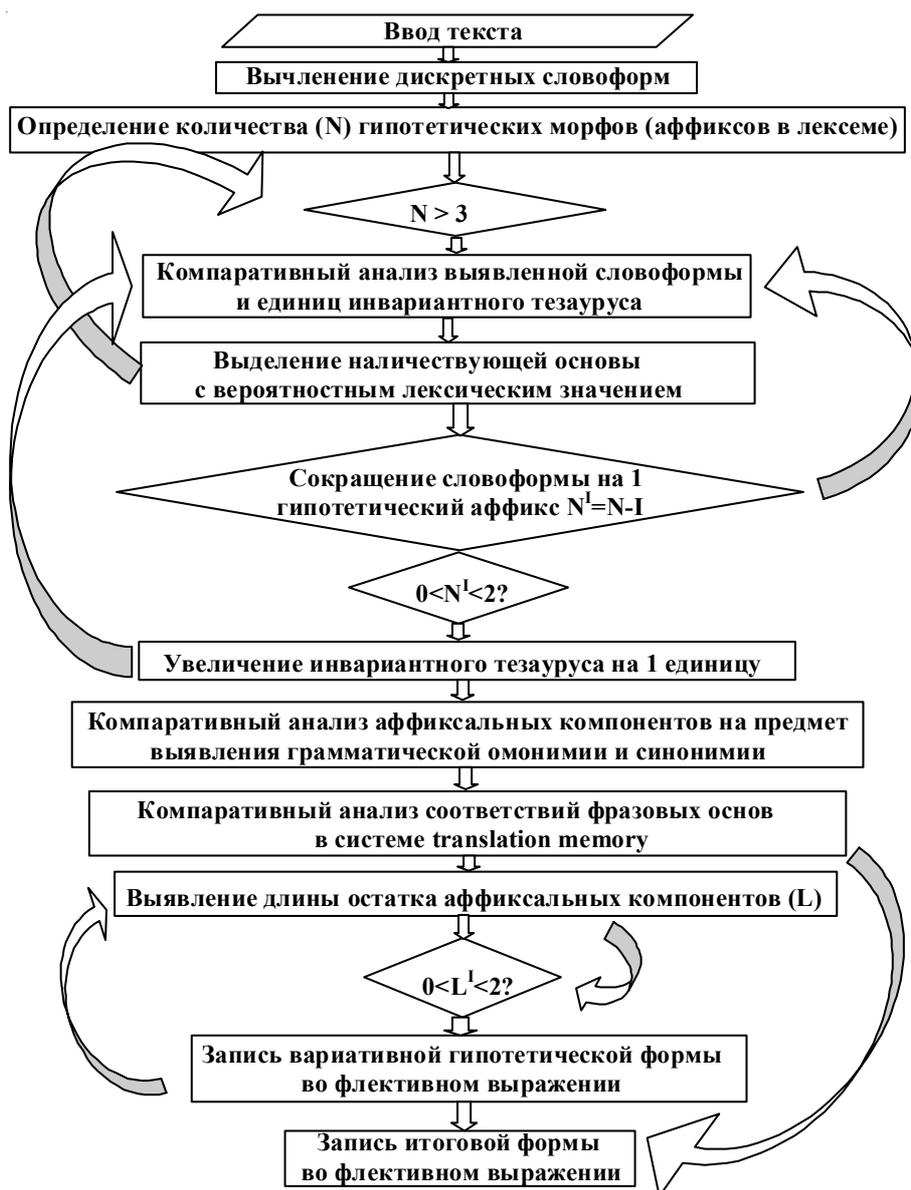


Рис. 2. Блок-схема лексико-морфологического анализа / синтеза при автоматизированной трансляции текста на агглютинативном языке с учетом когнитивного-коммуникативного синтаксиса

Fig. 2. Control-flow chart of lexico-morphological analysis/synthesis in automated text translation in an agglutinative language with cognitive-communicative syntax consideration

плавать), сводимые к единой основе типа /yüz/, которую невозможно отнести к той или иной части речи, делимитируются уже на втором шаге рекурсивного сравнения при сокращении основы на один гипотетический аффикс. Подтверждение происходит при компаративном анализе аффиксальных компонентов. Промежуточные результаты, полученные в процессе применения алгоритма, представленного в блок-схеме (рис. 2), можно для наглядности привести в следующей таблице (см. табл. 1).

Специфика инвариантного тезауруса общеупотребительных оборотов

Инвариантный тезаурус основ предполагает включение в него только общеупотребительных несвободных сочетаний, построенных по типу $N + V_f$ – данная модель выявляется в результате стандартного трансформационного анализа. На основании второго компонента имеющиеся общеупотребительные обороты классифицируются на три подгруппы. Первая из них формируется областью кон-

Таблица 1. Вариативные соответствия с учетом когнитивно-коммуникативного синтаксиса

Table 1. Variable translation equivalents with cognitive-communicative syntax consideration

Словоформа	Вариативные соответствия на уровне лексико-морфологического анализа	Вариативные соответствия при подключении когнитивно-коммуникативного синтаксиса
<i>Öğretmek</i>	<i>öğretim</i> преподавание <i>öğretim etmek</i> обучение + делать , сделать	<i>Öğretmek</i> преподавать, учить
<i>Hesap etmek</i>	Счет <i>hesap</i> , расчет + делать , сделать <i>hesap etmek</i>	<i>Hesap etmek</i> считать, подсчитывать
<i>Yüreklen dirmek</i>	Сердце <i>yürek</i> , душа + давать, отдавать <i>vermek</i>	<i>Yüreklen dirmek</i> ободрять, подбадривать <i>yürek</i> сущ., аффикс <i>len</i> присоединяется к существительным или прилагательным, преобразуя их в непереходные глаголы (при этом такие глаголы приобретают страдательное или возвратное значение), <i>dir</i> – аффикс присоединяется к сказуемому (как глагольному, так и именному) и придает предложению значение предположения, <i>mek</i> показатель инфинитива
<i>yoklamak</i>	Голова <i>baş</i> , головной + тянуть <i>çekmek</i> , таскать <i>taşımak</i>	Наведываться <i>ziyaret</i> сущ. <i>etmek</i> , навестить <i>yoklamak</i>

струкций со вторым компонентом – служебным глаголом активного деятеля *etmek* (*делать / сделать*, (*to do*)). Вторая создает пространство для сравнения неактивных глубинных ролей со служебным глаголом *olmak* (*быть*, (*to be*)) и т. д. Третья подгруппа, не входящая в инвариантный тезаурус, создает отдельную область ассоциированных образцов для сравнения и сохраняется как «запасной» ход трехчленного формантного анализа. Следует отметить, что она используется крайне редко в условиях присутствия в исходном высказывании осложняющих коммуникативный синтаксис компонентов.

Приведем некоторые глагольные единицы из данного инвариантного тезауруса, которые формируют подсистемы третьего уровня, в случае невозможности отнести их к тому или иному типу служебного второго компонента или же возможности формирования обеих форм, например, основы */al/*, */et/*, */ver/*, */al/*, */vur/*, */çık/*. При этом обороты, принадлежащие первой и второй подгруппам, подвержены рекурсивному сравнению длиной остатка аффиксальных компонентов на шаге записи гипотетической вариативной формы, либо пополняя новыми единицами инвариантный тезаурус, либо формируя третью группу «нерегулярных» соответствий как отдельное подмножество в translation memogy. Данное пополнение тезауруса основ не влияет на

увеличение операций алгоритма, а значит, остается в рамках трехчленного формантного анализа, то есть функционирование синтаксического блока базируется на исходных принципах. Таким образом, создание неконтролируемой рекурсивной отсылки, которая затруднила бы работу автоматизированного анализатора по созданию целевой фразы и ограничила бы возможности генерации адекватного генерализованного смысла, нивелируется.

Кроме того, в фонде translation memogy отдельно сохраняются все возможные вариативные соответствия первых элементов глагольных сочетаний в соответствии с группами классификации. Размещение в данном списке первых компонентов происходит в зависимости от выявленного количества (N) гипотетических аффиксов, соотношенных с формой именительного падежа единственного числа (см. табл. 2). Сохранение вариантов в оперативной памяти до формирования итогового высказывания на переводящем языке позволяет быстро вернуться к вариативному соответствию, которое было исключено на том или ином шаге автоматизированного анализа. Последующие шаги призваны верифицировать промежуточный результат и исключить первоначально прогнозируемый эквивалент из синтеза итоговой фразы на переводящем языке. Такая последовательность шагов (с обращением к оперативной памяти до сравнения с

Таблица 2. Фрагмент инвариантного тезауруса общеупотребительных оборотов

Table 2. Fragment of the invariant thesaurus of common constructs

Основа или 1-й компонент	Трансляционный вариант или генерализованное значение со вторым компонентом <i>et</i>
<i>azat etmek</i>	освободить <i>boşaltmak</i> (османский язык <i>azat etmek</i>)
<i>arzulamak</i> <i>hayal etmek</i>	желать <i>istemek, dilemek, arzu etmek, arzulamak</i> , мечтать <i>hayal etmek, hayal kurmak</i>
<i>teşebbüs etmek</i>	покушаться <i>teşebbüs etmek</i> , посягать <i>tecavüz</i>
<i>tahmin etmek</i>	полагать, предполагать <i>tahmin etmek</i> (догадываться)

инвариантным тезаурусом) дает возможность ускорить процесс обработки данных. Фундаментальные различия инициальных элементов разноструктурных языковых систем с коллекционными словоформами, снижающими влияние семантико-синтаксической прогностики на формирование «предсказательного» высказывания в различных видах флективных языков [Валентинова, Рыбаков, 2021], учитываются при компаративном анализе аффиксальных компонентов на предмет выявления грамматической омонимии и синонимии.

Глагольные значения определяются в процессе компаративного анализа выделения наличествующей основы с вероятностным (прогнозируемым на основе инвариантных валентностных (синтаксических) конструкций) лексическим значением, соответствующим, как правило, наименьшему количеству формантов в императивной форме.

Отдельные лексемы находят эквивалентное соответствие в той или иной привязке к глагольному значению в зависимости от релевантного варианта, трансляция общего содержания в данном случае имеет вариативный характер. Подобные вариации основываются на «принципе оптимального интерпретативного ожидания» [Бредихин, 2017, с. 420], который предполагает наивысшую степень узнаваемости и релевантности сегмента при условии его реализации в рамках общеупотребительного оборота. Когда употребление словоформы не является узуальным, она оказывается предельно усложненной для автоматизированного анализа, так как исключается из групповой классификации в силу зависимости от когнитивно-коммуникативного контекста. Это необходимо учитывать в сложных случаях при неоднозначности аффиксов в агглютинативных языках, поскольку в некото-

рых системах существуют амплификсы, способные к манифестации как грамматических, так и лексических значений (подробно о таких формантах см.: [Гирфанова, 2007, с. 83]). Эта сложность может быть успешно преодолена посредством создания отдельных подсистем классификации спецификаторов. Например, существительное *tanıdık* в инвариантном тезаурусе основ репрезентирует предметное значение «знакомый». Та же лексема в процессе компаративного анализа аффиксальных формантов в различных группах получает различные формы актуализации в промежуточном результате, в подмножестве *et* она интерпретируется глаголом *poznaкомитъ* с дополнительным значением транзитивности, который требует после себя дополнений в определенной форме, в подгруппе *ol* – глаголом *poznaкомитъся*, содержащим дополнительный компонент взаимно возвратного действия.

Принадлежность к третьей группе маркируется отсутствием форманта – *tanıdık etmek* «ознакомиться» / *tanış olmak* «быть знакомым, встречаться» / *tanışmak* «встретить» – что предполагает введение в инвариантный тезаурус нового элемента. На основе выявления таких сложных случаев пополняется не только translation memory, но и тезаурус инвариантных основ, что существенно улучшает работу автоматизированной системы.

Рассмотрим последовательную рекурсию со сравнением на различных этапах дедимитации и сокращения формантов. В случае обнаружения соответствия в рамках первичного рекурсивного сравнения *dans* (*танец / пляска*) по инвариантному тезаурусу данная лексема будет переводиться как существительное с процессуальным значением. Если то же слово обнаруживается в подгруп-

Таблица 3. Трансляционные вариативные соответствия (генерализованное значение / общеупотребительный оборот)

Table 3. Variable translation equivalents (generalized sense/common construct)

Исходная форма/основа	Соответствие в инвариантном тезаурусе основ	Трансляционное вариативное соответствие в общеупотребительном обороте	
		Позиционные изменения <i>et</i>	Позиционные изменения <i>al</i>
<i>tanış</i>	Знакомый <i>tandık</i>	<i>tanıtmak</i> познакомить	<i>tanudık olmak</i> быть знакомым
<i>hesap</i>	Счет / расчет <i>hesap</i>	<i>hesap etmek</i> считать	<i>hesap olmak</i> быть учетной записью / принимать в расчет
<i>dahil</i>	Внутренность <i>içeri</i>	<i>dahil etmek</i> включить / включать	–
<i>huzursuz/ rahat</i>	Суеулиый <i>telaşlı</i> / беспокойный <i>huzursuz</i>	<i>rahatsız etmek</i> беспокоить / тревожить	<i>rahatsız olmak</i> болеть / тревожить

пе *et* (деятельностное начало) и имеет форму *etmek* (*dans*), то оно определяется как глагольное соответствие и переводится как *танцевать / плясать* (*to dans*), то есть рекурсивное сравнение с перманентно пополняющимся инвариантным тезаурусом основ позволяет идентифицировать не только исходные лексические и грамматические значения, но и выбрать в трансляте соответствующие супертипические характеристики. Это дает возможность построить высказывание, адекватно передающее генерализованное значение с учетом конситуативных и контекстуальных условий.

Внеконтекстуальная интерпретация (самостоятельность значения) лексемы или ее принадлежность к общеупотребительному обороту фиксируется в рамках рекурсивного сравнения на основе компонентов когнитивно-коммуникативного синтаксиса совместно с алгоритмом инвариантного тезауруса.

Выводы

На основе анализа трансляционных соответствий в агглютинативных и флективных языках был разработан алгоритм поиска релевантных форм сохранения как лексического, так и грамматического значений сложных аффиксальных компонентов при производстве целевого высказывания средствами флективных языков. Алгоритм включает строго определенную последовательность шагов рекурсивного сравнения когнитивных и конситуативных элементов исходного высказывания с

вариативными формами, позволяющими пополнить инвариантный тезаурус или систему оперативной памяти *translation memory*. Всего алгоритм предполагает четырнадцать шагов с четырьмя рекурсивными и двумя рекурсивными отсылками, которые не имеют ограничений в повторяемости, но не ведут к увеличению количества операций.

В основе автоматизированной системы трансляции в парах разноструктурных языков (агглютинативный – флективный) лежит лексико-морфологический компаративный анализ, включающий алгоритм сравнения исходных текстовых форм с инвариантным тезаурусом основ. В процессе нивелировки барьеров синонимичности, омонимичности как основ, так и аффиксов в компаративный анализ включаются подмножества трех подгрупп распределения основ и дополнения инвариантного тезауруса. Окончательное снятие амфиболичности целевого высказывания производится в рамках введения в алгоритм анализа компонентов когнитивно-коммуникативного синтаксиса. В завершении рекурсивного спуска происходит сравнение исходных единиц с общеупотребительными оборотами в той или иной подгруппе на основе классификационного членения по позициям (например, *et*, *al*, нерегулярная подгруппа), после чего конечный элемент вводится в инвариантный тезаурус как набор необходимых и достаточных условий для использования его в трансляте. Именно последовательный рекурсивный спуск с возможностью сокращения формантных элементов и пополнения инвариантного тезауруса

основ и списка аффиксальных формантов дает возможность, не увеличивая количество прямых операций и оставаясь в рамках трех-членного формантного анализа, существенно повысить адекватность и эквивалентность целевой фразы во флективном оформлении.

СПИСОК ЛИТЕРАТУРЫ

- Бектаев К. Б., 1978. Статистико-информационная типология тюркского текста. Алма-Ата : Наука. 183 с.
- Большакова Е. И., Клышинский Э. С., Ландэ Д. В., Носков А. А., Пескова О. В., Ягунова Е. В., 2011. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика. М. : МИЭМ. 272 с.
- Бредихин С. Н., 2013. Лингвокультурологический аспект смыслопорождения на грамматическом уровне // Филологические науки. Вопросы теории и практики. № 3–1 (21). С. 29–33.
- Бредихин С. Н., 2017. Когнитивно-семантическая модель параметрированной актуализации фразовой основы // Когнитивные исследования языка. № 30. С. 419–422.
- Бредихин С. Н., Серебрякова С. В., 2016. Субъектно-объектная асимметрия при распознавании речи // Вопросы когнитивной лингвистики. № 4 (49). С. 114–121. DOI: 10.20916/1812-3228-2016-4-114-121
- Валентинова О. И., Рыбаков М. А., 2021. Логика детерминантного анализа агглютинативных и флективных языков (часть вторая) // Полилингвистичность и транскультурные практики. Т. 18, № 3. С. 234–244. DOI: 10.22363/2618-897X-2021-18-3-234-244
- Гирфанова А. Х., 2007. К проблеме слово- и формообразования в языках агглютинативного типа // Acta Linguistica Petropolitana. Труды института лингвистических исследований. Т. 3, № 1. С. 83–93.
- Дурст-Андерсен П. В., 1995. Ментальная грамматика и лингвистические супертипы // Вопросы языкознания. № 6. С. 30–42.
- Ксалов А. М., Гошкова Ф. М., Денисенко В. А., 2013. Разработка естественно-языкового интерфейса для мультиагентных систем, основанных на знаниях, на основе парсера агглютинативных языков // Известия Кабардино-Балкарского научного центра РАН. № 4 (54). С. 37–41.
- Орехов Б. В., Слободян Е. А., 2010. Проблемы автоматической морфологии агглютинативных языков и парсер башкирского языка // Информационные технологии и письменное наследие (EL'MANUSCRIPT-10) : материалы Междунар. науч. конф. (Ижевск, 28–31 октября 2010 г.). Уфа ; Ижевск : Вагант. С. 167–171.
- Пинес В. Я., 1974. Некоторые проблемы автоматического перевода и тюркские языки // Советская тюркология. № 3. С. 100–107.
- Рыбаков М. А., 2004. Внутренняя синтагматика словоформы и морфологические типы языков // Вестник Российского университета дружбы народов. Серия: Лингвистика. № 6. С. 64–70.
- Семенов А. Л., 2008. Современные информационные технологии и перевод. М. : Академия. 224 с.
- Bredikhin S. N., Babayants V. V., Pelevina I. I., 2021. A Comprehensive Cognitive-Perceptual Model of Analysis for Contextually Determined Components of a Conceptualized Term // E3S Web of Conferences. Vol. 273, № 11038. DOI: 10.1051/e3sconf/202127311038
- Bredikhin S. N., Serebriakov A. A., 2019. The Explicitness of the Deep Structure of Meaning in Prognostic Strategies // The European Proceedings of Social & Behavioural Sciences EpSBS. Vol. LVIII, № 294. P. 2549–2557. DOI: 10.15405/epsbs.2019.03.02.294
- Fodor J. D., 1978. Parsing Strategies and Constraints on Transformations // Linguistic Inquiry. № 9. P. 93–133.
- Hakkani-Tür D., Tur G., Celikyilmaz A., Chen Y.-N., Gao J., Deng L., Wang Y.-Y., 2016. Multi-Domain Joint Semantic Frame Parsing Using Bi-Directional RNN-LSTM // Understanding Speech Processing in Humans and Machines. № 1 (5). P. 715–719. DOI: 10.21437/Interspeech.2016-402
- Konieczny L., Scheepers Ch., Hemforth B., Strube G., 1994. Semantikorientierte Syntaxverarbeitung // Kognitive Linguistik. Opladen : Westdeutscher Verlag. P. 129–158. DOI: 10.1007/978-3-663-05399-6_6
- Proudian D., Pollard C., 1985. Parsing Head-Driven Phrase Structure Grammar // ACL'85: Proceedings of the 23rd Annual Meeting on Association for Computational Linguistics (July 1985). P. 167–171. DOI: 10.3115/981210.981231
- Tanaka M. N., Branigan H. P., Pickering M. J., 2011. The Production of Head-Initial and Head-Final Languages // Processing and Producing Head-Final Structures. Vol. 38. Dordrecht ; N. Y. : Springer. P. 113–129. DOI: 10.1007/978-90-481-9213-7_6

REFERENCES

- Bektaev K. B., 1978. *Statistiko-informatsionnaya tipologiya tyurkskogo teksta* [Statistical and Informational Typology of the Turkic Text]. Alma-Ata, Nauka Publ. 183 p.

- Bolshakova E.I., Klyshinskiy E.S., Lande D.V., Noskov A.A., Peskova O.V., Yagunova E.V., 2011. *Avtomaticheskaya obrabotka tekstov na estestvennom yazyke i kompyuternaya lingvistika* [Automatic Text Processing in Natural Language and Computational Linguistics]. Moscow, MIEM Publ. 272 p.
- Bredikhin S.N., 2013. Lingvokulturologicheskiy aspekt smysloporozhdeniya na grammaticheskom urovne [Linguo-Culturological Aspect of Sense Creation on Grammatical Level]. *Filologicheskie nauki. Voprosy teorii i praktiki* [Philology. Theory & Practice], no. 3–1 (21), pp. 29–33.
- Bredikhin S.N., 2017. Kognitivno-semanticheskaya model parametrizovannoy aktualizatsii frazovoy osnovy [Cognitive-Semantic Model of Parameterized Phrase Head Attachment]. *Kognitivnye issledovaniya yazyka* [Cognitive Studies of Language], no. 30, pp. 419–422.
- Bredikhin S.N., Serebryakova S.V., 2016. Subyektno-obyektnaya asimmetriya pri raspoznavanii rechi [Subject-Object Asymmetry in Speech Recognition]. *Voprosy kognitivnoy lingvistiki* [Issues of Cognitive Linguistics], no. 4 (49), pp. 114–121. DOI: 10.20916/1812-3228-2016-4-114-121
- Valentinova O.I., Rybakov M.A., 2021. Logika determinantnogo analiza agglutinativnykh i flektivnykh yazykov (chast vtoraya) [Logic of Determinative Analysis of Agglutinative and Inflectional Languages (Part 2)]. *Polilingvialnost i transkulturnye praktiki* [Polylinguality and Transcultural Practices], vol. 18, no. 3, pp. 234–244. DOI: 10.22363/2618-897X-2021-18-3-234-244
- Girfanova A.Kh., 2007. K probleme slovo- i formoobrazovaniya v yazykakh agglutinativnogo tipa [On the Problem of Word- and Form Formation in Agglutinative Languages]. *Acta Linguistica Petropolitana. Trudy instituta lingvisticheskikh issledovaniy* [Acta Linguistica Petropolitana. Transactions of the Institute for Linguistic Studies], vol. 3, no. 1, pp. 83–93.
- Durst-Andersen P.V., 1995. Mentalnaya grammatika i lingvisticheskie supertipy [Mental Grammar and Linguistic Supertypes]. *Voprosy yazykoznaniiya* [Topics in the Study of Language], no. 6, pp. 30–42.
- Ksalov A.M., Goshokova F.M., Denisenko V.A., 2013. Razrabotka estestvenno-yazykovogo interfeysa dlya multiagentnykh sistem, osnovannykh na znaniyakh, na osnove parsera agglutinativnykh yazykov [Development of a Natural-Language Interface for Knowledge Based Multi-Agent Systems Based on the Parsing of Agglutinative Languages]. *Izvestiya Kabardino-Balkarskogo nauchnogo tsentra RAN* [News of the Kabardino-Balkarian Scientific Center of RAS], no. 4 (54), pp. 37–41.
- Orekhov B.V., Slobodyan E.A., 2010. Problemy avtomaticheskoy morfologii agglutinativnykh yazykov i parser bashkirskogo yazyka [Problems of Automatic Morphology of Agglutinative Languages and Bashkir Language Parser]. *Informatsionnye tekhnologii i pismennoe nasledie (EL'MANUSCRIPT-10): materialy Mezhdunar. nauch. konf. (Izhevsk, 28–31 oktyabrya 2010 g.)* [Information Technologies and Written Heritage (EL'MANUSCRIPT-10). Proceedings of the International Scientific Conference (Izhevsk, October 28–31, 2010)]. Ufa, Izhevsk, Vagant Publ., pp. 167–171.
- Pines V.Ya., 1974. Nekotorye problemy avtomaticheskogo perevoda i tyurkskie yazyki [Some Problems of Automatic Translation and Turkic Languages]. *Sovetskaya tyurkologiya* [Soviet Turkology], no. 3, pp. 100–107.
- Rybakov M.A., 2004. Vnutrennyaya sintagmatika slovoformy i morfologicheskie tipy yazykov [Internal Syntagmatics of the Word Form and Morphological Types of Languages]. *Vestnik Rossiyskogo universiteta druzhby narodov. Seriya: Lingvistika* [Russian Journal of Linguistics], no. 6, pp. 64–70.
- Semenov A.L., 2008. *Sovremennye informatsionnye tekhnologii i perevod* [Modern Information Technologies and Translation]. Moscow, Akademiya Publ. 224 p.
- Bredikhin S.N., Babayants V.V., Pelevina I.I., 2021. A Comprehensive Cognitive-Perceptual Model of Analysis for Contextually Determined Components of a Conceptualized Term. *E3S Web of Conferences*, vol. 273, no. 11038. DOI: 10.1051/e3sconf/202127311038
- Bredikhin S.N., Serebriakov A.A., 2019. The Explicitness of the Deep Structure of Meaning in Prognostic Strategies. *The European Proceedings of Social & Behavioural Sciences EpSBS*, vol. LVIII, no. 294, pp. 2549–2557. DOI: 10.15405/epsbs.2019.03.02.294
- Fodor J.D., 1978. Parsing Strategies and Constrains on Transformations. *Linguistic Inquiry*, no. 9, pp. 93–133.
- Hakkani-Tür D., Tur G., Celikyilmaz A., Chen Y.-N., Gao J., Deng L., Wang Y.-Y., 2016. Multi-Domain Joint Semantic Frame Parsing Using Bi-Directional RNN-LSTM. *Understanding Speech Processing in Humans and Machines*, no. 1 (5), pp. 715–719. DOI: 10.21437/Interspeech.2016-402

Konieczny L., Scheepers Ch., Hemforth B., Strube G., 1994. Semantikorientierte Syntaxverarbeitung. *Kognitive Linguistik*. Opladen, Westdeutscher Verlag, pp. 129-158. DOI: 10.1007/978-3-663-05399-6_6

Proudian D., Pollard C., 1985. Parsing Head-Driven Phrase Structure Grammar. *ACL85: Proceedings of the 23rd Annual Meeting on Association for*

Computational Linguistics (July 1985), pp. 167-171. DOI: 10.3115/981210.981231

Tanaka M.N., Branigan H.P., Pickering M.J., 2011. The Production of Head-Initial and Head-Final Languages. *Processing and Producing Head-Final Structures*, vol. 38. Dordrecht, New York, Springer, pp. 113-129. DOI: 10.1007/978-90-481-9213-7_6

Information About the Authors

Sergey N. Bredikhin, Doctor of Sciences (Philology), Professor, Department of Translation Studies, North-Caucasus Federal University, Pushkina St, 1, 355017 Stavropol, Russia, bredichinsergey@yandex.ru, <https://orcid.org/0000-0002-2191-4982>

Stanislav G. Sidorenko, Candidate of Sciences (Philology), Associate Professor, Department of Translation Studies, North-Caucasus Federal University, Pushkina St, 1, 355017 Stavropol, Russia, samuels@mail.ru, <https://orcid.org/0000-0003-4963-869X>

Информация об авторах

Сергей Николаевич Бредихин, доктор филологических наук, профессор кафедры теории и практики перевода, Северо-Кавказский федеральный университет, ул. Пушкина, 1, 355017 г. Ставрополь, Российская Федерация, bredichinsergey@yandex.ru, <https://orcid.org/0000-0002-2191-4982>

Станислав Геннадьевич Сидоренко, кандидат филологических наук, доцент кафедры теории и практики перевода, Северо-Кавказский федеральный университет, ул. Пушкина, 1, 355017 г. Ставрополь, Россия, samuels@mail.ru, <https://orcid.org/0000-0003-4963-869X>