

УДК 811.161.1+811.512.161

ТЕОРЕТИЧЕСКИЕ ПРИНЦИПЫ ПРОГРАММНО-ОРИЕНТИРОВАННОГО ОПИСАНИЯ ГРАММАТИЧЕСКОЙ КАТЕГОРИИ ПАДЕЖА

Шеремет В. В.

*Таврический национальный университет им. В. И. Вернадского
E-mail: sheremetvitaliy@gmail.com*

Работа посвящена выявлению таких теоретических принципов описания категории падежа в русском языке и ее соответствий в турецком языке, которые позволили бы достичь эквивалентности при переводе падежа машиной. Поэтапное изучение историографии вопроса машинного эквивалентного перевода падежа позволило локализовать данную проблему, а также определить основы описания падежа и инструменты, при помощи которых реализация эквивалентности может быть осуществлена.

Ключевые слова: категория падежа, русский язык, турецкий язык, параллельный корпус, эквивалентный машинный перевод, системный подход, языкознание.

Постановка проблемы. Одним из признаков современной лингвистики является «машиноориентированность» ее инструментов – совмещение языкового материала и лингвистических знаний о нем в виртуальном пространстве баз данных, удобном для хранения информации и последующей ее обработки компьютером (например, в корпусе текстов). Подобный подход к организации языкового материала открывает большие перспективы в изучении языка. В этой связи чрезвычайно важной является специфика теоретического описания языка, которое эксплицируется в тесной взаимосвязи с текстом.

В современных прикладных исследованиях программная организация языковых элементов баз данных, в том числе и корпуса, осуществляется на основе реляционного принципа – отношения *функции* или *значения*. Соответственно, в описании языка, используемом в подобных проектах, находим зеркальное его отражение – структурный семасиологический подход, который характеризует изолированное рассмотрение языковых явлений и антропоцентризм. Под изолированным изучением языковых явлений мы подразумеваем наблюдение только одного из отношений, выделяемых в языке – функции (либо значения). Под антропоцентризмом – наличие в описании языкового явления подразумеваемого субъекта, мыслительная деятельность которого воссоздает понимание единиц речи из их функций.

На сегодняшний день использование этих подходов в прикладной лингвистике представляет собой серьезную проблему, т. к. они проявляются в недостатках систем машинного перевода и автоматической обработки текстов на естественном языке – наиболее востребованных на сегодняшний день отраслях лингвистики.

В особенности эта проблема касается области грамматики – структурообразующего явления, которое настолько абстрактно, что не поддается полному семасиологическому описанию. В свою очередь, антропоцентрические описания, расширяющие изолированную трактовку грамматики, подобно объяснению дательного падежа как маркирующего существительное или местоимение «со значением посетора в предложении с комплексным обозначением лица как целого, часть или при-

надлежность которого выступает в качестве локализованного объекта действия» [1], видятся невозможными для полноценной эксплуатации машиной. При подобном абстрагировании одного из ключевых элементов структуры словосочетания либо его функциональном представлении описание эквивалентности в доступном для машины языке сильно затрудняется. Несмотря на это, в лингвистике данная проблема остается вне зоны внимания ученых.

В этой связи **целью** нашего исследования была избрана выработка теоретических принципов, на базе которых стало бы достижимым понятное для машины описание категории падежа и ее эквивалентов в близкородственных и разнородных языках.

Для достижения цели решались следующие **задачи**:

1) были рассмотрены существующие подходы к фиксации эквивалентов падежных значений в Национальном корпусе русского языка (НКРЯ), параллельных корпусах, созданных на его базе, а также проанализирован ряд сопоставительных работ, посвященных категории падежа в близкородственных и разнородных языках.

2) на основе анализа принципов их работы был выработан альтернативный подход, теоретическая основа которого позволила бы достичь грамматической эквивалентности.

В ходе подготовки работы использовались общеполитические **методы** анализа, синтеза, а также методы описания, дистрибутивного анализа, сравнительно-сопоставительного анализа.

В НКРЯ падеж рассматривается в рамках структурного подхода на нескольких уровнях [2]. Так, в рассматриваемом корпусе текстов существует несколько «падежей»: формобразующий морфологический (словоизменяемая категория морфологии слова) и структурообразующий синтаксический (функциональная единица, элемент синтаксического отношения) [3; 4]. В первом случае падеж используется только для различения форм слова и не участвует непосредственно в смыслообразовании. Во втором случае – подразумевается, что существует субъект познания – человек, который осуществляет *понимание*, т. к. падеж в функциональном представлении не интегрирован в языковую систему, а является только лишь инструментом для построения синтаксических структур [5].

Работа с падежом в параллельных корпусах на базе НКРЯ ограничена возможностью его рассмотрения на уровне лексемы и предложения. Так, при сопоставительном анализе любого из падежей русского языка в предложении исходного текста выделяется лексема, содержащая соответствующую словоформу падежа, а в языке перевода предоставляется эквивалент только предложению, которое попадает в выдачу [6]. Исходя из этого, подобные эквиваленты могут быть использованы только человеком, т.к. для выполнения перевода необходим субъект, на основе мышления которого осуществилась бы обработка текста и трансляция на другой язык.

Сопоставительный анализ падежа в рамках метода интроспекции, а также в параллельных текстах, позволяет исследовать падеж в близкородственных и разнородных языках на более высоком уровне – синтаксическом, благодаря процедурам контекстного и трансформационного анализов, что позволяет найти не текстовое

соответствие, как в корпусе, а функциональное [7; 8; 9; 10; 11; 12]. Расширенная трактовка падежа, являющаяся *результатом* понимания того или иного контекста его употребления, отражается либо в широком определении падежной функции, учитывающем не все аспекты его понимания (подобно подходу в НКРЯ), либо в слишком абстрактной для прикладного использования форме.

Таким образом, на сегодняшний день, существующие описания падежа русского языка в корпусе и описания его соответствий в близкородственных и разнородных языках, осуществляемые на базе интроспективного сопоставления, сопоставления в параллельном корпусе, параллельных текстах, носят антропоцентрический характер, подразумевая осуществление процессов *понимания* и *перевода* человеком.

В современной прикладной лингвистике отмечается, что *понимание* машины существенно отличается от *понимания* человека и на данный момент остается недостижимым [13]. Именно поэтому работающие с машинным переводом ученые отмечают его «условную эквивалентность» [14], когда достигается максимальное соответствие в рамках узкой прагматической цели.

Обратившись непосредственно к понятию межъязыкового соответствия можно отметить, что понимание играет основополагающую роль для перевода, так как эквивалентность представляет собой соответствие смысла / понимания исходного текста и целевого текста [15; 16]. Термин «понимание», в свою очередь, обычно трактуется как интеграция нового понимания в уже существующее знание посредством опыта, интерпретации, предзнания [17]. В этой связи перспективным подходом к категории падежа видится такое описание, которое было бы машиноориентированным: кодировало бы понимание падежа в доступных для интерпретации программой средствах.

Подобным подходом является системный подход, отличительная черта которого заключается в его интегративности и ономазиологичности: явления рассматриваются не обособленно, а в системе релевантных для них языковых феноменов, при этом учитывается граница системы, ее связь с внеязыковым пространством [18], что изоморфно трактовке понимания.

Падеж, рассматриваемый в подобном русле, реализуется в совокупности всех релевантных для его смыслообразования признаков: частеречной принадлежности релевантных лексем, самих лексем, обладающих в ряде случаев специфическим управлением, семантического типа предиката, семантической ролью актанта и предиката, порядком слов и др. Это, в совокупности с указанием параметров ограничений на понимание падежа, подобно наличию фразеологических единиц, длины синтагмы, релевантной для его параметров, опосредовано обуславливающих падеж стилистических признаков и т. д., позволяет определить границы смысловой реализации падежа, указать его связь с внеязыковыми явлениями и определить параметры его смыслонаполнения.

Эквивалентность в плане переноса понимания из одного языка в другой язык реализовывается в навыке перевода, которым обладает человек [19]. В виртуальном пространстве данный навык может фиксироваться в параллельном корпусе, как явлении уже заведомо эквивалентном в своей совокупности.

В качестве проверки теории нами были составлены принципы разметки русского, украинского, турецкого и арабского языков. Результатом работы стал 21 параметр, кодирующий понимание падежа в русском языке. При анализе параллельных текстов данные параметры остаются релевантными и для анализа языков перевода (в нашем случае – украинского, турецкого, арабского), т.к. в них сохранено *понимание* текста. Изменения касаются только значений данных параметров. Например, в турецком языке предикативность передается не только лексемами, но и аффиксами (при именах аффикс «-dı»), также, в турецком языке подлежащее может передаваться при помощи аффикса.

Основой разметки послужили общепринятые обозначения частей речи (N – существительное, V – глагол. и т.д), падежей: Nom – именительный падеж, Gen – родительный падеж), наличия связи между релевантными элементами падежной системы (+) и эквивалентности явления в целевом языке (=). Вертикальной чертой (|) отделяются комплексы явлений, принадлежащих к разным модулям разметки. Косая черта (/) разделяет релевантные элементы одного модуля. В обозначении релевантных для падежа параметров используются выработанные нами условные обозначения (например, kon – имя конкретное, odush – имя одушевленное, SemRAgens – семантическая роль «агенса», SocDeyst – семантический тип предиката «социальное действие» и т.д.). При определении списка значений параметров нами использовались фундаментальные описания тематических групп имен, семантических типов предикатов, семантических ролей, составленные такими учеными, как Ю. Д. Апресян, Е. В. Падучева, а также описания языковых явлений в академической грамматике. При описании специфических значений языковых явлений в турецком языке использовался грамматический справочник А. Н. Кононова.

При составлении экспериментального корпуса на базе данных параметров можно отметить возможность тонкого анализа падежей. Так, наиболее простой в антропоцентричном сопоставительном анализе именительный падеж в ряде случаев действительно не показывает расхождений:

«Отец» | Nnom / Ed | Kon / Odush/ Nar | Them | Sub / SemRAgens | Pod + «Служил» | V / Act / Tr / Nevoz / Sov / NeKompl | Rhem | SocDeyst / Nekauz / Nephaz | Glav | Skaz | «отец служил» =

«Babam» | Nnom/ ed | Kon/ Odush/ Nar | Them | Sub / SemRAgens | Pod + «çalışmış» | V / Act / Tr / Nevoz / Sov / NeKompl | Rhem | SocDeyst / Nekauz / Nephaz | Glav | Skaz | «babam çalışmış».

Однако, можно привести пример, когда происходит ряд существенных изменений:

«Мы» | Nnom / Ed | Kon / **Odush** / Nar | Them | Sub / **SemRPac** | Pod + «Воспитывались» | V / Act / Tr / Nevoz / Sov / NeKompl | Rhem | **Vozd** / Nekauz / Nephaz | Glav | Skaz | «воспитывались мы (не по-настоящему)» =

«**eğitim yöntemi**» | Nnom / Ed | Kon / **Neodush** / Nar | Them | Sub **SemRThem** | Pod + «(**başkaydı**)» | **Npredaff** / Act / Tr / Nevoz /Sov / NeKompl | Rhem | **Byt** / Nekauz / Nephaz | Glav | Skaz | «eğitim yöntemi başkaydı».

Данные эквиваленты отличаются: частеречной принадлежностью предиката (глагол в русском языке, существительное с аффиксом предикативности в турецком языке), лексико-грамматическими свойствами актанта (одушевленный в русском

языке, неодушевленный в турецком). Также различаются семантические типы предиката (глагол воздействия в русском языке и бытийный аффикс в турецком языке) и семантические роли актанта (*паиенс* в русском языке и *тема* в турецком языке). Помимо этого отличается и лексическое значение («eđitim uđntemi» не является эквивалентом слова «мы»). Таким образом, машиноориентированное описание позволяет проводить более глубокий анализ падежного значения, чем антропоориентированный и предоставляет более точные эквиваленты.

Выводы и перспективы: 1) существующие подходы к изучению падежа носят структурный, семасиологический, антропоцентрический характер, что отражается в теоретическом описании данной категории и в специфике ее применения в машинном переводе, препятствуя достижению эквивалентности. Машинный перевод, базирующийся на описаниях подобного рода, требует корректировки человеком; 2) ключевыми для эквивалентного машинного перевода являются принципы системности, машиноориентированности, основным инструментом выступает корпус параллельных текстов;

Вопрос системного описания языка не исчерпывается рассмотрением только лишь теоретических вопросов системного корпусного изучения категории падежа и предполагает дальнейшее всестороннее изучение данного феномена языка на практике: создание полного машиноориентированного описания параметров падежа русского языка, их эквивалентов в близкородственных и разнородных языках, таких как украинский, турецкий, арабский, а также моделирование межъязыковых эквивалентов данной категории.

Литература

1. Золотова Г. А. Синтаксический словарь. Репертуар элементарных единиц русского синтаксиса / Золотова Г. А. – М. : УРСС, 2006. – С.126. – ISBN: 5–354–01147–7.
2. Национальный корпус русского языка [Электронный ресурс]. – Режим доступа: <http://www.ruscorpora.ru/>.
3. Зализняк А. А. Грамматический словарь русского языка: словоизменение / Зализняк А. А. – М. : Рус.яз., 1980. – 880 с.
4. Апресян Ю. Д. Теоретические проблемы русского синтаксиса: Взаимодействие грамматики и словаря / Апресян Ю. Д., И. М. Богуславский и др. – М. : Языки славянских культур, 2010. – 408 с. – ISBN : 978–5–9551–0386–0.
5. Там же [4], 24.
6. Там же [2].
7. Макаров А. К. Дативные отношения в скрытой грамматике : дис. ... канд. филол. наук : 10.02.04 «Германские языки» / Макаров Андрей Константинович – Воронеж, 2008. – 166 с.
8. Шведова И. Н. Функционирование дательного падежа в немецком и русском языках : дис. ... канд. филол. наук : 10.02.20 «Сравнительно-историческое, типологическое и сопоставительное языкознание» / Шведова Ирина Викторовна : Москва, 2011. – 175 с.
9. Сунь Шуан. Семантическая интерпретация падежных форм в системах русско–китайского машинного перевода : на примере творительного падежа : дис.... канд. филол. наук : 10.02.21 «Сравнительно–сопоставительное языкознание» / Сунь Шуан ; Моск. гос. ун–т им. М.В. Ломоносова. – Москва, 2009. – 198 с.
10. Мансурова О. Ю. Представление и передача пространственных отношений в многоязычной системе машинного перевода (на материале английского, русского и турецкого языков) : дис.... канд. филол. наук : 10.02.22 «Языки народов зарубежных стран Европы, Азии, Африки, аборигенов Америки и Австралии» / Мансурова Оксана Юрьевна ; Рос. ун. дружбы нар. – Москва, 2006. – 160 с.

11. Гаврилова Н. Н. Casus Instrumentalis в русском и английском языках : дис. ... канд. филол. наук : 10.02.21 «Прикладная и математическая лингвистика» / Гаврилова Александра Ивановна. – Москва, 2009. – 161 с.
12. Матченко Г. В. Генетивные отношения в английском языке : дис. ... канд. филол. наук : 10.02.04 «Германские языки» / Матченко Герман Валерьевич – Курск, 2005. – 131 с.
13. Шаляпина З. М., Канович М. И. Система русского синтеза Russian в двух разных коммуникативных задачах [Электронный ресурс] / З. М. Шаляпина, М. И. Канович // Понимание в коммуникации. Человек в информационном пространстве: материалы междунар. конф., 22-24 ноября 2012 г., Ярославль, Россия / ЯГПУ им. К.Д.Ушинского. – Режим доступа: http://yspu.org/conferences/the_person_in_information_field_2012/4/Kanovich_Shalyapina.pdf.
14. Леонтьева Н. Н. Автоматическое понимание текстов: системы, модели, ресурсы / Леонтьева Н. Н. – М. : Академия, 2006. – 304 с. – ISBN : 5-7695-1842-1.
15. Комиссаров В. Н. Теория перевода (лингвистические аспекты) / Комиссаров В. Н. – М. : Высш.шк., 1990. – 253 с. – ISBN : 5-06-001057-0.
16. Финкельберг Н. Д. Теория и технология перевода / Финкельберг Н. Д. – М. : Восточная книга, 2010. – 400 с.
17. Герменевтический круг [Электронный ресурс] / Малахов В. С. // Новая философская энциклопедия / гл. ред. В. С. Степин : [в 4 т]. – М. : Мысль, 2010. – ISBN : 978-2-244-01115-9. – Режим доступа: <http://iph.ras.ru/elib/0762.html>.
18. Мельников Г. П. Системная типология языков / Мельников Г. П. – М. Наука, 2003. – С. 7-8. – ISBN : 5-209-01017-1.
19. Там же [13].

Шеремет В. В. Теоретичні принципи програмово-орієнтованого опису граматичної категорії відмінку / В. В. Шеремет // Вчені записки Таврійського національного університету імені В. І. Вернадського. Серія «Філологія. Соціальні комунікації». – 2014. – Т. 27 (66), № 3. – С. 100–105.

Робота присвячена встановленню таких теоретичних засад опису граматичної категорії відмінку в російській мові та її еквівалентів у турецькій мові, які б дозволили досягнути еквівалентності у разі перекладу відмінка машиною. Поетапне вивчення проблеми машинного еквівалентного перекладу відмінку дозволило локалізувати її, а також визначити основи опису відмінку, та виявити інструменти, за допомогою яких реалізація еквівалентності може бути здійснена.

Ключові слова: категорія відмінку, російська мова, турецька мова, паралельний корпус, еквівалентний машинний переклад, системний підхід, мовознавство.

Sheremet V. V. Theoretical principles of the program-aimed description of the case grammatical category / V. V. Sheremet // Scientific Notes of Taurida V. I. Vernadsky National University. – Series: Philology. Social communications. – 2014. – Vol. 27 (66), No 3. – P. 100–105.

The work to the distinguishing of the such theoretical principles of the case grammatical category in the Russian language and its equivalents in the Turkish language description is devoted, that allowed to reach the equivalence in translation the case grammatical category by machine. Step-by-step investigation of the equivalent machine translation problem's historiography to locate it and determine base of case depicting, instruments by means of which equivalence realization is possible.

In the investigation the case grammatical category is under observation because of its abstractness. Description of such phenomena allows to judge about the perceptiveness of the approach in full equivalence achievement.

The main premise of the investigation became the social request to the equivalent machine translation of non-specialized texts and absence in the present works the theory, which will be able to answer this request.

Key words: the case category, the Russian language, the Turkish language, the parallel corpora, the equivalent machine translation, the system approach, linguistics.

Поступила в редакцію 03.03.2014 г.