

УДК 811.512

Д. П. Кирьянов, Б. В. Орехов**СЕТЕВОЙ ПОДХОД К ОПИСАНИЮ БАШКИРСКОЙ МОРФОЛОГИИ**

Статья посвящена современному, основанному на теории графов (в западной терминологии — сетей), подходу к описанию свойственной тюркским языкам морфологической агглютинативности. Хотя в целом сети использовались в теоретической лингвистике и ранее, морфология оказалась почти не затронута таким методом описания. Между тем, именно в случае агглютинативной морфологии он дает существенные преимущества, позволяя установить закономерности в сочетании аффиксов, выявить типичное и нетипичное поведение морфем и морфов, предложить лингвистически оправданное описание редких, но любопытных с точки зрения теории языка случаев контактного сочетания одинаковых морфем (но разных морфов).

Ключевые слова: теория графов, башкирский язык, морфология.

Настоящая работа посвящена применению метода сетей к морфологии башкирского языка. Мы опираемся на [29, с. 2]: «Любой математический объект, в котором есть вершины и связи между ними, может быть назван графом» (здесь и далее перевод наш). По замечанию из работы [17, с. 4] «<...> термины граф и сеть часто используются как эквивалентные. Один из способов различить их — использовать слово *граф* при обозначении формальной математической структуры, а слово *сеть* — при обозначении конкретного примера» (здесь и в примере ниже курсив наш). В дальнейшем мы будем в основном пользоваться термином «сеть», условившись, что это понятие взаимозаменяемо, в соответствии с цитатой выше, с понятием «граф». Таким образом, сеть состоит из двух множеств — множества ребер и множества вершин; вершины могут связываться при помощи ребер. В англоязычной терминологии сеть называется *complex*, когда речь идет о большой изменяющейся во времени сети, проявляющей нетривиальные топологические свойства, ср. «*complex networks*, т. е. сети, структура которых нерегулярна, сложна и динамически меняется во времени...» [19]. В русскоязычной терминологии нет устоявшегося перевода термина *complex networks*; можно встретить такие переводы, как «сложные сети» [6, с. 245] или «комплексные

Кирьянов Денис Павлович — магистрант (Национальный исследовательский университет «Высшая школа экономики», Москва); e-mail: nevmenandr@gmail.com

Орехов Борис Валерьевич — кандидат филологических наук, доцент (Национальный исследовательский университет «Высшая школа экономики», Москва); e-mail: nevmenandr@gmail.com

Работа выполнена при поддержке РГНФ (грант № 15-04-12030 «Система автоматического морфологического и синтаксического анализа для корпусов миноритарных тюркских языков России»)

© Кирьянов Д. П., Орехов Б. В., 2015

сети» [2]. Некоторые исследователи (см. [8, с. 278]) при переводе и пересказе англоязычных статей, в которых речь идет о complex networks, говорят о них просто как о работах по теории сетей, никак не уточняя, таким образом, собственно характер сетей. Мы, разделяя эту точку зрения, будем пользоваться термином «сети».

В настоящее время особую популярность приобретают методы, связанные с автоматической обработкой текстов, написанных на естественном языке. Стоит отметить, что именно сейчас для таких методов появляется материал: в оцифрованном виде появляется большое количество текстов на исследуемых (в т. ч. малых) языках, это позволяет применять к ним количественные методы. Интернет становится для языка самодокументирующим ресурсом [9, с. 197]: люди пишут тексты на естественном языке, и эти тексты мгновенно становятся доступны исследователям. Вышесказанное в большой степени касается и метода сетей как метода анализа лингвистических данных, (см. [34], [3], а также обзор [42]). Этим и обусловлена **актуальность** настоящей работы.

Сети неоднократно использовались в лингвистических исследованиях. Так, проект Принстонского университета Wordnet, который является огромной структурированной базой лексических единиц английского языка, представляет собой именно сеть, (см. [26]). Сетью же является [37] проект Framenet, посвященный исчислению и описанию фреймов (в терминологии, введённой в [27]). Использование метода сетей позволяет этим проектам показать иерархическую организацию семантики. Оба эти проекта не обращаются к морфологическому уровню языка, в то время как в рамках настоящей работы рассматривается в основном именно он, поэтому далее мы хотели бы подробно рассмотреть работы по морфологии, исследовательским методом которых являлся метод сетей. Прежде всего заметим, что в достаточно подробном обзоре [42], библиография которого насчитывает 62 пункта, нет ни одной ссылки на работу, которая бы использовала метод сетей в морфологии; по всей видимости, это свидетельствует о том, что к настоящему моменту теоретическая морфология и методика сетей далеки друг от друга.

Первой известной нам серией работ по морфологии с использованием сетей (этот проект даже называется по-английски соответствующим образом: network morphology) был ряд исследований, проведённых в университете Саррея в 90-х годах прошлого века, (см., [22], а также [20], [23], [24]), полный список можно найти в сети Интернет по адресу: http://www.surrey.ac.uk/englishandlanguages/research/smg/webresources/network_morphology_bibliography.htm. В 2012 году вышла монография [21], которая подытоживает и обобщает результаты, полученные в рамках этого подхода. Этот подход, первоначально разработанный Гревиллом Корбеттом для описания системы русского словоизменения, состоит в следующем: вершинами сети являются «базовые элементы» (фактически — или леммы, или грамматические классы практически в понимании [7]), способные хра-

нить факты («вершина — это именованная локация, в которой могут храниться один или более фактов» [22, с. 117]). Фактом, в свою очередь, авторы называют некоторую пару «атрибут — значение». Сеть имеет древовидную структуру, поэтому если в данном узле не указано значение некоторого атрибута, то это значение берётся из расположенного выше узла, связанного с данным; цепочка наследования свойств, отмечают авторы, может быть достаточно длинной. Таким образом, возможно, с точки зрения авторов, описать разные морфологические системы (помимо словоизменения в русском языке, они описали, например, систему именных классов языка арапеш); для этих целей даже был разработан язык программирования DATR (см. [25]). Алгоритм работает по следующему принципу: например, если нам нужно узнать все словоформы лексемы *закон*, то мы можем сделать это таким образом: у самого узла *закон* два атрибута: *<stem>* (словоизменительная основа слова) и *<gloss>* (перевод). При этом у него есть связанный с ним узел *Class I* уровня выше, у которого есть атрибуты форм слова типа *<nom sg>*, *<gen sg>* и так далее, со значениями, которые показывают, как образуется соответствующая словоформа: например, атрибут *<gen sg>* имеет значение *<stem> + <a>*. Таким образом, для образования форм этой лексемы нам следует взять *<stem>* слова *закон* и присоединить к нему окончания по правилам, перечисленным в «фактах», присущих первому классу. Отметим, что практически единственное преимущество метода сетей при использовании этого подхода — это возможность иерархической организации лексик и наглядной иллюстрации принципов наследования. Насколько нам известно, никакие иные характеристики сетей в работах Саррейской группы не фигурируют; заметим, что, поскольку в действительности узлы являются достаточно разнородными (это и лексемы, и самые разные грамматические классы), едва ли можно предполагать, что подсчёт каких-либо коэффициентов был бы осмысленным. Кроме того, мы не уверены, что такую сеть можно назвать комплексной.

В работе А. Ч. Пиперски [33] рассматривается применение теории графов для измерения сложности языков. Автор рассматривает именные группы как двудольные графы¹, одной долей которых являются морфемы каждой входящей в состав ИГ словоформы, а второй долей — лексические и грамматические значения, которые выражаются в рассматриваемых именных группах. В работе подсчитана сложность строения именных групп в трёх германских языках (английском, немецком и шведском) на основании числа вершин в обеих долях графа, числа рёбер и средней степени вершины (т. е. среднего количества вершин, с которыми связана вершина). Таким образом, в этой работе сеть используется не ради построения иерархии, а ради подсчёта некоторых простых коэффициентов.

¹ «Граф G называется двудольным, если множество V(G) его вершин можно разбить на два непересекающихся подмножества X и Y таким образом, что у всех рёбер графа G один конец содержится в X, а другой — в Y. При этом графы, порожденные множествами X и Y, называются долями графа G» [Мельников 2010: 50].

Наконец, рассмотрим наиболее близкую по методике к настоящей работе статью [34]. В этом исследовании обсуждается линейный порядок словообразовательных аффиксов в английском языке. Главные вопросы, которые пытаются решить авторы, могут быть сформулированы следующим образом: верно ли, что гипотетический линейный порядок этих аффиксов формирует жесткую иерархию и, если да, то в чём причины формирования именно такой иерархии? Для ответа на первый вопрос авторы строят ориентированную сеть (т. е. такую сеть, каждое ребро которой имеет направление), в которой ребро направлено из аффикса, располагающегося ближе к корню, к его соседу (например, для лексемы *militaryment* строится ребро *ary* → *ment*, а для лексемы *complementaryment* → *ary*), и определяют, является ли такая сеть направленной ациклической сетью, т. е. они выясняют, есть ли в построенной сети такой путь (переход по рёбрам), который начинается и заканчивается в одной и той же вершине; несколько огрубляя, в переводе на собственно лингвистический материал этот вопрос звучал бы как «верно ли, что есть такая пара аффиксов *a* и *b*, которая в некоторых словах встречается в порядке *a-b*, а в некоторых других — в порядке *b-a*?» (хотя, строго говоря, речь могла бы идти не только о парах). После того, как небольшое количество таких пар (а именно 10; например, это суффиксы — *less* и — *ness*) всё-таки находится (это означает, что существуют такие звенья иерархической цепи, где иерархия может нарушаться), авторы генерируют множество случайных графов с тем же количеством вершин и рёбер, что и в полученном ими графе (31 и 161 соответственно), и фиксируют вероятность наличия всего лишь 10 циклов в них. Получается, что она исчезающе мала, а значит, можно говорить о жёсткости иерархии и о скорее случайном характере исключений из неё; подробный лингвистический анализ исключений доказывает эту гипотезу.

Таким образом, в отличие от большинства обсуждаемых выше работ, в этом исследовании вопрос иерархической организации сети является хотя и в некотором смысле главным, но не единственным: во-первых, эта иерархия не самоочевидна, авторы должны доказать ее существование; во-вторых, авторы пользуются сетью как источником получения некоторых численных характеристик (в данном случае — числа циклов и последующего моделирования этого коэффициента у множества случайных графов), которые затем осмысливаются с позиций собственно лингвистического материала. За этой работой последовали некоторые другие, применявшие ту же методологию для определения иерархии порядка следования аффиксов в других языках и определения случайности характера исключений из иерархии (см. [36] про префиксы в русском языке, [32] про суффиксы в болгарском, [43] про суффиксы в итальянском языке и недавнюю статью [40], обобщающую работы, написанные в рамках этой методологии, в целом и отдельно рассматривающей суффиксы в русском языке). Нам неизвестны другие работы (кроме упоминаемых в последних абза-

цах), которые бы использовали сети в морфологии именно с такой позиции; отметим, однако, что основная цель этих работ скорее лежит в области психолингвистики: их авторы пытаются объяснить, почему сложившаяся иерархия устроена именно так, а сети используются лишь при определении иерархии и степени её жёсткости.

Несмотря на то, что понятие агглютинативности известно с XIX века, в научной среде нет однозначного консенсуса относительно того, какими именно чертами должен обладать язык, чтобы считаться агглютинативным. Например, в энциклопедическом словаре мы находим определение, основанное сразу на ряде признаков: «[для агглютинативных языков] <...> характерны развитая система словообразовательной и словоизменительной аффиксации, отсутствие фонетически не обусловленного алломорфизма, единый тип склонения и спряжения, грамматическая однозначность аффиксов, отсутствие значимых чередований» [15, с. 511].

Поскольку наше исследование является в первую очередь количественным исследованием агглютинативности, то далее мы обсудим литературу, посвящённую именно этой проблеме, а не агглютинативности вообще.

Едиственный параметр, которым пользуется при подсчёте индекса агглютинативности в своей хрестоматийной работе (по всей вероятности — первой работе на данную тему в количественном ключе) Джозеф Гринберг, — это наличие/отсутствие фузии: «Если оба морфа в конструкции являются автоматическими вариантами морфем, то такая конструкция называется агглютинативной. Индекс агглютинации — это отношение числа агглютинативных конструкций к общему числу морфных швов» [28, с. 185]. Таким образом, агглютинативная конструкция, по Гринбергу, — это такая конструкция, в которой при присоединении аффикса ни сам аффикс, ни основа слова не претерпевают никаких значимых фонологических изменений.

После работы Гринберга был проведён ряд количественных исследований в схожем ключе, в частности, в работе [31] предлагается модифицировать некоторые из индексов Гринберга; об индексах в языке и методике их подсчёта рассуждает и Г. Г. Сильницкий в работе [39], но в обеих этих работах индекс агглютинативности подсчитывается на основании всё того же единственного параметра — характера изменений на морфемном шве, причём изменений как основы, так и аффикса.

Немецкий типолог Мартин Хаспельмат провёл количественное исследование [30], одной из целей которого было узнать, есть ли корреляция между разными чертами, каждая из которых в разное время отмечалась в литературе (сам Хаспельмат ссылается на [16], [18] и ряд других работ) как ключевое свойство для определения морфологического типа языка (в том числе агглютинативного). Так, Хаспельмат, подытоживая обсуждения своих предшественников, изучал связь между тремя величинами (он отмечает, что хотя таких свойств на самом деле больше трёх, остальные по раз-

ным причинам либо не влияют самостоятельно на тип языка, либо являются слишком периферийными для того, чтобы считаться определяющими свойствами) в 30 языках, отличающихся генетически и распространённых в разных ареалах; 12 из них ранее описывались в литературе как агглютинативные:

– индекс кумуляции (Cumulation index): какой процент аффиксов в языке выражают более одного грамматического значения? Предполагается (см., например, [44, с. 175]), что в агглютинативных языках один аффикс выражает ровно одно грамматическое значение;

– индекс фузии (Alternation index): какой процент грамматических значений может быть выражен (или совыражен) посредством изменения внутри основы слова. Этот индекс похож на индекс агглютинативности Гринберга, но с тем существенным отличием, что рассматриваются изменения только в пределах основы слова. Предполагается, что в агглютинативных языках основа слова (почти) никогда не меняется, именно на этот признак агглютинативных языков обращали внимание лингвисты еще в XIX веке (см., например, [45]);

– индекс грамматического супплетивизма (Suppletion index). Принцип этого индекса состоит в следующем: во флективных языках есть разные парадигмы словоизменения, например, в русском языке у существительных выделяется несколько склонений (традиционно – 3, но возможны и другие точки зрения, см. [7]) а также подтипы внутри каждого склонения. Соответственно, если мы хотим ответить на вопрос типа «при помощи какого аффикса в русском языке выражается значение творительного падежа», то полный правильный ответ будет включать в себя несколько аффиксов (-ой, -ем и т. п.). Таким образом, в этом случае мы наблюдаем явление грамматического супплетивизма: одно и то же грамматическое значение может быть выражено разными аффиксами. Предполагается (см. [41]), что в агглютинативных языках индекс грамматического супплетивизма (равный числу граммем, в которых это явление фиксируется) должен стремиться к нулю: одно значение может быть выражено только одним аффиксом.

В результате своего исследования Хаспельмат пришёл к выводу, что эти параметры в действительности не коррелируют (и вышеприведённые нулевые гипотезы про морфологические явления в агглютинативных языках зачастую оказываются неверными); кроме того, языки, принадлежащие, согласно существующим описаниям, к разным морфологическим типам, могут получать очень близкие значения по рассматриваемым индексам. Таким образом, «после этого исследования, термины *агглютинация* и *фузия* сильно потеряли в своей легитимности (кроме тех случаев, когда они употребляются в техническом смысле, который расходится со стандартными контекстами)» ([30, с. 27]). Таким образом, вопрос о том, какие свойства определяют морфологический тип языка (и в том числе, его агглютинативный характер), во многом остается открытым.

Ещё одной характерной чертой агглютинативных языков можно назвать транскатегориальность. Транскатегориальность — это способность некоторой языковой единицы (обычно — морфемы или клитики) сочетаться с разными классами слов (частями речи). Как отмечает В. А. Плуныян (ср. также более подробное обсуждение этой проблемы в специальной статье того же автора [35]), «в лингвистической типологии принято связывать долю транскатегориальных морфем со степенью агглютинативности или аналитичности языка» [12, с. 101].

Башкирский язык принадлежит к кыпчакской ветке тюркской группы языков. Принятый в 1940 году алфавит башкирского языка создан на основе кириллицы с некоторыми дополнительными символами (см. [4, с. 12]). Языки тюркской группы имеют ярко выраженные морфологические особенности, в частности, их принято считать языками агглютинативного типа. Словоизменение (как и словообразование, насколько их вообще стоит различать в агглютинативном языке, см., например, [4, с. 89]: «между словообразованием и формообразованием <...> нет четких границ») осуществляется при помощи суффиксов. Префиксов в башкирском языке не существует (Н. К. Дмитриев [5, с. 44] на этом основании предлагает даже заменить термин «суффикс» на термин «аффикс» для башкирского языка вследствие отсутствия позиционной оппозиции), а порядок суффиксов жёстко фиксирован (впрочем, на периферии языка есть и исключения из привычного порядка; в разделе 3.4 мы подробно обсуждаем проблематику порядка суффиксов). Как и в других тюркских языках, широко развита алломорфия, у одного аффикса может быть до 16 алломорфов; в языке наблюдается сингармонизм по ряду и по огубленности (тембровый и лабиальный в терминологии [15, с. 445], подробнее о сингармонизме в тюркских языках см. [14]). С точки зрения нашего исследования важно отметить, что это означает, что, например, алломорф некоторого аффикса, содержащий гласный переднего ряда, практически никогда не может соседствовать в рамках одной словоформы с алломорфом другого аффикса, содержащим гласный непереднего ряда.

Из сказанного выше можно заключить, что основная цель настоящей работы — применение метода сетей для решения широкого спектра задач, которые ранее, за редким исключением, не решались этим методом ни в одном языке.

Нашими данными послужил корпус, составленный из текстов статей газеты «Йәшлек» (*‘молодость’*) за 2007–2014 гг. Суммарный объём корпуса — 5,8 млн словоупотреблений. Тексты газеты были размечены при помощи морфологического анализатора *bashmorph* (далее — парсера; разработчик — Б. В. Орехов, см. подробнее [11]). Каждому слову приписывалась морфологическая аннотация, но при этом грамматическая неоднозначность не снималась, в анализаторе отсутствует соответствующая функциональность. Структура разметки на выходе программы

устроена следующим образом: сначала приводится словоформа, затем — лемма, потом следует аффиксный состав, далее — перевод леммы на русский язык, после чего читатель находит глоссы для каждой морфемы исходной словоформы; в случае, если теоретически возможно несколько разборов, то они разделены знаком вертикальной черты. Символы оформления вывода программы в основном следуют формату, предусмотренному в морфологическом анализаторе для русского языка «Mystem» [38].

Пример (1) — это пример данных на выходе работы программы:

$\dot{y}ыл.\{\dot{y}ыл+\emptyset=\langle год \rangle=S=NOM, SG \mid \dot{y}ыл+l=\langle собирать \rangle=V=PASS\}$ (1)

Нашим материалом являлись все возможные разборы (так, из примера (1) мы использовали оба разбора). В ходе предварительной обработки корпуса мы извлекли из каждого возможного разбора только цепочки аффиксов. Далее мы построили сеть, вершинами в которой являлись аффиксы; ребро между аффиксами появлялось в том случае, если был найден хотя бы один разбор, в котором они следуют друг за другом и между ними нет никакого другого аффикса. У каждого ребра есть вес — количество примеров совместной встречаемости двух аффиксов.

Рассмотрим пример (2):

бала-лар-ым-ды (2)
child-PL-POSS.1SG-ACC
'моих детей'

При обработке такого разбора программой, генерирующей сеть, в графе строилось два ребра: *лар-ым* и *ым-ды*. Если таких рёбер до этого в графе не было, то вес каждого ребра был бы равен единице; если такие рёбра существовали до обработки этого разбора, то к весу каждого из них добавлялась единица.

При этом следует отметить несколько важных нюансов:

1. В отсутствие механизмов дизамбигуации очень остро стоит проблема нахождения грани между адекватным отражением реалий языка и таким «прескриптивным» набором правил анализатора, который бы не допускал большого количества ошибочных разборов: действительно, если допустить стопроцентную транскатегориальность и свободный порядок аффиксов, то высок риск получить большое количество неверных разборов как минимум из-за омонимии типа той, которая представлена морфом *де*, который может быть показателем как аккузатива, так и прошедшего времени; существует и ряд других проблем, которые мы не будем детально рассматривать, подробное их описание могло бы стать материалом для отдельного исследования. Если же максимально ограничить возможный набор аффиксов у разных лексем — и качественно, и в отношении линейного порядка — то велика вероятность упустить существующую в языке вариативность. Разработчик парсера решил эту проблему

следующим образом: цепочка аффиксов каждой лексемы рассматривается с конца, при этом для каждой части речи задана возможная цепочка аффиксов (поэтому, например, парсер не предлагает вариантов разбора вида именной суффикс + глагольный). В случае если при этом основа нашлась в словаре, а аффиксальная цепочка удовлетворяет заданным правилам, разбор считается принятым (об иерархии разборов см. ниже). Однако, отдавая себе отчёт в том, что реалии языка могут оказаться шире заданных правил, разработчик добавил в программу следующую опцию: в случае, если программа не выделила описанного выше разбора, но при этом аффиксы образуют некоторую другую последовательность с найденной в словаре основой, то тогда принятым считается такой разбор; остальные типы разборов находятся ниже него в иерархии. Не во всем соглашаясь с такой схемой приоритетов, мы разделяем ту позицию, что граница, о которой мы упоминали в начале настоящего абзаца, действительно очень шаткая и что такой выход, возможно, является оптимальным.

2. Парсер предлагает нулевое окончание (обозначается как \emptyset) в именительном падеже единственного числа и повелительном наклонении единственного числа.

3. В парсере существует трехуровневая иерархия разборов: разборы третьего уровня — это такие, при которых основа после разделения слова на аффикс и гипотетическую основу не обнаружена в словаре; разборы второго уровня — аналогичная ситуация, при которой гипотетических аффиксов выделено несколько; разборы первого уровня — «уверенные» разборы, при которых основа в словаре парсера нашлась. Если парсер предлагал разбор более высокого уровня, то остальные разборы отбрасывались.

4. Вначале был построен ненаправленный граф (т. е. рёбра не имели направлений).

В процессе работы мы построили несколько сетей. В первую очередь, необходимость построения нескольких сетей обусловлена тем, что некоторые характеристики актуальнее смотреть на такой сети, в которой каждый алломорф каждого аффикса представлен как самостоятельная вершина, а некоторые другие — на такой, в которой вершинами являются только морфемы, а алломорфы одной морфемы представлены как единое целое. Помимо этого, при анализе первой сети обнаружилось, что в некоторых случаях парсер предлагает невозможные разборы. Поскольку настоящая работа носит в том числе и прикладной характер, мы сознательно обсуждаем далее вопросы отражения сетью не только собственно языковых реалий (хотя этим вопросам мы, безусловно, отдаём предпочтение), но и те структурные недочёты работы парсера, которые с ее помощью удалось выявить.

В результате обработки данных мы получили следующие результаты. В сети 293 вершины; за вычетом искусственно введённого нами нуле-

вого окончания, можно говорить о том, что обнаружено 292 морфа. Эти 293 вершины связаны 3214 рёбрами; таким образом, средняя степень вершины — 21,94, а плотность сети (реальное количество рёбер относительно максимально возможного) — 0,075. Однако при переходе к другим характеристикам обнаружилось, что в сети имеются петли («Рёбра вида (а, а) или {а} называются петлями» [10, с. 16]): ребро от некоторого аффикса вело к нему самому. Это означает, что парсер разобрал некоторые словоформы таким образом, что некоторые аффиксы повторялись в словоформе, при этом непосредственно следуя друг за другом, что невозможно в башкирском языке. Рассмотрим пример (3):

Акмулла

{акмул?+л+а="?"=V=PASS, PRES|акму?+л+л+а="?"=V=PASS, PASS, PRES} (3)

Как мы видим, башкирское имя *Акмулла* разбирается как глагольная форма, второй из вариантов разбора которой включает в себя два показателя пассива — *л-*. Все примеры на 19 циклов, обнаруженных нами в построенной сети, были неверными разборами¹ (чаще всего в верном разборе этих словоформ должна была фигурировать основа длиннее той, в которой предлагался повтор аффиксов; эта основа, соответственно, оканчивалась на то же буквосочетание, которым далее выражался настоящий аффикс). В 11 из 19 случаев фигурировали однобуквенные аффиксы.

После этого мы поменяли логику работы парсера, чтобы устранить этот недочёт. Парсер стал работать следующим образом: в процессе работы программа собирает в памяти все возможные разборы, и на финальной стадии ранжирует их по вероятности. Наиболее вероятными признаются (1) разборы, которые завершились успешным нахождением основы в словаре. При этом не учитываются такие разборы, которые привели к нахождению присутствующей в словаре основы, но содержащие при этом однобуквенный аффикс. Если разборы (1) имеются в финальном наборе, остальные разборы отбрасываются.

Разборами типа (2) мы считаем такие разборы, при которых и основа, и аффиксы нашлись в словаре, но при этом аффиксы нарушают заданную программой последовательность своего линейного порядка.

Если разборов типа (1) и (2) не нашлось, то показываются разборы (3), удовлетворяющие таким условиям:

- либо основа нашлась в словаре после восстановления цепочки аффиксов, содержащих однобуквенный морф;
- либо основа в словаре не нашлась, но была восстановлена гипотетически после выделения нескольких (больше одного) аффиксов, ни один из которых не однобуквенный.

Если разборы (3) имеются в финальном наборе, остальные разборы отбрасываются.

¹ Мы благодарны А.А. Галлямову за экспертную оценку этих разборов.

Если разборов (3) не нашлось, то показываются разборы (4), где основа не нашлась в словаре, но была восстановлена гипотетически после выделения одного аффикса (одноаффиксной цепочки).

Если разборы (4) имеются в финальном наборе, остальные разборы отбрасываются.

Если разборы (1), (2), (3) и (4) отсутствуют в финальном наборе, то показываются разборы (5), содержащие не найденную в словаре гипотетическую основу, восстановленную после выделения аффиксной цепочки, содержащей хотя бы один однобуквенный аффикс.

При этом действует запрет на контактный повтор алломорфов одной морфемы, чтобы избежать разборов типа показанного в примере (3).

После этого мы переразметили исходный корпус и построили ещё одну сеть. В неё вошли только те разборы, которые содержали найденные в словаре основы, — предварительный анализ показал, что фильтрация по этому критерию значительно увеличивает качество анализируемого материала (следовательно, мы работали только с разборами типа (1) и (2) из классификации выше). Таким образом, нашим материалом послужили 4192760 разборов; заметим, что иногда у словоформы может быть несколько разборов, при которых основа находится в словаре, поэтому собственно словоупотреблений в нашем корпусе оказалось меньше: 2330595.

Эта сеть представлена на рисунке 1. Её базовые характеристики оказались следующими: 279 вершина, 3187 рёбер, средняя степень вершины — 22,84, плотность — 0,082. Средний вес ребра (т. е. средняя частотность одной пары морфов) — 1756.

Диаметр графа (самый длинный путь в графе) — 6; таким образом, есть большой соблазн сказать о том, что максимально возможная цепочка аффиксов в башкирском языке, по-видимому, равна 6. И действительно, длина самой длинной засвидетельствованной цепочки в нашем корпусе разборов оказалась равна 6, см. пример (4):

<i>Аша-ган</i>	<i>ризык-тар-быз-за</i>	<i>ла,</i>	<i>айырыуса</i>		
есть-PST	пицца-PL-POSS.1PL-LOC	PTCL	особенно		
<i>сит-тэн</i>	<i>кил-тер-ел-гэн-дэр-ен-дэ</i>		<i>нитрат-тар</i>	(4)	
крайний-ABL	прийти-CAUS-PASS-PST-PL-POSS.3-LOC		нитрат-PL		
<i>һәм</i>	<i>организм</i>	<i>өсөн</i>	<i>зыян-лы</i>	<i>элемент-тар</i>	<i>күп.</i>
CNJ	организм	для	вред-ADJR	элемент-PL	много

‘В пицце, которую мы едим, особенно заграничной (букв. «принесённой из-за границы»), много нитратов и вредных веществ’

Отметим, что в башкирско-русском словаре под редакцией К. З. Ахмерова помимо глагола *килеу* «приходить, прийти» [1, с. 257] есть и глагол *килтереу* «приносить, доставлять» [1, с. 259], т. е. каузатив от глагола *килеу* можно считать лексикализированным.

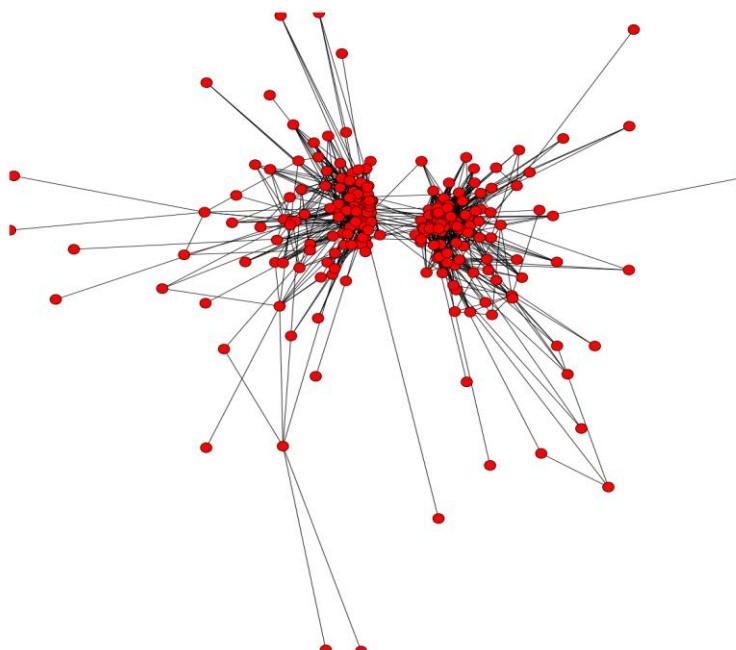


Рис. 1. Сеть совместной встречаемости аффиксов в башкирском языке

В случае трактовки формы *килтерелгәндәрәндә* как формы глагола *килтәреү* мы получаем цепочку из 5 аффиксов. Поскольку других цепочек из 6 аффиксов мы не зафиксировали, то, как кажется, можно констатировать, что цепочки длиннее, чем из 5 аффиксов, крайне редко встречаются в естественных текстах на башкирском языке. Примечательно, что в известном нам квантитативном исследовании (проведённом, правда, на материале корпуса значительно меньшего объёма) башкирского языка [13] автор пришёл к схожим выводам: «Словоизменений, содержащих более пяти аффиксов, не обнаружено» ([13, с. 14]). Хотя, например, Н. К. Дмитриев [5, с. 44] приводит пример словоформы (5), в которой содержится шесть аффиксов.

урман-дар-ыбыз-за-гы-лар-дан (5)
лес-PL-POSS.1PL-LOC-POSS.3-PL-ABL
'из тех, которые находятся в наших лесах' (перевод Н. К. Дмитриева)

Н. К. Дмитриев, однако, не приводит источника этой словоформы и её контекста: вполне возможно, что это не произведённый носителем текст, а сконструированная самим исследователем словоформа.

С точки зрения теории графов, можно заключить, что, возможно, диаметр графа может отражать длину максимальной цепочки аффиксов, но для того, чтобы утверждать это наверняка, нужна большая языковая выборка.

Таким образом, на данном этапе приложения метода сетей к агглютинативному материалу башкирского языка удалось выяснить, что значимыми для описания морфологии тюркского языка оказываются такие базовые характеристики сети, как число вершин, число рёбер, средний вес ребра и диаметр графа. В то же время средняя степень вершины и плотность графа не нашли своих лингвистически значимых эквивалентов и их встраивание в теоретическую сторону описания морфологии требует дальнейших исследований.

СПИСОК ЛИТЕРАТУРЫ

1. Башкирско-русский словарь / под ред. К. З. Ахмерова, Т. Г. Баишева, Г. Р. Каримовой, А. А. Юлдашева. М.: Гос. изд-во иностранных и национальных словарей, 1958. 804 с.
2. Болгова Е. В., Иванов С. В., Гринина Е. А., Слоот П. М. А., Бухановский А. В. Параллельные алгоритмы моделирования динамических процессов на комплексных сетях // Известия высших учебных заведений. Приборостроение. 2011. Вып. 10. С. 72–79.
3. Браславский П. И., Мухин М. Ю., Ляшевская О. Н., Бонч-Осмоловская А. А., Крижановский А. А., Егоров П. В. YARN: начало // Компьютерная лингвистика и интеллектуальные технологии: по материалам ежегодной Международной конференции «Диалог» (Бекасово, 29 мая – 2 июня 2013 г.). Электронная публикация. URL: http://www.dialog-21.ru/digests/dialog2013/materials/pdf/BraslavskiyP_YARN.pdf.
4. Грамматика современного башкирского литературного языка / под ред. А. А. Юлдашева. М.: Наука. 1981. 495 с.
5. Дмитриев Н. К. Грамматика башкирского языка. М. – Л.: Наука. 1948. 276 с.
6. Додонов А. Г., Ландэ Д. В., Путьтин В. Г. Компьютерные сети и аналитические исследования. Киев: ИПРИ НАН Украины, 2014. 486 с.
7. Зализняк А. А. Грамматический словарь русского языка. Словоизменение. М.: Русский язык, 1977. 880 с.
8. Малков А. С., Зинькина Ю. В., Коротаев А. В. К математическому моделированию степенных и сверхстепенных распределений в социальных системах // Моделирование и прогнозирование глобального, регионального и национального развития / отв. ред. А. А. Акаев, А. В. Коротаев, Г. Г. Малинецкий, С. Ю. Малков. М.: ЛИБРОКОМ, 2011. С. 277–304.
9. Махмутов З. А., Орехов Б. В. Методологические и практические аспекты изучения национальных интернетов в России // Русский язык и новые технологии: коллективная монография / сост. Г. Ч. Гусейнов; под ред. М. В. Ахметовой, В. И. Беликова. М., 2014. С. 189–199.
10. Мельников Ю. Б. Элементы теории графов: раздел электронного учебника для сопровождения лекции. 3-е изд., испр. и доп. Екатеринбург. 2010. URL: <http://lib.usue.ru/resource/free/12/MelnikovAlgebra4/00GraphTh.pdf>
11. Орехов Б. В. Проблемы морфологической разметки башкирских текстов // Труды Казанской школы по компьютерной и когнитивной лингвистике TEL-2014. Казань: Изд-во «Фэн» Академии наук РТ. 2014. С. 135–140.
12. Плунгян В. А. Введение в грамматическую семантику. Грамматические значения и грамматические системы языков мира. М.: РГГУ. 2011. 672 с.

13. Суразитдинов З. А. Моделирование грамматики башкирского языка. Словоизменительная система. Уфа: АН РБ, Гилем. 2006. 160 с.
14. Черкасский М. А. Тюркский вокализм и сингармонизм. М.: Наука 1965. 142 с.
15. Языкознание. Большой энциклопедический словарь / гл. ред. В. Н. Ярцева. 2е изд. М.: Большая Российская энциклопедия, 1998. 685 с.
16. Anderson S. R. Typological distinctions in word formation // Language typology and syntactic description. vol. III. Cambridge: CUP, 1985, pp. 3–56.
17. Bliss N. T., Schmidt M. T. Confronting the challenges of graphs and networks. // Lincoln laboratory journal, 2013, vol. 20, no. 1, pp. 4–9.
18. Bloomfield L. Language. University of Chicago Press. 1933. 564 p.
19. Boccaletti S., Latora V., Moreno Y., Chavez M., Hwang D-U. Complex networks: structure and dynamics // Physics Reports, 424, Issue 4–5, 2006, pp. 175–308.
20. Brown D., Corbett G., Fraser N., Hippisley A., Timberlake A. Russian Noun Stress and Network Morphology. // Linguistics 34, 1996, pp. 53–107.
21. Brown D., Hippisley A. Network Morphology: a Defaults-based Theory of Word Structure. Cambridge: CUP. (Cambridge Studies in Linguistics). 2012. 346 p.
22. Corbett G., Fraser N. Network morphology: a DATR account of Russian nominal inflection. // Journal of linguistics 29, 1993, pp. 113–142.
23. Corbett G., Fraser N. Gender assignment: a typology and a model. // Gunter Senft (ed.) Systems of Nominal Classification (Language, Culture and Cognition 4). Cambridge: CUP, 2000, pp. 293–325.
24. Erjavec T. Treatments of Slovene verb morphology in inheritance models. MSc thesis, University of Edinburgh. 1992.
25. Evans R., Gazdar G. Inference in DATR. // Proceedings of the fourth conference of the European Chapter of the Association for Computational linguistics. Manchester, 1989, pp. 66–71.
26. Fellbaum C. WordNet: An Electronic Lexical Database. Cambridge, MA: MIT Press, 1998.
27. Fillmore Ch. Scenes-and-frames semantics. // A. Zampolli, ed. Linguistic Structures Processing. Amsterdam: North-Holland, 1977, pp. 55–81.
28. Greenberg J. A quantitative approach to the morphological typology of language // International Journal of American linguistics, no. 26, 1960, pp. 178–194.
29. Gross J., Yellen J. Fundamentals of Graph theory. // Handbook of Graph Theory / J. L. Gross, J. Yellen (ed.). CRC Press, 2003, pp. 2–20.
30. Haspelmath M. An empirical test of the Agglutination Hypothesis. // Scalise, Sergio & Magni, Elisabetta & Bisetto, Antonietta (eds.). Universals of language today. (Studies in Natural Language and Linguistic Theory, 76.) Dordrecht: Springer, 2009, pp. 13–29.
31. Krupa V. On quantification in typology. Linguistics 12, 1965, pp. 31–36.
32. Manova S. Suffix Combinations in Bulgarian: Parsability and Hierarchy-based Ordering. // Morphology 20(1), 2010, pp. 267–296.
33. Piperski A. An application of graph theory to linguistic complexity. // Yearbook of the Poznan linguistic meeting, 1, 2014, pp. 89–102.
34. Plag I., Baayen R. H. Suffix ordering and morphological processing. // Language, 85, 2009, pp. 109–152.
35. Plungian V. Agglutination vs. flection. // Haspelmath, Martin et al. (eds) Language typology and language universals: An international handbook (HSK), vol. I. Berlin: de Gruyter, 2001, pp. 669–678.

36. Reynolds R. Out of Order?: Russian Prefixes, Complexity-based Ordering and Acyclicity. // University of Pennsylvania Working Papers in Linguistics. Volume 19, Issue 1: Proceedings of the 36th Annual Penn Linguistics Colloquium. Article 19. 2013
37. Ruppenhofer J. et al. Framenet II: Extended theory and practise. Berkeley, California: International Computer Science Institute. Manuscript. 2006.
38. Segalovich I. V. A fast morphological algorithm with unknown word guessing induced by a dictionary for a web search. // Proceedings of MLMTA-2003. Nevada. 2003, pp. 273–280.
39. Silnitsky G. Typological indices and language classes: a quantitative study. // Glottometrika 14. 1993, pp. 139-160.
40. Sims A., Parker J. Lexical processing and affix ordering: cross-linguistic prediction // Morphology, 25(2), 2015, pp. 143-182.
41. Skalička V. Das Erscheinungsbild der Sprachtypen // Typologische Studien / Skalička V. Braunschweig: Vieweg, 1979. ss. 21-58.
42. Solé R., Murtra B., Valverde S., Steels L. Language networks: their structure, function and evolution. // Complexity 15(6), 2010, pp. 20-26.
43. Talamo L. Suffix Combinations in Italian: Selectional Restrictions and Processing Constraints. Unpublished ms. 2011
44. Vance T. J. An introduction to Japanese phonology. Albany: SUNY Press, 1987, 226 pp.
45. Schlegel F. Ueber die Sprache und Weisheit der Indier: Ein Beitrag zur Begründung der Alterthumskunde. Heidelberg: Mohr & Zimmer. 1808. 324 pp.

* * *

Kirjanov Denis P., Orekhov Boris V.**NETWORK BASED APPROACH TO THE BASHKIR MORPHOLOGY DESCRIPTION**

(National Research University Higher School of Economics, Moscow)

This study introduces a complex networks-based approach to quantifying agglutination. This approach is one of the most powerful ways of model description but it has been rarely used for linguistic needs and there are very few papers where it is applied to morphology.

The Bashkir language belongs to the Turkic languages which are considered to be agglutinative. Although the notion of agglutination was introduced in the 19th century, there is no generally accepted definition of an agglutinative language. Different features were supposed to be necessarily present in an agglutinative language, however, there seems to be no correlation between them. In this study we discuss the data provided by our network and relevant for the notion of agglutination and transcategoriality.

We conducted our study on Bashkir newspaper texts containing 5.8 mln tokens overall. They were annotated with the program «Bashmorph». We built a network where nodes are affixes while edges represent cooccurrence of an affix pair. The network was built as weighted (based on the frequency of cooccurrences) and undirected. The network consists of 294 nodes and 3446 edges.

It turns out that several standard coefficients characterizing such a network help to quantify and describe certain characteristics of a language. In our case, most parameters correspond to agglutination. Namely, we discuss the meaning of assortativity coefficient, cliques number, maximal k-core, cluster coefficient and network density as well as some other data.

Keywords: networks, bashkir language, morphology

REFERENCES

1. *Bashkirsko-russkiy slovar'* (Bashkir-Russian dictionary), ed. K. Z. Akhmerova, T. G. Baisheva, G. R. Karimovoy, A. A. Yuldasheva. M.: Gos. izd-vo inostrannykh i natsional'nykh slovarey, 1958. 804 p.
2. Bolgova E. V., Ivanov S. V., Grinina E. A., Slood P. M. A., Bukhanovskiy A. V. Parallel algorithms of modeling of dynamic processes in complex networks [Parallel'nye algoritmy modelirovaniya dinamicheskikh protsessov na kom-pleksnykh setyakh], *Izvestiya vysshikh uchebnykh zavedeniy. Priborostroenie*, 2011., vol. 10, pp. 72–79.
3. Braslavskiy P. I., Mukhin M. Yu., Lyashevskaya O. N., Bonch-Osmolovskaya A. A., Krizhanovskiy A. A., Egorov P. V. YARN: start [YARN: nachalo], *Komp'yuternaya lingvis-tika i intellektual'nye tekhnologii: po materialam ezhegodnoy Mezhduna-rodnoy konferentsii «Dialog» (Bekasovo, 29 maya – 2 iyunya 2013 g.)*. Elek-tronnaya publikatsiya. URL: http://www.dialog-21.ru/digests/dialog2013/materials/pdf/BraslavskiyP_YARN.pdf.
4. *Grammatika sovremennogo bashkirskogo literaturnogo yazyka* (Grammar modern Bashkir literary language), ed. A. A. Yuldasheva. M.: Nauka. 1981. 495 p.
5. Dmitriev N. K. *Grammatika bashkirskogo yazyka* (Grammar Bashkir language), M. – L.: Nauka. 1948. 276 p.
6. Dodonov A. G., Lande D. V., Putyatin V. G. *Komp'yuternye seti i analiticheskie issledovaniya* (Computer networks and analyzes) Kiev: IPRI NAN Ukrainy, 2014. 486 p.
7. Zaliznyak A. A. *Grammaticheskiy slovar' russkogo yazyka. Slovoizmenenie* (Grammatical Dictionary of the Russian language. Inflection) M.: Russkiy yazyk, 1977. 880 s.
8. Malkov A. S., Zin'kina Yu. V., Korotaev A. V. For mathematical modeling of power and excess power distributions in social systems [K matematicheskomu modelirovaniyu stepennykh i sverkhstepennykh raspredeleniy v sotsial'nykh sistemakh], *Modelirovanie i prognozirovanie global'nogo, regional'nogo i natsional'nogo razvitiya* (Modelling and forecasting of the global, regional and national development), ed. A. A. Akaev, A. V. Korotaev, G. G. Malinetskiy, S. Yu. Malkov. M.: LIBROKOM, 2011, pp. 277–304.
9. Makhmutov Z. A., Orekhov B. V. Methodological and practical aspects of the study of national Internet in Russia [Metodologicheskie i prakticheskie aspekty izucheniya natsional'nykh internetov v Rossii], *Russkiy yazyk i novye tekhnologii: kollektivnaya monografiya* (Russian language and new technologies), comp. G. Ch. Guseynov; ed. M. V. Akhmetovoy, V. I. Belikova. M., 2014, pp. 189–199.
10. Mel'nikov Yu. B. *Elementy teorii grafov: razdel elektronnoy uchebnika dlya soprovozhdeniya lektsii* (Elements of graph theory: a section of the electronic textbook to accompany the lectures), Ekaterinburg, 2010. URL: <http://lib.usue.ru/resource/free/12/MelnikovAlgebra4/00GraphTh.pdf>
11. Orekhov B. V. Problems of morphological marking Bashkir texts [Problemy morfologicheskoy razmetki bashkirskikh tekstov], *Trudy Kazanskoy shkoly po komp'yuternoy i kognitivnoy lingvistike* (Proceedings of the Kazan School of Computer and Cognitive Linguistics) TEL-2014. Kazan': Izd-vo «Fen» Akademii nauk RT, 2014, pp. 135–140.
12. Plungyan V. A. *Vvedenie v grammaticheskuyu semantiku. Grammaticheskie znacheniya i grammaticheskie sistemy yazykov mira* (Introduction to grammatical semantics. Grammatical meanings and grammatical system of languages) M.: RGGU. 2011. 672 p.
13. Sirazitdinov Z. A. *Modelirovanie grammatiki bashkirskogo yazyka. Slovoizmenitel'naya sistema* (Simulation of the grammar of the Bashkir language. Inflectional system.), Ufa: AN RB, Gilem. 2006. 160 p.

14. Cherkasskiy M. A. *Tyurkskiy vokalizm i singarmonizm* (Turkic vowel and vowel harmony) M.: Nauka, 1965. 142 p.
15. *Yazykoznanie. Bol'shoy entsiklopedicheskiy slovar'* (Linguistics. Great Encyclopedic Dictionary), Chief Ed. V. N. Yartseva. M.: Bol'shaya Rossiyskaya entsiklopediya, 1998. 685 p.
16. Anderson S. R. Typological distinctions in word formation // *Language typology and syntactic description* / Shopen, Timothy (ed.). vol. III. Cambridge: CUP, 1985, pp. 3–56.
17. Bliss N. T., Schmidt M. T. Confronting the challenges of graphs and networks. // *Lincoln laboratory journal*, 2013, vol. 20, no. 1, pp. 4–9.
18. Bloomfield L. *Language*. University of Chicago Press. 1933. 564 p.
19. Boccaletti S., Latora V., Moreno Y., Chavez M., Hwang D-U. Complex networks: structure and dynamics // *Physics Reports*, 424, Issue 4–5, 2006, pp. 175–308.
20. Brown D., Corbett G., Fraser N., Hippisley A., Timberlake A. Russian Noun Stress and Network Morphology. // *Linguistics* 34, 1996, pp. 53–107.
21. Brown D., Hippisley A. *Network Morphology: a Defaults-based Theory of Word Structure*. Cambridge: CUP. (Cambridge Studies in Linguistics). 2012. 346 p.
22. Corbett G., Fraser N. Network morphology: a DATR account of Russian nominal inflection. // *Journal of linguistics* 29, 1993, pp. 113–142.
23. Corbett G., Fraser N. Gender assignment: a typology and a model. // Gunter Senft (ed.) *Systems of Nominal Classification* (Language, Culture and Cognition 4). Cambridge: CUP, 2000, pp. 293–325.
24. Erjavec T. *Treatments of Slovene verb morphology in inheritance models*. MSc thesis, University of Edinburgh. 1992.
25. Evans R., Gazdar G. Inference in DATR. // *Proceedings of the fourth conference of the European Chapter of the Association for Computational linguistics*. Manchester, 1989, pp. 66–71.
26. Fellbaum C. (ed.). *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press, 1998.
27. Fillmore Ch. Scenes-and-frames semantics. // A. Zampolli, ed. *Linguistic Structures Processing*. Amsterdam: North-Holland, 1977, pp. 55–81.
28. Greenberg J. A quantitative approach to the morphological typology of language // *International Journal of American linguistics*, no. 26, 1960, pp. 178–194.
29. Gross J., Yellen J. *Fundamentals of Graph theory*. // *Handbook of Graph Theory* / J. L. Gross, J. Yellen (ed.). CRC Press, 2003, pp. 2–20.
30. Haspelmath M. An empirical test of the Agglutination Hypothesis. // Scalise, Sergio & Magni, Elisabetta & Bisetto, Antonietta (eds.). *Universals of language today*. (Studies in Natural Language and Linguistic Theory, 76.) Dordrecht: Springer, 2009, pp. 13–29.
31. Krupa V. On quantification in typology. *Linguistics* 12, 1965, pp. 31–36.
32. Manova S. Suffix Combinations in Bulgarian: Parsability and Hierarchy-based Ordering. // *Morphology* 20(1), 2010, pp. 267–296.
33. Piperski A. An application of graph theory to linguistic complexity. // *Yearbook of the Poznan linguistic meeting*, 1, 2014, pp. 89–102.
34. Plag I., Baayen R. H. Suffix ordering and morphological processing. // *Language*, 85, 2009, pp. 109–152.
35. Plungian V. Agglutination vs. flection. // Haspelmath, Martin et al. (eds) *Language typology and language universals: An international handbook* (HSK), vol. I. Berlin: de Gruyter, 2001, pp. 669–678.

36. Reynolds R. Out of Order?: Russian Prefixes, Complexity-based Ordering and Acyclicity. // University of Pennsylvania Working Papers in Linguistics. Volume 19, Issue 1: Proceedings of the 36th Annual Penn Linguistics Colloquium. Article 19. 2013
37. Ruppenhofer J. et al. Framenet II: Extended theory and practise. Berkeley, California: International Computer Science Institute. Manuscript. 2006.
38. Segalovich I. V. A fast morphological algorithm with unknown word guessing induced by a dictionary for a web search. // Proceedings of MLMTA-2003. Nevada. 2003, pp. 273–280.
39. Silnitsky G. Typological indices and language classes: a quantitative study. // Glottometrika 14. 1993, pp. 139-160.
40. Sims A., Parker J. Lexical processing and affix ordering: cross-linguistic prediction // Morphology, 25(2), 2015, pp. 143-182.
41. Skalička V. Das Erscheinungsbild der Sprachtypen // Typologische Studien / Skalička V. Braunschweig: Vieweg, 1979. ss. 21-58.
42. Solé R., Murtra B., Valverde S., Steels L. Language networks: their structure, function and evolution. // Complexity 15(6), 2010, pp. 20-26.
43. Talamo L. Suffix Combinations in Italian: Selectional Restrictions and Processing Constraints. Unpublished ms. 2011
44. Vance T. J. An introduction to Japanese phonology. Albany: SUNY Press, 1987, 226 pp.
45. Schlegel F. Ueber die Sprache und Weisheit der Indier: Ein Beitrag zur Begründung der Alterthumskunde. Heidelberg: Mohr & Zimmer. 1808. 324 pp.

* * *