З. А. Сиразитдинов

О МОДЕЛИРОВАНИИ СЛОВОИЗМЕНИТЕЛЬНОЙ СИСТЕМЫ ИМЕННЫХ ЧАСТЕЙ РЕЧИ БАШКИРСКОГО ЯЗЫКА ПАРНЫМИ СОЧЕТАНИЯМИ

Башкирский язык относится к агглютинативным языкам, которые характеризуются низкой средней частотой встречаемости словоформ в текстах и большим разнообразием реализуемых от основы самих словоизменительных форм. В табл. 1 приведены средние частоты повторяемости словоформ по текстам ряда языков с сопоставимыми объемами. Данные, за исключением башкирского языка, взяты с работы К. Б. Бектаева [1, с. 47–48]. Показатели по башкирскому языку получены из корпусов башкирского языка (mbfl2.ru).

Таблица 1

Средняя частота повторяемости словоформ в текстах

Тип языка	Язык	Подъязык	Объем текста в словоформах	F (частота повторяемости словоформы)
Флективно-аналитический	Английский	Судостроение	50000	9,94
Флективно-синтетический	Болгарский	Публицистика	51894	5,2
Агглютинативный	Казахский	Публицистика	50000	3,23
	Башкирский	Публицистика (газеты “Йәшлек”, “Киске Өфө”)	57089 59799	3,33 3,48

Средняя частота повторяемости словоформы увеличивается при увеличении текста. Так в текстах английского языка (смешанные тексты всех подъязыков) объемом в 60 123 359 словоупотреблений этот показатель равен 234,56 [2], в текстах же башкирского языка (подъязык художественной литературы) объемом вдвое выше (12 043 546) средняя повторяемость словоформ всего лишь 30,56.

Моделирование словоизменительной системы башкирского языка показывает, что от именной основы возможна практическая реализация 611 словоизменительных форм, от глагольной основы – 665 форм [3, с. 105]. Данные близкие к нашими имеются и по казахскому языку (700 вариантов словоизменительных форм для казахского глагола) [4, с. 88].

Таким образом, формальное описание словоформы, составление моделей словоизменительной системы языка, которое является актуальной задачей прикладной лингвистики в целях разработки систем автоматической переработки текстов, многократно становится значимой для агглютинативных языков. Для решения таких задач нами была предложена псевдотензорная модель, которая дает хорошие результы при разработке системы автоматического морфологического разбора словоформы национального языка [5, с. 20–22].

В процессе работы вознакла идея представить более простую модель словоизменения через порядки расположения морфем. Существует грамматика порядков, теоретическое обоснование которой на основе упорядочения элементов языка по их позиции в правильных последовательностях принадлежит Глисону [6]. Предложенную идею он иллюстрирует на агглютинативном турецком языке. Есть работа по грамматике порядков относительно узбекского языка [7].

Несмотря на некоторые различия, в учебной и академической грамматике башкирского языка для именных частей речи выделяются 15 основных категорий словоизменительных аффиксов [8, 9]. Эти категории представлены в табл. 2.

Таблица 2

Категории словоизменения именных частей речи

1	категория множественного числа
2	категория падежного склонения
3	категория сказуемости
4	категория принадлежности
5	категория вопросительности
6	категория неопределенности (частица неопределенности)
7	категория усиления (усилительно-утвердительная частица)
8	категория притяжательности (с аффиксом **-дыҡы/-деке**)
9	категория уменьшительно-ласкательности (с аффиксом **-ҡай/-кәй**)
10	категория уподобления (с аффиксами **-дай/-дәй, -са/-сә**)
11	усложненная конструкция (с аффиксом **-тағы/-тәге**)
12	категория обладательности (аффиксом **-лы/-ле**)
13	категория лишительности (аффиксом **-һыҙ/-һеҙ**)
14	категория предельности (аффиксом **-ғаса/-гәсә**)
15	категория сравнительной степени (аффиксом **-раҡ/-рәк**)

По Глисону понятие порядка вводится следующим образом: порядок 1 включает все морфемы которые непосредственно следуют за корнем. Порядок 2 включает морфемы, которые могут следовать непосредственно за морфемой первого порядка или непосредственно за корнем. Порядок 3 состоит из морфем, следующих за корнем или за морфемами первого или второго порядка и т.д. [6, с. 164]. Для словоизменительных аффиксов именных частей речи башкирского языка в 1 порядок войдут все 15 аффиксов из таблицы 2. Порядок 2 будет также включать те 15 аффиксов, поскольку все они входят в 1 порядок, и повторение некоторых из них. То же самое произойдет с другими порядками. Применить грамматику порядков непосредственно по Глисону для системы словоизменения башкирского языка оказывается не простой задачей. Отметим, что и Глисон, и Ревзин с Юлдашевой иллюстрировали теорию порядков на примере смешанных словоизменительных, формообразовательных и словообразовательных форм языка. Конкретно только словоизменительная система языка в полном разнообразии всех форм ими не рассматривалась.

Для именных частей речи башкирского языка существуют определенные порядки следования словоизменительных морфем между собой. Так аффикс множественности может сочетаться со всеми аффиксами, при этом данный аффикс стоит перед всеми остальными аффиксами за некотрыми исключениями. Падежные аффиксы могут сочетаться со всеми аффиксами и частицами. При этом они употребляются перед частицами, но после аффиксов, за исключением аффикса сказуемости. Подробный порядок следования аффиксов представлен нами в одной из наших работ (5, с. 14–18). На основе анализа порядка следования аффиксов словизменения нами составлены парные сочетания именных аффиксов. Результат приведен в табл. 3.

Таблица 3

Парные сочетания именных словоизменительных аффиксов

1	2	8	10	12	4
1	3	9	1	12	5
1	4	9	2	12	6
1	5	9	3	12	7
1	6	9	4	12	8
1	7	9	5	12	9
1	8	9	6	12	10
1	11	9	7	12	11
1	13	9	8	12	15
1	14	9	11	13	1
2	3	9	12	13	2
2	5	9	13	13	3
2	6	9	14	13	4
2	7	10	1	13	5
2	15	10	2	13	6
3	5	10	3	13	7
3	6	10	4	13	8
3	7	10	5	13	9
4	2	10	6	13	10
4	3	10	7	13	11
4	5	10	8	13	15
4	6	10	11	14	5
4	7	11	1	14	6
4	8	11	2	14	7
4	10	11	4	15	1
4	11	11	5	15	2
4	14	11	6	15	3
4	15	11	7	15	4
5	6	11	8	15	5
8	1	11	10	15	6
8	2	11	15	15	7
8	3	11	35	15	8
8	5	12	1	15	10
8	6	12	2	15	35
8	7	12	3		

Как видно из табл. 3, парные сочетания образуют определенный порядок, хотя и не в понимании грамматики порядков Глисона, и могут быть использованы при моделировании в целях автоматического описания любой словоформы языка. Проиллюстрируем это на следующих примерах:

1. Словоформа *кешеләребезҙе* формально на основе списка словоизменительных аффиксов и словаря основ может быть описана как *кеш+е+ләр+е+бе+ҙ+ҙе* с моделью сочетаний аффиксов 4+1+4+3+2 и как *кеше+ләр+ебе+ҙ+ҙе* с моделью сочетаний аффиксов 1+4+3+2. Парные сочетания позволяют сразу же отбросить не реализуемый в языке первый вариант.

2. Словоформа *балыҡсыға* может быть формально разбита на модели *балыҡсы* + ға (2) и *балыҡ+сы+ға* (7+2). Задаваемая модель сочетаний (порядков) аффиксов именных частей речи исключает второй вариант.

Аналогичным образом работают и модели парных сочетаний глагольных форм.

При моделировании словоизменительных форм через парные сочетания необходимо учитывать и возможные исключения, типа не полного сочетания (аффикса сказуемости с аффиксом принадлежности: *не* реализуется сочетание принадлежности и сказуемости в одном лице) и некотрые другие.

Разработанная программа автоматического морфологического анализа башкирской словоформы на основе модели парных сочетаний и словаря основ показывает хорошие результаты. Отметим, качество и полнота анализа во многом определяется объемом словаря основ. На сегодня наш словарь включает более 80 000 единиц. Словарь основ создан на базе ранее изданных словарей и пополняется за счет издающихся в ИИЯЛ УНЦ РАН новых академических, учебных и отраслевых лексикографических разработок. Мы считаем, что предложенный подход может быть с успехом применен и для остальных агглютинативных языков народов России, представители которых сегодня заняты разработкой национальных корпусов с системой автоматической переработки и поиска информации [10; 11; 12].

ЛИТЕРАТУРА

1. *Бектаев, К. Б.* Статистико-информационная типология тюркского текста / К. Б. Бектаев. – Алма-Ата : Наука, 1978. – 182 с.
2. *Risland, H. D.* A Basic Writing Vocabulary of Elementary Scool Children / H. D. Risland. – New York, 1945.
3. *Сиразитдинов, З. А.* О моделировании словоизменительной системы и разработке программ автоматического морфологического анализа башкирского языка / З. А. Сиразитдинов, Б. З. Сиразитдинов // Современное казахское языкознание: актуальные вопросы прикладной лингвистики : матер. Междунар. науч.-теор. конф., посвящ. 75-летию проф. А. К. Жубанова. – Алматы, 2012. – С. 103–107.
4. *Жубанов, А. Х.* Основные принципы формализации содержания казахского текста / А. Х. Жубанов. – Алматы, 2002. – 250 с.
5. *Сиразитдинов, З. А.* Моделирование грамматики башкирского языка / З. А. Сиразитдинов. – Уфа: Гилем, 2006. – 159 с.
6. *Глисон, Г.* Введение в дескрептивную лингвистику. М.: Издательство иностранной литературы, 1959. – 485 с.
7. *Ревзин, И. И.* Грамматика порядков и ее использования / И. И. Ревзин, Г. Д. Юлдашева // Вопросы языкознания. – 1969. – № 1. – С. 42–56.
8. Грамматика современного башкирского литературного языка (под. ред. А. А. Юлдашева). – М. : Наука, 1981. – 795 с.
9. Хәҙерге башҡорт теле (педагогия институтының башланғыс кластар факульеты студенттары өсөн дәреслек). – Өфө, 1986. – 308 с.
10. *Куканова, В. В.* Словоизменительные типы в калмыцком языке в свете автоматической обработки текстов (на примере имени существительного) / В. В. Куканова // Вестн. Калмыц. ин-та гуманитар. исслед. РАН. – 2012. – № 2. – С. 168–177.

11. *Салчак, А. Я.* Электронный корпус тувинского языка: состояние, проблемы / А. Я. Салчак, А. В. Байыроол // Мир науки, культуры, образования. – Горно-Алтайск, 2013. – № 6. – С. 408–409.
12. *Бадмаева, Л. Д.* Бурятский языковой корпус: создание, проблемы и перспективы / Л. Д. Бадмаева // Вестн. Бурятского науч. центра Сибирского отд. РАН. – 2013. – № 2 (10). – С. 118–122.