N E W S
OF THE NATIONAL ACADEMY OF SCIENCES OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES
ISSN 1991-346X
Volume 6, Number 298 (2014), 74 – 80

FORMAL MODEL OF MORPHOLOGICAL RULES OF AGGLUTINATIVE LANGUAGES

A. A. Sharipbayev, G. T. Bekmanova, A. S. Mukanova, B. Zh. Yergesh

L. N. Gumilyov Eurasian national university, Astana, Kazakhstan.
E-mail: sharalt@mail.ru; gulmira-r@mail.ru; asel_ms@bk.ru; saturn_banu@mail.ru

Key words: ontology, morphological rules, semantic hypergraph, morphological analysis.
Abstract. The paper presents the ontological models of Kazakh language morphology in the form of semantic hypergraph.

УДК 004.89

ФОРМАЛЬНАЯ МОДЕЛЬ МОРФОЛОГИЧЕСКИХ ПРАВИЛ КАЗАХСКОГО ЯЗЫКА

А. А. Шарипбаев, Г. Т. Бекманова, А. С. Муканова, Б. Ж. Ергеш

Евразийский национальный университет им. Л. Н. Гумилева, Астана, Казахстан

Ключевые слова: онтология, морфологические правила, семантический гиперграф, морфологический анализ.
Аннотация. В работе приводятся онтологические модели морфологических правил в казахском языке в виде семантических гиперграфов.

Очевидная сложность обработки естественно-языковых процессов вызвана трудностью их формализации. Сложность заключается в невозможности словоизменения слов для какой-либо части речи по заданной траектории без предварительной обработки словаря начальных форм, поскольку существует зависимость словоизменения слова от его смысла, то есть от его семантического содержания. В связи с этим необходимо было выбрать формальные средства представления знаний, которые позволяют описать не только структуру, но и семантические признаки языковых единиц.

Одним из формальных средств представления знаний является язык семантических гиперграфов, в котором можно в зависимости от типов связей реализовывать классифицирующие, функциональные, ситуационные, структурные сети и сценарии. Семантический гиперграф является расширением семантических сетей, где естественным образом представляются *n*-арные отношения, которые позволяют задавать не только атрибуты объектов, но и представлять их структурные, «целостные» описания [2,3].

Гиперграф $\boldsymbol{H}$ определяется парой $(\boldsymbol{V}, \boldsymbol{R})$, где $V = \{v_i\}$ – множество вершин; $\boldsymbol{R} = \{r_j\}$ – множество ребер;i = 1, 2, …, n;j = 1, 2, ..., m; каждое ребро представляет собой пару из элементов множество V, т.е. $r = \{(v_{j_s}, v_{j_t})\}$, $j_s \neq j_t$, s, t – натуральные числа.

В настоящее время онтология является мощным и распространенным инструментом моделирования отношений между объектами различных предметных областей. Принято классифицировать онтологии по степени зависимости от задач или прикладной области, по модели

представления онтологических знаний и его выразительным возможностям и другим параметрам [4, 5]. Прикладные онтологии описывают концепты, которые зависят как от онтологии задач, так и от онтологии предметной области.

Прикладная онтология разрабатывается на основе общих принципов построения онтологий, но с учетом использования в качестве модели представления знаний семантических гиперграфов. Данный формализм позволят определить онтологию O в виде тройки: (V, R, K),где V – множество понятий проблемной среды (вершины гиперграфа), R – множество отношений между понятиями (дуги и ребра гиперграфа), а K– множество имен понятий и отношений в данной предметной области.

В данной статье предлагается использовать семантические гиперграфы для представления описания не только атрибутов, но и структуры морфологических единиц казахского языка с учетом их семантики.

Морфологический анализ в казахском языке. Казахский язык относится к тюркской группе языков и характеризуется большим числом словоформ для каждого слова, образованных путем добавления к его концу суффиксов и окончаний. Для него определен строгий порядок аффиксов. В начале к корню слова прибавляются суффиксы затем окончания множественности, притяжательные окончания, падежные окончания, окончания спряжения [6].

Процесс словоизменения основывается на детальном анализе начальной формы слова с целью выделения его морфологических признаков и считывания его семантических признаков из базы знаний. Далее определяется траектория словообразования и словоизменения, происходит сам процесс словообразования исловоизменения на основе семантического гиперграфа.

Морфологические признаки в казахском языке выделяются по следующему принципу. Определяется последняя буква начальной формы слова и относится к одной из категорий. В соответствии с этими признаками осуществляется добавление того или иного окончания. Морфологические признаки казахского языка представлены на рисунке 1.

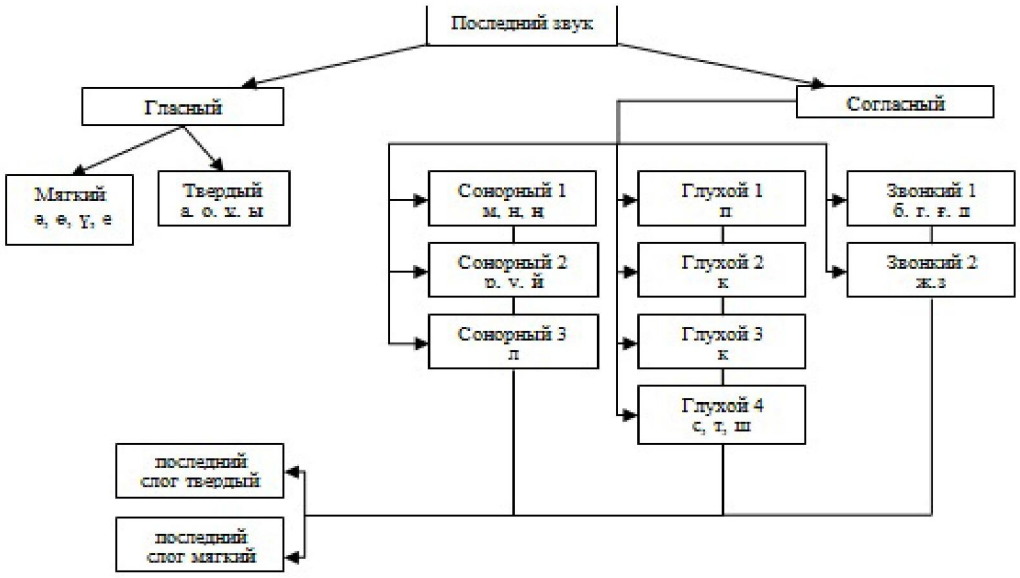

Рисунок 1 – Дерево морфологических признаков

С помощью семантического гиперграфа формализуются правила добавления суффиксов и окончаний к основам. На основе такого гиперграфа генерируются словоформы казахского языка и порождается структура словаря начальных форм в виде синхронизированного дерева.

Синхронизированное дерево представляется с помощью линейных скобочных записей в виде строк, содержащих символы, помечающие узлы дерева, а также открывающие и закрывающие круглые скобки. Между деревьями и их линейными скобочными записями существует взаимно однозначное соответствие. На рисунке 2 приводятся примеры деревьев и соответствующие им линейные скобочные записи.