

А.В. Венцов,  
кандидат биологических наук,  
программист-исследователь  
Института филологических  
исследований СПбГУ,  
Е.И. Риехакайнен,  
кандидат филологических наук,  
доцент кафедры общего языкознания  
им. Л.А. Вербицкой СПбГУ

### **О единицах ментального лексикона носителя русского языка и их морфологическом описании**

В процессе моделирования процедуры восприятия звучащей речи, неминуемо возникает вопрос о единицах ментального лексикона слушающего и о том, какая информация о языковых единицах должна быть отражена во внутреннем словаре, к которому человек обращается при этом. Поскольку распознавание речи начинается с обработки слуховой системой акустического сигнала, приходится допустить, что в ментальном лексиконе (точнее, в том его разделе, который мы называем перцептивным словарем; см. об этом подробнее в [Венцов 2007]) могут (и должны) быть представлены элементы с одинаковой «акустической» реализацией, но с разным грамматическим значением. Например, *в частности* как вводная конструкция и *в частности* – существительное с предлогом; *вначале* – наречие и *в начале* – именная группа; *заработай* – повелительная форма глагола и *за работой* – существительное с предлогом.

При аннотировании материалов Корпуса русской устной речи (<http://russpeech.spbu.ru/>), которые используются для тестирования алгоритмов распознавания естественного речевого сигнала, мы, развивая подход создателей Корпуса русского литературного языка (<http://narusco.ru/>), признаем существование в русском языке так называемых «составных слов», т.е. устойчивых сочетаний лексических единиц, «воспроизводимых в речи как единое целое» [Рогожникова 2003]. Подобные сочетания в речевом корпусе объединяются в целостную графическую единицу путем замены межсловных пробелов символом «=»: *то+=есть, потому+=что, в=ка+честве, должно+=быть* и т.п. В дальнейшем, при морфологической разметке, эти «словоформы» должны получать соответствующее их статусу во фразе морфологическое описание. При таком подходе сочетание *то есть* описывается как союз в отличие от принятого, например, в Национальном корпусе русского языка (<https://ruscorpora.ru/new/>) описания этого сочетания как последовательности местоимения и формы настоящего времени глагола *быть* или союза *то* и того же глагола. Чтобы различать в орфографической транскрипции «фонетические слова», представляющие собой просто именные группы, в их орфографической записи используется «объединяющий» символ «\_»: *в\_ча+стности, должно+\_бы+ть, в\_ка+честве* и т.п. Символ подчеркивания «\_» используется и при орфографическом описании звуковых единиц типа *так\_же, в\_начале, за\_работой, что\_бы, по\_этому* и им подобных.

В звучащей речи часто встречается еще один класс «составных» слов, возникающих при стяжении звуков на стыке словоформ. В их орфографическом описании также используется символ «\_». Можно выделить по крайней мере два класса образующихся в результате стяжения «фонетических» слов: регулярно появляющиеся в речи разных дикторов (высокочастотные) и возникающие у конкретного диктора в конкретной речевой ситуации (низкочастотные).

За основу описания морфологии единиц устной речи мы изначально взяли словарь морфологически размеченных словоформ, который был создан при работе над Корпусом русского литературного языка и использовался в программах полуавтоматического морфологического аннотирования текстов. В этом словаре каждая словоформа представлена всеми возможными в языке вариантами морфологического описания, но без указания частоты встречаемости этих вариантов. При формировании словаря учитывались материалы существующих грамматик и словарей русского языка, в частности Малого академического словаря под ред. А. П. Евгеньевой [МАС].

Работа с корпусом устной речи и попытки создания алгоритма восстановления редуцированных форм вызывают сомнения в оправданности использования подобного словаря в качестве модели ментального лексикона. Мы исходим из того, что носитель языка на основании анализа прозвучавшей словоформы выбирает в ментальном лексиконе ее значение. На сегодняшний день по-прежнему нет никаких сведений о способах кодирования значений в когнитивной системе носителей языка. Поэтому при функциональном моделировании процессов восприятия речи приходится использовать орфографическую запись в качестве маркеров значений соответствующих акустических реализаций, хотя не умеющие читать и писать взрослые носители языка, а также дети в период дописьменного освоения языка очевидным образом обходятся без орфографии при повседневном речевом общении. Тогда возникает вопрос: все ли морфологические формы-значения, включаемые лингвистами в словари и грамматики, необходимы носителю языка при восприятии звучащей речи?

Например, прошедшее время глагола и сослагательное наклонение морфологически описываются по-разному, хотя акустически не различаются. Получается, что сослагательность в процессе восприятия речи определяется только благодаря наличию в клаузе частицы *бы*, т.е. уже **после** лексической интерпретации всех элементов клаузы. Другой пример – словоизменительные грамматические категории зависимых от существительного элементов именной группы, которые слушающий может вывести, предварительно определив морфологические характеристики существительного. На наш взгляд, более полное представление о том, как именно должно выглядеть описание морфологической информации в ментальном лексиконе носителя русского языка, можно получить, если будет описана система синтаксических валентностей, реализующихся в естественной устной речи.

*Исследование выполняется при поддержке гранта РФФИ №19-012-00629.*

#### **Литература:**

*Венцов 2007* – Венцов А.В. Восприятие устной речи и ментальный лексикон // Русская языковая личность: Материалы шестой выездной школы-семинара. Череповец: ГОУ ВПО «Череповецкий государственный университет», 2007. С. 63–69.

*МАС* – Словарь русского языка: В 4-х т. / под ред. А. П. Евгеньевой. 4-е изд., стер. М.: Рус. яз.; Полиграфресурсы, 1999.

*Рогожникова 2003*–Рогожникова Р.П. Толковый словарь сочетаний, эквивалентных слову: около 1500 устойчивых сочетаний русского языка. М.: Астрель, АСТ, 2003.