

УДК 415.5:621.139

ПРИНЦИПЫ РАЗРАБОТКИ МОРФЕМНОГО СЛОВАРЯ РАСПОЗНАВАНИЯ

А.С.Асиновский, С.И.Богданов

Целесообразность использования информации о внутреннем строении слова в целях оптимизации автоматизированных систем анализа языкового текста признается все большим количеством исследователей. Возможности, которые появляются в связи с использованием сведений о морфемных структурах слова, требуют решения вопроса о выборе необходимой лингвистической базы, которая должна представлять собой корпус данных о строении слова, полученный на основе достаточно представительного материала. В идеале в качестве такой лингвистической базы мог бы выступить морфемный словарь русского языка, включающий все разновидности собственно морфемных структур.

Переорганизация такого словаря в конструируемый морфемный словарь распознавания заключалась бы в таком случае в переходе от списка определенным образом организованных (например, по гнездовому принципу) расчлененных на морфемы словоформ к системе использования информации разных лингвистических уровней, в структуре которой центральное положение (т.е. статус базовой единицы записи) отводится морфеме, компоненту собственно морфемной структуры.

Информация о составляющих внутрисловных структурах в блоке автоматического распознавания, который обеспечивает переход от текстовых морфов, соответствующих опорным фонетическим участкам, к словоформам распознаваемого текста, содержится в морфемном словаре распознавания.

В системе блоков автоматического распознавания информация о внутрисловных структурах, с одной стороны, обусловлена содержанием фонологического блока, а с другой — минимальной семантической моделью. Собственно морфемный словарь распознавания можно представить в виде набора морфем (минимальных компонентов структуры сло-

ва), упорядоченных по фонетическим, морфотактическим (прежде всего, позиционным) и морфологическим параметрам. Морфотактическая характеристика элемента разрабатываемого словаря с необходимостью обуславливает то, что в таком словаре оказываются представленными и возможные комбинации морфем, являющиеся промежуточным звеном соотношения элементов морфемной структуры и основы в целом. Существенно отметить, что выбор в качестве базовой единицы записи морфемы обеспечивает возможность наиболее полного системного представления выводимых словоформ.

Для перехода от текстовых морфем к словоформам, реализованным в распознаваемом тексте, морфемный словарь распознавания должен содержать информацию, во-первых, о том, как соотнести текстовой морф и его словарный коррелят, словарный вид морфемы, а во-вторых, как соотнести текстовую реализацию словоформы с возможными внутрисловными контекстами опорного морфа, обеспечивая тем самым полное фонетическое и морфологическое описание реализованных в тексте словоформ. При этом должна сохраняться информация о характере варианта морфемы, поскольку этот вариант используется не во всех внутрисловных контекстах, и далее должны быть рассмотрены только те словоформы, в которых реализуется распознанный текстовой морф.

Поступила в ред.-изд.отд.

2 апреля 1984 года