

**ПРОГРАММНЫЕ СРЕДСТВА МОДЕЛИРОВАНИЯ
СЛОВООБРАЗОВАНИЯ**

И. Е. Воронина

Воронежский государственный университет

Рассматривается реализация программных средств моделирования словообразовательных процессов и их диагностирования.

1. ВВЕДЕНИЕ

Все языковые уровни характеризуются наличием базовых элементов. Например, на комбинаторику букв и слогов накладываются ограничения сочетаемости морфем. Комбинаторика морфем ограничивается сочетаемостью знаков более высокого порядка — слов. Далее речь может идти о сочетаемости словосочетаний и предложений, после чего можно говорить об экстралингвистических композиционно-сюжетных ограничениях. При моделировании естественного языка следует учитывать направление укрупнения языковых объектов.

Таким образом, актуальна задача выявления и программного подтверждения сочетаемости языковых единиц. Выявлению правил может эффективно способствовать наличие программно реализованного набора инструментов, позволяющего автоматизировать заданный процесс, насколько это возможно.

Важную роль играет диагностический инструментарий, целью которого является обеспечение обоснования и поддержки принимаемых решений [1]. Диагностика не обязательно должна порождать количественные оценки с абсолютной степенью точности, поскольку имеет своей целью скорее обеспечение тенденций, подтверждение или опровержение хода исследовательского процесса.

Объект исследования — русское слово — представляет собой последовательность морфем, которые стыкуются друг с другом по определенным правилам. Основная цель разработанных программных и диагностических средств — выявление, подтверждение и фиксация этих правил.

**2. ИСПОЛЬЗОВАНИЕ ПРОГРАММНОГО
ИНСТРУМЕНТАРИЯ ДЛЯ
МОДЕЛИРОВАНИЯ И ИССЛЕДОВАНИЯ
СЛОВООБРАЗОВАТЕЛЬНЫХ
ПРОЦЕССОВ**

Программный инструментарий для моделирования и исследования словообразовательных процессов позволяет выполнить генерацию слова по заданному набору морфем с фильтрацией получаемого материала.

Разработанные программные средства исследования задач словообразования можно разделить на 2 категории:

- 1) средства моделирования словообразовательного процесса;
- 2) средства обучения решению задач словообразования.

Рассмотрим программные средства первой категории, которые предназначены исключительно для исследовательской работы.

На рис. 1 представлена общая схема обработки данных при проведении исследований.

Настройка базовых морфемных наборов — это сервис для задания материала генерации. Для каждой категории морфем предусмотрены возможности просмотра, выбора, сортировки и редактирования.

Выбор решает задачу сужения базовых морфемных наборов, участвующих в словообразовательном процессе. Это необходимо при проведении исследовательских работ по созданию проверки новых фильтров, а также в режиме объяснения (обучения). Выбор можно производить по образцу и перечислением; его можно отменить.

Выбор по образцу подразумевает выбор по фонетической структуре (образец формируется с помощью специальных обозначений, характеризующих фонетическую структуру). Кроме

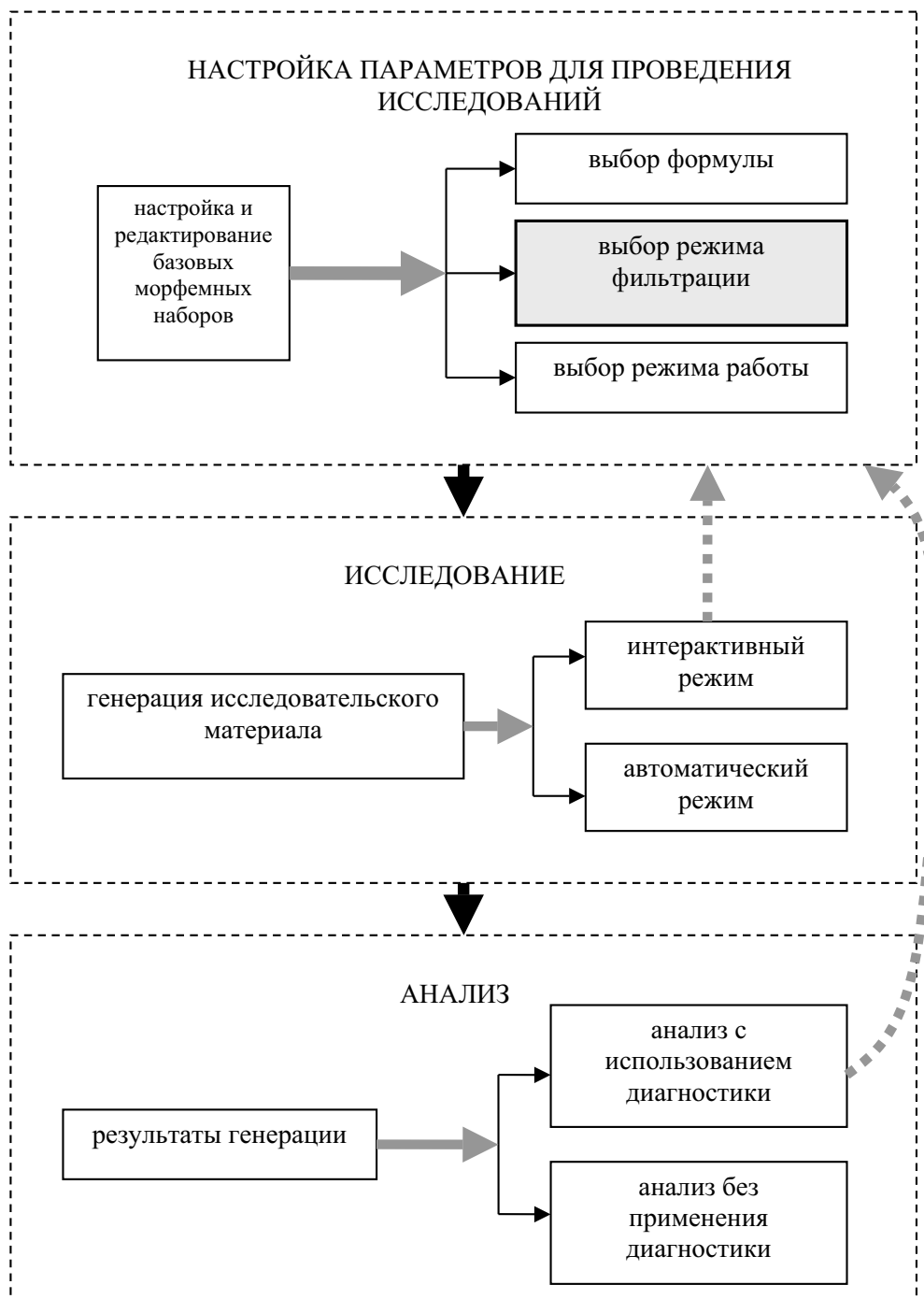


Рис. 1. Схема обработки данных

того, выбор может осуществляться по сочетаемости с окончанием (для корневых и суффиксальных морфем), по маске. В идеале хорошо было бы реализовать выбор по семантике, что само по себе является достаточно нетривиальной исследовательской задачей.

Выбор перечислением подразумевает прямое выделение морфем, которые должны войти в состав активного набора.

Редактирование осуществляется с учетом сочетаемости с окончанием для корневых и суффиксальных наборов.

Для проведения исследования пользователь задает формулу слова, согласно которой будет происходить генерация словарного материала из активных наборов морфем. Предусмотрена возможность фиксации морфем на соответствующих позициях, а также фильтрация сгенери-

рованного материала с использованием выбранных фильтров. Режимы работы программы могут быть следующими: режим объяснения срабатывания фильтра при получении отрицательного материала и автоматический режим.

Предоставляется возможность генерации слов из базовых наборов, когда морфемы выбираются случайным образом, или перебор может быть полным.

Предусмотрено сохранение сгенерированного материала: положительного (прошедшего фильтрацию) и отрицательного (не прошедшего какой-либо фильтр).

Полученный исследовательский материал анализируется. На этой стадии возможно применение диагностических средств. Анализ может осуществляться в процессе наблюдения за поведением словообразовательной системы при заданных параметрах, а также путем изучения полученного при генерации положительного и отрицательного материала.

Что касается второй категории программных средств, то спектр их действия более широк. Первое их функциональное назначение — обучение решению задач словообразования, второе — это инструмент моделирования и исследования словообразовательного процесса, инструмент для поиска и выявления закономерностей и накопления количественной информации. В этом смысле рассматриваемый набор инструментальных средств — необходимая часть поддержки средств первой категории.

Следует заметить, что обе функции (обучающая и исследовательско-диагностирующая) тесно связаны между собой благодаря тому, что в качестве методологической основы создания программного комплекса моделирования словообразовательного процесса выбран метод лингвистического эксперимента. Выполняя роль многоуровневого генератора отрицательного материала, программа выдает информацию, подводящую обучаемого к открытию и познанию правил словообразования, делая его активным участником процесса обучения, а значит, и исследователем.

Структура алгоритмического обеспечения программного комплекса моделирования словообразовательного процесса и обучения решению задач словообразования показана на рис. 2.

Блок управления обучением содержит комплекс средств для задания управляющих установок формирования процесса обучения.

Инициатива выбора установок может принадлежать как преподавателю, так и самому обучаемому, в зависимости от того, как спланированы занятия.

Блок моделирования словообразования включает в себя комплекс средств, осуществляющих генерацию слова по заданным установкам, причем работа может осуществляться и в режиме моделирования словообразовательных сетей.

В блоке формирования результатов формируется выходная информация. При этом, если пользователь (обучаемый) находится в режиме пассивной работы, то есть не занимается анализом материала, а запрашивает полную информацию о процессе словообразования (все правила фильтрации и, согласно им, полную картину прохождения фильтров каждым сгенерированным словом), то оказываются задействованными процедуры формирования выходной информации с полным объяснением).

По запросу пользователя могут быть активированы средства диагностики, которые будут оказывать влияние на процесс выбора новых управляющих установок.

Если имеет место исследовательский процесс, то блок управления обучением выполняет функции формирования управляющих установок, которые будут определять направление исследовательского процесса.

Подсистема моделирования словообразовательного процесса содержит модули фильтрации, модули диагностики фильтрационного процесса, а также сервисные модули, обслуживающие статистику словообразования, пользовательский интерфейс данного режима работы, модули обслуживания файлов результатов и модули поддержки генерации слова в режиме полного перебора или случайным образом.

Набор модулей фильтрации и диагностики является расширяемым, что подразумевает его дополнение в случае разработки и проверки действия новых фильтров. Модули фильтрации, собственно, и содержат саму систему фильтров, через которые пропускается сгенерированное слово. Модули диагностики отвечают за обработку и представление информации о результатах фильтрационного процесса.

Модули пользовательского интерфейса включают в себя инструментальные наборы процедур поддержки интерфейса и обеспечивают не только максимальную легкость работы с

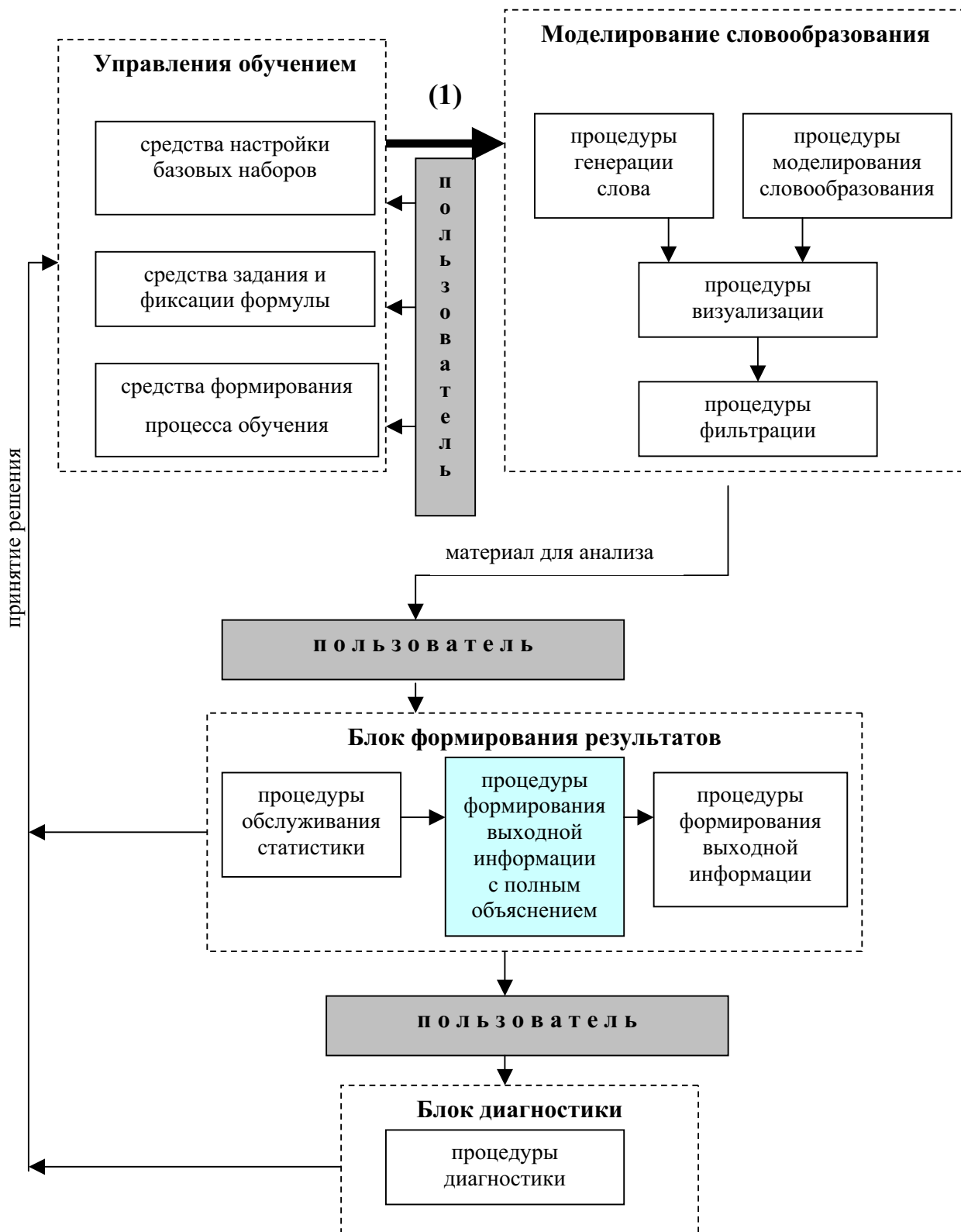


Рис. 2. Структура программных средств: (1) — управляющие установки процесса обучения

программным комплексом, но и берут на себя реализацию методической поддержки процесса обучения.

Подсистема настройки базовых морфемных наборов включает в себя группу модулей определения активных наборов морфем, которые будут принимать участие в словообразовательном процессе, модули поддержки пользовательского интерфейса данного режима функционирования, модули редактирования базового морфемного набора в режиме защиты от несанкционированного доступа, а также модули сортировки морфемных наборов по различным критериям. По сути дела, подсистема настройки базовых морфемных наборов — это набор сервисных средств по подготовке и заданию направления словообразовательного процесса.

Можно говорить и о том, что вышеупомянутые программные и диагностические средства могут быть использованы в качестве инструмента исследования словообразования других индоевропейских языков.

Для моделирования процесса порождения русского слова используется познавательная модель.

Рассмотрим подсистему моделирования словообразовательного процесса с позиции возможности ее адаптации к условиям замены русского языка на какой-либо другой язык.

Если полагать набор морфем переменной частью, то есть подразумевать, что слово может создаваться из морфем некоторого другого индоевропейского языка, отличного от русского, то процедуры все равно не претерпят практически никаких изменений. Что касается сервисных модулей, то замена одного комплекта морфемных наборов на другой не приведет ни к каким существенным изменениям, поскольку это модули, которые отвечают за вывод информации как на экран, так и в файл (или файлы) по желанию пользователя, а также отвечают за задание и фиксацию установок по выводу этой информации. Изменения претерпят лишь процедуры, отвечающие за выдачу на экран правил, зафиксированных в фильтрах, в силу того, что изменятся сами фильтры, которым будут соответствовать новые правила.

Только одна группа модулей полностью должна быть изменена. Это модули фильтрации. Имеются основания полагать, что в каждом языке существуют собственные законы, следуя

которым морфемы образуют то, что можно назвать словом данного языка. Точно так же, как и при рассмотрении русского слова, мы получаем инструмент исследования словообразовательного процесса. Конечно же, модуль статистики должен будет учитывать, с какой системой фильтров он имеет дело, поэтому процедуры, связанные с формированием информации, также необходимо заменить другими.

В подсистеме настройки базовых морфемных наборов изменению подвергнется выбор по образцу, поскольку заранее неизвестно, какие образцы понадобятся для направления процесса исследования. Ясно также, что выбор по семантике — явление уникальное для каждого языка, словообразование которого подвергается анализу. Это всегда отдельная большая задача.

Сервис для задания и фиксации формулы будет изменен лишь в том случае, когда предельная формула слова языка, подлежащего рассмотрению, отлична от аналогичной формулы русского слова.

Что касается **аппарата диагностических средств**, то идеи останутся теми же, что и в случае диагностики задач русского словообразования. В случае другой предельной формулы слова (например, она может отличаться по числу составляющих), изменится порождающая грамматика. Возможно, усложнение вычислений в случае влияния корневой части на префиксальную часть слова.

3. ЗАКЛЮЧЕНИЕ

Таким образом, мы приходим к выводу, что разработанный инструментарий в значительной степени и в существенной своей части обладает свойствами универсальности как средство диагностики исследуемого объекта. Эта универсальность обеспечивается широкими возможностями системы в области настройки на объект исследования в соответствии с его внутренними особенностями, связанными с фонетическим, морфологическим, морфемным, семантическим и синтаксическим своеобразием естественного языка, словообразовательная система которого подлежит моделированию.

Пополнение инструментального комплекса модулем выбора языка, обеспечивающим настройку системы на конкретный естественный язык, позволит под единым углом зрения рассмотреть словообразовательные системы близ-

ко- и дальнородственных языков и тем самым заложить основы для создания сопоставительного словообразования, как минимум, индоевропейских языков и получить новое знание в области дивергенции словообразовательных систем данных языков.

ЛИТЕРАТУРА

1. *Воронина И.Е.* Оценки сочетаемости структурных единиц в задачах формализации естественного языка / *И. Е. Воронина* // *Вестн. Воронеж. гос. ун-та. Сер. Системный анализ и информационные технологии.* — 2006. — С. 51—57.

*Статья принята к опубликованию
25 декабря 2006 г.*