

ставляет значительный интерес своим нетрадиционным подходом к рассмотрению вопроса о роли и функциях порядка слов в предложении и значимости самого фактора инверсии для выделения предложений особого типа. Ее можно рекомендовать не только специалистам по румынскому синтаксису, которые, несомненно, почерпнут из нее ценные сведения о предложениях как единицах текста, но и всем синтаксистам, занимающимся проблемами

актуального членения предложения и проблемами порядка слов в предложении. Она также безусловно полезна для тех, кто работает в области типологии, ибо подкрепляя свои тезисы, М. Ульрих пользуется материалами разных языков, а выводы ее, несомненно, обладают и общетипологической значимостью.

Кубрякова Е. С.

Автоматизация анализа научного текста. Киев: Наукова думка, 1984. 257 с.

Описываемая в рецензируемой монографии система автоматического анализа научно-реферативного текста разрабатывается в рамках создания системы автоматизации редакционно-издательских работ сотрудниками Отдела структурно-математической лингвистики Института языковедения АН УССР. Представленные в ней алгоритмы морфологического и синтаксического анализа могут рассматриваться как первые шаги на пути к разработке единой системы процедур автоматического анализа текста, характеризующихся преемственностью алгоритмов (результаты, получаемые на первых этапах анализа, служат исходными данными для последующих), поэтапным усложнением процедур, общим принципом использования дистрибутивно-статистических характеристик единиц текста, получаемых в результате предварительных исследований его структурных свойств. Описание процедур включает домашний анализ с целью определения формальных признаков, создания лингвистического алгоритма и его машинную (программную) реализацию.

Выбранные в качестве объекта анализа реферативные тексты представляют интерес как особый вид текстов, в которых находит отражение взаимодействие двух лингвистических систем (системы реферирования текста и лингвистической системы собственно реферата), и как разновидность текстов, которые в первую очередь должны подвергаться автоматической обработке, поскольку являются одним из наиболее оперативных средств информирования. Тематическая ограниченность исходных текстов (анализируются рефераты по теме «Программирование и теория математических машин») и стандартизация их формы (выявлению структурных особенностей реферативного текста посвящена монография этого же коллектива [1]) обеспечивают довольно высокий уровень формализации морфологического

и синтаксического анализа реферативных текстов.

Подсистема морфологического анализа (гл. I «Автоматизация МА») представлена в книге алгоритмами определения грамматических классов слов текста, определения словоизменительных характеристик существительного и глагола как основных претендентов на выполнение в тексте функции организующего ядра предложения и алгоритмом определения грамматических подклассов количественных числительных.

Предложенная авторами грамматическая классификация в основном совпадает с традиционным делением слов русского языка на части речи. Отличие обусловлено конкретными задачами создаваемой системы, а также стремлением отразить специфичность текстов рефератов на грамматическом уровне (в отдельные классы выделены аббревиатуры, символы, сокращения слов и словосочетаний, а также слова, написанные буквами нерусского алфавита). Определение грамматических классов слов должно осуществляться в два этапа. На первом на основе выведенных в результате предварительного анализа квазифлексий словам приспосабливаются коды грамматических классов, в том числе и дизъюнктивные (омонимичные), или код ненайденного класса. Здесь используется известная идея Г. Г. Белогова, в соответствии с которой по концу слова можно определить его грамматические характеристики. В отличие от других работ, опирающихся на эту идею, в данном исследовании грамматические характеристики приняты определять по квазифлексиям для всех слов текста (словарь основ в системе не предусмотрен). На втором этапе предусматривается анализ слов с дизъюнктивными кодами по их грамматическому окружению. Список квазифлексий снабжается частотными характеристиками распределения их в тексте. Приведены данные о частоте

квасифлексий различной длины. Результаты проверки работы алгоритма свидетельствуют о его достаточной эффективности: нераспознанными¹ остались 8% анализируемых слов. К сожалению, в книге нет данных о количестве случаев приписывания дизъюнктивных кодов. Уменьшение (снижение) процента неопознанных слов возможно за счет уточнения списка квазифлексий в плане уменьшения их длины, что, естественно, вызовет увеличение количества дизъюнктивных кодов и, следовательно, повысит роль (значимость) второго этапа. Как известно из выступлений авторов рецензируемого сборника на конференциях, именно в этом направлении сейчас и осуществляется исследование.

Принцип использования конечных буквосочетаний слов для определения грамматической информации последовательно проводится и в алгоритмах определения словоизменительных подклассов слов в пределах грамматического класса. Причем, как показано в разделе «Автоматическая идентификация глагольных форм», определение последних в реферативных текстах вообще возможно без обращения к грамматическому контексту. Обусловлено это, как показал автор раздела, двумя важными лингвистическими факторами: во-первых, тем, что в грамматической системе глагольных форм конечные аффиксы имеют по сравнению с аффиксами других грамматических классов невысокую степень омонимии, и, во-вторых, реферативный текст отдает предпочтение только некоторым из них, игнорируя, например, почти полностью формы повелительного наклонения и будущего простого, квазифлексии которых омонимичны квазифлексиям широко распространенным в рефератах форм настоящего времени глаголов несовершенного вида.

Для исследования степени стандартизованности функционирования глагольных форм, а также для определения статистических характеристик распределения их с учетом длины реферата, позиции предложений, содержащих эти формы, предлагается алгоритм, описанный в разделе 4: «Автоматический анализ степени расхождений в грамматических значениях глагольных форм в научно-реферативном тексте». Полученные данные важны для раскрытия закономерностей функционирования глагольных форм в зависимости от структуры текста как линейно развертывающегося целостного образования. В разделе «Алгоритмы автоматического определения словоизменительных подклассов существительных» показано, что однозначное определение рода в пределах класса существительных по списку квазифлексий осуществляется

в 99% случаев; словоизменительные же характеристики существительных могут быть определены по квазифлексиям для 57%, а для остальных 43% омонимия падежа (или падежа и числа) должна сниматься по грамматическому контексту.

Использование сочетания контекстного анализа и анализа квазифлексий при определении грамматических характеристик описано в разделе «Определение подклассов количественных числительных».

Исследования, направленные на изучение синтаксических связей различных грамматических классов слов, представлены в гл. II монографии, где дается алгоритм установления предикативного ядра предложения, состоящего из сказуемого и подлежащего; приводятся также алгоритмы определения беспредложных глагольно-именных связей, считающихся основными выразителями предикации, и алгоритмы установления связей предложения. Алгоритмам свойствен единственный принцип выбора в качестве дифференциальных признаков речевых характеристик функционирования синтаксических связей в текстах данного подязыка, выявленных в результате предварительного дистрибутивно-статистического анализа синтаксической структуры реферата. К этим характеристикам относятся данные о длине зон исследуемых связей, о цепочках классов, стоящих в тексте между участниками связи, о позиции их внутри зоны и по отношению к началу предложения, о комбинационных свойствах грамматических признаков связанных слов. Выделенные признаки снабжаются показателями частоты распространения их в текстах. Показатели используются при составлении алгоритмических правил и для определения их иерархии в алгоритмах. Важны они (как и данные, полученные по алгоритму классификации предложений, представленных в виде цепочек грамматических форм слов) и в теоретическом плане, поскольку раскрывают количественные соотношения языковых и речевых свойств описываемых явлений и могут рассматриваться в качестве стилистически-различительных признаков текстов исследуемого подязыка.

Особого внимания заслуживает раздел, посвященный анализу функционирования знаков препинания в научном тексте. Целесообразность разработки алгоритма анализа пунктуационных знаков обусловлена прежде всего практическими нуждами автоматической переработки текстовой информации, поскольку знаки препинания являются опорными точками сегментирования текста и формальными указателями синтаксических связей слов. В этой роли они используются и в представленных в монографии алгоритмах синтаксического анализа. Кроме того,

определение с помощью ЭВМ дистрибутивно-статистических характеристик знаков препинания в тексте (а именно такова цель алгоритма, описываемого в разделе «Анализ функционирования знаков препинания») на больших выборках даст возможность проверить и уточнить установленные стилеразличительные характеристики знаков препинания, выявить новые закономерности их функционирования, что очень важно для раскрытия структурных особенностей построения любого текста. Исходным для работы алгоритма является морфологическая информация. При этом предусматривается поэтапное увеличение ее использования: на первом этапе учитываются только грамматические классы окружения знака, на втором к информации о грамматическом классе присоединяются словоизменительные характеристики. Такое постепенное привлечение грамматической информации интересно с точки зрения исследования зависимости расстановки знаков препинания от определенных грамматических факторов.

Если первые две главы книги отличаются единым подходом поэтапного анализа текста, отказом от больших словарей, стремлением получить максимум информации из самого текста путем анализа графемной структуры слов и дистрибутивно-статистических характеристик элементов текста, то третья глава «Анализ фраз на естественном языке для диалого-

вой ИПС» находится в этом отношении в некоторой изоляции. (Изолированность ее, по-видимому, обусловлена тем, что в ней, в отличие от первой и второй глав, решаются вопросы логико-семантического анализа, требующие другого подхода.) Но она, как и две предыдущие главы, демонстрирует возможность формализации логико-семантического анализа текстов тематически ограниченного подязыка, обладающего большой степенью стандартизации.

Что же касается описания в монографии программной реализации алгоритмов, то все они, к сожалению, отличаются большой степенью разнообразия относительно как самой формы описания, так и использования различных языков программирования. Винаваты в этом, возможно, не авторы книги, а отсутствие ЭВМ в Институте языковедения АН УССР.

В целом монография, безусловно, представляет интерес для исследователей в области лингвистики текста и будет полезна для специалистов по автоматизированной обработке текстовой информации.

СПИСОК ЛИТЕРАТУРЫ

1. Закономерность структурной организации научно-реферативного текста. Киев, 1983.

Нелюбин Л. Л.

Никитина С. Е. Семантический анализ языка науки. На материале лингвистики. М.: Наука, 1987. 135 с.

Рецензируемая работа необычна во многих отношениях. С одной стороны, ее проблематика выходит далеко за рамки собственно языкознания и затрагивает целый ряд логико-философских проблем: соотношение теоретического и эмпирического, наблюдаемого и ненаблюдаемого в науке, отражение этих соотношений в семантике научного языка и т. п.; с другой стороны, материалом исследования служит лексика самой науки о языке, т. е. «святая святых» лингвистики. С одной стороны, в ней разбираются дискуссионные вопросы такой степени общности, которая, как правило, не предполагает немедленных конструктивных выходов и практических приложений; с другой стороны, автор, разрушая сложившиеся здесь стереотипы, представляет сугубо конкретные результаты предлагаемых решений (набор актуальных для лингвистической терминологии «сквозных» се-

мантических отношений, образцы формализованных тезаурусных описаний и т. п.). Наконец, научный характер повествования, с одной стороны, сочетается со свободной и с индивидуальной формой изложения, с другой.

Такой характер исследования в какой-то мере определяется не вполне привычным ракурсом изучения многих традиционных вопросов. Так, во введении перечисляются три основных исследуемых объекта: язык науки как общенаучное и лингвистическое понятие, тезаурус как лингвистический инструмент описания этого понятия и язык лингвистики как материал изучения. Выбор языка лингвистики в качестве материала анализа обусловлен бурным развитием этой науки, многообразием новых понятий, широким использованием лингвистических терминов в других науках, возможностью переноса методов семантического