РОССИЙСКАЯ АКАДЕМИЯ НАУК

ВСЕРОССИЙСКИЙ ИНСТИТУТ НАУЧНОЙ И ТЕХНИЧЕСКОЙ ИНФОРМАЦИИ
(ВИНИТИ)

НАУЧНО • ТЕХНИЧЕСКАЯ
ИНФОРМАЦИЯ

Серия 2. ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ И СИСТЕМЫ

ЕЖЕМЕСЯЧНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ СБОРНИК

Издается с 1961 г. № 6 Москва 2012

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

УДК 004.89

В. А. Лапшин

Вопросно-ответные системы: развитие и перспективы

Рассматриваются различные вопросы реализации вопросно-ответных систем. Приводится обзор наиболее популярных на данный момент реализаций таких систем. Типичная архитектура вопросно-ответной системы содержит модуль классификации вопросов. В работе рассматриваются различные методы реализации этого модуля.

Ключевые слова: *вопросно-ответные системы, DeepQA, Watson, классификация вопросов*

1. ВВЕДЕНИЕ

В последнее десятилетие активизировались исследования в области разработки так называемых *вопросно-ответных систем* (question-answering systems, или QA-системы). Вопросно-ответная система представляет собой программный модуль, позволяющий человеку вести с машиной диалог на естественном языке. Пользователь задает вопросы программной системе, а программная система печатает ответы, формируемые в виде осмысленных предложений.

Первые вопросно-ответные системы появились в 60-х годах прошлого века. Среди наиболее известных реализаций следует выделить системы BASEBALL и LUNAR. Система BASEBALL позволяла вести диалог с пользователем, интересующимся результатами соревнований бейсбольной лиги США за прошедший год. Система LUNAR отвечала на вопросы, связанные с геологическим анализом образцов пород, доставленных с лунной поверхности экспедициями программы «Аполлон». Обе системы были достаточно эффективно реализованы и представляли собой примеры вопросно-ответных систем, ориентированных

на конкретную предметную область. Например, система LUNAR, демонстрировавшаяся на конференции 1971 г., на которой обсуждались вопросы лунных исследований, позволяла получить ответы примерно на 90% всех вопросов, заданных данной системе.

Некоторые известные программные системы, разработанные в 60-х годах прошлого века, содержали в себе вопросно-ответные модули в виде подсистем. Так, программа ELIZA[1] содержала в качестве программного модуля вопросно-ответную систему, которая, собственно, и позволяла общаться с пользователем [1].

В 70-х и 80-х годах прошлого века было реализовано достаточно много вопросно-ответных систем, позволяющих вести диалог с пользователем в конкретной предметной области. Например, программный комплекс Unix Consultant отвечал на вопросы, связанные с операционной системой UNIX. Unix Consultant был основан на достаточно сложной и развитой *базе знаний*, содержащей информацию об операционной системе UNIX. Интерфейс к базе знаний был реализован в виде вопросно-ответной системы.

Все рассмотренные реализации вопросно-ответных систем позволяли отвечать на вопросы, связанные с конкретной предметной областью, т. е. являлись *узкоспециализированными* вопросно-ответными системами. В конце 90-х годов прошлого века, в связи с развитием Интернета и Веб, была осознана необходимость создания вопросно-ответных систем, не связанных с какой-либо предметной областью, – так называемых *открытых* вопросно-ответных систем. Такие системы позволяют вести диалог по всем областям знаний, например, на основе частично структурированных знаний, содержащихся в Веб.

В настоящее время имеется довольно большое количество реализаций вопросно-ответных систем. Заслуживает внимания реализация вопросно-ответной системы START [2]. Хороший обзор этой системы на русском языке был дан в [3].

Другой интересной системой, претендующей на статус открытой, является система Cyc [4]. История создания и развития этой системы насчитывает уже около 30 лет. За это время в системе разработано множество онтологий, как говорится, на все случаи жизни. Изначально представляющая собой объемную базу знаний, система Cyc предоставляет также интерфейс на английском языке и, таким образом, является вопросно-ответной системой. Более подробно система Cyc рассмотрена в [5, 6].

Среди наиболее известных реализаций открытых вопросно-ответных систем следует упомянуть системы OpenEphyra [7] и PIQUANT [8] (следующее поколение этой системы, известное под названием Watson, мы рассмотрим более детально в разделе 3).

Далее в нашей статье будут обсуждаться вопросы, связанные с реализацией открытых вопросно-ответных систем.

[1] ELIZA представляла собой симуляцию общения с врачом, основанную на методе так называемой *Рогерианской терапии*, при которой врач воздерживается от указаний пациенту, что тот должен делать, и лишь концентрирует свое внимание на создании обстановки взаимопонимания и доверительности.

2. ПРОГРАММА ИССЛЕДОВАНИЯ ОТКРЫТЫХ ВОПРОСНО-ОТВЕТНЫХ СИСТЕМ

В начале третьего тысячелетия группа исследователей в области текстового поиска разработала программу исследований, в которой были намечены основные этапы изучения особенностей построения открытых вопросно-ответных систем [9]. Перечислим ее основные пункты.

Типы вопросов

Для различных типов вопросов необходимы разные стратегии поиска ответов. Например, для вопроса *«Где родился Александр Пушкин?»* необходимо произвести поиск по базе знаний[2] на предмет нахождения даты рождения великого русского поэта. Для вопроса *«Почему Ваня приехал в Макдональдс сегодня на роликах?»* необходимо осуществить иную стратегию поиска. Вероятно, необходимо будет произвести логический вывод из фактов типа *«23 июля Маша и Ваня договорились встретиться возле Макдональдса, чтобы покататься на роликах»* или *«В это воскресенье Ваня ночевал у друга в районе метро Кунцевская, поэтому имел возможность покататься на его роликах»*.

Все возможные вопросы можно подразделить на классы в соответствии со стратегиями поиска ответов на них. Построение системы таких классов представляет собой довольно сложную задачу. Подробному обсуждению проблем, связанных с классификацией вопросов, будет посвящен одни из следующих разделов.

Обработка вопросов

Одну и ту же информацию можно выразить различными способами. Например, можно спросить *«Кто такой Александр Пушкин?»*, а можно спросить *«Как зовут человека, который написал поэму про Евгения Онегина?»*. Такие семантически схожие вопросы необходимо считать одними и теми же. Для этого требуется создать эффективные методы понимания и обработки семантики вопросов. Важно, чтобы программа распознавала эквивалентные по смыслу вопросы независимо от используемых слов, стиля, синтаксических взаимосвязей и идиом. Хотелось бы, чтобы вопросно-ответная система производила разделение сложных вопросов на простые и производила корректный анализ фраз, которые зависят от контекста, возможно, уточняя этот контекст у пользователя в процессе диалога.

Контекстные вопросы

Смысл вопроса обычно определяется не только его содержанием, но и контекстом, в котором этот вопрос задается. Контекст представляет собой информацию, которая запоминается в процессе диалога с пользователем системы и может служить для устранения различных неоднозначностей. Например, часто

[2] «База знаний» в данном контексте может означать все что угодно. Например, неструктурированный текст на русском языке, индексированный в инвертированный список слов в поисковой системе.

используемое для иллюстрации синтаксической омонимии выражение *«Эти типы стали есть в прокатном цехе»* можно уточнить на основе контекста диалога (скажем, если в процессе диалога обсуждались различные марки стали).

Источники знаний для вопросно-ответной системы

Для ответа на вопрос необходимо иметь доступ к какой-нибудь базе знаний, в которой содержится информация об ответе на этот вопрос. Иначе трудно, если вообще возможно, найти правильный ответ. Открытая вопросно-ответная система обычно работает с несколькими источниками знаний, в которых производит поиск ответов в зависимости от класса заданного вопроса.

Поиск ответов

Правильное выполнение процедуры поиска ответа на запрос зависит от сложности вопроса, его типа, контекста, качества доступных источников информации, метода поиска и т. п. Поэтому особое внимание следует обратить на изучение методов выделения правильных ответов на основе имеющихся баз знаний. Определение необходимых источников для выделения ответа зависит также и от правильной классификации вопроса.

Формулировка ответа

Ответ должен выглядеть естественно, в идеале – так, чтобы пользователь не догадывался, что ответ сгенерирован машиной. В некоторых случаях для этого достаточно простого поиска по базе данных. Например, если требуется найти некоторое имя (человека, название прибора или болезни), численное значение (денежный курс, длина, размер) или дату (*«Когда родился Александр Пушкин?»*). Но иногда приходится иметь дело со сложными запросами и здесь нужны особые алгоритмы поиска ответов из разных источников и последующего слияния этих ответов в один. Формулировку окончательного ответа следует проводить так, чтобы результат выглядел синтаксически естественно и представлял собой именно то, что искал пользователь.

Ответы на вопросы в реальном времени

Вопросно-ответная система должна предоставлять ответы на вопросы в реальном времени, т. е. в течение секунд. Это условие должно быть выполнено вне зависимости от длины и сложности вопроса, а также источника информации, по которому ведется поиск.

Многоязыковые запросы

Разработка систем для работы и поиска на различных языках (в том числе автоматический перевод).

Интерактивность

Информация, выдаваемая вопросно-ответной системой в качестве ответа, довольно часто не является исчерпывающей. Например, система может неверно определить класс вопроса или вообще неправильно разобрать вопрос. Поэтому необходимо реализовать возможность введения подсказок (коррекции) пользователей в систему. Процесс введения подсказок естественно реализовать в виде диалога (вопросно-ответной последовательности).

Механизм рассуждений (логического вывода)

Пользователи часто задают вопросы, ответы на которые не содержатся в доступных базах знаний. Для реализации поиска ответов на такие вопросы вопросно-ответная система должна иметь систему логического вывода на основе фактов, полученных из имеющихся источников информации.

Профили пользователей QA-систем

Сведения о пользователе (такие, как область интересов, стилистика формулировки вопросов и другая специфичная для пользователя информация) могли бы существенно улучшить качество работы системы.

Для того чтобы стимулировать исследования в перечисленных областях, была разработана программа организации конкурсов на конференции «Text Retrieval Conference» (TREC) [10] и сформированы соответствующие конкурсные задачи на конференциях TREC 2000 – 2005 гг. (TREC-9 – 14). Детали можно уточнить в [9]. Вообще, конкурсы, связанные с оценкой работы вопросно-ответной системы, проводились на конференции TREC с 1999 по 2007 г. Более подробную информацию можно найти на странице конференции TREC, посвященной дорожкам оценки качества вопросно-ответных систем [11].

3. ПРОЕКТ DeepQA

В середине прошлого десятилетия основным спонсором исследований в области разработки вопросно-ответных систем стала известная американская компания IBM [12]. В 2005 г. она основала проект под названием DeepQA. Это проект развитой открытой вопросно-ответной системы с расширяемой архитектурой, позволяющей адаптировать систему для работы в различных предметных областях.

Основным вызовом проекта DeepQA стало создание программной системы, позволяющей играть в популярную в США телевикторину под названием «Jeopardy!» (на российском телевидении она выходит под названием «Своя игра»). В игре принимают участие три игрока, которым предлагается доска, разделенная на 30 ячеек, – 6 строк по 5 ячеек. В ячейках находятся тексты вопросов, которые выбираются по запросу игроков. Изначально тексты вопросов скрыты, при выборе ячейки вопрос открывается и игроки имеют возможность в течение 5 секунд дать правильный ответ. Тот игрок, который такой ответ дал, получает на свой счет сумму, которой помечен данный вопрос. Каждая строка на доске имеет название, т. е. все вопросы данной строки принадлежат одной категории. Категории создаются для каждого выпуска телевикторины заново, т.е. множество категорий не фиксировано. Сама игра состоит из трех раундов и финала, в котором ведущий задает только один вопрос, на который каждый игрок обязан дать ответ, причем на ответ дается 30 секунд.

Программная система, позволяющая играть в «Jeopardy!», получила название Watson, в честь основателя компании IBM Томаса Уотсона. В начале 2011 г. система Watson провела несколько сеансов игры «Jeopardy!» с лучшими игроками среди людей за всю историю этой игры и выиграла во всех играх. Разработчики программы опубликовали опыт создания системы Watson (и проекта DeepQA в целом) в статье [13]. Далее мы обсудим архитектуру этой системы как наиболее удачного проекта за всю историю создания открытых вопросно-ответных систем.

* * *

Архитектура системы Watson изображена на рисунке, оригинал которого взят нами из статьи [13]. Основным принципом организации системы является ее открытость для добавления новых компонентов на всех стадиях работы. По сути, на каждом уровне обработки вопросов система реализует арбитров, которые выбирают наилучший вариант из результатов обработки вопроса различными алгоритмами.

Рассмотрим компоненты системы Watson подробнее.

1. Анализ вопроса

Анализ вопроса (Question Analysis) имеет несколько составляющих:

- **Классификация вопроса.** Классификация вопроса состоит в определении его категории. Всегда задано конечное множество категорий. На множестве категорий может быть установлено отношение иерархии (подробно об этом поговорим в разделе 4). В системе Watson для классификации задействовано несколько алгоритмов, большинство из которых имеют свои собственные наборы категорий.

- **Определение фокуса вопроса и лексического типа ответа:**

– *Фокусом вопроса* называется часть этого вопроса, которая, будучи замененной ответом, делает исходный вопрос осмысленным предложением. Например, для вопроса *«При бомбардировке электронами фосфор излучает электромагнитную энергию в этой форме»* фокусом будет являться словосочетание *«этой форме»*. При подстановке вместо него ответа *«свет»* получится осмысленное предложение *«При бомбардировке электронами фосфор излучает электромагнитную энергию в форме света»*. Определение фокуса вопроса важно как для его последующей обработки, так и для генерации ответа.

– *Лексическим типом ответа* в системе Watson называют слово или фразу, входящую в вопрос, которая характеризует тип ответа на этот вопрос, причем слово или фраза рассматривается без какой-либо семантической интерпретации. Например, в категории «Шахматы» для вопроса *«Для ускорения игры в 16 веке изобрели этот маневр, позволяющий делать несколько ходов без перемены цвета фигур»* ответом является *«рокировка»*, а лексическим типом ответа слово *«маневр»*. Лексические типы ответов используются в системе Watson в качестве предопределенных категорий. Если, например, лексический тип ответа данного вопроса *«страна»*, то можно сразу обратиться к географической базе знаний для поиска по словарю стран.

- **Выделение отношений.** Система Watson также пытается выделить из текста вопроса отношения между лексическими элементами. Например, вопрос *«Через эти области и национальные республики Российской Федерации протекает река Волга»* задает отношение вида (Волга, протекает, ?x). Для получения ответа на такой вопрос можно просто построить запрос к соответствующей онтологии. В процессе построения системы Watson разработчиками была проведена оценка сравнительного количества вопросов, из которых можно выделить отношения, среди всех вопросов телевикторины. Оказалось, что таких вопросов всего около двух процентов.

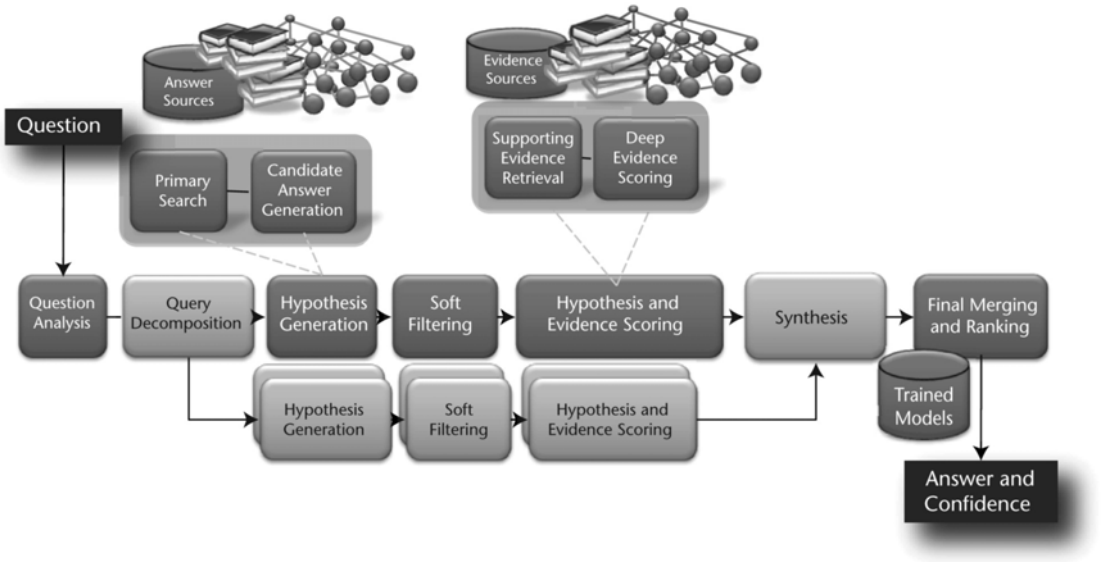


Рис. Архитектура открытой вопросно-ответной системы Watson

2. Декомпозиция вопросов

В системе Watson производится декомпозиция вопроса на более простые части. Учитывая, что вопросы, задаваемые в телевикторине «Jeopardy!», почти всегда имеют довольно сложную структуру, декомпозиция играет важную роль в процессе обработки вопроса системой. Для выделения частей вопроса в системе Watson используются алгоритмы, основанные на правилах, а также алгоритмы статистического машинного обучения. Каждая из выделенных частей вопроса обрабатывается отдельно. Алгоритмы обработки запускаются параллельно.

3. Генерация гипотез ответов

Для генерации возможных ответов производится поиск в источниках знаний. Это могут быть неструктурированные знания, например, обычные вебстраницы, слабоструктурированные знания, такие, как статьи из Википедии, а также структурированные знания, например, RDF-хранилища [14]. Тексты в системе Watson проиндексированы и находятся в обычном инвертированном индексе по словам. Но используется и дополнительная разметка синтаксических и семантических структур, выделенных в процессе анализа текста. Имеется также база данных, связанная с лексическими типами ответов, которые определяются на этапе анализа вопроса.

Процесс порождения гипотез подразделяется на две фазы: первичный поиск и генерация гипотез ответов.

В процессе первичного поиска производится обращение к различным источникам знаний. На этапе генерации гипотез из результатов первичного поиска производится преобразование этих результатов в формат ответа. Алгоритм этого преобразования специфичен для источника знаний. Например, для результатов, полученных поиском по «ориентированному на названия документов» индексу, в качестве результата выдается имя найденного документа. При нахождении значимых результатов в хранилище RDF-троек производится преобразование к выражению на естественном языке и т. д.

Если на данной стадии обработки вопроса не было получено подходящих гипотез, то дальнейшая работа становится бессмысленной и система сообщает о том, что результатов не найдено.

4. Мягкая фильтрация

После того как множество гипотез ответов для данного вопроса построено, производится так называемая «мягкая фильтрация» (Soft Filtering) этих гипотез. Мягкая фильтрация состоит в применении к гипотетическим ответам различных специализированных алгоритмов, которые не требуют значимых ресурсов для своего исполнения. Все эти алгоритмы могут исполняться параллельно, что позволяет значительно снизить время обработки гипотез. Задачей данной стадии обработки ответов является отсеивание большинства кандидатов в ответы с тем, чтобы к оставшимся гипотезам были применены более ресурсоемкие алгоритмы обработки.

В качестве примера алгоритма, запускаемого на этапе мягкой фильтрации, можно привести алгоритм, определяющий вероятность принадлежности гипотетического ответа лексическому типу ответа, определенному для данного вопроса на предыдущей стадии обработки.

5. Оценка обоснованности

Гипотетические ответы, прошедшие сито мягкой фильтрации, передаются на обработку алгоритмам оценки обоснованности ответов (Evidence Scoring). На этой достаточно ресурсоемкой стадии производится обращение к различным источникам знаний с целью доказательства того факта, что данная гипотеза действительно является ответом на переданный на обработку вопрос.

Эта стадия подразделяется на две фазы: фазу получения обоснований и фазу оценки этих обоснований.

На этапе получения обоснований производится обращение к различным источникам знаний. Например, к исходному вопросу добавляется текст гипотетического ответа и производится поиск по полученному таким образом запросу в источнике знаний. Иначе говоря, производится повторный поиск ответа в контексте заданного вопроса. Возможно также обращение к другим источникам, таким, как RDF-хранилища. На этом этапе гипотезы ответов получают обоснования из различных источников.

На этапе оценки обоснований производится комплексная оценка доказательств, полученных для данного гипотетического ответа на этапе получения обоснований. Алгоритмы оценки определяют согласованность этих обоснований. Архитектура проекта DeepQA позволяет легко добавлять различные алгоритмы оценки для разных множеств типов обоснований. Реализация DeepQA предоставляет специализированный интерфейс для алгоритмов оценки обоснований и унифицированный алгоритм работы с этими оценками. В оригинальной статье [13] приведен пример обработки вопроса *«He was presidentially pardoned on September 8, 1974»*, который иллюстрирует работу системы на различных алгоритмах оценки.

6. Окончательное слияние и ранжирование

Обычный поиск по документам возвращает в качестве результатов набор документов, которые, предположительно, наиболее релевантны заданному запросу. В игре «Jeopardy!» необходимо возвращать ответ в виде осмысленного предложения. Задачей данной стадии обработки вопроса является слияние похожих гипотез в осмысленные ответы и последующее ранжирование результатов слияния в порядке убывания их оценки, полученной на предыдущей стадии обработки.

На этапе слияния гипотез ответов производится слияние похожих гипотез в единое целое. Один и тот же ответ может быть выражен в тексте различными способами. Результирующие гипотезы нормализуются, чтобы результаты работы системы выглядели осмысленно для человека.

На этапе ранжирования производится окончательная оценка достоверности ответов на поставленный вопрос. В системе Watson для реализации ранжирования используются алгоритмы машинного обучения: оценки из различных источников обрабатываются на основе модели, полученной в результате обучения на некоторой обучающей выборке вопросов и ответов.

* * *

В системе Watson используется реализация распределенной обработки документов UIMA (Unstructured Information Management Architecture) [15]. Это система с открытым исходным кодом, позволяющая сохранять результаты обработки документов в поисковом индексе с их дополнительной разметкой синтаксическими и семантическими отношениями. Компоненты системы могу быть расположены на разных компьютерах. Таким образом, в случае необходимости система Watson может быть легко расширена дополнительными ресурсами.

Для реализации распределенного индекса предварительно обработанного корпуса источников данных в Watson используется система Hadoop [16]. Алгоритмы добавления аннотаций, которые вставляет в индексируемые данные UIMA, легко адаптируются к модели параллельных вычислений map-reduce, предоставляемой системой Hadoop.

Первые реализации системы Watson выполнялись на обычном настольном компьютере с двухъядерным процессором. Обработка одного вопроса занимала около двух часов. Современная реализация системы исполняется на кластере с 2,5 тыс. ядер. Это позволяет многим алгоритмам обработки выполняться одновременно. Время обработки запроса занимает от 3 до 5 секунд.

Как уже было сказано, архитектура DeepQA задумана таким образом, чтобы легко адаптироваться для работы с различными предметными областями. Так, в 2009 г. было произведено тестирование системы Watson на дорожке вопросов конференции TREC [11]. Были проведены эксперименты с адаптированной к дорожке вопросов системой Watson, а также с системой в том состоянии, в котором она была подготовлена к игре «Jeopardy!». В неподготовленном состоянии система показала результат примерно в 35% точности, адаптированная система показала результат в 60%, что является лучшим результатом за всю историю конференций TREC.

В начале 2012 г. система Watson стала применяться для решения различных практических задач. Пресса указывает по крайней мере на одно такое внедрение – создание вопросно-ответной системы для финансового холдинга Citigroup. Система Watson будет общаться с работниками холдинга в режиме диалога, запрашивая у пользователей необходимую дополнительную информацию и выдавая ответы. Корпорация IBM предоставляет систему Watson и в виде интернет-сервиса, т. е. сервиса облачных вычислений, общение с которым производится посредством сети Интернет. Для нового сервиса даже придумано название: WAAS (Watson as service).

4. КЛАССИФИКАЦИЯ ВОПРОСОВ

Как уже говорилось в предыдущем разделе, одним из важнейших модулей вопросно-ответной системы является модуль классификации вопроса. Задача этого модуля состоит в том, чтобы соотнести вопрос с одним из предопределенных классов (категорий). Такое соотнесение позволяет определить источник информации, в котором необходимо искать ответ на данный вопрос, а также выбрать алгоритм, с помощью которого этот поиск будет производиться.

Множество категорий, используемых для классификации вопросов, можно назвать онтологией [17]. Онтология вопросов обычно представляет собой иерархическую структуру – таксономию, понятия верхнего уровня которой имеют довольно специфический вид, основанный на функциональном предназначении вопроса.

Задача выделения классов запросов впервые была осознана в работе [18], вышедшей во второй половине 70-х годов прошлого века. Ее автор, Венди Ленерт, представила публике достаточно подробное исследование, в котором была разработана модель представления семантики текстовых выражений на основе теории так называемых «понятийных зависимостей» (Conceptual Dependency).

В. Ленерт предложила 13 категорий, в которые попадают все возможные вопросы, которые может задать пользователь. Эти категории, вместе с примерами вопросов, которые им принадлежат, приведены в таблице.

Категории вопросов В. Ленерт

Категория вопроса	Примеры вопросов
Причина (Causal Antecedent)	Почему Василий приехал в Москву?
Цель (Goal Orientation)	Для чего Иван взял эту книгу?
Возможность (Enablement)	Что должен сделать Олег, чтобы уехать?
Контроль (Verification)	Иван уехал?
Дизъюнкция (Disjunctive)	Были ли здесь Ося или Киса?
Процедура (Instrumental/Procedural)	Каким образом Василий добрался до Москвы?
Дополнение (Concept Completion)	Что ел Иван?
Ожидание (Expectation)	Почему Иван не приехал в Москву?
Суждение (Judgmental)	Что должен сделать Василий, чтобы Маша не уехала?
Квантификация (Quantification)	Сколько людей собралось на этом стадионе?
Свойство (Feature Specification)	Какого цвета глаза у Ивана?
Запрос (Request)	Не передадите мне соль?

Система классов, введенная В. Ленерт, была расширена Артуром Гессером в работе [19]. К существующим 13 категориям были добавлены еще пять: Определение (Definition), Пример (Example), Интерпретация (Interpretation), Отсутствие (Assertion) и Сравнение (Comparison). Позднее, в работе [20], было определено 16 категорий.

Приведенная классификация является довольно общей и пригодна только для первичной обработки вопроса, например, для выбора алгоритма обработки запроса. Для выбора источников данных, по которым будет производиться поиск ответов, необходимо определить более конкретные категории. Таким образом, необходима более развитая понятийная система. Спецификации таких систем называют онтологиями [17]. Иначе говоря, необходима онтология для классификации вопросов, подаваемых на вход вопросно-ответной системы. В системах классификации обычно используют онтологии специального вида, в которых имеется только одно отношение между классами – отношение иерархии. Онтологии с классами, упорядоченными в иерархическую структуру, называют *таксономиями*. В дальнейшем будем предполагать, что классы, используемые для классификации вопросов в вопросно-ответных системах, организованы в таксономии.

В работе [21] была определена таксономия вопросов, использующихся в дорожке TREC-10. В этой таксономии (см. *Приложение*) определено 6 категорий верхнего уровня (так называемых «грязных» категорий) и 50 «хороших» классов, по которым проводилась классификация.

В качестве алгоритмов классификации вопросов обычно применяются классификаторы на основе правил и статистические классификаторы. Классификаторы на основе правил используют достаточно простые правила для классификации: например, если вопрос начинается со слов «Какова скорость вращения», то этот вопрос наверняка относится к категории «скорость». В уже упоминавшейся нами работе [21] был использован статистический классификатор. Вообще, имеется большое количество работ на тему классификации вопросов, хороший обзор можно найти в [22].

5. ЗАКЛЮЧЕНИЕ

Данный обзор имеет своей целью представить текущее состояние исследований в области вопросно-ответных систем. Работа в этом направлении шла достаточно интенсивно все прошедшее десятилетие и продолжает идти в настоящее время. В начале XXI в. эта работа была мотивирована в основном программой, сформулированной в [9]. В соответствии с данной программой на конференциях TREC с 2000 по 2005 г. предлагались конкурсные задачи.

После окончания программы [9] исследования в области вопросно-ответных систем продолжились под патронажем компании IBM. Значительным успехом в реализации открытых вопросно-ответных систем выглядит создание системы Watson [13] и выигрыш этой системой телевикторины «Jeopardy!» в начале 2011 г. Текущее состояние развития вопросно-ответных систем позволяет использовать их в качестве полноценных экспертных систем, способных общаться с пользователями на естественном языке.

В настоящее время интенсивно идет развитие и узкоспециализированных вопросно-ответных систем. Например, не так давно запущенный проект WolframAlpha [23] представляет собой обширную базу знаний, предоставляющую интерфейс на естественном языке. С позиций, изложенных в данной работе, система WolframAlpha является специализи-рованной вопросно-ответной системой.

На наш взгляд, очевидно, что дальнейшее развитие открытых вопросно-ответных систем будет основываться на создании средств развитого анализа естественных языков с целью выделения фактов из текстов на этих языках[3]. Структура таких фактов должна быть описана в виде онтологии, соотнесение вопроса с конкретной онтологией должно выполняться модулем классификации вопросов. С этой точки зрения представляется, что развитая открытая вопросно-ответная система будет являться симбиозом узкоспециализированной вопросно-ответной системы и системы выделения знаний из слабоструктурированных некатегоризированных текстов.

СПИСОК ЛИТЕРАТУРЫ

1. ELIZA [Электрон. ресурс]. – URL: http://en.wikipedia.org/wiki/ELIZA (дата обращения: 07.03.2012).
2. START. Natural Language Question Answering System [сайт]. – URL: http://start.csail.mit.edu (дата обращения: 11.03.2012).
3. Никитин А., Райков П. Вопросно-ответные системы [Электрон. ресурс]. – URL: http://yury.name/internet/06ia-seminar.ppt (дата обращения: 11.03.2012).
4. Cycorp [сайт]. – URL: http://www.cyc.com (дата обращения: 11.03.2012).
5. Лапшин В. Система Cyc и ее библиотека онтологий. Ч. 1 // Искусственный интеллект и принятие решений. – 2010. – № 2.
6. Лапшин В. Система Cyc и ее библиотека онтологий. Ч. 2 // Искусственный интеллект и принятие решений. – 2010. – № 3.
7. OpenEphyra [Электрон. ресурс]. – URL: http://sourceforge.net/projects/openephyra (дата обращения: 14.03.2012).
8. Prager J. M., Chu-Carroll J., Czuba K. A Multi-Strategy, Multi-Question Approach to Question

[3] С этой точки зрения интересна серия статей, представленных в [24, глава 6: Перспективы развития вопросно-ответных систем]. Наибольшее внимание практически во всех статьях этой серии уделено развитию анализаторов естественных языков.

Answering / ed. by M. Maybury. – Menlo Park, CA : AAAI Press, 2004.

9. Burger J., Cardie C., Chaudhri V. et al. Issues, Tasks and Program Structures to Roadmap Research in Question & Answering (Q&A) [Электрон. ресурс]. – 2000. – URL: http://www-nlpir.nist.gov/projects/duc/roadmapping.html
10. Text REtrieval Conference (TREC) [сайт]. – URL: http://trec.nist.gov (дата обращения: 07.03.2012).
11. Question Answering Track [Электрон. ресурс]. – URL: http://trec.nist.gov/data/qamain.html (дата обращения: 07.03.2012).
12. IBM [сайт]. – URL: http://www.ibm.com (дата обращения: 13.03.2012).
13. Ferrucci D. A., Brown E. W., Chu-Carroll J. et al. Building Watson: An Overview of the DeepQA Project // AI Magazine. – 2010. – Vol. 31, № 3. – P. 59 – 79.
14. RDF [сайт]. – URL: http://www.w3.org/RDF (дата обращения: 14.03.2012).
15. Ferrucci D. A., Lally A. UIMA: An Architectural Approach to Unstructured Information Processing in the Corporate Research Environment // Natural Language Engineering. – 2004. – Vol. 3–4, № 10. – P. 327 – 348.
16. Hadoop [сайт]. – URL: http://hadoop.apache.org (дата обращения: 14.03.2012).
17. Лапшин В. А. Онтологии в компьютерных системах. – М. : Научный мир, 2010.
18. Lehnert W. The Process of Question Answering : PhD Dissertation // Research report. No. 88. – Yale University, 1977.
19. Graesser A., Person N. K. Question asking during tutoring // American Educational Research Journal. – 1994. – № 31. – P. 104 – 137.
20. Graesser A., Rus V., Cai Z. Question classification schemes // Proceedings of the 1st Workshop on Question Generation [Электрон. ресурс]. – 2008. – P. 8 – 9. – URL: http://141.225.40.110/16-GraesserEtAl-QG08.pdf
21. Li X., Roth D. Learning question classifiers // Proceedings of the 19th international conference on Computational linguistics. – Stroudsburg, PA, USA : Association for Computational Linguistics, 2002. – Vol. 1: COLING '02. – P. 1 – 7.
22. Sundblad H. Question classification in question answering systems : Ph. D. thesis. – Linkopings universitet, Department of Computer and Information Science, 2007.
23. WolframAlpha [сайт]. – URL: http://www.wolframalpha.com (дата обращения: 11.03.2012).
24. Strzalkowski T., Harabagiu S. Advances in Open Domain Question Answering (Text, Speech and Language Technology). – Secaucus, NJ, USA : Springer-Verlag New York, 2006.

ПРИЛОЖЕНИЕ

Таксономия вопросов TREC-10

Категория вопроса	Определение
ABBREVIATION	аббревиатуры
abb	аббревиатуры
exp	сложные аббревиатуры (выражения)
ENTITY	сущности
animal	животные
body	organs of body
color	цвета
creative	открытия, книги и т. д.
currency	названия валют
dis.med.	болезни и медицина
event	события
food	пища
instrument	музыкальные инструменты
lang	языки
letter	буквы (такие, как а–я)
other	остальные сущности
plant	растения
product	продукты
religion	религии
sport	спорт
substance	элементы и содержимое
symbol	символы и знаки
technique	подходы и методы
term	эквивалентные термины
vehicle	автомобили
word	слова со специальными свойствами
DESCRIPTION	описания и абстрактные понятия
definition	определения
description	описания
manner	способы действий
reason	причины
HUMAN	люди
group	группы и организации
ind	индивиды
title	имена персон
description	описания персон

Категория вопроса	Определение
LOCATION	местонахождения
city	города
country	страны
mountain	горы
other	остальные местонахождения
state	страны
NUMERIC	численные величины
code	почтовые и другие коды
count	числа
date	даты
distance	меры измерения
money	цены
order	ранги
other	остальные числа
period	период времени
percent	проценты
speed	скорость
temp	температура
size	размеры, площади и объемы
weight	вес

Материал поступил в редакцию 16.03.12.

Сведения об авторе

ЛАПШИН Владимир Анатольевич – кандидат физико-математических наук, руководитель отдела обработки запросов на естественном языке компании ABBYY, Москва

E-mail: mefrill@yandex.ru

УДК 81’322

В. А. Яцко

Проблемы и алгоритмы клаузальной декомпозиции текста

Рассматриваются особенности различных видов декомпозиции, применяемых в процессе предварительной обработки текста. Показываются лингвистические проблемы клаузальной декомпозиции, обусловленные трансформацией семантического, коммуникативного и модального планов текста. Описываются грамматика и алгоритмы, необходимые для проведения клаузальной декомпозиции.

Ключевые слова: *декомпозиция, алгоритмы декомпозиции, предикативная структура, клаузы, токены*

1. ВВЕДЕНИЕ

Одним из фундаментальных алгоритмов, применяемых практически во всех программах автоматического анализа текста, является декомпозиция – разбивка текста на составляющие его единицы. В зависимости от целей анализа в качестве таких единиц могут выступать: отдельные символы; морфемы и другие части слов; слова; словосочетания; предложения; группы предложений; абзацы; сверхфразовые единства. В соответствии с уровнями языковой системы можно выделить следующие виды декомпозиции: графемическая, морфологическая, лексическая, синтаксическая, дискурсивная. На выходе у программы, которая выполняет декомпозицию, будут списки, содержащие соответствующие единицы текста: символы; стеммы или леммы; токены; предикативные единицы (как правило, словосочетания и предложения); сегменты или отрезки текста. Алгоритмы, по которым выполняется декомпозиция, обычно получают название по имени соответствующей единицы текста (см. табл. 1).

Заметим, что в настоящее время нет общепринятой стандартной терминологии для обозначения различных методов и алгоритмов разбивки текста. Дж. Солтон, один из ведущих учёных в области автоматического анализа текста, использовал термин «декомпозиция» для обозначения сегментации текста на уровне абзацев [1]. Мы предлагаем использовать этот термин как обобщающий, родовой для обозначения различных видов разбивки текста на разных уровнях языковой системы.

Таблица 1

Алгоритмы и программы декомпозиции текста

Единица текста	Программа	Алгоритм	Уровни языковой системы
Символ	Оптическое распознавание символов (ОРС)	Рекогнайзер	Графемический
Стемма	Стеммер	Стемминг	Морфологический
Лемма	Лемматайзер	Лемматизация	
Токен	Токенайзер	Токенизация	Лексический
Стоп-слово	Фильтр	Фильтрация стоп-слов	
Словосочетание (chunk)	Чанкер	Чанкинг	Синтаксический
Клауза	Клауз-сплитер	Клауз-сплитинг	
Предложение	Синтаксический сплитер	Сплитинг	
Сегмент	Сегментатор	Сегментация	Дискурсивный

Можно выделить следующие особенности алгоритмов декомпозиции текста:

1. Алгоритмы автоматической декомпозиции опираются на лингвистические представления о структуре текста, однако единицы декомпозиции существенно отличаются от соответствующих лингвистических единиц. Это различие можно пояснить на примере такого алгоритма, как токенизация. Токены часто совпадают по форме со словами, однако интерпретация термина «токен» в области автоматической обработки текста имеет мало общего с его интерпретацией в теоретической лингвистике. Под токеном может пониматься: последовательность символов, отграниченная пробелами слева и справа; последовательность символов, слева от которой – пробел, а справа – знак пунктуации; последовательность символов, указанная в файле данных, в котором, например, могут содержаться списки сокращений или стоп-слов [2]. Если в лингвистике слово интерпретируется как совокупность знака, значения и обозначаемого объекта [3], то правила для распознавания токенов основываются на графическом представлении символов, а их значение игнорируется.

2. Алгоритмы декомпозиции выполняются на этапе предварительной обработки текста, под которой понимается трансформация входного текста, необходимая для выполнения основного алгоритма системы. Специфика основного алгоритма состоит в том, что данные, полученные в результате его выполнения, выдаются пользователю. Данные, получаемые после выполнения декомпозиции, не выводятся пользователю, а служат исходными для выполнения последующих алгоритмов. В качестве примера можно привести аннотирование тегами частей речи. Поскольку теги частей речи приписываются токенам, аннотирование может выполняться только после токенизации. Само аннотирование также может быть этапом предварительной обработки (например, в системах автоматической классификации текстов) либо выступать в качестве основного алгоритма при создании текстовых корпусов различных типов. Исключением являются алгоритмы оптического распознавания символов (ОРС), которые обычно применяются в качестве основного алгоритма, поскольку результаты распознавания непосредственно выводятся пользователю.

3. Алгоритмы декомпозиции могут применяться в строгой или произвольной последовательности, что необходимо учитывать при проектировании систем автоматического анализа текста в том случае, если применяется несколько таких алгоритмов. ОРС, стемминг, лемматизация, фильтрация стоп-слов, чанкинг выполняются на основе токенизации, поскольку для выделения стемм, лемм, слов, словосочетаний сначала следует распознать токены. Соответственно, токенизация является одним из наиболее фундаментальных алгоритмов декомпозиции. Возможны различные варианты последовательности токенизации, сплитинга и сегментации, если эти алгоритмы выполняются независимо друг от друга. В предложенной нами дедукционно-инверсионной архитектуре декомпозиции [4] установлена строгая последовательность их выполнения: вначале текст разбивается на абзацы, далее в абзацах находятся токены, затем генерируются предложения. Осуществляется последовательный переход от единицы более высокого уровня (абзаца) к единице более низкого уровня (токену), затем опять к единице более высокого уровня (предложению).

4. Алгоритмы декомпозиции характеризуются разным уровнем сложности, который зависит от структуры данных, используемых для лингвистической поддержки. Могут использоваться файлы данных, базы данных, базы знаний. Базы знаний, как мы полагаем, можно разделить на два вида: текстографические и документографические. Текстографические базы знаний включают грамматические правила, необходимые для распознавания структур единиц текста. Такие правила нужны, например, для определения структуры словосочетаний в процессе чанкинга. На входе у чанкера – текст, на выходе – список словосочетаний. Наиболее распространены чанкеры, распознающие именные словосочетания с управляющим существительным (*NP*), поскольку данный тип словосочетаний адекватно отражает смысл текста. Для чанкинга применяются линейные грамматики, содержащие правила типа $NP \rightarrow Det\ N$, где задаётся состав и последовательность компонентов фразы, т.е. в данном примере *N Det* – недопустимая последовательность. Документографические базы знаний содержат правила распознавания параметров документа, которые не соотносятся с лингвистической структурой текста. К ним могут относиться параметры форматирования документа, такие, как отступы, пустые строки, размеры абзацев, строк. Документографическая база знаний была создана нами для поддержки дедукционно-инверсионной декомпозиции текста. Разумеется, могут использоваться и базы знаний смешанного типа.

Файлы данных являются необходимыми для поддержки стемминга, лемматизации, токенизации, фильтрации стоп-слов. Они могут содержать списки суффиксов и окончаний, списки сокращений и стоп-слов. Однако и эти алгоритмы могут усложняться введением дополнительных компонентов либо сочетаться друг с другом, что требует создания базы данных. Фильтрация стоп-слов сама по себе требует файла со списком стоп-слов, но она выполняется на основе токенизации, для реализации которой необходимы файлы со списками сокращений и устойчивых сочетаний, т. е. для проведения фильтрации стоп-слов необходима база данных, включающая три файла.

Заметим, что усложнение алгоритмов, введение дополнительных компонентов отрицательно влияют на быстродействие систем автоматической обработки текста. Ошибки, допущенные при выполнении даже одного из алгоритмов декомпозиции текста, приводят к ошибкам в выполнении всех последующих алгоритмов, что отрицательно сказывается на эффективности функционирования сис-

темы в целом. Отсюда – актуальность изучения алгоритмов декомпозиции текста.

Из алгоритмов, указанных в табл. 1, наименее изученным является клаузальный сплитинг текста, проблемам которого и посвящена данная статья.

2. ЛИНГВИСТИЧЕСКИЕ АСПЕКТЫ КЛАУЗАЛЬНОЙ ДЕКОМПОЗИЦИИ

Под *клаузой* мы понимаем выделяемую по формальным признакам предикативную структуру, выражающую отдельное суждение. Понятие клаузы является операциональным понятием, используемым в области автоматической обработки текста, которому в теоретической лингвистике соответствует понятие пропозиции. «Пропозиция» является абстрактным понятием, которое также соотносится с понятием суждения. По мнению Ю. С. Степанова, «термин “пропозиция” может одновременно употребляться вместо термина “суждение”, когда не требуется тонких различий» [5, с. 30]; ср. также определение Е. В. Падучевой: «Суждение – 1) смысл предложения, которое может быть оценено как истинное или ложное; то же, что и пропозиция…» [6, с. 499].

Пропозиция обычно рассматривается как семантический инвариант, общий для всех членов модальной и коммуникативной парадигм предложения и производных от предложения конструкций. Соответственно, следует разграничивать семантический (пропозициональный), коммуникативный и модальный планы текста [7, с. 43 – 89]. Вместе с тем понятие пропозиции следует отграничивать от понятия предложения-высказывания как единицы речи, которая часто является полипредикативной структурой, выражающей несколько суждений. Выявление пропозиционального содержания считается одной из главных задач грамматики текста.

«Приведение всех синтаксических структур, встречающихся в тексте, к элементарным пропозициям, имеющим форму простой (неосложненной) субъектно-предикатной структуры, является средством установления логических контекстуальных отношений между высказываниями и представляет собой одну из основных задач грамматики текста», – считает О. И. Москальская [8]. С этим утверждением нельзя не согласиться, поскольку прежде чем устанавливать логические, семантические, логико-семантические отношения между суждениями следует вначале распознать предикативные синтаксические структуры, с помощью которых эти суждения выражаются. Такими структурами и являются клаузы.

Разбивка на клаузы имеет непосредственное значение для моделирования логико-семантической структуры текста в системах интеллектуального анализа, так как позволяет выявить иерархическую структуру на основе зависимостей между клаузами, выражающими отдельные суждения. В [9] описывается такая структура, созданная на основе отношений между ядерными и сателлитными отрезками (spans) текста, под которыми понимаются клаузы, а также некоторые словосочетания, использующиеся в заголовках. Отрезкам текста приписываются весовые коэффициенты в зависимости от глубины дерева зависимостей, т. е. в зависимости от количества подчиняемых узлов. Отрезки, получившие наибольшие весовые коэффициенты, выбираются в реферат. В [10] предлагается применить предварительную разбивку на клаузы с целью упрощения и повышения эффективности двуязычного перевода: сначала переводятся отдельные клаузы, а затем из них составляется перевод сложного предложения. В [11] приводится описание фактографической поисковой системы, которая в ответ на запрос выдаёт предикативные единицы, соответствующие клаузам.

Проблема разбивки текста на единицы, выражающие отдельное суждение, важна и для таких областей, как исчисление высказываний, теория искусственного интеллекта, теория логического вывода [12]. В булевой алгебре, как известно, выделяются простые и сложные высказывания и истинность последних зависит от истинности первых. В качестве типичного примера сложного высказывания часто приводятся высказывания типа *Это утро теплое и ясное*, при этом считается, что оно может быть разбито на два простых с отношением конъюнкции: *Это утро ясное* & *Это утро тёплое* [13]. Однако с лингвистической точки зрения сложность таких высказываний далеко не очевидный факт, поскольку данные высказывания в рамках системного синтаксиса интерпретируются как простые распространённые, а не сложные. Интерпретация простых распространённых предложений, включающих сочинительные словосочетания, представляет достаточно сложную проблему и требует глубокого лингвистического анализа. Л. Блумфилд словосочетания этого типа относил к эндоцентрическим, отличая их от экзоцентрических, на основе которых не могут быть развёрнуты отдельные предложения. Например, в предложении *Tom and Mary ran away* словосочетание *Tom and Mary* является эндоцентрическим, а словосочетание *ran away* – экзоцентрическим [14].

На наш взгляд, при интерпретации таких словосочетаний следует учитывать три взаимосвязанных фактора: позицию словосочетания в структуре высказывания; морфологический статус предикативных имён; влияние развёртывания высказывания на изменение контекста. Как показывает приводимый ниже анализ, развёртывание высказываний на основе эндоцентрического словосочетания нецелесообразно, если оно занимает субъектную позицию. В этом случае словосочетание можно рассматривать как один объект, которому приписывается общий признак. Например, сочинительное словосочетание в предложении *Пётр и Иван уехали на каникулы* предполагает пресуппозицию совместности (уехали вместе) или одновременности (уехали в одно время) выполнения действия. При развёртывании высказываний на основе этого эндоцентрического словосочетания (*Пётр уехал на каникулы*. *Иван уехал на каникулы.*) данная пресуппозиция снимается, что ведёт к изменению смысла текста. То же самое относится к приводимому Л. Блумфилдом примеру *Tom and Mary ran away.* В некоторых случаях такое развёртывание вообще невозможно, ср.: *Немецкий и узбекский языки разносистемны. – *Немецкий язык разносистемен.*

**Узбекский язык разносистемен.* В данном случае предикативное прилагательное имеет валентность на субъект во множественном числе.

Не меньшую сложность представляет анализ сочинительных словосочетаний в позиции предиката. Ср.: *I saw Tom and Mary → I saw Tom. I saw Mary.* Очевидно, что сочинительная конструкция предполагает одновременность событий, в то время как при её разбивке на отдельные высказывания выражается мысль о последовательности во времени (сначала увидел Тома, потом увидел Мэри), что изменяет смысл текста. Если предикат выражен составным именным сказуемым, в котором используется сочинительная конструкция с предикативными прилагательными, то развертывание на их основе отдельных высказываний приводит к изменению коммуникативного и стилистического планов текста. Ср.: (1) *Этот день ясный и тёплый* → (2) *Этот день ясный, этот день тёплый.* → (3) *Этот день ясный. Этот день тёплый.* Различие между этими вариантами состоит, во-первых, в степени узуальности: использование сочинительного словосочетания в простом предложении (1) является наиболее узуальным, а развёртывание в отдельных высказываниях (3) – наиболее окказиональным. Во-вторых – в степени категоричности: если (1) выражает констатацию факта, характерную для описательно-повествовательного контекста, то (3) выражает категоричное утверждение, характерное для аргументативного контекста, спора, дискуссии.

Аналогичные выводы можно сделать и относительно высказываний с другими типами союзов и соответствующих булевых функций.

Не всегда очевидна и возможность декомпозиции предложений, которые в системном синтаксисе считаются сложными. Это относится к сложным предложениям с ограничительными придаточными, вводными частями и вставными конструкциями. Во вводной части может выражаться модус или речевое действие. Ср.: (4) *I think he is right*; (5) *He is the doctor who works at our hospital.* Разбивка (4) приводит к аномальности из-за незаконченности высказывания: **I think. He is right.* Разбивка (5) приводит к несвязности текста: *He is the doctor. The doctor works at our hospital.*

Можно выделить следующие виды синтаксических структур, которые теоретически допускают развертывание высказываний на основе отдельных клауз:

1. Структуры с сочинительными, противительными, разделительными союзами, которые, как считается в исчислении высказываний, по своему значению соответствуют булевым функциями и могут соединять части сложного высказывания.

2. Структуры с полупредикативными конструкциями, причастными, деепричастными оборотами, которые в некоторых английских грамматиках рассматриваются как неличные клаузы (non finite clauses) [15], а предложения с этими конструкциями считаются сложными.

3. Структуры с предикативными конструкциями – ограничительными придаточными и вводными частями.

4. Структуры с предикативными конструкциями – придаточными описательными, частями сложного предложения.

Развертывание на основе всех этих структур отдельных высказываний может привести к несвязности и даже бессмысленности текста. Ср. отрывок из известного рассказа О. Генри «The Ransom of Red Chief» и его развернутый вариант:

«Bill and me had a joint capital of about six hundred dollars, and we needed just two thousand dollars more to pull off a fraudulent town-lot scheme in Western Illinois with…. Philoprogenitiveness, says we, is strong in semi-rural communities therefore, and for other reasons, a kidnapping project ought to do better there than in the radius of newspapers that send reporters out in plain clothes to stir up talk about such things. We knew that Summit couldn't get after us with anything stronger than constables and, maybe, some lackadaisical bloodhounds and a diatribe or two in the Weekly Farmers' Budget».

Bill had a joint capital of about six hundred dollars. I had a joint capital of about six hundred dollars. And we needed just two thousand dollars more to pull off a fraudulent town-lot scheme in Western Illinois with. Philoprogenitiveness is strong in semi-rural communities. Says we. Therefore, and for other reasons, a kidnapping project ought to do better there than in the radius of newspapers. The radius of newspapers send reporters out in plain clothes to stir up talk about such things. We knew. Summit couldn't get after us with anything stronger than constables. And, may be, Summit couldn't get after us with anything stronger than lackadaisical bloodhounds. And, may be, Summit couldn't get after us with anything stronger than a diatribe or two in the Weekly Farmers' Budget. And, may be, Summit couldn't get after us with anything stronger than two diatribes in the Weekly Farmers' Budget.

Мы считаем, что выбор того или иного варианта декомпозиции обусловливается целями конкретного проекта по автоматическому анализу текста. Наиболее очевидной является разбивка на отдельные высказывания сложносочинённых предложений и предложений с придаточными описательными. Однако и в этом случае нарушается коммуникативная структура текста, поскольку повышение синтаксического статуса высказывания, представление в форме отдельного предложения изменяет распределение тем текста. Оптимальной является виртуальная разбивка на клаузы на этапе предварительной обработки с последующей выдачей пользователю оригинального текста. В нашем проекте по разрешению анафоры в английских текстах разбивка на клаузы проводилась виртуально с целью вычисления расстояния от местоимений до возможных антецедентов. На выходе получался оригинальный текст, в котором местоимения были заменены кореферентными именами. Ниже описываются грамматика и алгоритмы разбивки на клаузы, которые были разработаны в процессе реализации этого проекта.

3. ГРАММАТИКА И АЛГОРИТМЫ

В соответствии с задачами проекта клаузсплитинг проводился для всех сложных утвердительных, отрицательных и восклицательных предложений, включая придаточные ограничительные. Вопросительные предложения не рассматривались, поскольку данный тип предложений, как правило,

не информативен и выражает запрос на получение информации. Под клаузой в английском предложении понималась последовательность именного (NP) и глагольного (VP) словосочетаний. Была разработана грамматика, включающая правила распознавания этих словосочетаний.

Поскольку структура именных словосочетаний достаточно разнообразна (см. табл. 2), для её распознавания были разработаны оптимизирующие правила:

1) Правило подстановки именного компонента (теги NN, NNS, PN). В соответствии с этим правилом структура с одним именным компонентом, указанная в табл. 2, также является действительной и для других именных компонентов. Например, структура PN POS NN предполагает, что могут также существовать структуры: PN POS NNS, NNS POS NNS, NN POS NN, PN POS PN.

2) Правило расширения однородными членами. В соответствии с этим правилом количество повторов одного и того же тега не ограничивается. Наряду со структурой Det Det NNS, указанной в табл. 2, могут существовать структуры: Det Det NNS NNS, Det Det Det NNS NNS, Det Det Det NNS NNS NNS.

Таблица 2

Структура именных словосочетаний (NP)

Структура NP в тегах частей речи	Значение тегов	Пример
NNS	Plural noun	Fruits
PNP	Personal pronoun	I/they/we
PN	Proper noun	John
NN NN	NN = singular noun	Stone wall
Det N	Det = {articles, demonstrative and indefinite pronouns}	My/his/John's car
Det Det NNS		All the cars
Det PNS NNS	PNS = possessive pronoun	All our cars
Det NN NN		A stone wall
AJ NNS	AJ = adjective	Nice cars
AJ AJ NNS		Nice, speedy cars
AJ AJ CC AJ NNS	CC = coordinating conjunction	Nice, speedy, and spacious cars
AV AJ NNS	AV = adverb	Very badly broken cars
AV AV AJ NNS		Very badly broken cars
Det AJ NN		A nice car
Det AV AJ NN		A broken car
PN NNS		Mercedes cars
PN PN NNS		Mercedes Benz cars
Det PN NN		A Mercedes car
PN CC PN		Nokia and Sony
PN CC PN NNS		Nokia and Sony corporations
PN POS NN	POS = Possessive marker	Mary's car
PN CC PN POS NN		John and Mary's car
NN PRP CRD NNS	PRP = preposition CRD = cardinal numeral	Distance of 70 miles
PN CRD NN		Mercedes 600 car
NNS VBN PRP NNS	VBN = past participle verb	Cars broken to pieces
Det NN VBN PRP N		The car broken to pieces
Det NNS VBG AV	VBG = gerund verb	The cars driving there
Det NN TO VBI	TO = particle VBI = infinitive verb	The work to do
Det NN TO VBI AV		The work to do tomorrow

Структура глагольных словосочетаний определялась по следующим правилам:

1) Началом глагольного словосочетания являются токены с тегами VBP (present simple verb), либо VBZ ("s" present simple verb), либо VBD (past simple verb), либо VBM (modal verb).

2) Концом глагольного словосочетания могут быть токены с тегами VBI (infinitive verb), либо VBN, либо VBG.

3) Между токеном, с которого начинается словосочетание, и токеном, которым оно заканчивается, может располагаться (в любой позиции) токен с тегом TO (particle), либо один или более идущих подряд токенов с тегом AV (general adverb). Этим тегом также обозначалась частица NOT, что позволяло распознавать отрицательные предложения.

Примеры глагольных словосочетаний, которые могут распознаваться по этим правилам, приведены в табл. 3.

Таблица 3

Образцы структур глагольных словосочетаний (VP)

Предложение	Структура предложения	Структура VP
I was waiting	I [NP] was waiting [VP]	VBD VBG
The office has already opened	The office [NP] has already opened [VP]	VBZ AV VBN
The students will not come	The students [NP] will not come [VP]	VBM AV VBI
The director didn't appear	The director [NP] didn't appear [VP]	VBD AV VBI
They were impressed	They [NP] were impressed [VP]	VBD VBN
They were very greatly impressed	They [NP] were very greatly impressed [VP]	VBD AV AV VBN

Можно отметить следующие особенности разработанной нами грамматики.

В соответствии с целью проекта распознавались только глагольные и именные словосочетания на основе линейной последовательности тегов частей речи. Адъективные, предложные, адвербиальные словосочетания не выделялись в качестве отдельных

видов фраз и включались либо в именное, либо в глагольное словосочетание. Неличные формы глагола (причастия и инфинитивы) могут входить в состав как NP, так и VP, в зависимости от их позиции. В составе глагольной фразы они всегда используются после личного или модального глагола, а в составе именной – после существительного или собственного имени. Аннотирование тегами частей речи проводилось стохастическим теггером, который был разработан нами ранее [4] и в котором используется такой же набор тегов, как и в Американском национальном корпусе [16]. Помимо теггера частей речи в процессе предварительной обработки текста использовались разработанные нами ранее модули лексической и синтаксической декомпозиции, описанные в [4].

Разбивка на клаузы проводилась по следующим правилам и алгоритмам:

1. Распознавание сложных предложений

Разбивка на клаузы проводилась для сложных предложений, простые предложения игнорировались. Под сложным понималось предложение, включающее более одной последовательности NP VP.

Находится глагол, обозначенный тегом VB/VD/VBZ/VBM. Слева от него находится именной компонент с тегом NN/NNS/PN/PNP. Образуется NP. Далее смотрится часть предложения справа. Ищется VB/VD/VBZ/VBM. Слева от него ищется NN/NNS/PN/PNP. Если находится ещё одна последовательность NP VP, то определяется, что предложение сложное и нужно проводить декомпозицию. Далее ищется ещё одна пара NP VP. В зависимости от того, сколько найдено таких пар, определяется количество клауз: если две пары – две клаузы, если три пары – три клаузы и т. д.

2. Определение границ клауз

В состав первого NP в предложении включаются все слова с начала предложения. В состав последнего VP включаются все слова до конца предложения.

Условным началом второй и последующих клауз является именной компонент. Если слева от него находится слово из специального списка, то разделитель передвигается к началу этого слова, которое включается в состав NP и, соответственно, клаузы. Если подряд идут два или несколько слов из специального списка, то началом является последнее слева слово. Специальный список включает слова, которыми не может заканчиваться клауза. К ним относятся артикли, указательные и относительные местоимения, некоторые прилагательные, союзные слова. Например: *My best friend John* [NP] *eats apples with pleasure* [VP] || *because he* [NP] *likes them very much* [VP].

Вначале находится первое VP *eats*, которое соотносится с NP *John*. Далее находится VP *likes*, которое соотносится с NP *he*. Определяется, что предложение сложное и состоит из двух клауз. В состав первой клаузы включаются все слова, которые занимают позицию перед *John*. В состав второй клаузы включаются все слова, которые занимают позиции после *likes* до конца предложения. Перед *he* ставится условная граница клауз. Далее смотрятся слова слева от *he* и обнаруживается слово из специального списка *because*. Разделитель переносится и ставится перед *because*.

Как показали предварительные экспертные оценки, по описанным правилам и алгоритмам правильно разбивается на клаузы 92 % предложений.

4. ЗАКЛЮЧЕНИЕ

В представленной статье были описаны оригинальные грамматика и алгоритмы клаузальной декомпозиции текста, которые имеют существенное значение для всех областей, связанных с автоматическим анализом текста. Клаузальная декомпозиция составляет основу для выполнения алгоритмов разрешения анафоры, моделирования логико-семантической структуры текста. Замена местоимений референтными именами, которая проводится в результате разрешения анафоры, повышает эффективность информационного поиска и автоматического реферирования, а также вопросно-ответных систем и систем интеллектуального анализа текста. Разбивка текста на клаузы позволяет устанавливать логико-семантические отношения между суждениями, раскрывать иерархическую структуру текста, что имеет непосредственное отношение к разработке систем искусственного интеллекта.

Несмотря на очевидную значимость клаузальной декомпозиции нам удалось найти только две работы, непосредственно затрагивающие эту тему: упомянутую выше работу по применению разбивки на клаузы в машинном переводе, а также работу по интеллектуальному анализу биомедицинских текстов [17]. Можно предположить, что недостаточная изученность алгоритмов клаузальной декомпозиции объясняется сложностью их разработки и применения.

Основной проблемой, как мы полагаем, является тесная взаимосвязь различных планов текста. Разбивка текста на клаузы как самостоятельные единицы семантического плана приводит к трансформации коммуникативного и модального планов, что может привести к несвязности и аномальности текста. Модальный план может быть интегрирован в семантический, что усложняет процесс декомпозиции. Ср., например, *He must be right*. В данном случае модальный глагол не только выполняет предикативную функцию, но и выражает модус предположения (*Perhaps he is right*). Для вычленения семантического плана требуется заменить модальный глагол и инфинитив на соответствующую форму глагола *be*, а для сохранения модальности высказывания – соотнести значение модального глагола со сходным по значению вводным словом. Если учесть, что *must* – многозначный глагол, который может использоваться в одной форме и в настоящем, и в прошедшем времени, то такие трансформации представляют собой достаточно нетривиальную задачу.

Разработка грамматики – также одна из сложных проблем. Очевидно, что грамматика должна быть адаптирована к типологии данного языка, учитывать универсалии словопорядка (SVO, SOV и т. п.). В разработанном нами варианте линейной грамматики эти особенности учитывались при делении клаузы на именные и глагольные словосочетания,

однако вполне возможен вариант более дробного членения с учётом фраз других типов, а также на основе иерархических структур.

Можно предположить, что исследования клаузальной структуры и декомпозиции текста со временем составят отдельную область исследований, а данная статья внесёт вклад в её развитие.

СПИСОК ЛИТЕРАТУРЫ

1. Salton J., Allan J. Automatic text decomposition and structuring // RIAO 1994: Proceedings of the Conference for Intelligent Multimedia Information Retrieval Systems and Management. – New York : CID, 1994. – P. 6 – 20.
2. Fox C. Information retrieval. Chapter 7. Lexical analysis and stoplists [Электрон. ресурс] / AT&T Bell Laboratories. – Holmdel, 2002. – URL: http://orion.lcg.ufrj.br/Dr.Dobbs/books/book5/chap07.htm
3. Cahill A. Proper Meaning Superstition. I. A. Richards [Электрон. ресурс] / University of Colorado at Boulder. – 1998. – URL: http://www.colorado.edu/Communication/meta-discourses/Papers/App_Papers/Cahill.htm
4. Яцко В. А., Стариков М. С., Ларченко Е. В., Вишняков Т. Н. Алгоритмы предварительной обработки текста: декомпозиция, аннотирование, морфологический анализ // НТИ. Сер. 2. – 2009. – № 11. – С. 8 – 18.
5. Степанов Ю. С. Имена. Предикаты. Предложения. – М. : Наука, 1981. – 360 с.
6. Падучева Е. В. Суждение // Лингвистический энциклопедический словарь. – М., 1990. – С. 499.
7. Яцко В. А. Рассуждение как тип научной речи. – Абакан : Изд-во Хакасского гос. ун-та им. Н. Ф. Катанова, 1998. – 182 с.
8. Москальская О. И. Грамматика текста. – М. : Высшая школа, 1981. – 183 с.
9. Marcu D. Discourse trees are good indicators of importance in text // Advances in automatic text summarization. – Cambridge ; London, 1999. – P. 123 – 136.
10. Manchanda S., Gupta D., Bhusal A. et al. Language independent Lexicon Building Tool [Электрон. ресурс]. – 2009. – URL: http://www.cfilt.iitb.ac.in/wordnet/webhwn/IndoWordnetPapers/09_iwn_Language%20independent%20Lexicon%20building%20tool.doc
11. Marchisio G., Dhillon N., Liang J. et al. A case study in natural language based Web search // Natural language processing and text mining / eds. Kao A., Poteet S. – London, 2007. – P. 69 – 90.
12. Barland I., Greiner J. Propositional Logic: inference rules [Электрон. ресурс]. – 2009. – URL: http://cnx.org/content/m10718/latest
13. Шауцукова Л. З. Информатика 10 – 11. Гл. 5. Логические основы компьютеров [Электрон. ресурс]. – М. : Просвещение, 2000. – URL: http://book.kbsu.ru/theory/chapter5/1_5.html
14. Блумфилд Л. Язык. – М. : Прогресс, 1968. – 317 с.
15. Byrd P. Complex sentences [Электрон. ресурс] / Georgia State University. – 2009. – URL: http://www2.gsu.edu/~eslhpb/grammar/lecture_10/complex.html
16. American National Corpus [Электрон. ресурс]. – 2002 – 2010. – URL: http://www.americannationalcorpus.org/
17. Chidambaram D. Processing complex sentences for information extraction : A thesis presented in partial fulfillment of the requirements for the degree master of science [Электрон. ресурс] / Arizona State University. – 2005. – URL: http://www.public.asu.edu/~cbaral/thesis/deepthi04.pdf

Материал поступил в редакцию 09.04.12.

Сведения об авторе

ЯЦКО Вячеслав Александрович – доктор филологических наук, профессор Хакасского государственного университета им. Н. Ф. Катанова, г. Абакан

E-mail: viatcheslav-yatsko@rambler.ru

УДК 81’367 : 811.531

Е. Л. Рудницкая

Структура предложения в корейском как языке типа SOV

Рассматривается корейский материал с точки зрения положений и структур генеративной грамматики, с точки зрения типологических языковых универсалий. Анализируются основные синтаксические характеристики корейского предложения: базовый порядок слов, структура именной группы, глагольной группы, другие особенности порядка слов в предложении. В целом, корейский считается неконфигурационным языком, и грамматика корейского недостаточно полно описывается исходя из положений Теории падежа или Связывания генеративной грамматики. Однако корейский – язык с в большой степени фиксированным, а не свободным порядком слов, так что Теорию передвижения генеративной грамматики до некоторой степени можно приложить к корейскому языку.

Ключевые слова: *порядок слов, ветвление влево, генеративная грамматика, корейский язык, структура предложения/группы, передвижение*

1. ТИПОЛОГИЧЕСКИЕ УНИВЕРСАЛИИ ПОРЯДКА СЛОВ

Порядок слов – одна из наиболее изученных и одновременно наиболее сложных областей синтаксиса. Это связано с тем, что исследователи не только пытаются описать закономерности порядка членов предложения (или элементов внутри именной группы (NP), сказуемого/глагольной группы (VP) и т. д.). Также важно выявить: 1) закономерности порядка слов, соблюдающиеся не в отдельно взятом языке, а во всех или хотя бы во многих языках; 2) связь порядка слов с грамматической и коммуникативной структурой предложения. Эти задачи в рамках типологии были поставлены Дж. Гринбергом в его работах по языковым универсалиям, например в [1]. Универсалии порядка слов также разрабатывались С. Куно [2] и другими исследователями, см. работу [3]. Порядок слов в предложении в данных работах обозначается через порядок подлежащего (S), глагола (V) и прямого дополнения (O) [1]. Глоссы и сокращения, а также обозначения составляющих приводятся в *Приложении*.

В языках мира возможны все варианты последовательности данных основных членов предложения, см. [3, с. 81]. Например, в тагальском языке (филиппинские языки) базовым считается порядок VOS [4, с. 223]. Однако, говоря о корейском и вообще об алтайских языках, можно говорить только о порядке SOV (для сравнения, в европейских языках порядок другой – SVO). В обоих случаях подлежащее стоит в начале предложения, а порядок дополнения и глагола разный. Этот факт привел многих исследователей к рассмотрению позиции в предложении подлежащего отдельно от взаимного расположения глагола и прямого дополнения (VO и OV, см. [2, с. 3–5; 5, с. 95–96; 6; 7, с. 482–483[2]]). Пример (1) иллюстрирует указанные различия в порядке слов:

(1) Kim sensayng-nim-i **phyenci-lul** (O)
Ким учитель-ГОН-ИМ письмо-ВИН
ssu-si-ess-ta (V) [корейский]
писать-ГОН-ПРОШ-ИЗЪЯВ
“Господин Ким **написал** (V) **письмо** (O)” [русский]

2. ПОРЯДОК СЛОВ В КОРЕЙСКОМ ПРЕДЛОЖЕНИИ: ВЕТВЛЕНИЕ ВЛЕВО

2.1. Можно ли говорить о параметре “право/левосторонней вершины”?

При рассмотрении позиции глагола и прямого дополнения в сказуемом (или глагольной группе), учитывая то, что тут глагол – главный член, а прямое дополнение – зависимый, в рамках грамматики непосредственных составляющих можно говорить о ветвлении вправо и влево (внутри глагольной группы) – см. работы [5–7]. Сравним представление глагола и его зависимого прямого дополнения в рамках структуры зависимостей и в рамках структуры непосредственных составляющих – схем (2а-б) и (3а-б), основанных на примере (1). Только в (3а-б) можно говорить о направлении ветвления. В (2а-б) предикат – главное слово – *ssu-si-ess-ta*/*написал* подчиняет оба актанта (подлежащее и прямое дополнение).

Структура составляющих до определенной степени учитывает (базовый) порядок слов, и поэтому порядок узлов в (3а-б) (NP_1, VP [NP_2, V]/NP_1, VP [V, NP_2]) отражает реальный порядок слов. В структуре непосредственных составляющих “зависимое” понимается в чисто синтаксическом смысле, и поэтому

[1] Под порядком слов понимается тот порядок, который в том или ином языке считается базовым (см. п. 2.2.)

[2] В работе [7] в качестве языка OV приведен один из кавказских языков – аварский.

подлежащее (NP_1) не считается зависимым (дополнением, или комплементом) глагола (V). Так что комплементом глагола является только дополнение. Это позволяет ввести бинарное противопоставление – право-ветвящийся (в котором вершина стоит перед комплементом) и лево-ветвящийся (в котором вершина стоит после комплемента) язык.

(2) а. ssu-si-ess-ta

[Kim sensayng-nim-i] phyenci-lul

б. написал

[господин Ким] письмо

(3) а. S

NP_1 VP

(подлежащее)

[Kim sensayng] NP_2 V

(дополнение) (глагол, вершина)

chayk ssu-si-ess-ta

б. S

NP_1 VP

(подлежащее)

[господин Ким] V NP_2

(глагол, вершина) (дополнение)

написал книгу

Нельзя ввести такой важный параметр, характеризующий язык, как "порядок вершины и комплемента", учитывая только данные о глаголе и прямом дополнении. В работах Дж. Гринберга [1] и многих других авторов обсуждается также порядок следования главного слова (вершины) и зависимого (комплемента) в основных словосочетаниях (конструкциях). Основные конструкции: существительное и его модификаторы/генитивная группа (в NP), предлог/послелог и существительное (в PredP/PostP) и др. Говорить о синтаксическом свойстве правого или левого ветвления применительно к языку можно только если во всех (в том числе в указанных) конструкциях порядок вершины и комплемента такой же, как в VP.

Как будет показано ниже (пп. 2.2, 2.3), корейский (как японский и алтайские языки) можно считать лево-ветвящимися. В работе [3, с. 90] отмечено, что в языках типа SOV более последовательно соблюдается левое ветвление во всех типах конструкций, чем в языках типа SVO – правое ветвление. Например, в романских языках (SVO) прилагательное (определение, или модификатор (Mod), в именной группе – выражающее актант или характеристику существительного – вершины (N)) почти всегда постпозитивно, т. е. стоит после существительного (4а-б)[3]. В русском генитивное дополнение (Gen) также постпозитивно, ср. перевод (5); в английском (5) предложная группа, соответствующая русскому генитивному дополнению, также постпозитивна.

[итальянский]

(4) а. L' **invasion** (N) [8, с. 86, 89]

АРТ.ЖЕН.ЕД вторжение.ЖЕН.ЕД

italiana (Mod) dell' Albania

итальянский.ПРИЛ.ЖЕН.ЕД ПРЕДЛ Албания

[ср. *L' **italiana** (Mod) **invasion** (N) dell' Albania]

"**Вторжение Италии** в Албанию"

[букв: **вторжение** (N) **итальянское** (Mod)…]

б. Le **aggressioni** (N)[8, с. 86, 89]

АРТ.ЖЕН.МНОЖ военные_действия.МНОЖ

brutali (Mod) vanno

грубый.МНОЖ идти.ВСПОМ.МНОЖ

severamente condannate

сурово осуждаться.ПРИЧ.МНОЖ

[ср. *Le **brutali** (Mod) **aggressioni** (N)…]

"**Грубые** (Mod) [**военные действия**] (N) будут сурово осуждаться"

(5) John's **(*of Peter)** (Gen) [английский]

Иван.ПОСС ПРЕДЛ Петр

murder **of Peter** (Gen)

убийство ПРЕДЛ Петр

ср. "**(*Петра)** (Gen) убийство **Петра** (Gen) Иваном"

[букв. **Иваново убийство**…]

Для корейского языка подобная проблема не встает. Как будет показано в п. 2.3, во всех перечисленных выше конструкциях вершина следует за ее комплементом или модификатором. Чисто агглютинативный характер корейского языка, упрощая задачу последовательного построения структуры непосредственных составляющих для корейского предложения, осложняет другой аспект грамматики – деление предложения на слова[4]. Эта проблема (и связанные с ней проблемы выделения и интерпретации некоторых грамматических категорий) почти не встает в европейских языках.

2.2. Базовый порядок слов. Сравнение с немецким и с языками Сибири

Во многих теоретических и учебных работах и конкретных исследованиях – [3, 5, 7], особенно [6, с. 13] – упоминается проблема базового порядка слов. Легче всего определить базовый порядок слов в языках с достаточно строго фиксированным порядком слов (английский). В таких языках допускаются отклонения от базового порядка слов в определенных контекстах, которые поддаются строгому описанию, например, общий и частный вопрос[5].

[3] Дж. Чинкве [8] и многие другие авторы все же пытаются доказать, что романские языки последовательно право-ветвящиеся (см. п. 2.2.) Р. Кейн [9] и другие авторы предлагают сложную схему вывода того порядка слов, который мы наблюдаем в языках типа SOV, из "базовых" структур типа SVO.

[4] См. работы [10–12] и многие другие: корейские авторы (как и японские) считают по крайней мере именные падежные показатели отдельными словами (частицами) и пишут их отдельно при делении текста на слова. См. п. 2.3.3.

[5] Эти отклонения описываются с помощью трансформаций, или передвижений: при образовании общего вопроса происходит инверсия подлежащего и вспомогательного глагола (subject-AUX inversion) или вставляется вспомогательный глагол *do*. См. п. 3.1. В частном вопросе, кроме того, вопросительное слово перемещается в начало предложения: ***You are*** *ready* → ***Are you*** *ready? He hears well* → ***Does*** *he hear well?* ***You are*** *reading [****what****]* → ***What are you*** *doing?*

В языках со свободным порядком слов бывает довольно трудно доказать базовый статус одного из типов порядка слов. Так Дж. Бейлин [13] доказывает, что в русском языке базовый порядок слов SVO, а не VSO (как предлагает Т. Кинг [14]). См. работу [15, с. 17–43], в которой проблема свободного порядка слов рассматривается на материале венгерского с учетом категорий актуального членения.

В языках типа SOV, как уже отмечалось, порядок слов обычно строго фиксирован, поэтому базовый статус этого порядка слов (в японском, корейском, турецком) почти не требует доказательств. Для иллюстрации рассмотрим немецкий и некоторые алтайские языки.

Считается, что базовым порядком слов в немецком языке является порядок слов в придаточном предложении, а не в главном. В примере (6а) показан этот порядок слов (SOV). В главном предложении (6б) дополнение следует за глаголом (порядок SVO). Хотя из примера (6б) можно было бы сделать вывод об отсутствии единого базового порядка в немецком языке, все же базовым считается порядок SOV из (6а). Это связано с тем, что в главном предложении главный или вспомогательный глагол всегда является вторым (Verb-second), а перед ним может стоять любой член предложения, ср. (6в)[6] с обстоятельством (Adv) в начальной позиции – подлежащее в (6в) должно стоять после глагола.

[немецкий]

(6) а. Ich weiβ [daβ (C) **der** **Mann** (S) [3, с. 83]
я знаю что АРТ.ИМ мужчина
den **Jungen** (O) sah (V)**]**
АРТ.ВИН мальчик видел
“Я знаю, что мужчина (S) видел (V) мальчика (O)”

б. …Der Mann (S) (*__den__ **Jungen** (O)) [3, с. 83]
АРТ.ИМ мужчина АРТ.ВИН мальчик
sah (V) **den** **Jungen** (O)
видел АРТ.ВИН мальчик
“Мужчина (S) видел (V) мальчика (O)”

в. **Gestern** (Adv) (*__Peter__ (S))[изм. 9, с. 28]
вчера Петр Петр
tanzte **Peter** (S)…
танцевал Петр
“**Вчера** (Adv) **Петр** (S) танцевал” (…лучше, чем всегда)

В корейском и в главном, и в придаточном предложении одинаковый порядок слов SOV, при этом косвенное дополнение (IO) обычно стоит перед прямым (DO), ср. примеры (7) и (8).

[корейский]

(7) а. *Ku* *yeca-eykey* (IO) **chayk-ul** (DO)
этот женщина-ДАТ книга-ВИН
cwu-ess-ta (V)
дать-ПРОШ-ИЗЪЯВ

б. *Ku yeca-eykey* (IO) ***cwu-ess-ta*** (V) **chayk-ul** (DO)

в. **Chayk-ul** (DO) ***cwu-ess-ta*** (V) *ku yeca-eykey* (IO)
“(Я) дал (V) этой женщине (IO) книгу (DO)”

(8) а. [*Ku yeca-eykey* (IO) **chayk-ul** (DO)
этот женщина-ДАТ книга-ВИН
cwu-ess-ki (V)] ttaymwuney…
дать-ПРОШ-ИЗЪЯВ причина.ВСПОМ

б. *Ku yeca-eykey* (IO)***cwu-ess-ki*** (V) **chayk-ul** (DO) ttaymwuney…

в. **Chayk-ul** (DO) ***cwu-ess-ki*** (V) *ku yeca-eykey* (IO) ttaymwuney…
“…потому что (я) дал (V) этой женщине (IO) книгу (DO)…”

В работе [17] показано, что основные закономерности порядка слов в придаточном предложении в немецком, с одной стороны, и в главном предложении в корейском, с другой, сходны. Таким образом, более однозначный корейский материал (одинаковый порядок SOV и в главном, и в придаточном предложении) – аргумент в пользу постулирования SOV как строго соблюдаемого базового порядка слов в немецком.

В заключение сравним корейский, как язык со строгим порядком слов SOV (strict SOV) с теми языками типа SOV, в которых этот порядок соблюдается не так строго (хотя является базовым). Подобные отклонения есть, в частности, в некоторых языках Сибири: тюркских, монгольских, тунгусо-маньчжурских языках. Например, в эвенкийском (тунгусо-маньчжурская группа) возможна постпозиция по отношению к глаголу (V) относительного предложения (Rel) с внешней вершиной (O) (пример (9а)), и даже постпозиция O и Rel одновременно [18, с. 137] (пример (9б)), что недопустимо в корейском [12, с. 296]. Ср. корейские примеры (10а-г) с относительным придаточным Rel и вершиной N. Rel может быть отделено от N другими модификаторами (Mod), но не может стоять после N.

[эвенкийский]

(9) а. Asā-l(S) **sulakī-wa**(O) [18, с. 137]
женщина-МНОЖ лиса-ВИН
ičə-rə-∅ (V)
видеть-ПРОШ-3.МНОЖ
[mō-wa tūkty-d’ərī-wə] (Rel)
дерево-ВИН влезать-ПРИЧ.ОДНОВР-ВИН
“Женщина увидела **лису** (O), **[(которая) влезала на дерево]** (Rel)” **[SOV-Rel]**

б. Bū (S) ičə-rə-w(V) [18, с. 137]
мы видеть-ПРОШ-1.МНОЖ
[agi-dū baka-na-l-wa-tyn] (Rel)
лес-ДАТ находить-ПРИЧ.ПФ-МНОЖ-ВИН-3.МНОЖ
oro-r-wo (O)
олень-МНОЖ-ВИН
“Мы (S) видели (V) **оленей** (O), **[(которых) (они) нашли в лесу]** (Rel)” **[SVO-Rel]**

[6] Согласно анализу в рамках генеративной грамматики, например в работе [16], в главном предложении вершина V (глагол) передвигается в вершину C (союз, комплементайзер), а перед V ставится аргумент или обстоятельство (6б-в). В придаточном предложении (6а) вершина C занята союзом и поэтому V не может передвинуться в C.

[корейский]

(10) а.Ku (Mod) **[sa-i cci-n]**(Rel)[12, с. 296]
этот плоть-ИМ толстеть-ПРИЧ
twu mali-uy(Mod) **mal** [N]
два КЛАСС-РОД лошадь
“Эти две (Mod) **растолстевшие/толстые** (Rel) **лошади** (N)” **[Mod-Rel-Mod-N]**

б. **[SA-I CCI-N]** (Rel)ku (Mod) twumali-uy(Mod) **mal** [N] **[Rel-Mod-Mod-N]**

в. Ku (Mod) twumali-uy(Mod) **[sa-i cci-n]** (Rel)**mal** [N] **[Mod-Mod-Rel-N]**

г. *Ku (Mod) twumali-uy(Mod) **mal** [N] **[sa-i cci-n]** (Rel) [***Mod-Mod-N-Rel]**

В разговорной речи или в нарративном тексте в эвенкийском языке, особенно при фокусировании глагола (V) или прямого дополнения (O), порядок SOV (в (11а)) легко нарушается[7]. Так, в (11а-е)) допускаются все порядки, указанные в работе [3, с. 81][8]. Данные работы [18, с. 1] в примере (11а-е) подтверждаются материалами [19]. В корейском в разговорной речи возможно нарушение порядка OV только с паузой после V [20, с. 99–100]. Такая пауза допускает интерпретацию с нулевым местоимением **pro**[9] в позиции O перед V (**Ø** в (12а)), кореферентного именной группе O, составляющей отдельное именное предложение (afterthought) – ср. (12а). После глагола может стоять не только прямое (DO), но также косвенное дополнение (IO) и подлежащее (12б), или несколько аргументов сразу (12в). Постпозитивная составляющая образует отдельное предложение (вероятно, с эллипсисом V и S, DO, IO и т.д.)

[эвенкийский]

(11) а. bəjūmimnī (S) **mōtȳ–wa** (O) [18, с. 129]
охотник лось-ВИН
wā–rə–n (V)
убить-ПРОШ-3.ЕД
“Охотник (S) убил (V) лося (O)”

б. ***WĀ–RƏ–N***(V)bəjūmimnī (S) **mōtȳ–wa** (O)

в. ***WĀ–RƏ–N***(V)**mōtȳ–wa** (O) bəjūmimnī (S)

г. bəjūmimnī (S) ***WĀ–RƏ–N***(V)**mōtȳ–wa** (O)

д. **MŌTȲ–WA** (O) bəjūmimnī (S) ***wā–rə–n*** (V)

е. **MŌTȲ–WA** (O) ***wā–rə–n*** (V) bəjūmimnī (S)
[корейский] [20, с. 99-100]

(12) а. *Yong-i* (S) Mia-eykey (IO)***ponay-ss-e-yo*** (V)…
Йонг-ИМ Миа-ДАТ посылать-ПРОШ-ИНФ-ВЕЖЛ
Kulim-yepse-lul. (DO)
открытка-ВИН
“Йонг **послал** (**ее**) Миа … **Открытку**.”
[S-IO-~~(DO)~~-V… -DO]

б. Kulim-yepse-lul (DO) Mia-eykey (IO)
ponay-ss-e-yo (V)… **Yong-i** (S)
“**(Он)** открытку Миа **послал** … **Йонг**.**”**
[~~(S)~~-DO-IO-V… -S]

в. **Ponay-ss-e-yo** (V)…
Yong-i (S) Mia-eykey(IO) kulim-yepse-lul(DO)
“**(Он) (ей) (ее) послал** … **Йонг Миа открытку**.”
[~~(S)-(IO)-(DO)-~~V… -S-IO-DO]

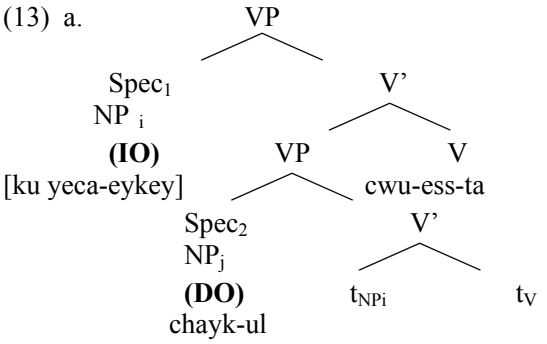
Таким образом, мы показали, что корейский – язык типа SOV, и этот порядок слов соблюдается строго, в отличие от многих других языков типа SOV[10].

2.3. Ветвление влево: разные типы составляющих в корейском

2.3.1. Глагольная группа. Как уже говорилось, прямое дополнение всегда предшествует глаголу и при отсутствии дополнительных контекстных факторов (ср. (12а) и (12б-в)) стоит после косвенных дополнений и обстоятельств – см. пример (7а) в п. 2.2., повторенный ниже. Один из способов построения структуры VP с несколькими дополнениями (дитранзитивных, DITRANS), например, для примера (7а) – конструкция из нескольких VP (“многослойная” VP, VP-Shell, анализ Р. Ларсона, см. работу [23] применительно к японскому материалу), которая позволяет придерживаться принципа “бинарного членения”. В (13а) показана структура (7а)[11].

(7) а. *Ku yeca-**eykey*** (IO)**chayk-ul** (DO)
этот женщина-ДАТ книга-ВИН
cwu-ess-ta (V)
дать-ПРОШ-ИЗЪЯВ
“(Я) дал (V) этой женщине (IO) книгу (DO)”

(13) а.



[7] Фокусирование слова мы обозначаем через его написание заглавными буквами.

[8] Такой “свободный” порядок слов в контексте коммуникативного выделения в (11б-е), скорее всего, объясняется влиянием русского языка (с его свободным порядком слов) на в прошлом более строгий порядок SOV в эвенкийском.

[9] В корейском аргументные именные группы опускаются (**Ø/ pro**), если их можно восстановить из контекста.

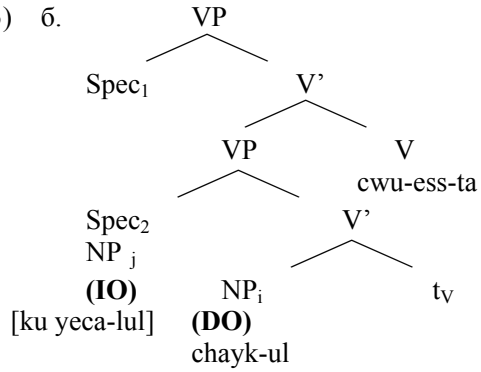
[10] Необходимо заметить, что все языки, которые обсуждались в этом пункте, аккузативные (именительному падежу в переходной конструкции соответствует семантическая роль ‘агенс’, а винительному – ‘пациенс’ (или ‘тема’)). Как показано в работе [21], классификация языков по признаку порядка S, O и V недостаточна при привлечении языков с другим падежным кодированием ядерных актантов (эргативных, активных и т. д.).

[11] В данной структуре вершина V *cwu-ess-ta*, передвигаясь из “нижнего” в “верхний” слой VP, присваивает винительный падеж DO *chayk-ul*. У IO [*ku yeca-eykey*] косвенный, а не прямой/структурный падеж (дательный); этот падеж присваивается вместе с семантической ролью. IO – комплемент V в “нижнем” слое VP, а для достижения препозиции по отношению к DO передвигается в “верхний” SpecVP. Важно, что в исходной структуре, до передвижения IO, DO иерархически выше, чем IO.

Интересно, что у многих дитранзитивных глаголов, в том числе с подъемом посессора, допускается показатель и дательного, и винительного у косвенного дополнения. В предложении могут быть две именных группы в винительном подряд. Так, вместо *yeca-**eykey*** ‘женщина-ДАТ’ в (7а) возможно *yeca-**lul*** ‘женщина-ВИН’: *yeca-**lul** chayk-**ul** cwu-ess-ta* ‘… дал женщине книгу’. В этом случае структура (7а’) – это (13б).

(7) а’. *Ku* *yeca-**lul*** (IO) **chayk-ul** (DO)
этот женщина-ВИН книга-ВИН
cwu-ess-ta (V)
дать-ПРОШ-ИЗЪЯВ
“(Я) дал (V) этой женщине (IO) книгу (DO)”

(13) б.



“Многослойная” VP позволяет представить распределение приоритетов дополнений при дитранзитивных глаголах. Как и в русском, в корейском [24, с. 43 и след.] в конструкции с дативным IO [*ku yeca-eykey*] в (7а) при пассивизации DO *chayk-ul* становится подлежащим. Таким образом, в структуре VP необходимо представлять DO, IO и другие дополнения как иерархически упорядоченные.

Некоторые глаголы [24, 25, 12, с. 370] **требуют** винительного падежа у косвенного дополнения-посессора, например *cap-ta* ‘хватать’, ср. (14а). Эти глаголы допускают так называемый “адверсативный пассив”, в котором номинативное подлежащее – косвенное дополнение, а прямое дополнение может сохранять[12] винительный падеж (14б):

(14) а. Nay-ka(S) **ai-lul**(IO) **son-ul**(DO) [24, с. 49]
я-ИМ ребенок-**ВИН** рука-**ВИН**
cap-ass-ta (V)
хватать-ПРОШ-ИЗЪЯВ
“Я (S) схватил (V) ребенка (IO) за руку” [букв. “…**ребенка** [ВИН](IO) **руку** [ВИН](DO)”]

б. **Ai-ka**(IO) *son-i*/**son-ul**(DO) ∅ (S) [24, с. 49]
ребенок-**ИМ** *рука-ИМ*/**ВИН**
cap-**hi**-ess-ta (V)
хватать-**ПАСС**-ПРОШ-ИЗЪЯВ
“Ребенок (IO) был схвачен (V) за руку” [букв. “…был схвачен (V) **руку** [ВИН] (DO)”]

Возможность такого пассива показывает, что (например, в (14а)) IO *ai-lul* ‘ребенок-ВИН’ (поссессор) иерархически выше (в частности, в структуре непосредственных составляющих), чем DO *son-ul* ‘рука-ВИН’[13].

Другой аргумент в пользу более высокой позиции *ai-lul* ‘ребенок-ВИН’ (IO), чем *son-ul* ‘рука-ВИН’ (DO) в (15а) – невозможность релятивизации *son* (DO) в подобных конструкциях, см. работу [26: 195] и примеры (15а-б) из работы [26].

(15) а. Bill-i(S) Mary-**lul**(IO)son-**ul**(DO) [25, с. 195]
Билл-ИМ Мэри-**ВИН** рука-**ВИН**
cap-ass-ta (V)
хватать-ПРОШ-ИЗЪЯВ
“Билл (S) схватил (V) Мэри (IO) за руку” [букв. “…**Мэри** [ВИН] (IO) **руку** [ВИН] (DO)”]

б. *[Bill-i(S) Mary-lul(IO) Ø cap-un (V)](Rel)
Билл-ИМ Мэри-**ВИН** хватать-ПРОШ-ИЗЪЯВ
son (DO)
рука “Рука, за которую Билл схватил Мэри”

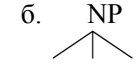
Согласно работе [27], DO стоит выше, чем IO в иерархии по доступности релятивизации, так что (15б) показывает несоответствие этой иерархии IO и DO в конструкциях с подъемом посессора типа (14а), в которых посессору присваивается винительный. Действительно ли IO в таких конструкциях иерархически выше, чем IO? Считается [28, 29], что IO – посессор при подъеме как бы “вытесняет” DO из позиции прямого дополнения, “отбирая” у DO часть синтаксических приоритетов DO. Таким образом, и возможность адверсативного пассива в (14б), и бинарная структура в (13а-б), в которой DO выше, чем IO, или наоборот (если оба имени стоят в винительном), кажется обоснованной.

2.3.2. Именная группа. Именная группа (NP) состоит из вершины N (noun, существительное) и модификаторов имени (Mod, указательного или притяжательного местоимения, прилагательного, причастия или причастного оборота), которые в корейском всегда стоят перед N. Например, (16б) – структура (16а).

[12] В работе [24] предлагается, что второй вариант (14б) с *son-**i*** ‘рука-**ИМ**’ (DO в именительном) представляет не адверсативный, а обычный пассив (с согласованием DO и IO по падежу). В обоих случаях мы имеем дело с исходной структурой, близкой к (13б) – см. [23]. Отличие заключается в том, что в вершине V “верхнего” слоя “многослойной” VP стоит показатель пассива *-hi-*. При обычном пассиве с *son-**i*** V, как правило, теряет возможность присваивать винительный. В случае обычного пассива V (*cap-ass-ta*) передвигается в “верхнюю” VP и, под влиянием *-hi-*, теряет возможность присваивать винительный. В результате и *ai-ka*, и *son-i* присваивается именительный.
В случае адверсативного пассива, V не утрачивает способность присваивать винительный (*son-**ul*** ‘рука-**ВИН**’ в (14б)). Это связано с тем, что V не передвигается из “нижней” в “верхнюю” позицию вершины VP (то есть в вершину V, в которой стоит показатель *-hi*). Пассивизация затрагивает только “верхнюю” часть VP с вершиой V *-hi-* ‘ПАСС’. Таким образом, V присваивает винительный только DO *son-**ul*** ‘рука-ВИН’, но не IO *ai* ‘ребенок’; *ai* передвигается в позицию подлежащего и в этой позиции *ai* присваивается именительный – однако, см. п. 2.3.2 и сноску 14. Заметим, что, в соответствии с гипотезой о порождении подлежащего внутри VP, *ai* в (13б) стоит в позиции SpecVP – позиции подлежащего (внешнего аргумента) – см. сноску 17.

[13] В корейской адверсативной конструкции между IO-подлежащим и DO-дополнением – в (14б) *ai-ka* ‘ребенок-ИМ’ и *son-ul* ‘рука-ВИН’ – должны быть отношения неотъемлемой принадлежности: *ai-**ui** son* ‘рука-РОД ребенок’ “рука ребенка”. Такие ограничения связаны с тем, что подлежащее такого пассива интерпретируется как субъект, которому совершенное действие наносит вред (*укусить* кого-л. (IO) за что-л., *наступить* кому-л. (IO) на что-л. и т. д.). В японском адверсативный пассив возможен при более широком спектре отношений связи между IO и DO: *Zirono musuko* ‘Зиро-РОД сын’ “сын Зиро”. Подробнее см. [25].

(16) а. **Ku kin** maktayki
этот длинный палка
“Эта длинная палка”
б. NP



Mod_1 Mod_2 N
ku kin maktayki

Если сравнить VP и NP в корейском, VP лучше поддается “внутреннему структурированию”. Модификаторы в NP могут следовать в любом порядке в зависимости от факторов контекста и актуального членения – см. [12] и (10а-г) выше. Поэтому мы, в отличие от (13), не придерживаемся принципа “бинарного членения” в (16б) – структуре непосредственных составляющих для NP.

Другой немаловажный признак, отличающий NP от VP – бóльшая просодическая и синтаксическая самостоятельность именных грамматических показателей (падежных) по сравнению с глагольными – см. [10, 11, 12, 22]. В этих работах падежные показатели считаются частицами (отдельными словами), а глагольные – аффиксами (связанными морфемами).

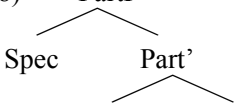
Следствием более автономного статуса именных показателей является то, что в разговорной речи допускается опущение падежных аффиксов/частиц, что невозможно для глагольных аффиксов. Ср. (17а-б) и (17в):

(17) а. Jon-**Ø** pap-**Ø** mek-ess-ni? [11, с. 333]
Джон-ИМ рис/еда-ВИН кушать-ПРОШ-ВОПР
“Поел ли Джон?”
б. **А**: Jon-i nwukwu-hako nol-ass-ni? [11, с. 336]
Джон-ИМ кто-СОВМ играть-ПРОШ-ВОПР
“С кем играл Иван?”
Б: Mali-hako/Mali-**Ø**
Мэри-СОВМ/Мэри
“С Мэри”
в. **А**: Jon-i mwues-ul hay-ss-ni? [11, с. 336]
Иван-ИМ что-то-ВИН делать-ПРОШ-ВОПР
“Что делал Иван?”
Б: Pap mek-ess-e yo/ ***mek**/ ***mek-ess**
рис кушать-ПРОШ-ИЗЪЯВ/* кушать/* кушать-ПРОШ
“Обедал”

Эти и другие факты, приведенные в работах [11, 22, 30] показывают, что имя и его грамматические показатели трудно рассматривать как одно слово, состоящее из корня и связанных служебных морфем. Кроме того, если падежный показатель – связанная морфема, то, согласно принципам генеративной грамматики, соответствующий падеж должен присваиваться некоторой, например, глагольной вершиной (ср. V_1 и V_2 в (13)). Как показано, в частности, в работе [31], механизм присваивания падежа, принятый в стандартной генеративной грамматике, трудно применить к корейскому материалу[14].

В работах [20, 33] предлагается рассматривать имя с падежным показателем (прямого падежа) как синтаксическую группу, вершиной которой является частица – падежный показатель, а не как слово – см. (18). При таком подходе структура предложения несколько усложняется (PartP вместо NP), однако будет учтен статус падежного маркера как отдельного слова.

(18) PartP



Spec Part’
NP Part
Mia -ka ‘Миа-ИМ’ “Миа”

Таким образом, структура NP, вершиной которой является имя N (т.е. группа “существительное, его зависимые и его грамматические показатели”) в меньшей степени поддается описанию в терминах стандартной бинарной структуры непосредственных составляющих, чем глагольная группа VP. Однако параметр ветвления влево (или конечной позиции вершины) в NP не нарушается: вершина в NP всегда на последнем месте.

2.3.3. Послелоги, частицы, наречия. В работах [1, 3, с. 84–85] отмечается, что в языках с последовательно соблюдающимся порядком SVO преобладают предлоги, а в языках с порядком SOV – послелоги. Поскольку предлог/послелог – вершина предложной/послеложной группы (PrP/PostP), в составляющих этого типа то же направление ветвления, что и в глагольной группе (в PrP ветвление вправо, а в PostP – влево) соблюдается более последовательно, чем в именной группе[15].

Русский (английский, романские) и корейский (японский), особенно хорошо иллюстрируют это противопоставление: в первых есть только предлоги, а во вторых – только послелоги. Так, русским предлогам в корейском как правило соответствуют послелоги: *-pwuthe* ‘с/от’, *-kkaci* ‘до’, *-pota* ‘чем’, *-chelem* ‘как’ [34, с. 147–149]; *-tele* ‘кому-л’, *-hanthey(se)* ‘кому-л, (от) кого-л’, *-hamkkey* ‘вместе с’ и т. д. [35, с. 217–218].

(19) 10si-eyse/ 10si(-eyse)**-pwuthe** [34, с. 147]
10_час-МЕСТ/10_час**(-МЕСТ)-с/от**
12si-**kkaci** kongpwuha-ca
12_час-**до** учиться-ПРИГЛ
“Давай будем заниматься **с (от)** 10 часов **до** 12 часов”

Многие исследователи считают послелоги отдельными словами, так же как и падежные частицы (см. обзор в работах [34, 22, с. 60–69]). Наиболее спорным является вопрос о том, принадлежат ли падежные показатели-частицы (см. выше) и послелоги к одной и той же категории “служебных слов и частиц, стоящих после имени”. Семантика послелогов и падежных частиц квази-синонимична: так, послелоги *-hanthey, -tele* во многих случаях эквивалентны показателю дательного падежа *-eykey* (и в некоторых работах даже называются падежными показателями). В примере (19) показано, что послелог *pwuthe* ‘от’ может дублировать

[14] Исходя из проблем, упоминаемых в работе [31], и из материала конструкций с сентенциальными актантами, корейские авторы, например [32], считают, что именительный в корейском – падеж “по умолчанию” (default), т. е. что этот падеж присваивается именной группе, которой не присвоено никакого падежа. Другими словами, именительный не присваивается именной группе какой-либо синтаксической вершиной.

[15] Ср. постпозитивные определения и генитивную группу в примерах (4)-(5) в п. 2.1.

показатель *-eyse* ‘МЕСТ’, одним из значений которого является аблативное (‘удаление от’).

Синтаксические свойства послелогов в сравнении с падежными частицами подробно изучены в работах [34, 36]. Общие выводы, которые делают Й.-М.Й Чхо, П. Селлз, И. Чхои-Джонин, такие:

1) в цепочке грамматических показателей имени (20а) послелог обычно стоит после падежного показателя косвенного падежа (**(Post-2)** в (20а));

2) частицы прямых падежей и ограничительные частицы стоят практически в одной и той же позиции **(Delim-1/Delim-2)**, после показателей косвенных падежей и послелогов. Возможный относительный порядок показателей косвенных и прямых падежей, а также ограничительных частиц, показан в (20б-в).

(20) а. [изм. 22, с. 23]; на основе [36]

(Stem)	***(Hon)***	***(Plur)***	***(Post-1)***	***(Post-2)***
	-nim-	**-tul-**	**-ey-, -eykey-, -(u)lo-**	**-se-, -pwuthe**

(Delim-1)	***(Delim-2)***	*[(Cop)/(Mood/Quot)]*
-man-/-kkaci-/-cocha-	**-(n)un/-i/-ka/-to...**	**[(-i)/(-ta) (-ko)..]**

б. pap-*man*-**i**/ *pap-**i**-*man* [22, с. 21]
еда-*только*-**ИМ**/ *еда-**ИМ**-*только*
“Только еда”
[**(Delim-1-Delim-2)**; ***(Delim-2-Delim-1)**]

в. Hankwuk-**ey**-*man*/*Hankwuk-*man*-**ey** [22, с. 21]
Корея-**МЕСТ**-*только*/*Корея-*только*-**МЕСТ**
“Только в Корее”
[**(Post-1-Delim1)**; ***(Delim-1-Post-1)**]

В русском есть не полностью грамматикализованные предлоги (например, *благодаря* (чему-л.)); в корейском степень грамматикализации послелогов также колеблется. Эти колебания довольно значительны – от частицы (типа приведенных в примере (19) *pwuthe* и *kkaci*) до служебного существительного в (21а-б) или деепричастия (21в), которые составляют самостоятельное слово и сохраняют именные или глагольные свойства.

(21) а. kihan **cen-ey** [35, с. 216]
срок **до-НАПР** “до срока”
б. *kihan-**ey-cen**
в. ...caki pwuha-**lul** **sikhy-ese** [37, с. 59]
себя подчиненный-**ВИН** **принуждать-ДЕЕПР**
[Yucengswu sikkwu-tul-ul samwusil-lo
[Йу_Чонг_Су родственник-МНОЖ-ВИН контора-ИНСТР
teylye o-la] ha-ko...
принести.ИНФ прийти-ПОВ] говорить-ДЕЕПР
“(Он) сказал **своим подчиннным** [привести родственников Йучонгсу в контору] и …”

Отыменные послелоги могут присоединять падежные окончания (*cen-ey* ‘до-НАПР’ в (21а)) и не могут стоять после существительного и его падежного окончания, ср. (21б). Для полностью грамматикализованного послелога (*pwuthe* ‘с/от’ в (19)) позиция после падежного показателя имени нормальна. В (21в) показано, что значение приказа выражается глаголом *(mal)ha-ta* ‘говорить’ и деепричастной формой глагола *sikhi-ta* ‘принуждать’, который управляет винительным падежом (*sikwu-tul-**ul*** ‘родственник-МНОЖ-**ВИН**’). Отыменные послелоги и особенно глагольные формы (по [35], (N-ВИН) *sikhy-ese* - отглагольный послелог) обычно считаются отдельными словами и пишутся отдельно от зависимого от этих послелогов имени. Отнесение отглагольных послелогов к послелогам представляется вообще сомнительным.

Таким образом, послелоги и частицы в корейском языке всегда следуют после того имени, к которому они относятся. Послелоги обычно считаются вершиной группы послелога (PostP). Если считать частицу (падежную) вершиной группы PartP, как это показано в (18), обе эти конструкции являются ветвящимися влево, а вершина в них стоит на последнем месте (как и в VP и NP).

Наречия не представляют большого интереса для структуры предложения в корейском. Наречия делятся примерно на те же группы, что и в европейских языках [37]: эпистемические/модальные, обстоятельственные, образа действия. Позиция этих наречий относительно свободна. Так, и эпистемические/модальные наречия, и наречия образа действия могут стоять в любой части предложения, включая абсолютное начало и конец [38, с. 168–168, 177]. Наречия являются адъюнктами[16], т. е. в структуре сочетания с наречием нельзя четко определить, что является вершиной, а что дополнением/комплементом. Поэтому такие структуры не играют существенной роли для направления ветвления в корейском языке.

2.4. Позиция топика и подлежащего: приоритетные признаки подлежащего в применении к подлежащему или топику

Вернемся к схемам (3а-б) в п. 2.1. Как уже говорилось, в (3а) показано, что корейский – язык, ветвящийся влево, а русский – язык, ветвящийся вправо. Другими словами, вершина (V) в корейском стоит после дополнения (O), а в русском – перед дополнением. Однако подлежащее (S) в обоих случаях стоит перед OV/VO. Хотя подлежащее – не прямое/косвенное дополнение V, такая иерархия все же предполагает более высокий синтаксический статус подлежащего по сравнению с (прямым) дополнением[17].

[16] То есть не являются аргументами или комплементами каких-то составляющих, а “вставками” в структуру непосредственных составляющих. См. [7, с. 563].

[17] В работе [39] предлагается теория передвижения подлежащего (S) из (исходной) позиции SpecVP (в схемах (13а-б) в п. 2.3.1 – из позиции $Spec_1VP$ “многослойной” VP) в позицию SpecIP. SpecIP – позиция подлежащего; IP – более современное название для S из схемы (3а) в п. 2.1. Подлежащее в (3а) – левостороннее зависимое S, что в схеме с IP соответствовало бы SpecIP – ср. структуру VP с позицией $Spec_{1/2}VP$, ветвящейся влево.

Введение такого передвижения учитывает то, что подлежащее – аргумент глагола, и ему в позиции $Spec_1VP$ присваивается семантическая роль – например, ‘агенс’ или ‘экспериенцер’. После того, как подлежащее передвигается в позицию SpecIP, ему присваивается именительный падеж. Данный анализ получил название VP-internal hypothesis (гипотеза о порождении подлежащего внутри VP). Эту гипотезу можно применить к корейскому – в корейском именительный падеж часто присваивается подлежащим деепричастных оборотов или инфинитивов, которые можно рассматривать как “усеченные”/ малые клаузы, в которых нет S/ IP. По поводу именительного падежа в корейском см. также сноску 14.

Данная особенность порядка слов связана с тем, что в языках типа SVO и SOV "фокус эмпатии", т. е. именная группа, сочетающая ряд определенных приоритетных признаков подлежащего, таких, как топикальность/тематичность, определенность, одушевленность и т. д. (см. [40, 41, 7, с. 342–343]), находится на левой периферии предложения. В корейском проблема определения подлежащего и его структурной позиции затрудняется наличием отдельного показателя топика (*-(n)un*), так что в корейском можно скорее говорить о двух основных квазиподлежащих – именной группы в именительном (N-*i/-ka*) и именной группы – топика (N-*(n)un*). См. работы [42, 22, с. 70–90] и др. Помимо того, что показатель топика *-(n)un* может присоединяться к NP-подлежащему вместо показателя именительного *-i/-ka*, возможны предложения с последовательностью "NP-ТОП + NP-ИМ" (22а) – широко известные "предложения с двумя подлежащими" (22б):

(22) а. Waikhikhi-**nun** kyengchi-**ka** coh-ta [12, с. 290]
Ваикики-ТОП пейзаж-ИМ хороший-ИЗЪЯВ
"В Ваикики хороший пейзаж" ('Ваикики характеризуется хорошим пейзажем')

б. Chelswu-**ka** Mia-**ka** mwusep-ta
Чольсу-**им** Миа-**им** страшно-ИЗЪЯВ
"(Именно) Чольсу боится Миа" ('Миа страшна (именно) для Чольсу')

В целом, именная группа в **именительном** характеризуется основными **грамматическими** признаками подлежащего. Это именительный падеж (который также присваивается подлежащему пассива), возможность контролировать гоноративное согласование, сочинительное сокращение. **Топик** в большей степени характеризуется теми признаками подлежащего (в типологическом смысле), которые связаны с его **выделенностью в дискурсе и тематичностью**[18]. Это начальная позиция, определенный статус именной группы – топика, контроль текстовой анафоры. Интересно, что именно группа с показателем топика в конструкциях типа (22а) подвергается синтаксическому подъему[19].

С формальной точки зрения, если вводить отдельную позицию в структуре непосредственных составляющих для составляющей-топика, основная (базовая) структура предложения в корейском будет более сложной, чем, например, в английском. В эту структуру придется включить как минимум некоторые позиции так называемой "левой периферии" (Left Periphery) по работе [43]: специальную позицию для составляющей-топика (Top(ic)P(hrase)). Такой вариант предлагает Ч. Ли ([44] и другие работы). Альтернативный подход, в большей степени ориентированный на европейские языки, предполагает, что топик не входит в базовую структуру предложения, а является адъюнктом.

3. ВОЗМОЖНЫЕ ИЗМЕНЕНИЯ ПОРЯДКА СЛОВ

Как уже говорилось в п. 2.2, одна из задач синтаксиса и типологии – выявление основного, "базового" порядка слов в конкретных языках. Есть определенные типы конструкций, в которых базовый порядок слов нарушается. Помимо "локальных" изменений порядка слов (см. п. 2.2, пример (12а-б)), во многих языках есть специальные конструкции, в которых какое-то отклонение от базового порядка слов грамматикализовано. Это, например, пассивная конструкция (п. 2.3.1), общий и частный вопрос в английском и других европейских языках, выдвижение в начало предложения (топикализация) и др. Согласно стандартной версии генеративной грамматики, такие конструкции описываются через так называемые "передвижения" (до некоторой степени соответствующие "трансформациям" в более ранней терминологии, см. [7, с. 574–586]). Примеры передвижений в русском вопросе показаны в (23а-б).

Важная часть теории передвижений состоит в определении синтаксических ограничений на те или иные передвижения – выявление контекстов (так называемых "островов", см. [7, с. 543–552]), в которых эти передвижения невозможны – ср. (23а) и (24а-б). В (24а-б) показано, что частный вопрос в европейских языках нельзя поставить к дополнению таких типов придаточных предложений, как обстоятельственные (24а) или относительные (24б). Вопросительное слово в (24а-б) должно было бы передвинуться в начало предложение из "острова".

(23) а. Иван купил **книгу**
→ **Что**$_1$ Иван купил t_1? [частный вопрос]

б. Иван **купил** книгу
→ **Купил**$_1$ ли Иван t_1книгу? [общий вопрос]

(24) а. Иван опоздал, так как **что-то** потерял →
*__Что__$_1$ Иван опоздал, так как потерял t_1?
[подр. "… так как потерял это$_1$", "остров адъюнктной (обстоятельственной) группы"]

б. Я видел девочку, которая **что-то** искала →
*__Что__$_1$ я видел девочку, которая искала t_1?
[подр. "…которая искала это$_1$", "остров сложной именной группы" – относительное предложение]

Рассмотрим, как можно применить аппарат передвижений и синтаксических "островов" к корейскому.

3.1. Построение вопроса

В китайском, японском, корейском и многих других языках трансформация вопроса отсутствует, т. е. вопросительное слово не выносится вперед при частном вопросе, ср. (25). При этом частный вопрос к члену придаточного обстоятельственного и относительного предложения (так называемых "островов для передвижения") возможен, как показано в (26а-б) – ср. (24а-б).

[18] Важно заметить, что показатель топика не только именной: например, он может свободно присоединяться к наречиям и другим обстоятельственным группам. Если учитывать этот факт, морфологическим признаком грамматического подлежащего все же следует считать показатель именительного, а не топика.

[19] В придаточном предложении (откуда происходит подъем) показатель *-(n)un* (ТОП) обычно заменяется на *-i/-ka* (ИМ). Так, если бы (22а) было придаточным предложением, в нем была бы конструкция с "двумя именительными": *Waikhikhi-ka* вместо *Waikhikhi-nun*.

(25) Ne-nun **nwukwu-lul** cohaha-ni? [20, с. 84–85]
ты-ТОП кто-ВИН любить-ВОПР
"**Кого** ты любишь?" [букв., разг. "Ты **кого** любишь?"] (другая возможная интерпретация: "Ты любишь кого-нибудь?")

(26) а. Minswu-nun [$_{CP}$ Senhi-ka **mwues$_1$-ul** [45, с. 51]
Минсу-ТОП Сони-ИМ что-ВИН
sa-se] hwakana-ss-ni?
купить-ДЕЕПР.СЛ огорчиться-ПРОШ/ПФ-ВОПР
"**Что$_1$** это, Минсу огорчился, так как Сони (это$_1$) купила?"; [букв. "*Минсу огорчилась, так как Сони **что$_1$** купила?"]; "остров адъюнктной (обстоятельственной) группы" (другая возможная интерпретация: "Минсу огорчилась, так как Сони что-то купила")

б. Minswu-ka [$_{NP}$[$_{RC}$ **nwukwu$_1$-ka** ssu-n] [45, с. 51]
Минсу-ИМ кто-ИМ писать-ПРИЧ
chayk-ul] sa-ss-ni?
книга-ВИН купить-ПРОШ-ВОПР
"**Кто$_1$** это, что Минсу$_2$ купил книгу, которую он$_1$ написал?" [букв. "*Минсу купил книгу, которую **кто$_1$** написал?"]; "остров сложной именной группы" (другая возможная интерпретация: "Минсу купил книгу, которую кто-то написал")

Можно ли говорить о трансформации вопроса в корейском, аналогичной этой трансформации в европейских языках? Видимо, аналогичной трансформации нет, поскольку нейтральным при вопросе является такой порядок слов, при котором вопросительное слово не выносится в начало предложения[20], что, вероятно, связано с возможностью интерпретации местоимения (*nwukwu-lul* 'кто-ВИН' в (25)) не только как вопросительного, но и как неопределенного.

C.–T.–J. Huang в работе [46] предлагает считать, что образование вопроса в китайском (и корейском/японском) происходит не на синтаксическом, а на логическом уровне (так называемый уровень логической формы (Logical Form))[21]. В русском подобная стратегия образования вопроса невозможна (см. сноску 20). Например, хотя предложения (24а-б) с нарушением "островных" ограничений грамматически неправильны, возвращение вопросительного слова в исходную позицию в (27а-б) также неправильно. Значит, вопрос в европейских языках и в корейском (японском, китайском) строится по разным синтаксическим правилам.

(27) а. *Иван опоздал, так как **ЧТО$_1$** потерял?
[только в контексте переспроса]
б. *Я видел девочку, которая **ЧТО$_1$** искала?
[только в контексте переспроса]

Основные "островные" ограничения вырабатывались в большой степени на основе трансформации вопроса. Раз в корейском (японском, китайском) нет трансформации вопроса, действуют ли в этих языках "островные" ограничения? Данной проблеме посвящено большое количество исследований, и разные исследователи приходят к различным выводам.

3.2. Топикализация

Топикализация в стандартной терминологии – вынос вперед слова, которое при базовом порядке слов не должно стоять в начале предложения, однако стоит в начале предложения в связи с определенной коммуникативной нагрузкой этой группы. Например, именная группа *эту книгу* в (28а) стоит в начале предложения, так является темой текста/разговора:

(28) а. **Эту книгу$_1$**, я знаю, что ты мне когда-то подарил **t_1/Ø$_1$**
б. ← Я знаю, что ты мне когда-то подарил **эту книгу$_{TOP}$**

Можно ли рассматривать такой вынос вперед, связанный с коммуникативными факторами[22], передвижением из исходной грамматической позиции **t_1** в (28б)? Или коммуникативные факторы предопределяют порядок слов наравне с грамматическими, так что *эту книгу* порождается в начальной позиции, а исходная позиция пустая или заполнена нулевым местоимением, кореферентным вынесенной группе (**Ø$_1$** вместо **t_1**)?

Ответ на этот вопрос зависит от того, можно ли найти ограничения на начальную позицию именной группы при топикализации, аналогичные "островным" ограничениям (см. (24а-б)).

Оказывается, что в одних языках такие ограничения обнаруживаются, а в других ограничений или вообще нет, или эти ограничения действуют лишь частично. С этим связано обилие точек зрения на топикализацию, представленное в сборнике [6]. Так, на основе материала немецкого языка Г. Мюллер и В. Штернефельд делают вывод, что вынесение слова в начало предложения (топикализация) – передвижение такого же типа, как передвижение вопросительного слова при образовании частного вопроса. Дж. Байер и Дж. Корнфильт, также на материале немецкого, показывают, что топикализацию нельзя рас-

[20] В корейском вынос вопросительного слова в начало предложения в возможен в особых контекстах как топикализация (п. 3.2), но не обязателен. Согласно данным Н. А. Зевахиной (дискуссия), в придаточном частном вопросе начальная позиция вопросительного слова намного более частотна, чем в главном.
В языках, в которых постулируется трансформация вопроса, вынос вопросительного слова в начало предложения или обязателен (английский), или по крайней мере желателен при соблюдении литературной нормы. Ср. рус. разг. *Ты **что** сказал?* с постановкой ***что*** перед глаголом и едва ли возможное только в контексте переспроса *Ты сказал… **ЧТО**?* – см. (27а-б). В последнем примере ***что*** стоит в позиции после глагола, наиболее нейтральной при базовом порядке слов SVO, который нейтрален для русского языка.

[21] Согласно работе [47, с. 189], вопрос к невынесенному вперед адъюнкту/обстоятельству, стоящему внутри "острова" сложной именной группы, невозможен (в отличие от вопроса к аргументу) – это говорит в пользу возможности постулирования "абстрактного" передвижения вопросительного слова на уровне логической формы, на которое накладываются хоть какие-то ограничения. Однако в работе [48, с. 175–176] приводятся почти аналогичные примеры, в которых релятивизация адъюнкта/обстоятельства из того же острова возможна. Существование "абстрактного" передвижения в целом труднее обосновать, чем существование синтаксического передвижения – так что эта проблема остается в стадии дискуссии.

[22] Здесь может иметься в виду как статус (контрастного) топика, так и (контрастного) фокуса. Мы употребляем термин "топикализация" в смысле "вынесение в начало предложения". По теории Л. Рицци [43], и NP-топик, и NP-фокус при передвижении передвигаются в начальную позицию (позицию на "левой периферии" (Left Periphery)).

сматривать как аналогичное образованию частного вопроса передвижение[23].

В работе С.-Х. Ли [48] термин "топикализация" понимается несколько по-другому, с учетом того, что в корейском есть морфологический показатель топика. Топикализация по [48] – вынесение именной группы в начало предложения и замена исходного грамматического показателя при этой группе на показатель топика (ср. п. 2.4). В этом случае, в соответствии с материалом из работы [48], передвижения нет.

(29) а. **[I chayk-un]$_1$** [ai-tul-i [48, с. 79]
этот книга-ТОП ребенок-МНОЖ-ИМ
[Meyli-ka [Inho-ka **Ø$_1$**/*$\mathbf{t_1}$ sse-ss-ta-nun]
Мэри-ИМ Инхо-ИМ написать-ПРОШ-ИЗЪЯВ-ПРИЧ
somwun-ul phettuly-ess-ta-ko]
слух-ВИН распространить-ПРОШ-ИЗЪЯВ-ЦТ
cwucangha-n-ta
утверждать-НАСТ-ИЗЪЯВ
"Эта книга$_1$… дети утверждают, что Мэри распространила слух, что (ее$_1$) написал Инхо"

б. **[Con-un]$_1$** [ai-tul-i [sensayng-nim-i
Джон-ТОП ребенок-МНОЖ-ИМ учитель-ГОН-ИМ
[ku-lul]$_1$ miweha-n-ta-ko] sayngkakha-n-ta
он-ВИН ненавидеть-НАСТ-ИЗЪЯВ-ЦТдумать-НАСТ-ИЗЪЯВ
"Джон$_1$… дети думают, что учитель его$_1$ ненавидит"

В примере (29а) нарушается "островное" ограничение "сложной именной группы"[24] (*…somwun-ul phettuly-ess-ta…* "распространила слух, что…"). Кроме того, если бы именная группа – топик [*i chayk-un*]$_1$ 'книга-ТОП' все-таки была бы передвинута из исходной позиции $\mathbf{t_1}$, топик выдвигался бы из трех простых предложений сразу ("дети утверждают…", "Мэри распространила слух…", "Инхо написал…"), что даже в английском встречается довольно редко.

Такое "длинное" (дистантное) передвижение с нарушением островного ограничения в (29а), в соответствии с правилами передвижения, должно быть невозможно. Поэтому С.-Х. Ли считает, что [*i chayk-un*]$_1$ 'книга-ТОП' порождается в специальной дополнительной группе топика TopP (см. п. 2.4). В позиции аргумента внутри глагольной группы, соответствующего именной группе – топику **[*i chayk-un*]$_1$** 'книга-ТОП' из позиции SpecTopP, стоит не "след" передвижения $\mathbf{t_1}$, а нулевое местоимение **PRO/Ø$_1$** с тем же индексом (кореферентное **[*i chayk-un*]$_1$**). В (29б) показано, что в этой позиции возможно не только нулевое, но и указательное/личное местоимение в винительном падеже (***ku-lul*** 'тот/он-ВИН'), что недопустимо исходя из определения передвижения[25].

Для сравнения приводится как пример начальная позиция именной группы, сохраняющая при этом свои исходные морфологические характеристики (падежный показатель, а не показатель топика) – см. (30а-б). По мнению С.-Х. Ли, примеры (30а-б) грамматически неправильны, что доказывает, что именные группы ***i chayk-ul*** 'этот книга-ВИН' и ***Con-ul*** 'Джон-ВИН' обе передвигаются из позиций дополнений глаголов *sse-ta* 'написать' и *miweha-ta* 'ненавидеть' (позиция "следа" $\mathbf{t_1}$ в (30а); позиция, в которой стоит не след $\mathbf{t_1}$, а местоимение ***ku-lul*$_1$** 'он-ВИН', в (30б)) – ср. (29б). В примере (30а) передвижение нарушает ограничение "сложной именной группы" (см. сноску 24); оба примера (30а-б) грамматически неправильны.

(30) а.?*[**I chayk-ul**]$_1$ [ai-tul-i [48, с. 79]
этот книга-ВИН ребенок-МНОЖ-ИМ
[Meyli-ka [Inho-ka $\mathbf{t_1}$/*Ø$_1$ sse-ss-ta-nun]
Мэри-ИМ Инхо-ИМ написать-ПРОШ-ИЗЪЯВ-ПРИЧ
somwun-ul phettuly-ess-ta-ko]
слух-ВИН распространить-ПРОШ-ИЗЪЯВ-ЦТ
cwucangha-n-ta
утверждать-НАСТ-ИЗЪЯВ
"Дети утверждают, что Мэри распространила слух, что ЭТУ КНИГУ$_1$ написал Инхо" [букв. *"Эту книгу$_1$, дети утверждают, что Мэри распространила слух, что написал Инхо"]

б. *[**Con-ul**]$_1$ [ai-tul-i [sensayng-nim-i
Джон-ВИН ребенок-МНОЖ-ИМ учитель-ГОН-ИМ
[ku-lul]$_1$ miweha-n-ta-ko]
он-ВИН ненавидеть-НАСТ-ИЗЪЯВ-ЦТ
sayngkakha-n-ta
думать-НАСТ-ИЗЪЯВ
"Дети думают, что учитель ненавидит ДЖОНА$_1$"]
[букв. "Джона$_1$… дети думают, что учитель ненавидит"]

Наши данные в примерах (31а-б) более-менее подтверждают данные С.-Х. Ли, хотя наши суждения о степени грамматичности менее строгие: так, с нашей точки зрения, начальная позиция ***ku chayk-ul*** в (31а) более приемлема, чем у ***i chayk-ul*** в аналогичном примере (30а)[26]. В (31б) при добавлении к ограниче-

[23] Ср. [7, с. 643–644], где дается более общее объяснение проблемы вынесения именной группы в начало предложения и свободного порядка слов.

[24] Ограничение "сложной именной группы" считается наиболее строгим из всех "островных" ограничений. Поэтому считается, что для постулирования передвижения хотя бы это ограничение должно соблюдаться. "Сложная именная группа" – относительное придаточное с внешней вершиной (см. (24б), (26б), (31а-б)) или сентенциальное дополнение существительного (*…слух, что…* в (29а) и (30а)). У этих двух конструкций почти одинаковая синтаксическая структура – имя, модификатором или комплементом которого является предложение. Ограничение "сложной именной группы" состоит в том, что передвижение из клаузы (сентенциальной структуры), некоторым образом синтаксически зависимой от именной группы запрещается.

[25] При определении передвижения на английском материале за основу брался постулат, что "след" **t** не может быть ненулевым, то есть "след" не может быть местоимением. Поэтому как только в аргументной позиции (из которой могло бы произойти передвижение) появляется местоимение, передвижение исключено и считается, что NP, стоящая в начале предложения, там и порождается. Как русский пример приведем предложение, в котором в аргументной позиции стоит местоимение (ср. (30б)). Если бы именная группа, стоящая в начале предложения, очутилась там в результате передвижения из аргументной позиции, было бы нарушено островное ограничение:
(i) **Иван$_1$**, [я знаю человека,[которого *(**он$_1$**) ненавидит]]

[26] Приемлемость (31а) скорее всего связана с тем, что *kes-ul* 'вещь.СЛУЖ-ВИН' в (31а) – служебное имя, в отличие от *somwun-ul* 'слух-ВИН' в (30а) и от *chayk-ul* 'книга-ВИН' в (31б). *Kes-ul* в результате грамматикализации не является полноценной вершиной причастной клаузы, и поэтому "остров сложной именной группы" в (31а) в некотором смысле "неполноценный"; он не создает строгого ограничения на топикализацию.

нию “сложной именной группы” – передвижению из относительного придаточного – еще одного ограничения (ограничения на вынесение подлежащего из сентенциального дополнения или обстоятельства) грамматичность понижается: вынесение подлежащего ***Mia-ka*** ‘Миа-ИМ’ из относительного придаточного [**t**$_1$ *ku pwun-kkey* **t**$_2$ *sa tuli-n*] “(которую) (она) купила тому человеку” невозможно.

(31) а.$^{?}$[**Ku chayk-ul**]$_2$ [Minho-nun[Mia-ka **t**$_2$
этот книга-ВИН Минхо-ТОП Миа-ИМ
sa-nun] kes-ul
купить-ПРИЧ.НАСТ вещь.СЛУЖ-ВИН
po-ass-ta]
видеть-ПРОШ-ИЗЪЯВ
“ЭТУ КНИГУ$_2$, Минхо видел, как Миа (ее$_2$) покупает” [*ku chayk* – фокус. Ослабленный “остров сложной именной группы”]

б. ***Mia-ka**$_1$, [na-nun[**t**$_1$ ku pwun-kkey **t**$_2$
Миа-ИМ$_1$ я-ТОП тот человек.ГОН-ДАТ.ГОН
sa tuli-n
купить.ДЕЕПР дать.СЛУЖ-ПРИЧ
[chayk-ul]$_2$ ilk-ess-ta]
книга-ВИН читать-ПРОШ-ИЗЪЯВ
“МИА$_1$, я читал книгу, которую (она$_1$) купила этому уважаемому человеку” [*Mia* – фокус. “Остров сложной именной группы”, и также ограничение на вынесение подлежащего из сентенциального дополнения или модификатора]

Таким образом, мы показали, что в корейском языке передвижение (коммуникативно выделенных) именных групп в начало предложения можно постулировать. Корейский накладывает меньше ограничений на подобные передвижения, чем, например, английский или немецкий языки[27]. Однако при наличии наиболее строгих ограничений передвижение все же запрещено. Проблему представляют причины более свободных (но все же присутствующих) ограничений на передвижения в корейском (японском и многих других языках).

3.3. Плавающие кванторы

Еще один тип передвижения, который в корейском можно сравнить с похожей конструкцией в европейских языках – передвижение именной группы, создающее конструкцию с “плавающими кванторами” (floating quantifiers). Плавающий квантор – это такой квантор-модификатор (типа *все, каждый*), который после определенного передвижения стоит не рядом с определяемым существительным, а отделен от него словом или цепочкой слов. В примере (32) показан классический анализ плавающих кванторов в английском языке [39; 7, с. 295].

При том порядке слов, который имеет место в предложении (32а), нарушается то общее правило, что модификатор/определение в английском предшествует существительному и входит с ним в одну составляющую: (*black*) *bird* (**black*) ‘черная птица’; *black bird must (be there)* и **bird must black*/**black must bird (be there)* ‘черная птица должна (быть там)’. В (32а) квантор AdvP *all* ‘все’ следует за именной группой *the boys* ‘мальчики’, а не предшествует *the boys*; кроме того, *the boys* отделяется от *all* словом – вспомогательным глаголом *have* ‘иметь’. Ср. русский пример (32б).

(32) а. [$_{IP}$ [$_{DP}$ ***The boys***]$_1$ ***have*** [$_{VP}$ [$_{AdvP}$ ***all***]
[$_{VP}$[t$_1$[$_{V'}$*seen the film*]]]] (строка 1) ←
← [$_{IP}$ [$_{I'}$ I[*have*] [$_{VP}$ [$_{AdvP}$ *all*] [$_{VP}$ [$_{DP}$ *the boys*]]
[$_{V'}$ *seen the film*]]] (строка 2)
“Мальчики все посмотрели фильм”

б. *За неделю до начала учебы* [$_{DP}$ ***ребята***]$_1$
начинают уже
[$_{VP}$ [$_{AdvP}$ ***все***] [$_{VP}$ [t$_1$ [$_{V'}$ *покупать учебники*]]]

Чтобы сохранить стандартную структуру именной группы (AdvP-DP) в “исходной”, порождаемой форме предложения (см. строку 2 в (32а)), постулируется передвижение. В строке 2 *all* – модификатор (AdvP) – порождается как адъюнкт VP (наречный модификатор). DP *the boys* ‘мальчики’ передвигается из позиции SpecVP (второй в VP), где эта DP порождается в строке 2, в позицию SpecIP (первую в IP) в строке 1.

В корейском (и японском) языках [50–52] плавающими кванторами могут быть не только универсальные кванторы (типа *motwu* ‘все’), но и любые числительные в конструкции с классификатором. Нейтральный вариант конструкции с классификатором предполагает постпозицию группы “числительное + классификатор” (*yel kwen(-ul)* ‘десять КЛАСС(-ВИН)’ в (33а-в)) по отношению к смысловому существительному (*chayk(-ul)* ‘книга(-ВИН)’ в (33а-в)). Как видно из этих примеров, группе классификатора (ClassP) присваивается падеж (например, винительный), показатель которого присоединяется к смысловому существительному (N) или/и к классификатору (Class)[28].

(33) а. [$_{ClassP}$ ***chayk***$_N$ *yel* ***kwen-ul***$_{Class}$] [20]
книга десять КЛАСС-ВИН
б. [$_{ClassP}$ ***chayk-ul***$_N$ *yel* ***kwen***$_{Class}$]
книга-ВИН десять КЛАСС
в. [$_{ClassP}$ ***chayk-ul***$_N$ *yel* ***kwen-ul***$_{Class}$]
книга-ВИН десять КЛАСС-ВИН
“Десять книг [ВИН]”

В работах, посвященных конструкции с классификатором в японском и корейском языках, упоминается возможность передвижения смыслового существительного в начало предложения (ср. (31а-б)) или просто влево с разрывом группы классификатора[29]. В работе [52] H. Ko предлагает два различных анализа для таких передвижений, ориентированных на модели падежного маркирования (33б) и (33в)[30] – ср. (34а) и (34б). При этом топикализуются, соответственно, смысловое существительное-подлежащее (*haksayng-tul-i* ‘студент-МНОЖ-ИМ’) и смысловое существительное-дополнение (*maykwu-lul* ‘пиво-ВИН’).

[27] Также связь между антецедентом и рефлексивом более свободна в корейском, чем в европейских языках, см. работу [49].

[28] Возможность присоединения падежного показателя к классификатору (*kwen*) в (33а-в) говорит о его именной природе, см. [51–53].

[29] Ср. феномен “свободного порядка слов”, п. 3.2, сноска 23.

[30] Для конструкции типа (33а) модель с плавающим квантором не характерна.

(34) а. ***Haksayng-tul-i*$_1$ sakwa-lul** [52, с. 178]
студент-МНОЖ-ИМ яблоко-ВИН
[t_1 ***sey myeng**(-i)*] *mek-ess-ta*
три КЛАСС-ИМ кушать-ПРОШ-ИЗЪЯВ
“Три студента съели яблоки”

б. ***Maykwu-lul*$_1$ John-i** [52, с. 32]
пиво-ВИН Иван-ИМ
[t_1 ***sey myeng***] *masi-ess-ta*
три КЛАСС пить-ПРОШ-ИЗЪЯВ
“Джон выпил три бутылки пива”

С учетом типологических параллелей[31] и именных свойств классификатора в корейском языке, нам кажется наиболее подходящим не адвербиальный анализ плавающего квантора (как в английском, см. (32а)), а именной. Плавающий квантор рассматривается как именная группа, и одновременно часть более крупной комплексной именной группы, или как вторичный предикат с именными свойствами. Такой анализ подтверждается материалом других алтайских языков, см. анализ плавающих кванторов в татарском языке в работе [54].

4. ВЫВОДЫ

Мы рассмотрели основные параметры структуры предложения в корейском как языке типа SOV, проиллюстрировали базовые структурные свойства глагольной и именной групп, группы послелога. Также были рассмотрены особенности позиций подлежащего и топика в корейском и некоторые нарушения базового порядка слов – топикализация, конструкция с плавающими кванторами.

С одной стороны, в корейском языке строго соблюдается модель SOV, корейский нельзя считать языком со свободным порядком слов. С другой стороны, корейский нельзя считать конфигурационным языком в смысле генеративной грамматики, поскольку многие постулаты, которые касаются порядка слов и допустимых передвижений (и не только их, а, например, правил присваивания прямого/структурного падежа) в корейском языке не выполняются.

СПИСОК ЛИТЕРАТУРЫ

1. Greenberg J. H. Some universals of grammar with particular reference to the order of the meaningful elements // Universals of language / ed. Greenberg, J. H. – Second edition. – Cambridge (Mass) : MIT Press, 1966.

2. Kuno S. The structure of the Japanese language. – Cambrudge (Mass.) : MIT Press, 1973.

3. Comrie B. Language universals and linguistic typology. – Chicago : The University of Chicago Press, 1981.

4. Шкарбан Л. И. Грамматический строй тагальского языка. – М. : Изд. фирма «Восточная литература», 1995.

5. Haegeman L. Introduction to government and binding theory. – Second edition. – Cambridge (Mass.) : Blackwell Publishers, 1994. – P. 95 – 97.

6. Studies on scrambling: Movement and non-movement approaches to free word-order phenomena / eds. Koster J., van Reimsdijk H. – Berlin ; New York : Mouton de Gruyter, 1994.

7. Тестелец Я. Г. Введение в общий синтаксис. – М. : РГГУ, 2001.

8. Cinque G. On the evidence for partial N-movement in the Romance DP // Paths towards universal grammar / eds. Cinque G., Koster J., Pollock J.-Y., Rizzi L., Zanuttini R. – Washington, D.C. : Georgetown University, 1994. – P. 85 – 110.

9. Kayne R. The antisymmetry of syntax. – Cambridge (Mass.) : MIT Press, 1994.

10. Martin S. E. A reference grammar of Korean. A complete guide to the grammar and history of the Korean language. – Rutlend-Tokyo, 1992.

11. Yoon J. H.-S. Nominal, verbal and cross-categorial affixation in Korean // Journal of East Asian Linguistics. – 1995. – Vol. 4, №4. – P. 325 – 356.

12. Sohn Ho-Min. The Korean Language. – Cambridge (Mass.) : Cambridge University Press, 1999.

13. Bailyn J. A configurational approach to Russian free-word order : PhD Dissertation. – Cornell University, 1995.

14. King T. Configuring topic and focus in Russian : PhD Dissertation. – Stanford University, 1993.

15. Kiss K. É. Configurationality in Hungarian. – Budapest : Akadémiai Kiadó, 1987.

16. den Besten H. On the interaction of root transformations and lexical derivative rules // On the formal syntax of Westgermania / ed. Abraham W. – Amsterdam : John Benjamins, 1977. – P. 47 – 131.

17. Lee Y.-S., Santorini B. Towards resolving Webelhuth’s paradox: evidence from German and Korean / eds. Koster J., van Reimsdijk H. – 1994. – P. 257 – 300.

18. Nedjalkov I. Evenki. – London ; New-York : Routledge, 1997.

19. Казакевич О. А., Митрофанова Н. К., Рудницкая Е. Л. Истории жизни автохтонного населения Сибири: публикации глоссированных текстов (публикация 2-я) // Вестник РГГУ. Сер. «Языкознание». – 2009. – Т. 11, № 6 – С. 286 – 323.

20. Chang S.-J. Korean. – Amsterdam ; Philadelphia : John Benjamins, 1996.

21. Кибрик А. Е. Типы базисных конструкций предложения // Константы и переменные языка. – СПб. : Алетейя, 2003. – С. 126 – 132.

22. Рудницкая Е. Л. Спорные вопросы корейской грамматики: теоретические проблемы и методы их решения. – М. : Восточная литература, 2010.

[31] См. работу [53], в которой рассматривается конструкция с плавающими кванторами именного типа в немецком языке.

23. Hoji H. On the syntactic properties of passive morpheme in Japanese // Japanese/Korean linguistics / eds. Choi S., Clancy P. – 1993. – Vol. 3. – P. 137 – 153.

24. Maling J., Kim S.-W. Case assignment in the inaliable possession construction in Korean // Journal of East Asian linguistics. – 1992. – Vol. 1, №1. – P. 37 – 68.

25. Oshima D. Y. Adversity and Korean/Japanese passives: a constructional analogy // Journal of East Asian Linguistics. – 2006. – Vol. 15, №2. – P. 137 – 166.

26. Gerdts D. B., Jhang S.-E. Case mismatches in Korean // Harvard Studies in Korean Linguistics VI: Proceedings of the 1995 Harvard International Symposium on Korean Linguistics / eds. Kuno S. et al. – Seoul : Hanshin Publishing Company, 1995. – P. 187 – 202.

27. Keenan E. L., Comrie B. Noun phrase accessability and universal grammar // Linguistic inquiry. – 1977. – Vol. 8, №1. – P. 63 – 99.

28. Mirto I. M. The syntax of the meronymic constructions. – Pisa : ETS, 1998.

29. External possession / eds. Payne D. L., Barshi I. – Amsterdam ; Philadelphia, 1999.

30. Lapointe S. G. Comments on Cho and Sells. "A lexical account of inflectional suffixes in Korean" // Journal of East Asian Linguistics. – 1996. – Vol. 5. – P. 73 – 100.

31. Рудницкая Е. Л. Применимость формализма генеративной грамматики к языкам «с выдвижением топика и подлежащего» на примере корейского // Теоретическая и прикладная лингвистика: пути развития (к 100-летию со дня рождения В. А. Звегинцева) : материалы конференции / сост.: А. Е. Кибрик, А. В. Архипов, А. А. Бонч-Осмоловская, И. М. Кобозева, О. Ф. Кривнова, А. И. Кузнецова. – М. : Изд-во Моск. ун-та, 2010. – С. 127 – 130.

32. Yoo E.-J. Case marking in auxiliary verb constructions // The Proceedings of the 9th International Conference on HPSG / eds. Kim J.-B., Wechsler S. – Stanford University, 2003.

33. Yoon J. Non-morphological determination of nominal particle ordering in Korean // Clitic and Affix Combinations: Theoretical Perspectives / eds. Heggie L., Ordonez F. – John Benjamins, 2005. – P. 239 – 282.

34. Choi-Jonin I. Particles and postpositions in Korean // Adpositions. Pragmatic, semantic and syntactic perspectives / eds. Kurzon D., Adler S. ; University of Haifa. – 2008. – Vol. 74. – P. 133 – 170. – (Typological Studies of Language).

35. Холодович А. А. Очерк грамматики корейского языка. – М. : Изд-во лит-ры на ин. яз., 1954.

36. Cho, Young-Me Y., Sells P. A lexical account of inflectional suffixes in Korean // Journal of East Asian Linguistics. – 1995. – Vol. 4, №2. – P. 119 – 174.

37. Никольский Л. Б. Служебные слова в корейском языке. – М. : Изд-во вост. лит., 1962.

38. Lee J.-E. Korean adverbs // Working papers in linguistics / Univ. of Venice. – 1999. – Vol. 9, №1–2.

39. Sportiche D. A theory of floating quantifiers and its corollaries for constituent structure // Linguistic inquiry. – 1988. – Vol. 19. – P. 425 – 449.

40. Keenan E. Towards a universal definition of "subject" // Subject and Topic / ed. Ch. Li. – New York, 1976. – P. 303 – 333. (Перевод на рус. яз. см. в сб.: Новое в лингвистике. 1982. №11.)

41. Li Ch. N., Thompson S. A. Subject and Topic: A New Typology of Language // Subject and Topic / ed. Ch. N. Li. – New York : Academic Press, 1976. – P. 457 – 489.

42. Yoon J. H. The distribution of subject properties in multiple subject construction // Proceeding of the Japanese-Korean linguistics conference / eds Takubo Y., Kinuhata T., Grzelak Sz., Nagai K. ; Univ. of Chicago. – 2009. – Vol. 3.

43. Rizzi L. The fine structure of the left periphery // Elements of grammar / ed. Haegeman L. – Dordrecht, 1997. – P. 281 – 337.

44. Lee Ch. Contrastive topic and proposition structure // Asymmetry in grammar / ed. Di Sciullo A.-M. – Amsterdam, 2002.

45. Shin J.-Y. Wh-constructions in Korean: a lexical account // Toronto working papers in linguistics. – 2005. – Vol. 25. – P. 48 – 57.

46. Huang C.-T.-J. Logical relations in Chinese and the theory of grammar : PhD. – MIT, 1982.

47. Park M.-K., Sohn K.-W. Floating quantifiers, scrambling and the ECP // Proceedings of the Japanese/Korean Linguistics Conference / eds. S. Choi, P. Clancy ; CSLI Publications. – 1993. – №3. – P. 187 – 203.

48. Lee S.-H. A lexical analysis of select unbounded dependency constructions in Korean : PhD Dissertation. – The Ohio Styete University, 2004.

49. Sells P. Aspects of logophoricity // Linguistic inquiry. – 1987. – Vol. 18, №3. – P. 445 – 479.

50. Muromatsu K. On the syntax of Classifiers : PhD Dissertation. – Univ. of Maryland, 1998.

51. Park M.-K., Sohn K.-W. Floating quantifiers, scrambling and the ECP // Proceedings of the Japanese/Korean Linguistics Conference / eds. S. Choi, P. Clancy ; CSLI Publications. – 1993. – №3. – P. 187 – 203.

52. Ko H.-J. Syntactic edges and linearization : PhD Dissertation. – MIT, 2005.

53. Merchant J. Object scrambling and quantifier float in German // Proceedings of NELS 27 / GSLA, UMass Amherst. – 1996. – P. 179 – 193.

54. Grahchenkov P. Floating quantifiers in Tatar // Proceedings of WAFL-3 (Investigations into Formal Altaic Linguistics) / ed. S. Tatevosov. – Moscow : MSU, 2009. – P. 193 – 207.

ПРИЛОЖЕНИЕ

Глоссы и сокращения

АРТ — артикль
БУД — будущее время
ВЕЖЛ — показатель вежливого регистра
ВОПР — вопросительное наклонение
ВСПОМ — вспомогательный (глагол)
ВИН — винительный падеж
ГОН — показатель гоноратива
ДАТ — дательный пажеж
ДЕЕПР — показатель деепричастия
ЕД — единственное число
ЖЕН — женский род
ИЗЪЯВ — изъявительное наклонение
ИМ — именительный падеж
ИНСТР — инструментальный падеж
ИНФ — инфинитив
КЛАСС — классификатор
МЕСТ — местный падеж
МНОЖ — множественное число
НАПР — направительно-местный падеж
НАСТ — настоящее время
ОДНОВР — одновременность
ПАСС— показатель пассива
ПФ — перфектив
ПОВ — повелительное наклонение
ПОСС — поссессивный
ПРЕДЛ — предлог
ПРИГЛ — пригласительное наклонение
ПРИЛ — прилагательное
ПРИЧ — причастие
ПРОШ — прошедшее время
РОД — родительный падеж
СЛ — (деепричастие) следствия
СЛУЖ — служебное (существительное)
СОВМ — совместный падеж
ТОП — показатель топика
ЦТ — показатель цитатива
Adv — наречие
Cop — позиция связки в удлиненной словоформе
Delim — позиция ограничительной частицы в именной словоформе
DITRANS — дитранзитивный (глагол)
DO — прямое дополнение
Gen — генитив/генитивное дополнение
Hon — позиция гоноратива
IO — косвенный объект
Mod — модификатор
O — объект (дополнение)
Plur — позиция показателя множественного числа
Post — позиция показателя косвенного падежа
PRO/Ø — нулевое местоимение-аргумент
Rel — относительное придаточное предложение
S — субъект (подлежащее), см. также IP ниже
Stem — позиция корня/основы
t — “след” (trace), обозначающий исходную позицию (аргумента) при передвижении
1 — первое лицо
3 — третье лицо
букв. — буквально
подр. — подразумеваемое

Обозначения составляющих и их компонентов

Adv — вершина группы наречия
AdvP — группа наречия
C — вершина группы союза (комплементайзера, complementizer)
CP — группа союза (комплементайзера, complementizer)
Class — вершина группы классификатора
ClassP — группа классификатора
D — вершина группы детерминатора
DP — группа детерминатора
I — вершина группы флексии
IP — группа флексии (вся клауза, также обозначается как S)
N — вершина именной группы
NP — именная группа
Part — вершина группы частицы
PartP — группа частицы
Part’ — промежуточный уровень группы частицы (штриховая группа)
Pr — вершина предложной группы
PrP — предложная группа
Post — вершина послеложной группы
PostP — послеложная группа
Spec — спецификатор (любой группы)
SpecNP — спецификатор именной группы
SUB — составляющая-субъект (малой) клаузы
Top — вершина группы топика
TopP — группа топика
V — глагол (и вершина группы глагола)
VP — глагольная группа
V’ — промежуточный уровень глагольной группы (штриховая группа)

Материал поступил в редакцию 13.03.12.

Сведения об авторе

РУДНИЦКАЯ Елена Леонидовна – кандидат филологических наук, старший научный сотрудник Института востоковедения РАН, Москва

E-mail: erudnitskaya@gmail.com

Москва, ВИНИТИ РАН
28–30 ноября 2012 года

***8-я Международная конференция* «НТИ-2012»,**
посвященная 60-летию ВИНИТИ РАН

«АКТУАЛЬНЫЕ ПРОБЛЕМЫ ИНФОРМАЦИОННОГО ОБЕСПЕЧЕНИЯ НАУКИ, АНАЛИТИЧЕСКОЙ И ИННОВАЦИОННОЙ ДЕЯТЕЛЬНОСТИ»

Для участия **в «НТИ-2012»** приглашаются специалисты в области информационных технологий, телекоммуникаций, создатели и потребители информационных продуктов и услуг, ученые и специалисты РАН, вузовской и отраслевой науки, работники информационных центров и библиотек, служб распространения информационных продуктов и услуг.

Доклады или тезисы докладов направлять в Оргкомитет конференции «НТИ-2012», которые будут опубликованы в специальном сборнике.

Планируется проведение пленарных заседаний, круглых столов и работа по секциям.

Адрес: Россия, 125190, Москва, ул. Усиевича, 20, Всероссийский институт научной и технической информации (ВИНИТИ РАН), ОНИИР, Оргкомитет **«НТИ-2012».**
Тел: (495) 155-44-22, 155-44-29, 152-64-41,
Факс: (495) 152-54-92, 943-00-60
E-mail: conf@viniti.ru , market@viniti.ru
http://www.viniti.ru

БАЗА ДАННЫХ ВИНИТИ РАН

ВИНИТИ предлагает к использованию через WWW-сервер (**http://www.viniti.ru**) крупнейшую Федеральную базу отечественных и зарубежных публикаций по естественным, точным и техническим наукам. БД ВИНИТИ РАН генерируется с 1981 г., обновляется ежемесячно, пополнение составляет около 1 млн документов в год. БД ВИНИТИ представлена ретроспективными тематическими фрагментами и единой политематической БД (ретроспектива с 2001 г.), объединяющей все тематические фрагменты БД ВИНИТИ.

БД ВИНИТИ РАН в сети INTERNET

Сервер ВИНИТИ – http://www.viniti.ru – обеспечивает on-line доступ к Базе данных ВИНИТИ РАН круглосуточно без выходных.

На основе БД ВИНИТИ РАН предоставляются следующие услуги:

- Диалоговый поиск научно-технической информации **в режиме on-line**;
- **Демо-версия**, позволяющая ознакомиться с основными функциями поисковой системы, составом данных, формами представления документов и получить навыки работы с системой;
- **Поисковые эксперты** ВИНИТИ выполнят тематический поиск по разовым или постоянным запросам, а также окажут **консультационные услуги**.

БД ВИНИТИ РАН на CD-ROM

Любые наборы тематических фрагментов БД ВИНИТИ или их разделов могут быть предоставлены **на CD-ROM в поисковой системе (ИПС) "Сокол"**, обеспечивающей все поисковые функции, доступные в режиме on-line:

- Поиск можно вести в годовом или ретроспективном массиве (за несколько лет сразу) в одном или нескольких тематических фрагментах .
- Поиск по словам и любым словосочетаниям из заглавия, реферата, ключевых слов.
- Использование года, языка, рубрик, шифров тематических разделов БД для уточнения поиска.
- Поиск по словарю, выполняющему функции многоаспектного указателя, в том числе авторского, предметного, источников, индексов МПК, номеров патентных документов и депонированных рукописей и т.д.
- Возможность запоминания запросов для последующего использования и/или редактирования их.
- Чтение документов не только как в РЖ (последовательный просмотр документов одного номера за другим), но и чтение документов нужных тематических фрагментов (разделов) по оглавлению за весь период заказанной ретроспективы.

ИПС "Сокол" является прикладной программой Microsoft Windows.

Любые наборы тематических фрагментов БД ВИНИТИ или их разделов могут быть подготовлены **в коммуникативных форматах** ISO-2709, МЕКОФ, txt на любых видах электронных носителей.

Продукты предоставляются на договорной основе.

Информационная служба БД ВИНИТИ: 125190, Москва, ул. Усиевича 20, ВИНИТИ
Телефон: (499) 155-45-01, 155-45-02, **Факс:** (499) 152-62-31 **e-mail:** csbd@viniti.ru