А
П-71
б-ка ИРЯ

ИНСТИТУТ РУССКОГО ЯЗЫКА АН СССР

Проблемная группа по экспериментальной
и прикладной лингвистике

ПРЕДВАРИТЕЛЬНЫЕ ПУБЛИКАЦИИ

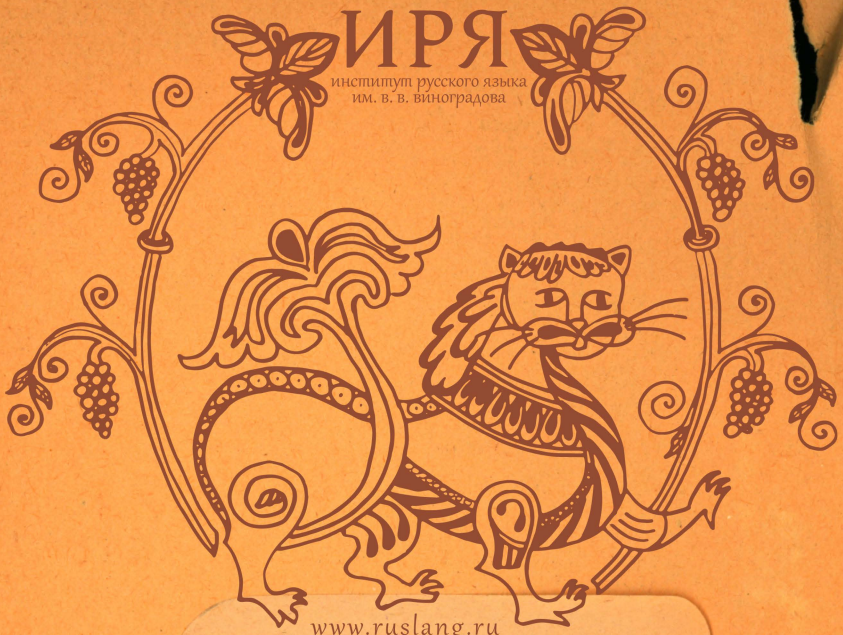
Выпуск 67

И. А. МЕЛЬЧУК

ИССЛЕДОВАНИЯ ПО АВТОМАТИЧЕСКОМУ ПЕРЕВОДУ в 1970—1974 гг.

Москва — 1975

ИРЯ
институт русского языка
им. в. в. виноградова



www.ruslang.ru

70325

ИНСТИТУТ РУССКОГО ЯЗЫКА АН СССР
Проблемная группа по экспериментальной и прикладной лингвистике

ПРЕДВАРИТЕЛЬНЫЕ ПУБЛИКАЦИИ

Выпуск 67

И.А.Мельчук

ИССЛЕДОВАНИЯ ПО АВТОМАТИЧЕСКОМУ ПЕРЕВОДУ В 1970 - 1974 гг.

Москва - 1975

4
П-71 67
БИБЛИОТЕКА
ИН-ТА РУС ОГО ЯЗЫКА
ИН-ТА ЯЗЫКОЗНАНИЯ АН СССР
70325

Рамки обзора: хронологические и предметные

Настоящий обзор был написан в конце 1974 г., чем и определяется верхняя граница рассматриваемого периода. Нижняя же его граница установлена на основе трех следующих соображений:

(i) Привлекаемый отрезок времени не должен быть слишком большим в силу ограниченности объема обзора; 5 лет представляется подходящим сроком.

(ii) В отечественной литературе уже имеются обзоры по автоматическому переводу (в дальнейшем - АП), охватывающие эпоху вплоть до 1970 г.: [61], [46], [63], [80]; см. также [149], [165], [145] и [146]. Кроме того, изданы два тома критико-библиографического справочника по АП, также доведенные до 1970 г.: [70] и [71].

(iii) Желательно, чтобы начало выбранного для обзора отрезка времени более или менее соответствовало началу естественно выделяемого периода в развитии интересующей нас дисциплины. На границе 60-х и 70-х гг. в исследованиях по АП как раз начинается новый период, длящийся и сейчас (см. ниже, стр. 5).

Таким образом, все три соображения указывают на 1970 гг. в качестве нижней временной границы обзора.

Основным материалом обзора послужили работы, прореферированные в соответствующих разделах реферативных журналов "Информатика" и "Математика"; кроме того, в ряде случаев привлечены публикации, не представленные в этих РЖ по случайным причинам. Оказалось, однако, что отразить в нашем обзоре всю литературу, учтенную в обоих РЖ за 1970-1974 гг., невозможно. Мы произвели (достаточно субъективный, по-видимому) отбор, руководствуясь следующими нередко конфликтующими критериями:

- Близость работы по тематике к АП в собственном смысле слова; ряд весьма интересных работ сознательно оставлен вне поля зрения, поскольку они принадлежат скорее к формальной лингвистике, информационному поиску и т.п. Иначе говоря, в обзоре принят достаточно узкий взгляд на АП. Впрочем, вопрос о том, что относится к АП, а что - нет, далеко не прост; см. об этом ниже, стр. 7.

- Научная содержательность работы.

- Представительность работы для определенного направления исследований; мы стремились отразить по мере возможности максимальное количество известных нам подходов, деятельность разных коллективов и т.п.

Читатель этого обзора должен постоянно иметь в виду, что обзор написан лингвистом (ср. [145] - [146]) и с точки зрения лингвистики. Как известно, АП - это научная дисциплина, имеющая по крайней мере три в одинаковой степени существенных аспекта:

а) лингвистический аспект - описание конкретных языков для целей АП и разработка общей теории подобных описаний;

) формально-математический и математико-алгоритмический аспект - создание и исследование формализмов, на основе которых строятся как лингвистические описания и теории, так и описания процедур, непосредственно осуществляющих преобразование текста на одном языке в разнозначный ему текст на другом языке;

в) машинно-математический аспект - реализация систем АП на ЭВМ и разработка теории подобных реализаций. Названные аспекты в действительности тесно переплетены, так что их разделение в высшей степени условно. Тем не менее, имеет смысл утверждение, что настоящее изложение - в силу профессиональной подготовки и интересов автора - сосредоточено преимущественно на лингвистическом аспекте АП.

Обзор рассчитан на подготовленного читателя, знакомого с основными понятиями в сфере АП; хорошо известные факты и термины используются нами без специальных пояснений. Необходимые справки можно навести в [70], [71] и [1].

Укажем, наконец, два ранее опубликованные обзора, в которых рассматриваются многие вопросы, касающиеся состояния АП в интересующий нас здесь период: [184] и [198], а также сборники [1], [101], [104], [124] и [166], посвященные проблеме "ЭВМ и естественный язык" или даже прямо АП; в них можно найти много данных, относящихся непосредственно к нашему обзору.

Немного истории: периодизация АП до 1970 г.

Развитие АП до 1970 г. естественно разбивается на три периода (см., в этой связи, [46] и [43]):

I. 1949 г. (знаменитый меморандум У.Уивера) - январь 1954 г. (Джорджтаунский эксперимент - фактическое рождение АП): зачатие и утробное развитие АП.

II. 1954 г. - конец 50-х гг.: детство и отрочество АП. Это эпоха так называемых систем АП первого поколения: бинарные алгоритмы (перевод с одного конкретного языка на другой с зависимым анализом) без отделения лингвистического описания, т.е. сведений о привлекаемых языках ("грамматики"), от программ, реализующих преобразование текста на одном языке в текст на другом языке (от "механизма"); перевод еще весьма близок к пословному.

III. 60-ые годы: бурная юность АП. Именно в данный период формируются основные положения этой дисциплины. Создаются многочисленные системы АП второго поколения, для которых характерны независимость анализа и синтеза, а также установка на максимальное разделение "грамматики" и "механизма". При таких системах возникает необходимость в формальном представлении перерабатываемых текстов, и в связи с этим (хотя далеко не только в связи с этим) в рамках АП начинают активно осваиваться идеи и методы современной теоретической лингвистики. Перевод в этот период осуществляется, как правило, через сопоставление синтаксических структур фраз входного и выходного языков ("синтаксический перевод"); в центре внимания оказываются проблемы синтаксического анализа и преобразования синтаксических структур.

Современное представление о лингвистическом аспекте АП.

С начала 70-х гг. начинается, как уже указывалось, IV-ый, современный период развития АП - эпоха зрелости. Выдвигаются проекты систем АП третьего поколения, в которых учтен достаточно богатый опыт истекших двадцати лет и которые в то же

время исходят из нового, складывающегося на наших глазах представления об АП как особой научной проблеме. Это представление опирается на три следующих взаимосвязанных положения[1].

1. Принципиальное решение проблемы АП (т.е. АП достаточно высокого качества, рассчитанный на достаточно широкий круг текстов) практически недостижимо без существенного использования семантики ("семантический перевод"); современные системы АП должны уметь обращаться со смыслом текста так же формально, как они обращаются с синтаксическими структурами[2]. Это означает, что необходимо иметь специальное семантическое представление (СемП) текста Т и что анализ/синтез при АП суть операции « Т $\Rightarrow$ СемП(Т) » и « СемП(Т) $\Rightarrow$ Т » соответственно; тогда система АП оказывается формальным способом задания соответствия между текстами и смыслами.

2. Взгляд на естественный язык как на определенное соответствие между (бесконечным) множеством текстов и (бесконечным) множеством смыслов был выдвинут в теоретической лингвистике достаточно давно. Со второй половины 60-х гг. известны попытки реализовать этот взгляд в виде лингвистических моделей типа "Смысл $\Leftrightarrow$ Текст"; см., в частности, [67]. Модель "Смысл $\Leftrightarrow$ Текст" (далее - МСТ) - это сложно организованная система правил, задающая соответствие между текстами некоторого конкретного языка и их смыслами, т.е. всевозможными СемП. Тогда получается, что система АП между языками L_1 и L_2 есть не что иное, как МСТ для L_1 и МСТ для L_2, реализованные на определенной ЭВМ; перевод в таком случае должен идти через СемП входного текста - непосредственно или с необходимыми преобразованиями этого СемП в СемП выходного текста.

Сказанное означает, что задача АП в его лингвистическом аспекте, или, если так можно выразиться, задача АП-лингвисти-

---

[1] Данное представление о природе и задачах современных исследований принимается (или, во всяком случае, полностью принимается) не всеми специалистами - см., например, ниже, стр. . Тем не менее, на наш взгляд, оно является фактически доминирующим, а теоретически - наиболее перспективным.

[2] Этот тезис неоднократно высказывался и ранее; см., например статью [31], сыгравшую важную роль в развитии АП и вообще лингвистики в СССР.

ки и задача "чистой" лингвистики по существу совпадают. В результате получается, что многие лингвистические исследования, выполняемые сейчас в тех или иных коллективах АП, не имеют в себе ничего специфически машинно-переводческого – это просто формальные описания языка в плане соотнесения текстов и смыслов. Более того, чем лучше подобные АП-лингвистические исследования, чем выше их ценность для АП, тем меньше в них машинно-переводческой специфики; эта парадоксальная на первый взгляд особенность объясняется теперь уже практически общепринятым принципом максимального разделения "грамматики" и "механизма". В то же время целый ряд чисто лингвистических работ оказывается имеющим непосредственное отношение к АП; можно утверждать, опять-таки несколько парадоксально, что к АП имеет большее или меньшее отношение любая достаточно формализованная лингвистическая работа, направленная, хотя бы в конечном счете, на описание соответствия между текстами и смыслами. Подобное слияние АП (точнее – АП-лингвистики) и лингвистики на почве семантичности является весьма характерной чертой нынешнего представления об АП.

Что касается отбора работ для нашего обзора, то указанное слияние АП-лингвистики и лингвистики понятным образом усложняет его и делает его еще более субъективным. В РЖ реферируется большое количество публикаций по теоретической и математической лингвистике, формально не связанных с АП. Автору обзора часто приходилось решать, достаточно ли связана с проблематикой АП та или иная работа, только на основе своей собственной интуиции. (Относительно сказанного в пункте 2 см. подробнее [63], стр. 16-20.)

3. Переход от текста к формальному изображению его содержания и обратно, т.е. получение по тексту его СемП а по СемП – соответствующего текста оказывается центральной операцией не только для АП, но и для прочих видов автоматической обработки текстов (АОТ), которая, в свою очередь, связана и с широким кругом задач, известных под названием "Искусственный интеллект" (речевое управление механизмами; решение задач, формулируемых на естественном языке; моделирование различных аспектов разумного поведения; и т.п.), и с проблемой создания так называемых

диалоговых систем (различные ИПС и АСУ, рассчитанные на работу с естественным языком; программирование на естественном языке), и с распознаванием устной речи. Вообще, какие бы манипуляции с содержанием текстов ни ожидались от ЭВМ, она всегда должна уметь прежде всего построить по тексту его СемП, т.е. выполнить анализ текста; любые последующие действия предпринимаются на основе СемП, а окончательные результаты могут снова оформляться в виде легко понятного человеку текста, т.е. ЭВМ выполняет синтез текста по предварительно построенному СемП. Тем самым, АП перестает быть особой, достаточно автономной задачей, стоящей как бы наравне с прочими задачами АОТ; он оказывается первоочередной и притом, возможно, наиболее сложной подзадачей во всех этих последних. В результате возникает естественное стремление – создать вместо конкретной системы АП сразу нечто более общее – систему, которую можно было бы назвать "языковым процессором" и которая обеспечивала бы соответствие между текстом и формальным изображением его содержания, т.е. между текстом и его СемП. А подобная система есть не что иное, как программно реализованная модель "Смысл $\Leftrightarrow$ Текст", о которой говорилось выше.

Неслучайно поэтому, что львиная доля исследований в различных сферах АОТ – во всяком случае, в той части, где они касаются соответствия "тексты $\Leftrightarrow$ смыслы", – практически неотличимы от исследований по АП. (Это опять-таки осложняет выделение в массе публикаций по проблемам типа "язык и ЭВМ" работ по собственно АП.)

Итак, три главные черты современных представлений об АП, на наш взгляд, таковы:

– признание центральной (а возможно – и решающей) роли семантики в АП;

– трактовка системы АП как машинной реализации лингвистической системы "Текст $\Rightarrow$ Смысл $\Rightarrow$ Текст";

– включение АП в более широкие рамки АОТ.

NB : принятие указанного представления об АП еще не означает его успешной реализации в разрабатываемых ныне переводных системах. Так, даже среди систем АП третьего поколения нам неизвестны системы, предполагающие перевод через "настоящее" СемП. Речь здесь пока идет лишь об отчетливой тенденции к этому, о выдви-

жении идеала – и о стремлении хотя бы частично приблизиться к нему. (Впрочем, в некоторых языковых процессорах – см. ниже, стр. 40 – переход от текста к его СемП и обратно уже осуществляется и сейчас, так что "семантический идеал" АП представляется не таким уж безнадежно далеким.)

Иная точка зрения на АП в начале 70-х гг. излагается Р.Г. Пиотровским в статьях [70] и [78], где отстаивается ориентация на достаточно грубый АП со значительной долей постредактирования. Эта ориентация лежит в основе рабочей программы группы "Статистика речи": см. [71], с. 409. Главный упор здесь делается на создании частотных автоматических словарей и упрощенных алгоритмов частичного грамматического анализа; поэтому речь идет о "вероятностном отраслевом МП". Контуры возможной системы соответствующего типа намечены А.В.Зубовым в [35].

Строение обзора.

В соответствии со сказанным, наш обзор организован следующим образом:

1) Прежде всего, кратко характеризуются те известные автору обзора цельные системы АП в собственном смысле слова, работа над которыми была начата до 1970 гг. и продолжалась в 1970-1974 гг.; это, преимущественно, системы "классических" типов (полуторного и второго поколения).

2) Затем описываются некоторые системы "нового" (= "неклассического") типа, в том числе – и находящиеся на стадии проектирования. Сюда относятся не только системы АП как такового, но и системы более общие – языковые процессоры типа "Смысл⇐⇒Текст" (см. выше, стр. 8).

3) Далее говорится о таких системах АОТ (в частности, о некоторых ИПС и так называемых моделях искусственного интеллекта), в которых в качестве составных частей существенным образом используются компоненты "Текст ⇒ Смысл" или "Смысл ⇒ Текст".

4) Наконец, рассматриваются частные проблемы собственно АП, т.е. проблемы перехода от текста к смыслу и наоборот. Эти проблемы группируются в соответствии со схемой, которая стала уже вполне традиционной:

А. Анализ (от текста к смыслу) - морфологический, синтаксический, семантический.

Б. Представление текста (т.е. запись результатов анализа и исходных данных синтеза).

В. Осуществление собственно перевода (результаты анализа текста на входном языке преобразуются в исходные данные синтеза текста на выходном языке).

Г. Синтез (от смысла к тексту) - семантический, синтаксический, морфологический.

Д. Автоматические словари.

Условные сокращения.

В тексте обзора нам придется широко пользоваться сокращениями; приведем здесь перечень основных сокращений, используемых по всему обзору:

А-	- автоматический (например, АСинт = автоматический синтез; и т.д.);
Ан	- анализ;
АОТ	- автоматическая обработка текстов;
АП	- автоматический перевод;
АСУ	- автоматизированная система управления;
Г-	- глубинный/глубинно- (ГМП = глубинно-морфологическое представление; и т.д.);
ИПС	- информационно-поисковая система;
ИЯ	- информационный язык;
КС-	- контекстно-свободная (грамматика);
М-	- морфологический;
МСТ	- модель "Смысл $\Leftrightarrow$ Текст";
О	- отношение (ГСО = глубинно-синтаксическое отношение; и т.д.);
П-	- поверхностный/поверхностно- (не в конечном положении: ПСС = поверхностно-синтаксическая структура);
П	- представление (в конечном положении: ПСП = поверхностно-синтаксическое представление);
С-	- синтаксический (не в конечном положении: СС = синтаксическая структура);
С	- структура (в конечном положении);
Сем	- семантический (СемП = семантическое представление);

Синт – синтез;
ТГ – трансформационная грамматика;
ЭВМ – электронно-вычислительная машина;
ЯП – язык-посредник.

Прочие – "окказиональные" – сокращения вводятся по мере надобности и поясняются непосредственно при появлении.

Системы собственно АП "классических" типов.

Среди систем АП полуторного и второго поколения, над которыми велась работа в рассматриваемый период (системы первого поколения в это время автору неизвестны), необходимо упомянуть по крайней мере следующие четыре:

1) Система Ин-та прикладной математики (ИПМ) АН СССР (Москва).
2) Система Монреальского ун-та (Канада).
3) Система Саарского ун-та (ФРГ).
4) Система Центр. НИИ патентной информации (ЦНИИПИ; Москва).

1) Система французско-русского АП математических текстов "ФР-Д" (О.С.Кулагина) реализована на ЭВМ БЭСМ-4 [44].

В основу системы "ФР-Д" положен синтаксический анализ фильтрового типа, выдающий для входной (французской) фразы дерево зависимостей (или несколько деревьев – в случае синтаксической неоднозначности фразы); подробнее см. ниже, раздел "Синтаксический анализ". Найденное дерево французской фразы подвергается ряду преобразований (см. раздел "Преобразование синтаксических деревьев") для получения допустимых в русском языке конструкций. В течение нескольких последних лет система "ФР-Д" эксплуатируется в экспериментальном режиме; вот образцы переводов, получаемых на ЭВМ с ее помощью:

Revenons à notre equation $\frac{dy}{dx} = f(x, y)$ (E), **f étant rationnelle en y.**	Возвратимся к нашему уравнению $\frac{dy}{dx} = f(x, y)$ (E), когда/, где f рационален (-a, -o) относительно/в y.
Les singularités supérieures, comme nous venons de le voir,	а) Верхние особенности, как/так как мы только что это увидели,

sont fixes, mais, en général, les pôles et les points critiques algébriques sont variables d'une intégrale à l'autre.

фиксированы, но, вообще говоря, полюсы и критические алгебраические точки переменны от интеграла до другого/другой /другого.

б) Верхние особенности, как/так как мы приходим из этого увидеть, фиксированны, но... [дальше - как в варианте а)].

L'intégrale de l'équation (E), devenant égale à $\underset{\sim}{y}_0$ pour $\underset{\sim}{x}=\underset{\sim}{x}_0$, aurait en ce point un point critique algébrique, puisque $\frac{d\underset{\sim}{y}}{dx}$ est infinie pour (x_0, y_0).

Интеграл уравнения (E), становясь равной становящийся равным $\underset{\sim}{y}_0$ для $\underset{\sim}{x} = \underset{\sim}{x}_0$, будет иметь в этой точке алгебраическую критическую точку, так как $\frac{d\underset{\sim}{y}}{dx}$ бесконечен (-а, -о) для (x_0, y_0).

Car nous sommes encore dans le cas du coefficient différentiel devenant infini.

Так как мы имеем еще случай дифференциального коэффициента, становящегося бесконечным.

(Пример взят из [44], с. 59-60.)

При некоторых модификациях (прежде всего, при условии расширения и усовершенствования словаря "ФР-Д" может лечь в основу промышленной системы АП.

Различные подробности относительно "ФР-Д" можно найти в [37], [39], [41] и [45].

2) Система англо-французского АП политических текстов (парламентские выступления, правительственные отчеты и т.п.) разрабатывается в Монреальском ун-те (Р.Киттредж) - во многом на основе идей и принципов Гренобльской системы русско-французского АП (см. ниже, стр. 14). В теоретическом плане Монреальская система - это, видимо, одна из наиболее прогрессивных среди ныне действующих систем АП. Ее наиболее характерные особенности таковы ([188], [74]):

- Максимальная расчлененность (английская часть полностью отделена от французской; лингвистическое описание обоих языков

отделено от собственно процедур машинной обработки; описание языка разбито на независимые уровни).

- Высокая степень формализованности лингвистических данных: сведения об английском и французском языках записываются только посредством формализма так называемых W-грамматик (W-грамматика - это, грубо говоря, КС-грамматика Г плюс множество схем трансформационных правил, в которые вместо вхождения нетерминального символа А можно подставлять терминальные цепочки, выводимые в Г из А); соответствия между синтаксическими структурами английских и французских фраз описываются особым формализмом Q-систем (см. раздел "Преобразование синтаксических деревьев"); и т.п. Это позволило монреальцам построить машинную систему переработки текстов (= Q-преобразователь), полностью не зависимую от содержания лингвистических описаний, а ориентированную только на форму этих последних.

- Использование большого словаря современного типа, куда легко включается богатая семантическая информация и который допускает неограниченное расширение и усовершенствование.

- Принципиальная семантическая направленность. В последнее время монреальцы исследуют возможность использовать в процессе АП семантические графы, в частности - накапливать единый семантический граф текста и пользоваться им при разрешении неоднозначностей (см. ниже, стр. 47). Проблеме построения семантических графов текста посвящены работы Т.Хофманна (см. раздел "Проблемы формального представления текста").

3) Система русско-немецкого АП Саарского ун-та (г. Саарбрюккен, ФРГ; Г.Маас, А.Роткегель, Г.Циммерман) разрабатывается одновременно и в тесном взаимодействии с системой автоматического анализа немецких текстов [174]. Насколько можно судить по имеющимся публикациям, обе системы в своей синтаксической части относятся к полуторному поколению (применяются синтаксические анализаторы достаточно устаревшего типа), а в своей словарной части - ко второму. Это связано с тем, что главный упор здесь делается на автоматический словарь.

4) Система англо-русского АП текстов описаний изобретений к патентам (ЦНИИ патентной информации, Москва; А.Л.Василевский,

Ю.М.Эмдина и др., [11], [12], [96]) относится к системам полуторного поколения; эта система не была реализована на ЭВМ.

Наряду с четырьмя названными системами, следовало бы охарактеризовать и Гренобльскую систему русско-французского АП, лучшую - и по идеям, и по реализации - систему АП второго поколения в мире (см. о ней подробный обзор [62]). Однако первый вариант этой системы был, в основном, закончен к 1970 г.; система была опробована в целом ряде машинных экспериментов (всего было переведено текстов общим объемом свыше 400 тыс. слов), и работа над ее первым вариантом прекратилась. Соответствующие публикации появились до 1970 г., так что Гренобльская система в первом варианте оказалась вне нижней хронологической границы нашего обзора. С начала 70-х гг. в Гренобле ведется работа над вторым вариантом системы, который относится скорее к третьему поколению систем АП, и потому о нем пойдет речь в следующем разделе.

Кроме упомянутых "больших" систем АП, известны (благодаря весьма кратким публикациям) еще семь действующих или разрабатываемых (в 1970-1974 гг.) систем АП:

- Англо-французский АП геологических текстов (Саскачеванский ун-т, Саскатун, Канада; [114]). Это система полуторного (или даже первого) поколения; перевод близок к пословному и нуждается в существенном постредактировании.

- Русско-английский АП общенаучных текстов (Информационный Центр ЕВРАТОМ'а - CETIS - в г. Испра, Италия); система рассчитана на дешевую массовую продукцию очень низкого качества [170].

- Система русско-английского АП SYSTRAN (Управление зарубежной техники ВВС США) [148]. Качество АП характеризуется следующим экспериментом 1971 г.: постредактор изменил в переводе каждую фразу, затронув 35% слов; он работал со скоростью 400 слов/час, тогда как переводчик-человек перевел тот же текст набело со скоростью 450 слов/час.

- Система "промышленного" АП LOGOS I(LOGOS Development Corp., США), выполняющая технические переводы с русского языка на английский и с английского на вьетнамский (см. [4], с. 200; и [186]). В связи с чисто коммерческим характером системы какие-либо точные сведения о ней в печати отсутствуют. Однако заключе-

ние статьи [186] о качестве выдаваемых системой англо-вьетнамских переводов таково: "Сырой текст, выданный машиной (за исключением наименее технических разделов переведенной инструкции), по-видимому, ни на что не пригоден" (с. 44).

- Система итальянско-английского АП полуторного поколения (Кардиффский ун-т, Англия; [126]): словарь на 1500 словоформ, упрощенный САн, минимальные изменения порядка слов в выходном тексте.

- Система англо-русского АП второго поколения, "проходившая в 1967-1970 гг. (в первом своем варианте) комплексную экспериментальную проверку" [50], с. 75 (ИЦ ЛГУ). Полное описание системы нам неизвестно, однако она упоминается в нескольких статьях ([48], [53], [89]) и представляет значительный интерес, так как в ней воплощен целый ряд наиболее прогрессивных принципов и методов (например, отчетливое разделение "грамматики" и "механизма", первичного "грубого" и последующего "тонкого" анализов и т.п.).

- Система русско-грузинского АП второго поколения (Ин-т систем управления АН Груз.ССР) [6], [7], [8].

Наконец, в многочисленных публикациях группы "Статистика речи" (см. библиографию [9]) описан ряд фрагментарных компонентов систем "отраслевого вероятностного АП" с немецкого, английского, французского и испанского языков на русский (см., в частности, [10], [13], [15], [24], [28] и т.д.); однако описание какой-либо завершенной целостной (тем более - действующей) системы АП нам неизвестно.

Системы АП третьего поколения и языковые процессоры типа "Смысл <=> Текст".

В этом разделе мы опишем вкратце следующие шесть систем:

1) Система MIND корпорации "РЭНД" (Санта-Моника, США).

2) Система АРАП (Ин-т Языкознания АН СССР и Моск. Гос. Пед. ин-т иностр. языков им. М.Тореза).

3) Гренобльская система русско-французского АП (Франция).

4) **Preference Semantic System** (Стэнфордский ун-т, США).

5) Португальско-английская автоматическая грамматика Дж. Уайета.

6) Система проверки трансформационных грамматик Дж.Фридман.

1) Система **MIND** (= **Management of Information through Natural Discourse**) , созданная под руководством М.Кэя и поставленная на ЭВМ в конце 1971 г. в **RAND Corporation** ([148] , [108]) , не является системой АП в узком смысле: это языковой процессор, т.е. универсальная система обработки текстов на любом языке, описанном грамматиками в стандартной форме. **MIND** состоит из семи основных компонентов: (1) морфологический анализатор, (2) синтаксический анализатор (выдающий для фразы ее глубинную структуру в смысле современной трансформационной теории), (3) преобразователь структур (система трансформаций, соотносящих глубинные структуры фраз входного языка с глубинными структурами семантически эквивалентных фраз выходного языка), (4) синтаксический синтезатор, (5) морфологический синтезатор, (6) автоматический словарь с богатой системой обслуживания и (7) **"Disambiguator"** – механизм разрешения неоднозначностей во входном тексте посредством формулирования вопросов на входном языке, обращаемых системой к однозычному предредактору (предполагается работа в режиме диалога). Все перечисленные компоненты не зависят ни друг от друга, ни от входного – выходного языков. Система **MIND** испытывалась на переводе технических руководств с английского языка на корейский, и результаты АП признаются в [148] приемлемыми.

2) В основу проекта системы АРАГ (= Англо-Русский Автоматический Перевод) была положена установка на создание, учет и использование максимума известных фактов и результатов исследований по АП в 60-е гг. [47] ; и хотя этот проект пока все еще не осуществлен, он интересен, в частности, благодаря своей теоретико-лингвистической ориентированности. Система АРАП была задумана в русле подхода "Смысл ⇔ Текст" [67] и опиралась непосредственно на идеи и предложения статьи [32] . Это многоуровневая система, представляющая собой (наподобие известных моделей С.Лэма, П.Сгалла, Д.Хейса) систему преобразователей, выполняющих переходы от более поверхностных представлений текста к более глубинным

(анализ) и обратно (синтез); преобразователи не зависят друг от друга и связаны только через входные/выходные представления текста. Как и в системе MIND , такие преобразователи можно использовать отдельно, например, в составе других систем АОТ.

Представляется целесообразным отметить две следующие особенности системы АРАП:

- Стремление обеспечить множественность синтеза и анализа, учитывая, таким образом, характернейшую для естественных языков многозначность соответствия между текстами и смыслами. В составе АРАП'а предусмотрена система перифразирования [32], получающая по одной ГСС все синонимичные ей и сводящая все синонимичные ГСС к каноническому виду - к базовой ГСС, которая удобна и для перехода к другому языку, и для преобразования в СемП (когда таковое будет введено в процедуру АП).

- Специфическое глубинно-синтаксическое представление (ГСП) фраз (см., в частности, [67], с. 33, 141 и сл.). Благодаря весьма абстрактным ГСО и использованию лексических функций принятое для системы АРАП ГСП оказывается достаточно обобщенным представлением фраз, которое, как предполагается, может служить хорошим мостиком при переходе между английским и русским языками (более удобным, по нашему мнению, чем глубинная структура стандартной трансформационной теории).

Необходимые лингвистические сведения задаются для системы АРАП посредством формальных грамматик, которым авторы стремились придать максимально универсальную форму - так, чтобы они были пригодны для описания любых явлений самых разных языков (см., в частности, [30] - грамматика морфологического синтеза для русского языка, [19] - фрагмент грамматики глубинно-синтаксического синтеза для русского языка, [68] - фрагмент грамматики ПСАн - и [69] - перечень ПСО для английского языка). Кроме того, очень большое количество лингвистических сведений должно быть собрано в (английском и русском) словарях системы - так называемых толково-комбинаторных словарях (о них см. подробнее ниже, раздел "Автоматические словари"). В настоящее время ведется работа главным образом над словарем системы (см. [95], а также цикл статей в сб. "Маш. пер. и прикл. лингвистика", вып.

www.ruslang.ru
БИБЛИОТЕКА ИН-ТА РУССКОГО ЯЗЫКА ИН-ТИ ЯЗЫКОЗНАНИЯ АН СССР

17, 1974) и над модифицированным семантико-синтаксическим анализатором [94].

3) Система русско-французского АП Гренобльского ун-та, как уже отмечалось, в 1971-1974 гг. подверглась полной перестройке. При этом общая стратегия и принципы, лежащие в ее основе, остались по существу теми же; не изменилась сколько-нибудь заметно и лингвистическая концепция ее авторов (см., в частности, [195], а также [71], с. 55-57). Однако в рассматриваемый период все основные установки были развиты значительно отчетливее и полнее, а главное - началась их полномасштабная реализация. Вообще, в центре внимания оказалась оптимальная машинная реализация - такая, которая обеспечивала бы, с одной стороны, дешевую, быструю и надежную переработку текстов (при условии, что система располагает хорошими грамматиками), а с другой - максимальное удобство работы с системой (как для лингвистов, так и для математиков). Чтобы добиться всего этого, процедуры АП, предусмотренные в Гренобльской системе, были подразделены на два класса: более поверхностные, типа "цепочка ⟷ цепочка" и "цепочка ⟷ дерево", которые могут быть выполнены конечно-автоматными преобразователями (или их композициями), и более глубинные, типа "дерево ⟷ дерево", для которых требуются более мощные (в математическом смысле) формализмы. Для реализации процедур первого класса создана формальная система **ATEF (Analyse de Textes en Etats Finis)** - автоматный преобразователь А, который работает с грамматикой естественного языка, написанной на языке **ATEF** [120]. Такая грамматика содержит морфы (префиксы, конечные/неконечные основы, конечные/неконечные суффиксы), описание их сочетаемости и правила объединения их означаемых в означаемое словоформы; при этом могут учитываться данные сразу о шести словоформах - рабочая, 4 предшествующие и 1 последующая. Преобразователь А выполняет морфологический анализ (разрезание каждой словоформы всеми возможными способами; предварительное разрешение неоднозначности - на основе сочетаемости внутри словоформы и в пределах микроконтекста; вычисление окончательной грамматической характеристики словоформы) и начальный синтаксический анализ (построение условного дерево текста в целом: фиктивному элементу "ТЕКСТ" под-

чиняются узлы Φ_1 , отвечающие фразам текста, каждой Φ_1 подчинены узлы Φ_{1j} , отвечающие омофразам j фразы Φ_1 , а каждой Φ_{1j} подчинены узлы, представляющие вершины разных деревьев зависимостей для данной омофразы). Дальнейшая работа с таким деревом, то есть АСАн и АСемАн, - это процедуры второго класса, которые выполняются специальным преобразователем деревьев, называемым "СЕТА" [121] . Его существенной особенностью является то, что все правила используемых им грамматик могут применяться к дереву текста одновременно в разных местах, т.е. преобразователь СЕТА обеспечивает параллельную обработку; это существенно ускоряет и упрощает анализ. В группе АП Гренобльского ун-та активно исследуются также новые формализмы и способы записи лингвистической информации для лучшей организации анализа текста и более гибкого варьирования режимов анализа; ср., в частности, так называемые сети переходов и разрывов (**réseaux de transitions et de ruptions**) , предлагаемые К.Буате [113] в качестве эффективного средства анализа простых рекурсивных языков (сети Буате родственны сетевым грамматикам У.Вудса, см. ниже, стр. 34). В докладе Буате на Пизанской конференции 1973 г. [111] дан очерк строения Гренобльской системы на ее современном этапе и, в то же время, изложен проект ее развития в ближайшие годы. Система будет состоять из девяти независимых блоков (= модулей): (1) морфологический анализатор, включающий и словарь (на основе формализма АТКГ); (2) "алгограмматика" - сетевая грамматика, обеспечивающая поверхностно-синтаксический анализ; (3) преобразователь деревьев, задающий соответствие между ПСС и ГСС; (4) "переводчик", преобразующий структуру входной фразы в структуру выходной, причем, возможно, по частям (не дожидаясь завершения анализа всей фразы); (5) синтезатор выходного текста; (6) блок самообучения (следит за изменениями, которые вносит в систему пользователь; учитывает частотность тех или иных явлений в текстах для соответствующих перестроек в системе; собирает и анализирует ошибки и т.п.); (7) блок запоминания (учитывает и хранит те результаты обработки предыдущих n фраз, которые могут понадобиться при обработке фразы n + 1 ; (8) блок диалогового режима, обеспечивающий прямое общение с системой; (9) управляющий блок - монитор, через который исследователь мо-

жет воздействовать на систему, заставляя ее работать в разных режимах: пословный перевод, перевод по синтаксическим группам, перевод целыми фразами и т.п. Блоки (1), (3), (4) и (5) были готовы в 1973 г.; остальные блоки разрабатываются в настоящее время.

4) **Preference semantic system (PSS)** , разработанная Й.Уилксом в Лаборатории искусственного интеллекта Стэнфордского ун-та и реализованная на ЭВМ **PDP** 6/10 ([204],[205]), это, по существу, система типа "Текст ⇒ Смысл", строящая по английской фразе ее СемП, (подробнее о семантическом анализе в **PSS** см. ниже, стр. 41). Система **PSS** включена в систему АП с английского языка на французский: в работе [141] описывается преобразование того СемП, которое получается в результате анализа, во французский текст.

5) Система Дж.Уайета - португальско-английская автоматическая трансформационная грамматика (ПААТГ; [214]), - вообще говоря, не является ни системой АП, ни процессором типа "Смысл ⇔ ⇔ Текст". Тем не менее, мы сочли целесообразным сказать о ней здесь несколько слов, поскольку специфический подход, выдвигаемый Дж.Уайетом, может представлять интерес для разработки и, особенно, отладки подобных процессоров.

ПААТГ состоит из собственно грамматики и алгоритма, выполняющего выводы в этой грамматике (выводятся пары фраз - португальская + английская, - являющихся переводами друг друга). Собственно грамматика включает базовый (НС-)компонент и трансформационный компонент; первый насчитывает 569 схем правил, содержащих в правых частях одновременно португальские и эквивалентные им английские символы, а второй - 109 унарных (образование пассивных, отрицательных и вопросительных форм) и 227 бинарных (образование сочиненных и сравнительных конструкций, сложноподчиненных предложений и т.п.) трансформаций. Алгоритм вывода фраз запрограммирован на языке **FAP** (≈ 18000 команд) для ЭВМ серии IBM 70; программа состоит из четырех частей: I. Порождение ядерных фраз; II. Унарные трансформации; III. Бинарные трансформации; IV. Набор стандартных подпрограмм, используемых основными программами (в том числе - подпрограммы порождения именных и глагольных групп).

Было проведено 50 машинных экспериментов по порождению пар равнозначных португальских и английских фраз; получаются грамматически правильные, хотя во многих случаях и бессмысленные фразы (для повышения их естественности существительным приписываются в словаре признаки типа [+ человек], [+ предмет], [+ абстрактное], а глаголам - сочетаемостные ограничения, учитывающие эти признаки).

6) Оговорка, сделанная в предыдущем пункте, необходима и здесь: система Дж.Фридман, о которой пойдет речь, также не является ни системой АП, ни процессором типа "Смысл ⇐⇒ Текст"; это так называемый тестер для трансформационных грамматик (ТГ). Дело, однако, в том, что подобные тестеры, обеспечивающие независимую экспериментальную проверку лингвистических описаний, оказываются совершенно необходимым инструментом при разработке любых достаточно богатых и сложных систем АОТ. Поэтому целесообразно кратко охарактеризовать тестер Дж.Фридман, являющийся типичным (и наиболее известным) представителем систем этого типа.

Тестер Дж.Фридман [130] - это система программ, реализованная на ЭВМ IBM 360/67 в ВЦ Стэнфордского и Мичиганского ун-тов (США). Система принимает любую конкретную ТГ (КС-правила плюс трансформационный компонент в стандартной форме) и выдает фразы, выводимые в данной ТГ. Тем самым система обеспечивает проверку ТГ, оказываясь очень эффективным средством их разработки и совершенствования, в особенности полезным при создании тех систем АП и АОТ, в которых для описания привлекаемых языков используется формализм ТГ. Тестер Фридман был успешно опробован на нескольких разных ТГ современного английского языка, а также на ТГ для древнеанглийского, французского и суахили.

Среди других аналогичных систем мы отметим еще тестер для ТГ Л.Гросса, реализованный на вычислительной системе CTSS в Массачузетском Технологическом Ин-те [134]; на Международной конференции по вычислительной лингвистике в Пизе (1973) сообщалось еще о трех тестерах: А.Ржиги - С.Маховой (ЧССР; тестер для порождающего компонента модели чешского языка П.Сгалла - своеобразной КС-грамматики, строящей СемП в смысле этой модели, т.е. опреде-

ленным образом организованные цепочки семоглифов; разрабатывается в ВЦ Пражского ун-та); И.Батори (ФРГ) и С.Бойсверта – А.Дюга – Д.Беланже (Канада; оба последние тестера – для ТГ) [47], с. 205-206.

Отдельно следует упомянуть еще систему АП с английского языка на испанский, португальский, немецкий, французский и японский языки, разрабатываемую в ун-те им. Б.Янга (Прово, шт. Юта, США) [163]; основанная на теории "юнкционной грамматики" (многоуровневая модель с очень наивным представлением о семантике), эта система предполагает перевод через своеобразное СемП. Система реализована в виде экспериментального варианта на ЭВМ IBM 360/50; неоднозначности входного текста разрешаются путем обращения к однозначному предредактору (ср. DISAMBIGUATOR М.Кэя, стр. 16), так что здесь основной упор делается на синтез.

Системы АОТ, включающие значительные АП-компоненты

Количество созданных в 1970-74 гг. систем АОТ, в которых в той или иной форме осуществляются переходы "Текст ⇒ Смысл" и "Смысл ⇒ Текст" очень велико (ср. прекрасные обзоры Р.Симмонса [183] и Д.Боброу [109], относящиеся к 1964-1969 гг.); мы выбрали лишь шесть таких систем АОТ, где сугубо лингвистические компоненты достаточно развиты. Это следующие системы:

1) Система Р.Шенка, моделирующая понимание текста (**Language Understanding System – LUS**).

2) Так называемый "робот" Т.Винограда.

3) Система понимания устной речи Д.Уокера (**Speech understanding system – SUS**).

4) ИПС **LUNAR** У.Вудса.

5) Семантическая модель Ш.Клейна.

6) ИПС **TITUS-II** М.Джкро.

1) Р.Шенк выдвинул оригинальную теорию понимания текстов на естественном языке [180], сущность которой заключается в использовании особого СемП текста – семантического квазиграфа, позволяющего довольно тонко изображать содержание текстов независимо от их языковой формы (о СемП Шенка см. ниже, стр. 48). Переход

"Текст ⇒ Смысл" выполняется с помощью словаря, набора синтаксических правил и набора семантических правил. В словаре каждому слову сопоставлено его СемП, выполненное в терминах особых семантических единиц четырех сортов: PP - имена вещей, ACT - имена событий, PA - характеристики вещей и AA - характеристики событий, и семантических связей между этими единицами. Пара семантических единиц, связанных семантической связью определенного типа, называется концептуализацией; концептуализация (семантический аналог предикативной синтагмы) является главной единицей в СемП Шенка (как отдельных слов, так и целых высказываний) - семантический квазиграф есть не что иное, как сеть концептуализаций.

Семантические единицы подвергаются дальнейшей классификации. Так, все ACT разбиваются на классы, для каждого из которых фиксируется набор семантических валентностей (= "падежей", по Шенку); например, слова типа трогать относятся к классу PACT (названия физических воздействий) и требуют объектного и инструментального "падежей".

Синтаксических правил - около 30; ими задаются основные синтаксические конструкции английского языка. Синтаксический анализ играет, по мнению Р.Шенка, чисто вспомогательную роль, и во многих случаях без него можно вообще обойтись. Главное средство анализа, или понимания, текста, т.е. главное средство перехода к СемП, - это семантические правила. Их - 19, и они указывают, какие семантические единицы могут находиться в каких семантических отношениях друг к другу, например: 1) PP ⟷ ACT ("вещь PP - субъект события ACT": Он спит), 2) PP ⟷ PA ("вещи PP предицируется атрибут PA": Он глуп), ... 11) CACT ←O— ↕ ("Объектом ←O— сообщения CACT является целая концептуализация ↕ ": ... заявил, что Маша уехала), 12) SACT ←O— ↕ ("Объектом ←O— психического акта SACT является целая концептуализация ↕ ": ... хочу, чтобы ты пришла ко мне), ..., 19) PP ↕LOC ("Местом ↕LOC концептуализации ⟷ является вещь PP": Маша лежит на диване). В процессе перехода от текста к его СемП должен использоваться не только широкий контекст (весь предшествующий текст), но и большое количество энциклопедических, психологических и культурных сведений. Для этого система LUS должна иметь обширную "семантическую па-

мять", где все сведения о мире, известном системе, записываются в виде квазиграфа Шенка, а также механизм логической обработки - для построения умозаключений, которые Р.Шенк считает совершенно необходимыми в процессе семантического анализа текста. (См. подробный обзор и изложение всех основных работ Р.Шенка и его сотрудников в [65] и [66]; там же приведены все необходимые ссылки.)

На основе теории Шенка были построены две машинные системы - SPINOZA I и SPINOZA II , осуществляющие семантический анализ английских фраз, а также система MARGIE , которая помимо анализа, выполняет перифразирование входных высказываний и делает на их основе всевозможные умозаключения. Пример работы системы MARGIE : входная фраза - John advised Mary to sell Bill a banana; заключения, выведенные системой, - Mary considered giving Bill a banana; John possibly believes that Mary would benefit from Bill giving her money, Mary began thinking about Mary sells Bill a banana benefits Mary [грамматически неправильная фраза; это объясняется несовершенством механизма поверхностного синтеза]; вопрос, заданный системой, - Does Bill want to eat a banana?

2) "Робот" Т.Винограда ([207]и[208]) представляет собой систему типа "вопрос - ответ", работающую в режиме письменного диалога. В ее памяти хранится описание (на несложном формальном языке, устроенном, как язык исчисления предикатов) некоторого условного "мира" - набора разноцветных кубиков, параллелепипедов и пирамидок, произвольно расположенных в пространстве; это описание образует банк данных системы, которая способна отвечать на английские вопросы, относящиеся к этому банку данных, и изменять банк данных в соответствии с получаемыми приказами. Для большей наглядности содержимое памяти может выводиться на дисплей в виде картинки, изображающей "расположение" названных предметов. (Таким образом, никакого робота в действительности не имеется, так что название системы плохо отражает ее природу.) Образец диалога системы (Р) с экспериментатором (Э; в переводе на русский язык):

Э. Найди большой красный кубик.

Р. Хорошо.

Э. Возьми теперь пирамидку.

Р. Не понимаю, о какой пирамидке ты говоришь.

Э. Я имею в виду ту, которая стоит на красном кубике, который ты нашел.

Р. Хорошо, понял.

Э. Возьми эту пирамидку и положи ее в коробку. Что теперь в коробке?

Р. Синяя пирамидка и синий кубик.

Для поддержания подобных диалогов система сначала переводит входной вопрос или приказ на свой внутренний язык (т.е. на язык, на котором записан ее банк данных); это - семантический анализ. Затем она отыскивает запрашиваемую информацию или выполняет требуемое изменение в своем "мире" и конструирует ответ на внутреннем языке, т.е. проводит логическую переработку информации. Наконец, готовый ответ переводится на английский язык, что и составляет семантический синтез. Анализ и синтез выполняются на основе словаря, где английским словам приписана грамматическая информация и толкование (или СемП) на внутреннем языке системы; кроме того, имеется английская грамматика, оформленная как набор подпрограмм, каждая из которых обрабатывает один тип составляющих. При анализе сначала строится НС-дерево входной фразы; затем проверяется семантическая согласованность составляющих (на основе признаков, приписанных словам в словаре); в случае семантической несогласованности полученное дерево считается неправильным, и таким образом разрешается ряд синтаксических неоднозначностей. Наконец, с помощью правил "семантического проецирования" (типа правил Катца - Фодора) строится СемП всей входной фразы. При синтезе указанный путь проходится в обратном направлении.

3) Система SUS , разрабатываемая под руководством Д.Уокера [199], представляет собой такую модификацию системы Винограда, что становится возможным понимание устных вопросов и приказов. К системе Винограда добавлены три компонента:

- предварительный акустический анализ входных фраз;
- порождающая грамматика, которая на основании всех данных, имеющихся у системы, предсказывает цепочки английских слов, наиболее вероятных в данном контексте;
- механизм сравнения предсказанной цепочки и результатов

акустического анализа (этот механизм управляет порождением новых цепочек и одновременно уточнением акустической интерпретации входного высказывания).

SUS частично реализована на ЭВМ PDP-10.

4) Фактографическая ИПС LUNAR У.Вудса ([211]-[212]) имеет в качестве банка данных трехмерную матрицу $\langle i, j, k \rangle$, в которой указано содержание определенных химических элементов (i) в различных образцах (j) лунных пород в разное (с момента взятия образцов) время (k), а также ссылки на статьи, где были опубликованы результаты соответствующих анализов; ИПС выдает любые данные, непосредственно содержащиеся в матрице, различные их суммы и отношения, средние величины, сведения о классах образцов с определенными свойствами и т.п. - в ответ на запросы, формулируемые в письменном виде на английском языке. Запросы подвергаются синтаксическому анализу на основе сетевой грамматики (см. ниже, стр.), а затем семантическому анализу - переводу на формальный язык типа языка исчисления предикатов; СемП запроса позволяет системе отыскать нужную информацию. LUNAR реализована на ЭВМ PDP-10 и демонстрировалась в действии на II Ежегодной конференции по селенологии (Хьюстон, февраль 1971 г.). Из 111 запросов, предложенных участниками конференции, 86 были правильно проанализированы системой и на них были выданы ответы.

В настоящее время ведутся работы по созданию блока понимания устной английской речи [213], который должен позволить системе LUNAR отвечать и на устные вопросы. В связи с этим следует подчеркнуть, что системы понимания устной речи, включающие существенный семантический компонент (и, как правило, объединяемые с той или иной ИПС, так что они могут базироваться на ее семантике) стали весьма популярны в начале 70-х гг. (см. [167]).

5) Универсальная автоматизированная модель языка (УАМЯ) типа "Смысл $\Leftrightarrow$ Текст" Ш.Клейна [153] - это система программ, обеспечивающая переход от СемП английских фраз (СемП задается в виде набора "семантических троек", см. ниже, стр. 48) к фразам. Соответствие между СемП фразы и ее поверхностной НС-структурой описывается правилами вида $\langle R \| ST, M \rangle$, где R - обычное КС-правило, ST - семантическая тройка (два объекта CE_1 и CE_2, связан-

ные семантическим отношением СО; один из объектов может быть пустым), а М – соответствие между компонентами правила R и троек ST , например: 1) S → NP + VP || CE_1 – CO: NP = CE_1, VP = CO; 2) NP_1 → Adj + NP_2 || CE_1 – атрибут – CE_2; NP_2 = CE_1, Adj = CE_2 ; и т.п. Кроме того, имеется набор синтаксических трансформаций и морфологических правил, превращающих НС-структуру в реальную фразу. УАЯМ обратима и способна осуществлять как анализ, так и синтез текста. Ш.Клейн использует эту систему в качестве компонента двух других экспериментальных систем, работающих с естественным (= английским) языком: в системе, моделирующей диахроническое развитие языка, а также изменения языков в результате контактов и многоязычия [151] ; и в системе, моделирующей сочинение детективных новелл [154]. Очевидно, что УАЯМ Клейна могла быть легко применена и для АП; однако подобные попытки нам пока неизвестны.

6) ИПС TITUS-II интересна исключительно как типичный образец системы, в которой информационный поиск комбинируется с частичным АП; в 1970-1974 гг. было построено довольно много таких систем.

Система TITUS-II(Traitement de l'Information Textile, Universelle et Selectionnée) разработана и реализована в Национальном Ин-те Текстиля под руководством Ж.Дюкро (Париж). Система осуществляет информационный поиск на большом массиве рефератов по технологии текстильного производства и переводит эти рефераты на английский, французский, немецкий и испанский языки [162]. Рефераты составляются человеком и записываются им на специальном ИЯ (который выступает при переводе рефератов в качестве ЯП). Словарь этого ИЯ представляет собой набор дескрипторов и слов-связок; и то, и другое – пучки лексических соответствий между четырьмя названными языками. Синтаксис ИЯ – "общая часть" синтаксиса тех же языков (17 стандартных синтагм, встречающихся во всех этих языках). TITUS-II , отыскав нужный реферат на ИЯ (запросы также формулируются на ИЯ человеком, обычно – пользователем), переводит его на любой из выходных языков, т.е. выполняет, так сказать, его синтез; анализа же в этой системе нет. Система TITUS-II запрограммирована на языке ассемблера IBM-Sub2 и работает со следую-

щей скоростью: поиск по сложному запросу (несколько дескрипторов) ≈ 5-10 сек, перевод полного реферата ≈ 3 сек; в систему поступает до 18 тыс. рефератов в год.

Частные вопросы АП (конкретные этапы переходов "Текст ⇒ Смысл" и "Смысл ⇒ Текст").

А. Анализ.

1. Морфологический анализ.

К 1970 г. задача АМАн была практически решена. Это позволило исследователям сосредоточить усилия на оптимизации систем АМАн как в лингвистическом отношении (разработка универсальных формализованных описаний), так и в плане их машинной реализации.

Мы охарактеризуем сначала две работы, посвященные морфологии русского языка.

В брошюре [30] предлагается полное формальное описание русского именного словоизменения, ориентированное на машинный синтез и анализ соответствующих словоформ. Особенность этого описания состоит в том, что оно построено по некоторой универсальной схеме (выдвинутой в рамках модели "Смысл ⇔ Текст" [67], которая пригодна для описания морфологии языков самого разного строения (эта схема была проверена на материале испанского, венгерского, английского, татарского и алюторского[1] языков). Морфологическое описание, или модель морфологии данного языка, трактуется как система из четырех наборов правил: 1) правила, задающие соответствие между морфологическими характеристиками (= значениями морфологических переменных) и морфемами (ПМП словоформы ⇔ набор морфем словоформы), 2) правила соответствия между морфемами и их представляющими морфами (набор морфем ⇔ цепочка представляющих морф), 3) морфонологические и фонологические правила (цепочка представляющих морф ⇔ цепочка фонем), 4) графико-орфографические/фонетические правила (цепочка фонем ⇔ цепочка букв/фонетических знаков).

[1] Один из языков чукотско-камчатской семьи.

Для таких моделей может быть разработан стандартный программный механизм, учитывающий только форму правил; этот механизм был бы пригоден для морфологического анализа/синтеза словоформ любого языка (разумеется, при условии, что морфология этого языка будет описана в необходимой форме). Данная модель была запрограммирована для ЭВМ БЭСМ-6 и экспериментально опробована [40].

А.-Л.Согвалл построила и реализовала на ЭВМ IBM 370/155 систему морфологического анализа для русского текста [177], в которой предусмотрена обработка чередований (полон - полн-, разобрать - разбирать, чертить - черчу и т.п.), благодаря чему в словаре подавляющее большинство лексем представлено одной основой. Любопытной лингвистической особенностью системы Согвалл является выделение во всех глагольных словоформах основообразующего (= "тематического") элемента: j/Ø (чита-j-у/чита-Ø-л), уj/ова (команд-уj-ет/команд-ова-л), оj/ы (кр-оj-у/кр-ы-л) и т.п. (всего 39 подобных пар), что обеспечивает компактное и естественное представление глагольных основ. В ходе эксперимента система безошибочно проанализировала более тысячи словоформ менее чем за три минуты.

С точки зрения достаточно общего и гибкого формализма, допускающего эффективную машинную реализацию и весьма удобного для лингвистов, особенно интересен морфологический анализатор, основанный на системе ATEF (Гренобльская группа; см. выше, стр.18).

Автоматическая модель польской морфологии (анализ и синтез словоформ) реализована на ЭВМ GIER (Ин-т Математических Машин АН ПНР, Варшава); она войдет в состав диалоговой системы MARYSIA , разрабатываемой в настоящее время [107].

Кроме того, укажем еще 6 систем АМАн для различных языков:

- Английский язык: 1) [48] . Учитывается не только словоизменение, но и регулярное словообразование; предусмотрена обработка не найденных в словаре слов; обеспечено распознавание чередований в основах (усечение конечного -e: come - com-ing ; удвоение конечной согласной: stop - stopp-ed и т.п.).

2) [137] . Эта система АМАн написана в виде формальной грамматики, использующей формализм q-систем (см. ниже, стр.51).

- французский язык: 3) [39] (на основе системы АМАн Гренобльской группы); для не найденных в словаре слов предусматривается также определение их грамматических характеристик по семи последним буквам.

- Венгерский язык: 4) [102]; система реализована на ЭВМ IBM 360.

- Немецкий язык: 5) [23] ; учитывается и регулярное словообразование.

6) [182]; система реализована и работает со скоростью 1000 словоформ в 17 секунд.

Наконец, заслуживают упоминания работы, посвященные двум специальным проблемам АМАн: анализу сложных слов и "содержательному" кодированию букв.

В книге [196] для АМАн голландских сложных слов предлагается указывать в словаре при каждой морфе ее способность входить в сложное слово в качестве n-го компонента; в статье [181] исследуется аналогичная проблематика для немецких сложных слов (особое внимание уделено интерфиксам и смысловым связям между компонентами сложных слов).

Стремясь упростить обработку чередований при АМАн (на материале польского языка), Я.Бень [106] выдвинул идею кодировать алфавитные знаки так, чтобы буквы и диграфы, входящие в один чередовательный ряд (например, d-dz-dź-dż; c-ć-k-t [syt-y - syć-i] ; a-e-o-ó ; и т.п.), имели в своем коде общий бит, выделяющий их среди прочих букв; тогда чередованные основы типа miast-o 'город' / mieśc-e 'городе' будут легче совмещаться, т.е. распознаваться как варианты одной основы. Сходная задача обсуждается в статьях Г.Б.Чикоидзе ([91] и [92]), который предложил кодировать грузинские буквы по различительным признакам соответствующих фонем, и в статье М.Бокша [110], предложившей "содержательные" коды для чередующихся букв в румынском языке.

2. Синтаксический анализ.

Проблема автоматического синтаксического анализа (в дальнейшем - АСАн) оставалась весьма популярной в 1970-1974 гг., и ей посвящено большое количество работ. Для лучшей обозримости мы разобьем их на две группы:

- Работы более общего характера, рассматривающие синтаксический анализ в целом и уделяющие существенное внимание методам анализирования.

- Работы более лингвистического характера, сосредоточенные на описании синтаксиса того или иного конкретного языка (выполненное для целей АСАн).

Работы первой группы удобно, в свою очередь, классифицировать по характеру организации системы АСАн, что, разумеется, тесно связано с методом анализирования. При этом получается три типа систем (I-III):

I. Системы АСАн "доклассического" направления (относящиеся к первому или полуторному поколению систем АП). Это так называемые "ползучие" алгоритмы, представляющие собой последовательности правил весьма разнообразной природы, вводимых зачастую ad hoc. Грамматика, т.е. описание фактов языка, здесь по существу не отделена (или плохо отделена) от программного механизма, реализующего процедуру анализирования; естественно, что никакой формализм для записи грамматики в таких системах АСАн не используется, равно как и какие-либо общие методы анализирования. Несмотря на их научную бесперспективность, подобные системы еще существовали и даже разрабатывались и в рассматриваемый период (хотя и в небольшом числе). В качестве примера можно сослаться на [161], где для АСАн русского текста не предполагается какая-либо единая грамматика, а вводятся отдельные подпрограммы обнаружения именных и глагольных групп, установления связей между выделенными группами, разрешения падежной неоднозначности и т.п. Приблизительно таковы же по содержанию многочисленные публикации сотрудников группы "Статистика речи": например, [28] и [56]; см. еще ссылки на стр. 15 .

АСАн совсем простых испанских фраз (подлежащее + сказуемое + дополнение) реализован на ЭВМ IBM 7090 в ВЦ Мадридского ун-та; в случае неправильности входной фразы - например, несогласование артикля с существительным) программа обнаруживает и указывает ошибку [129].

II. Системы АСАн "классического" направления; все они - типичные представители второго поколения. Классическими мы называем

широко распространенные и наиболее освоенные во всех отношениях системы, основанные на описании языка в терминах формальной НС-грамматики (при этом синтаксическая структура понимается как структура(составляющих) и использовании какого-либо готового алгоритма анализирования текста в таких грамматиках (особенно часто используется здесь алгоритм Дж.Кока и его различные модификации). Такая склонность авторов классических анализаторов к НС-грамматикам не случайна; она объясняется по крайней мере тремя следующими факторами. Во-первых, благодаря Н.Хомскому и его школе, НС-грамматики весьма популярны в современной лингвистике и поэтому для многих языков существуют более или менее подробные фрагменты таких грамматик; в то же время, уже более пятнадцати лет активно исследуются различные лингвистические проблемы, связанные с представлением сведений о синтаксисе естественного языка посредством именно НС-грамматики (в особенности много здесь сделано для английского языка). Во-вторых, формальные свойства НС-грамматик, в том числе - их различных разновидностей, задаваемых ими языков, и соответствующих распознающих алгоритмов также хорошо изучены математической лингвистикой, теорией автоматов и другими относящимися сюда дисциплинами (см. [17]). И, наконец, в-третьих (возможно даже, что именно это соображение является наиболее веским), НС-грамматики (а точнее - их важный подкласс: КС-грамматики) оказались эффективным и повсеместно принятым средством описания синтаксиса языков программирования ("бэкусова нормальная форма" и т.п.).

Естественно поэтому, что большинство синтаксических анализаторов, создаваемых не только для АП, но и для разных других систем АОТ, в первую очередь - для ИПС, строится по классической схеме: НС-грамматика + какой-либо стандартный распознающий алгоритм. Типичными образцами таких анализаторов являются системы, описанные, например, в [169] (небольшая КС-грамматика чешского языка; алгоритм Й.Копривы типа "сверху вниз, справа налево"; машинный эксперимент в 1968 г. на ЭВМ SAAB-D21 - анализ 12 фраз), в [144] (КС-грамматика французского языка, насчитывающая 104 правила, 17 типов составляющих и 23 терминальных класса слов; универсальный алгоритм типа "сверху вниз"; анали-

затор реализован на ЭВМ IBM-360) и в [140] (система REL = Rapidly Extensible Language , обеспечивающая общение с ЭВМ на слегка стандартизованном английском языке; КС-грамматика английского языка, насчитывающая свыше 300 правил; анализатор реализован на IBM 370/155). Публиковались и новые варианты универсальных алгоритмов для работы с КС-грамматиками, например, [159] или [171]; в последней статье предложен способ сведения НС-правил специального вида (контекстные условия в дереве носят только локальный характер) к КС-грамматике, так что делается возможным КС-распознавание соответствующего языка.

Подробный и очень четкий обзор анализаторов классического типа дан С.Р.Петриком [172]; в [172] указываются также основные типы грамматик, используемых в настоящее время для АСАн.

В рамках классической схемы предлагались ее различные модификации, ведущие к упрощению анализаторов и увеличению их мощности, повышению скорости анализа, улучшению изображения структуры и т.п. Мы укажем здесь следующие типы подобных модификаций:

(i) Введение в систему АСАн трансформационного компонента, за счет которого расширяется класс анализируемых фраз и лучше отображаются семантические связи между компонентами фраз. Впрочем, высказывалось и такое мнение (Хениш-Достерт и Томпсон, см. [140]), что использование аппарата трансформаций чрезмерно усложняет машинную обработку текста и что поэтому необходимый эффект выгоднее получать за счет расширения набора НС-правил и даже усложнения их формата.

(ii) Использование синтаксических признаков - индексов при нетерминальных символах в НС-правилах, что существенно сокращает число правил и одновременно - число применений правил в процессе анализа (так, в системе REL [140] благодаря признакам число применений правил при анализе двух несложных английских фраз сократилось приблизительно в 250 раз).

(iii) Присоединение к анализатору специального механизма, способного при переборе вариантов выделить те фрагменты синтаксической структуры, которые из содержательных соображений долж-

ны быть признаны достоверными; эти фрагменты фиксируются (= "замораживаются") и в дальнейшей процедуре анализа трактуются как цельные блоки, т.е. повторно не анализируются. Благодаря этому существенно сокращается время анализа.

(iv) В словарь и грамматику анализатора добавляются сведения, позволяющие устанавливать с м ы с л о в ы е связи между глаголом-сказуемым и именными группами во фразе, т.е., выражаясь терминами Ч.Филлмора, определять падежные роли именных групп ("субъект действия", "инструмент", "локализация", "адресат" и т.п.); так сделано, например, в анализаторе системы REL [140]. Синтаксическая структура, содержащая указания на падежные роли именных групп, несет больше полезной информации о фразе и оказывается гораздо удобнее для перевода на другой естественный или информационный язык. Впрочем, введение в СС падежных ролей означает, по существу, выход за пределы "классических" НС-анализаторов.

Д. "Неклассические" системы АСАн второго поколения. Сюда относится ряд работ, в которых исследуются синтаксические модели, отличные от НС-грамматики, и применяются другие методы анализирования, нежели стандартные алгоритмы коковского типа. Мы рассмотрим шесть "неклассических" анализаторов:

1) Сетевой анализ У.А.Вудса .
2) Матричный анализ О.С.Кулагиной.
3) Упрощенный анализ Г.С.Цейтина.
4) Цепочечный анализ З.Хэрриса - Н.Сейгер.
5) Q-анализатор Р.Киттредж - А.Кольмероэра.
6) Графовый анализ М.Вейлеча.

1) В статье [210] У.Вудс предложил в качестве формального средства описания синтаксиса естественного языка для целей АСАн особый механизм - "сетевую грамматику"(**transition network grammar**) . Сетевая грамматика - это конечный связный ориентированный граф, вершины которого отвечают классам слов и словосочетаний, дуги (= переходы) - составляющим, а пути в графе - допустимым последовательностям составляющих. Сетевая грамматика работает с набором регистров, в которых формируются фрагменты получаемой синтаксической структуры (СС) фразы; эта СС имеет вид сложного списка, в котором для каждой составляющей указаны ее собственные составляющие

и т.п. Грамматика Будса легко допускает переупорядочение, элиминирование и добавление составляющих, благодаря чему с ее помощью можно получать сразу ГСС, минуя отдельный этап получения ПИС. Для повышения надежности и эффективности анализа дугам сетевой грамматики сопоставляются добавочные наборы условий (="функции"), которые запрещают/разрешают/предписывают установление той или иной конкретной связи. Алгоритм АСАн на основе сетевой грамматики очень прост: он строится из десятка стандартных подпрограмм (подача на вход очередного слова, проверка наличия у рабочего слова заданных свойств, проверка содержимого заданного регистра и т.п.). Анализатор Будса (для английского языка), реализованный на ЭВМ SDS-940, был весьма популярен в 1970-1974 гг.: он использовался в качестве готового компонента в целом ряде систем АОТ (ср., в частности, сетевую грамматику К.Буате, стр.19).

2) Анализатор О.С.Кулагиной [41] , разработанный в рамках системы французско-русского АП "ФР-П", осуществляет АСАн французских фраз, выдавая результат в виде дерева зависимостей. Грамматика анализатора представляет собой список конфигураций, или синтагм, - пар символов классов слов (возможно, с индексами тех или иных подклассов), для которых указано, при каких условиях (имеющих место во фразе) соответствующие слова могут быть связаны зависимостью определенного типа. На основании списка конфигураций анализатор строит матрицу потенциальных связей; для фразы длиной в n слов нужна матрица $n \times n$, где в клетке $\langle i, j \rangle$ ставится единица, если слово i принципиально (с точки зрения только списка конфигураций) может зависеть от слова j , и 0 - в противном случае (если $\langle i, j \rangle = 1$, то отдельно еще указывается тип зависимости, которая связывает слово i с его хозяином j). Далее, у анализатора имеется перечень условий, при которых та или иная потенциальная связь может реализоваться в конкретной фразе. Это условия типа проективности (с указанием случаев разрешенной непроективности), обязательной насыщенности определенных валентностей, несовместимости или неразделимости тех или иных связей и т.п. Все условия используются анализатором как система ф и л ь т р о в , через которую пропускается матрица, содержащая весь набор гипотетических связей; те и только те связи, которые удовлет-

воряют всем фильтрам, признаются правильными, а остальные элиминируются. Таким образом, анализатор Кулагиной принадлежит к системам эксплицитно фильтрового типа, каковые, на наш взгляд, являются на данном этапе наиболее перспективными.

О.С.Кулагиной предложена весьма эффективная и удобная методика фильтровки гипотетических связей - "чистка матрицы", состоящая в том, что фильтры излагаются на языке допустимых/недопустимых комбинаций единиц в матрице и "лишние" единицы выбрасываются; матрица, оставшаяся после чистки, содержит все допустимые СС анализируемой фразы и только такие СС (при условии, конечно, лингвистической правильности списка конфигураций и набора фильтров).

Особо следует подчеркнуть, что метод Кулагиной делает ненужным какое бы то ни было л о г и ч е с к о е у п о р я д о ч е н и е правил используемой грамматики (= конфигураций) или задание л о г и ч е с к и определенной п о с л е д о в а т е л ь н о с т и обработки фразы (то или иное практически удобное упорядочение может, разумеется, иметь место). Таким образом, анализатор Кулагиной является, по существу, асинхронным.

В данной связи следует упомянуть специальную работу Р.Каплана [147], где также предлагается сходный метод асинхронного (параллельного) осуществления всех операций при АСАн. Грамматика задается в виде сетевой грамматики У.Вудса (см. выше). Для анализируемой фразы строится граф потенциальной связности (соответствующий матрице Кулагиной), который затем совмещается с сетевой грамматикой, играющей здесь роль фильтров.

Важное преимущество систем указанного типа состоит, в частности, в том, что они позволяют анализировать неполные и даже лишь частично правильные фразы.

3) Одна из основных идей Г.С.Цейтина [89] состоит в разделении процедуры АСАн на два этапа: во время предварительного АСАн сознательно не учитываются наиболее сложные случаи, так что сначала получается "грубая" СС фразы (в ней могут быть лишние связи, т.е. дизъюнкции связей вместо конкретных связей, и еще не разрешенные неоднозначности, но в ней не должно быть неправильных решений); затем проводится расширенный АСАн, дающий "тонкую" структуру. Такая двухэтапность позволяет упростить анализатор в целом и сократить время его работы. В [53] описан машинный эксперимент по

предварительному АСАн русских текстов. В эксперименте использовалась упрощенная "минирамматика" зависимостей (202 несложных правила, типа следующего: $V_{мод} \longleftarrow V_{инф}$, т.е. "инфинитив управляет влево модальным глаголом") и не слишком богатая словарная информация (10 классов слов, 11 признаков синтаксических подклассов, 12 семантических признаков типа "модальность" и 28 грамматических признаков). Из 80 фраз (рефераты из РЖ по сварке) 70 были проанализированы верно, т.е. для них были получены предполагаемые "грубые" СС (содержащие информацию о том, какие зависимости встречаются хотя бы в одной из допустимых "тонких" СС и какие омонимы оказываются выбранными хотя бы в одной из таких структур). Скорость анализа - $\approx$ 30 сек/фраза.

4) Цепочечный анализ был предложен З.Хэррисом [138] и затем активно разрабатывался для целей АП группой под руководством Н.Сейгер. Цепочечный анализатор (ЦА) для английского языка был неоднократно реализован на ЭВМ различных марок и проверен в серии многочисленных экспериментов; см. о нем стр. 182-213 в книге [1]. Здесь мы укажем только следующее существенное усовершенствование, внесенное в ЦА в начале 70-х гг. [133]. Как известно, одним из центральных компонентов всякого ЦА является список ограничений - требований, накладываемых на допустимые комбинации элементарных цепочек (простейших двух- и трехчленных конструкций, в терминах которых ЦА изображает синтаксическую структуру); эти ограничения суть не что иное, как фильтры, отбирающие правильные СС среди множества гипотетических. Полнота и точность списка ограничений во многом определяет качество и скорость работы ЦА. Однако запись ограничений в предписанном жестком формате - весьма трудоемкая задача, которая отнимает значительную часть усилий лингвистов, составляющих грамматику для ЦА.

В связи с этим авторы последнего варианта ЦА разработали механизм, позволяющий записывать ограничения непосредственно на английском языке (высказываниями с достаточно простой синтаксической структурой); это дает лингвисту возможность целиком сосредоточиться на содержательной стороне дела и вовсе не заботиться о соблюдении формата. Новый ЦА сам анализирует предлагаемый ему на английском языке список его собственных ограничений (что - благодаря

их относительной стандартности - осуществляется легко), а проанализированные ограничения автоматически превращаются в стандартные подпрограммы, которые в дальнейшем и проверяют соблюдение ограничений при анализе текстов. Таким образом, здесь сделан, на наш взгляд, очень существенный шаг: попытка моделировать обучение человека языку с помощью уже знакомого ему фрагмента того же самого языка. Думается, что в перспективных системах АП и вообще в системах АОТ соответствующий компонент абсолютно необходим. (Он имеется, в частности, в системе REL [140].)

Будет, возможно, нелишним отметить еще одну любопытную частность: в новом ЦА предусмотрен механизм (реализованный посредством магазинной памяти), который обеспечивает автоматическую проверку одних и тех же ограничений для всех однородных членов и для параллельных членов симметричных конструкций с эллипсисом.

Достаточно полная цепочечная грамматика английского языка приведена в [98].

Цепочечная грамматика разрабатывалась и для французского языка: [178], а также [179], где дается относительно полный вариант такой грамматики; эта грамматика опробована в машинных экспериментах на ЭВМ IBM 360/91.

5) АСАн на основе так называемых q-систем (в рамках Монреальской системы англо-французского АП) описан Р.Киттреджем в [150]. Приводится полная грамматика, записанная в виде двух q-систем (см. ниже, стр.51), в соответствии с которой получается ГСС английской фразы.

6) В работах М.Вейлеча ([191] и [192]) намечается конструкция специализированного вычислительного устройства, предназначаемого для АСАн текстов на естественном языке. Синтаксические структуры фразы должны представляться так называемым Q-графом* [193]. Это связный ориентированный граф с циклами, притом многоуровневый: любой его подграф (= составляющая) может быть изображен одной вершиной высшего уровня; полный путь в Q-графе задает правильную СС фразы. Электронный анализатор Вейлеча имеет специальные блоки памяти, особо удобные для хранения грамматик, задающих Q-графы, и для переработки Q-графов, а так-

* Q-графы Вейлеча не имеют ничего общего с Монреальскими q-системами.

же встроенный механизм анализирования; этим обеспечивается высокая скорость анализа и устраняется необходимость в программировании анализаторов.

На базе рассмотрения публикаций по вопросам АСАн за 1970-1974 гг. можно попытаться выявить те требования, которым, как нам представляется, должны удовлетворять "идеальные" синтаксические анализаторы третьего поколения. Эти требования, в частности, таковы:

(i) Формальное задание (например, посредством порождающих грамматик) входного и выходного представлений фразы, с которыми имеет дело анализатор.

(ii) Достаточная глубина выходного представления: оно должно быть удобно для перехода к СемП или к другому языку. Отсюда вытекает, по-видимому, что в нем должны обязательно отражаться о т н о ш е н и я между словами (что и сделано в очень многих современных синтаксических анализаторах). Кроме того, отсюда вытекает также необходимость различения поверхностно- и глубинно-синтаксического представлений, связанных преобразованием: прямой переход от результатов морфологического анализа фразы к глубинному изображению ее содержания (пусть еще и не семантическому) оказывается слишком громоздким.

(iii) Описание синтаксиса привлекаемого языка посредством формальной грамматики, достаточно универсальной по своему характеру. Форма ее правил должна быть такой "богатой" и гибкой, чтобы допускать удобную и наглядную фиксацию любых синтаксических фактов. В этом смысле наиболее широко применяемые сейчас грамматики Хомского представляются мало пригодными.

(iv) Механизм анализирования, гарантирующий получение всех правильных и только правильных СС (при условии правильности грамматики) и не накладывающий никаких ограничений на порядок правил грамматики или на порядок их применения; анализ должен быть асинхронным.

(v) Экономность анализа: необходимы общие процедуры, обеспечивающие разумное эшелонирование анализа ("грубый", потом "тонкий"), сохранение ранее выявленных безусловно правильных фрагментов структуры и т.п.

(vi) Эффективная система обслуживания, обеспечивающая контроль над внутренним состоянием анализатора при всех его модификациях.

(№ : все эти требования естественно предъявлять и ко всем прочим компонентам современных систем АП.)

В заключение данного раздела мы перечислим некоторые работы, посвященные описанию синтаксиса конкретных языков для целей АСАн ("работы второй группы", см. стр. 31); их представляется удобным группировать по языкам.

- А н г л и й с к и й язык: исследования, посвященные строению и способам обнаружения во фразе именных групп [79] , [97] , обработке предложных групп [82], некоторых абсолютных конструкций и аналитических форм глагола [86] , членению текста на синтаксические группы [15] , а также функциям знаков препинания [55] и отысканию антецедентов местоимений-заместителей [90] .

- Р у с с к и й язык: работа по синтаксису не-языковых выражений ("формул") [27] и подробное описание синтаксических форм и функций наречий [155] ; разбор основных типов синтаксической неоднозначности [187] ; полный алгоритм анализа, ориентированный на тексты ограниченной тематики [36] .

- Н е м е ц к и й язык: работы по синтаксису отрицания [201] и обстоятельственных групп [173] ,[189] ; синтаксическая неоднозначность [176] .

- Ф р а н ц у з с к и й язык: классификация слов [100] , в частности, - предлогов [99] , для целей АСАн; описание различных типов вопросительных конструкций [115] .

- Наконец, вопрос о непроективности в ч е ш с к о м языке рассмотрен в [190] .

Необходимо подчеркнуть, что здесь указаны далеко не все работы соответствующей тематики, а лишь их небольшая часть.

3. Семантический анализ.

Мы кратко охарактеризуем три системы АСемАн:

1) Семантический анализатор Й.Уилкса.
2) Семантический анализатор Р.Шенка.
3) Естественно-семантический анализ Н.Н.Леонтьевой.

1) Анализатор Й.Уилкса ([203],[204],[205]) - система CSD (Computable Semantic Derivation), или PSS (Preference Semantic system) , - не предполагает предварительного морфологического или синтаксического анализа: СемП строится непосредственно по входному тексту, лишь расчлененному по союзам и знакам препинания на сегменты (= потенциальные элементарные высказывания). CSD использует словарь словоформ. Смысл каждого слова описывается "семантической формулой" СемФ - такой цепочкой семантических элементов (типа ЧЕЛОВЕК, ЖИВОЕ, КАУЗАЦИЯ, ...) с заданной на ней НС-структурой, в которой для каждой составляющей указан главный элемент этой составляющей - самый правый в ней элемент. Например, СемФ для слова стрелять имеет вид
((ЧЕЛОВЕК СУБЪЕКТ) ((ЖИВОЕ ОБЪЕКТ) ((ПОРАЗИТЬ ЦЕЛЬ) ((ВЕЩЬ ПЕРЕМЕЩАТЬСЯ) КАУЗАЦИЯ)))),
что означает: 'Человек-субъект каузирует перемещение вещи [т.е. пули] с целью поразить живой объект'.

Кроме того, для каждого слова указываются семантические признаки (= элементы) других слов, способных выполнять при данном слове те или иные семантические роли (ср. семантические условия заполнения валентностей у Н.Н.Леонтьевой, см. ниже, стр.42).

CSD осуществляет анализ посредством набора шаблонов (templates) - типовых трехчленных элементарных высказываний (т.е. простейших семантических синтагм). На анализируемый сегмент текста (в котором словам уже приписаны их СемФ) поочередно накладываются шаблоны - так, чтобы обнаружить в тексте три главных элемента каких-либо СемФ, совпадающие с одним из шаблонов; к этой предварительной семантической конфигурации применяются стандартные процедуры, устанавливающие связи между вошедшими в нее СемФ и прочими СемФ данного текстового сегмента. Затем с помощью особых "надшаблонов" устанавливаются межсегментные и анафорические связи. В результате получается СемП, о котором [illegible] говорится ниже, стр.49 ; здесь мы еще отметим, что в случаях неоднозначности текста всегда выбирается более плотное СемП, т.е. СемП с наибольшей повторяемостью семантических элементов (отсюда и название Preference Semantics) . Система [(= PSS)] CSD была реализована на ЭВМ IBM 360; в книге [204] приводятся результаты машинного семантического анализа десяти небольших текстов - газетных и

философских. (Представляется необходимым отметить, что АСемАн Уилкса является непосредственным продолжением работ Кэмбриджского Лингвистического кружка, основанным, в первую очередь, на идеях М.Мастерман (см. [71], с. 339-345.)

2) Семантический анализатор Р.Шенка, используемый в составе системы MARGIE (Meaning Analysis, Response Generation, and Inference in English) , также не предполагает отдельного этапа синтаксического анализа ([66], с. 41-42). Он построен из стандартных операторов трех типов: (i) операторы, работающие с СемП, в том числе - поиск узла по заданной роли, вставка СемП слова в СемП фразы и т.п.; (ii) операторы проверки синтаксических связей, в том числе - поиск сильноуправляемых групп, выделение именных групп и т.п.; (iii) операторы обращения к семантической памяти, в том числе - проверка того, обладает ли данный объект данным свойством и т.п. Управляющая программа обеспечивает последовательный перебор всех значений всех слов анализируемой фразы и для каждого очередного значения выполняет инструкции, содержащиеся в его словарной статье. Так, глаголу advise в словаре сопоставлены три инструкции: одна задает СемП этого слова, вторая предписывает искать по фразе группу to + V_{inf} и вставить СемП этой группы в определенную точку в СемП слова advise , а третья гласит, что ближайшее справа обозначение человека также должно быть вставлено в определенную точку конструируемого СемП всей фразы, ср. John advised Mary to go to bed with him . Таким образом, вся необходимая для анализа информация дается в словаре - в виде инструкций при слове.

3) Естественно-семантический анализ (ЕСА) Н.Н.Леонтьевой (см. [51] , где указаны и более ранние работы Леонтьевой, а также [52]) рассчитан на использование в составе ИПС любой степени сложности; ЕСА обеспечивает перевод отдельных фраз на ИЯ ЭХО (см. о нем ниже, стр.48). ЕСА существенным образом опирается на словарь, в котором каждому русскому слову приписано 5 характеристик: перевод слова на язык ЭХО, т.е. дескриптор; семантический признак этого дескриптора; набор его валентностей (ср. падежи Ч.Филлмора); семантические, синтаксические и морфологические условия заполнения

валентностей; и сведения о несовместимости определенных заполнений тех или иных валентностей (три последние характеристики – это, по существу, модель управления толково-комбинаторного словаря, см. стр.61). Анализ состоит в последовательном рассмотрении всех пар слов фразы и проверке (на основе их словарных информаций) возможности связать их семантическими отношениями. При этом анализ разбит на пять этапов, каждый из которых осуществляется отдельным блоком анализатора; эти этапы примерно соответствуют этапам, традиционно выделявшимся в синтаксических анализаторах первого поколения: на этапе I связываются S с препозитивными A, на этапе II – S с контактным атрибутом (S + $S_{род}$, S + Prep + S) , на этапе III – $S_{им}$ с $V_{личн}$, S с постпозитивным распространенным определением и т.п., на этапе IV обрабатываются слабоуправляемые предложные группы, а на этапе V – однородные члены. ЕСА может выдавать несколько гипотетических анализов фразы; разрешение этой омонимии предполагается осуществлять в дальнейшем в ходе построения приведенного СемП целого текста. Блоки I-IV ЕСА были запрограммированы на ЭВМ "Минск-22" и проверены в эксперименте по переводу на ЭХО двадцати заголовков рефератов из РЖ.

Кроме трех только что охарактеризованных систем АСемАн, известен еще ряд семантических анализаторов, входящих в более общие системы АОЯТ; в особенности многочисленны работы по АСемАн в сфере информационного поиска (например, [26]) и, шире, в той области исследований, которую называют искусственным интеллектом (см., в частности, обзор [164]).

Б. Представление текста.

Что касается проблемы формального изображения результатов ААн и исходных данных для АСинт, то в данной связи наиболее ценными результатами исследований 1970-1974 гг. являются, пожалуй, два следующих:

– осознание необходимости отчетливо разграничивать синтаксическое и семантическое представления текста, а в синтаксическом представлении различать, в свою очередь, глубинный и поверхностный уровни;

– осознание того факта, что представление текста на данном уровне не обязательно гомогенно: оно само может складываться из нескольких различных по своей природе компонентов, характеризующих разные аспекты текста, рассматриваемого на интересующем нас уровне.

Начнем с синтаксического представления. Здесь полезно рассмотреть следующие четыре предложения, выдвинутые в 1970–1974 гг.:

1) Системы синтаксических групп А.В.Гладкого.
2) "Грамматическая структура" фразы Ю.Кунце.
3) Синтаксическое представление в МСТ.
4) Синтаксическое представление Т.Д.Корельской – Е.В.Падучевой.

1) В статье А.В.Гладкого [16] строится комбинированный аппарат изображения ПСС, представляющий собой обобщение понятий "дерево составляющих" и "дерево зависимостей" – системы синтаксических групп ССГ. Синтаксической единицей у Гладкого является не только слово, но и группа слов (= словосочетание), притом, быть может, разрывная. Формализм Гладкого позволяет:

(i) одинаково удобно изображать как те синтаксические связи, для которых направление релевантно, так и по существу ненаправленные связи (например, будем называть; газета и журнал);

(ii) легко изображать связи между группами слов (которые, в свою очередь, могут иметь внутреннюю структуру);

(iii) избежать в ССГ непроективности (с точки зрения формализма Гладкого проективность – в обобщенном смысле – оказывается универсальным свойством, не знающим исключений);

(iv) описать некоторые до тех пор не отмечавшиеся случаи синтаксической омонимии; например, Он не работал весь вечер = 1) 'Он (не работал)' ⟶ (весь вечер), 2) 'Он не ⟵ (работал весь вечер)'.

Примеры представления ССГ русских фраз посредством ССГ:

$Я_1$ $буду_2$ $там_3$ $работать_4$ $завтра_5$: {2, 4} = A; A⟶1, A⟶3, A⟶5;

$Он_1$ $не_2$ $прочел_3$ $ни_4$ $газету_5$, $ни_6$ $журнал_7$: {2, 3} = A, {4, 5, 6,

7} = B , {4,6} = C, A → 1, A → B, 3 → 2;

O_1 докладе$_2$ дала$_3$ отзыв$_4$ комиссия$_5$: {1, 2} = A, {3, 4} = B,
B → 5, B → A, 3 → 4.

2) Ю.Кунце рассматривает грамматическую структуру фразы как состоящую из четырех компонентов: 1) парадигматика (словоизменительные грамматические значения слов); 2) сочетаемостные ограничения (сведения об управлении и т.п.); 3) локальное словорасположение (постпозиция/препозиция слова, связанного с данным); 4) собственно синтаксическая структура - отношение подчинения между словами (не зависящее от прочих компонентов). Ю.Кунце сформулировал критерии тождества грамматической структуры двух фраз и эквивалентности фраз по парадигматике, сочетаемостным ограничениям и порядку слов, а также предложил критерий правильности выделения типов зависимостей: пусть имеется два дерева зависимостей T_1 и T_2, причем $t_1 \subset T_1$ висит на ветви $\xrightarrow{P}$ и $t_2 \subset T_2$ тоже висит на $\xrightarrow{P}$; если P - "хорошо" выделенный тип, то для любых t_1 и t_2 при замене t_1 на t_2 (или наоборот) должно получаться дерево, соответствующее правильной фразе (при этом в t_2 и в остающейся части дерева T_1 можно производить любые изменения в парадигматике, сочетаемостных ограничениях и порядке слов) [158].

3) В рамках модели "Смысл ⇔ Текст" [67] предлагается рассматривать сложный объект, называемый синтаксическим представлением СП (фразы); различаются глубинное СП (= ГСП), изображающее строение фраз максимально обобщенно, в универсальных терминах, и поверхностное СП (= ПСП), изображающее строение фраз максимально конкретно, как можно ближе к их реальному виду. И то и другое СП состоят из четырех компонентов (глубинных - Г- и поверхностных - П-): синтаксическая структура, описывающая зависимости между словами; коммуникативная структура КС, описывающая членение фразы на тему - рему, данное - новое, выделенное - невыделенное; просодическая структура ПрС, описывающая просодику фразы; и анафорическая структура АнафС, описывающая кореферентность тех или иных узлов.

ГСС - это неупорядоченное* дерево зависимостей, узлы кото-

* Порядок слов является синтаксическим средством (наподобие показателей согласования и управления), и потому он не должен отображаться в СС.

рого помечены символами глубинных лексем (обычные знаменательные лексемы, идиомы, лексические функции), а ветви - символами девяти универсальных глубинно-синтаксических отношений; ПСС - такое же дерево, но ее узлы помечены символами всех реальных лексем представляемой фразы и только ими, а ветви - символами конкретных синтаксических конструкций данного языка (их - около пятидесяти). ГКС и ПКС очень сходны и различаются, в основном, тем, что в ПКС отображаются служебные слова и конструкции - показатели актуального членения. ГПрС задает только семантически нагруженную просодику (вопрос, восхищение, угроза, ...), и ПрС - всю просодику, в том числе - и синтаксически обусловленные просодические явления. В ГАнафС указывается кореферентность глубинных лексем как таковых, в ПАнафС указываются местоимения-заменители.

Охарактеризованное здесь СП предполагается, в частности, в качестве средства фиксации результатов анализа и исходных данных синтеза в системе АРАП, см. выше, стр. 17 .

4) Аналогичное понимание СП как сложного объекта выдвинуто Т.Д.Корельской и Е.В.Падучевой [38]; см. также [73], с. 71-72. СП трактуется ими как состоящее из трех компонентов: это синтаксическая структура (дерево зависимостей с размеченными ветвями, близкое к ГСС в МСТ, поскольку, в частности, в нем не фигурируют служебные слова), анафорическая структура и линейная структура (= сведения о порядке слов, которые, таким образом, здесь включаются в состав СП).

С е м а н т и ч е с к о е п р е д с т а в л е н и е.

В 1970-1974 гг. как в рамках АП-лингвистики, так и в смежных областях было предложено много разных способов и методов изображения смыслового содержания текстов. Важнейшие (или, во всяком случае, наиболее известные) из этих способов сводятся к следующим четырем подходам:

1) Семантическая сеть и ее всевозможные разновидности (Р. Симмонс, Т.Хоффман, Ш.Клейн, Д.Хейс, Н.Н.Леонтьева, З.Ф.Скороходько и др.).

2) Семантический квазиграф (Р.Шенк).

3) Семантические формулы (Й.Уилкс).

4) Семантический граф (МСТ).

1) Семантическая сеть (**semantic net(work)**) – одно из наиболее популярных в настоящее время средств представления содержания высказываний. Обычно семантическая сеть – это набор лексем рассматриваемого естественного языка (снабженных семантически наполненными морфологическими характеристиками) и связывающие их бинарные СемО типа быть субъектом, быть объектом, быть частью, быть местом, потому что и т.п. (обычно используется до нескольких десятков СемО). Таким образом, смысл передается здесь как узлами (= лексемы), так и дугами (= СемО) сети.

Техника обращения с семантической сетью хорошо иллюстрируется в статье Р.Симмонса и Дж.Слокема [185]. Сеть записывается в виде системы троек $\langle C_1, R, X \rangle$, где C_1 – номер некоторого узла, R – перечень всех СемО, которые связывают данный узел со всеми прочими узлами данной сети, со словарем и с морфологическими характеристиками, а X – перечень всех правых членов отношений из R. Например, сеть для фразы **John saw Mary** имеет следующий вид:

C_1	TOK	see	C_2	TOK	John	C_3	TOK	Mary
	TIM	PAST		NBR	SING		NBR	SING ;
	DAT	C_2		DAT^{*}	C_1		OBJ^{*}	C_1
	OBJ	C_3						

здесь **TOK** – отношение между узлом и словарной статьей, **TIM** – между узлом и грамматическим временем, **NBR** – между узлом и и грамматическим числом; DAT и OBJ – 'иметь адресатом' и 'иметь объектом', соответственно, а DAT^{*} и OBJ^{*} – конверсные отношения ('быть адресатом/объектом').

Т.Хофманн [142],[143] предложил использовать семантическую сеть текста в качестве важного вспомогательного средства при АП следующим образом. Перевод осуществляется в основном через ГСС фраз; однако после получения ГСС для входной фразы n эта ГСС интерпретируется, т.е. преобразуется в семантическую сеть, каковая присоединяется к сети, полученной ранее для всех предшествующих (n – 1) фраз. Всякий раз, когда требуется при-

влечение широкого контекста (отыскание антецедента местоимения, разрешение омонимии и т.п.), система АП обращается "за правкой" к полной сети предшествующего текста. (Аналогичная идея была в свое время выдвинута Ф.А.Дрейзиным [29].)

Ш.Клейн [152] также представляет содержание высказываний семантической сетью, вершины которой помечены именами объектов - семантических единиц СемЕ, а дуги - именами бинарных СемО или унарных свойств (которые Клейн называет, для общности, тоже СемО). Сеть записывается в ЗУ ЭВМ в виде системы семантических троек (**semantic triples**) : $СемЕ_1$ - СемО - $СемЕ_2$, СемЕ - СемО, $СемЕ_1$ - $СемО_1$ - ($СемЕ_2$ - $СемО_2$ - $СемЕ_3$) и т.п. Для каждого СемП специальной таблицей задаются "отношения транзитивности" между определенными парами вершин; например, если X - в Y-е, а Y - в Z-е, то X - в Z-е.

Язык ЭХО Н.Н.Леонтьевой, язык **RX**-кодов Э.Ф.Скороходько [84], да и вообще все семантические языки, различающие объекты (единицы) и отношения и передающие их разными средствами, фактически все относятся к тому же классу: предлагаемое в них СемП высказываний есть семантическая сеть.

Ряд общих соображений о семантических сетях см. в [139].

2) Семантический квазиграф Р.Шенка близок к семантической сети (так как и в нем противопоставляются объекты и отношения), но имеет, в то же время, по крайней мере три специфических особенности:

- В нем используются не только бинарные (типа 'быть субъектом', 'быть атрибутом'), но и тернарные ('X переходит от Y-а к Z-у') и даже кватернарные ('состояние X-а по параметру Y изменилось с Z_1 на Z_2') СемО.

- Главный упор делается Шенком на м и н и м и з а ц и ю набора СемО; по Шенку, их всего 10.

- Узлам квазиграфа приписываются "глубинные" части речи (аналоги **V, S, A** и **Adv**), и на этой основе формулируются правила образования квазиграфов: утверждения о том, какие классы семантических единиц (т.е. единицы какой "части речи") могут участвовать в каких СемО. Таких правил - около 20, и они составляют ядро семантического анализатора в системе Р.Шенка.

Кроме того, Шенк выдвигает и задачу семантического разложения слов – изображения их смысла в терминах элементарных смысловых компонентов (см. [66], с. 35-39), чего не было еще в большинстве "сетевых" семантических языков (но не во всех! Так, в языке RX-кодов Э.Ф.Скороходько глубокое семантическое разложение слов предполагалось с самого начала).

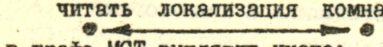
Таким образом, СемП Р.Шенка оказывается семантичнее стандартных семантических сетей, которые часто не так уж далеко уходят от СП с семантически проинтерпретированными связями.

3) Семантические формулы Й.Уилкса (см. выше, стр.41) еще более семантичны в том отношении, что они изображают смысл высказываний безотносительно к членению на слова и безотносительно к противопоставлению "объекты vs. отношения". СемП Уилкса действительно ориентировано на представление самого смысла, освобожденного от конкретно-языковой формы; это не удивительно, если учесть, что оно восходит к глубокому пионерскому учению М.Мастерман о семантике естественных языков (см. выше, стр.42).

4) Семантические графы СемГ в МСТ ([67], с. 62-69) должны быть максимально семантичными представлениями высказываний. Эти СемГ суть естественное развитие принципов, сформулированных в [31], а затем более детально разработанных А.К.Колковским.

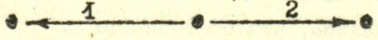
Укажем здесь три существенные особенности семантических графов в МСТ:

– Вся содержательная информация кодируется в СемГ пометами исключительно при узлах (= вершинах); стрелки (= дуги) несут лишь чисто дифференциальные, не знаменательные пометы, позволяющие различать разные "места" одного и того же предиката, но ничего не значащие сами по себе. То, что в стандартной семантической сети записывается как



(т.е. 'читать в комнате'),

в графе МСТ выглядит иначе:



Благодаря этому, в МСТ нет таких асемантичных отношений, как

'быть субъектом' или 'быть объектом' (каковые по-просту являются проекцией в семантику поверхностно-синтаксических отношений); в с е отношения (предикаты, свойства) представляются здесь одинаковым образом: узлами. Это различие, формально не слишком существенное (помету на дуге всегда можно тривиальным образом превратить в помету на новом узле) оказывается содержательно очень важным: фактически оно имплицирует иную семантическую концепцию.

– Семантическими единицами в СемГ являются не слова, а смысловые атомы, т.е. э л е м е н т а р н ы е порции смысла. Поскольку в СемГ нет слов, невозможна и семантическая характеризация связей между словами: связи в СемГ имеют место между столь элементарными смыслами, что на долю связей остается сугубо "синтаксическая", организующая роль. (Ср. неинтерпретированные валентные связи, соединяющие атомы в химических структурных формулах.) Синтаксическая организация СемП – вещь исключительно важная, но при достаточной семантичности СемП ее синтаксис сам уже не может быть семантичным.

– Смысл любых высказываний (слов, словосочетаний, фраз, ...) представляется одинаково и совершенно безотносительно к языковому оформлению этих высказываний.

Мы рассмотрели здесь не все существующие способы представления смыслового содержания текстов; краткий обзор таких способов, применяемых в рамках АОТ, можно найти в лекциях Т.Винограда [209], с. 8–13, где упомянуты способ концептуализаций У.Чейфа, "сценарии" Р.Эйбелсона, метод демонов М.Чарняка и кольцевые структуры Д.МакДермотта.

Вне тематических рамок нашего обзора осталась глубокая семантическая теория А.Вежбицкой [202] и предложенное ею СемП – lingua mentalis , семантическая запись П.Сгалла [83] , а также ряд других теорий, выдвинутых в лингвистике и психологии.

В. Собственно перевод.

Выше уже отмечалось, что большинство систем АП, разрабатывавшихся в рассматриваемый период, относится ко второму поколению, т.е. эти системы являются по преимуществу синтаксическими: пере-

вод осуществляется в них через СС фраз. Тем самым, перевод в узком смысле слова оказывается в значительной степени сведенным к преобразованию синтаксических структур, т.е. деревьев некоторых специальных видов. Этот факт, а также бурное развитие теории синтаксических трансформаций в теоретической лингвистике вызвали к жизни в рамках исследований по АП новую проблему, которую удобно обозначить как п р е о б р а з о в а н и е д е р е в ь е в.

В то же время по-прежнему оставалась очень острой задача в ы б о р а л е к с и ч е с к и х э к в и в а л е н т о в при переводе. Соответственно, мы выделим в данном разделе два подраздела: преобразование деревьев и выбор лексических эквивалентов.

П р е о б р а з о в а н и е д е р е в ь е в.

Отметим, прежде всего, что преобразование деревьев выполняется в процессе АП далеко не только при переходе между языками, т.е. не только в качестве преобразования СС входной фразы в СС выходной: аналогичные преобразования могут иметь место и внутри одного языка, при анализе и синтезе (переход от ПСС к ГСС или наоборот). Тем не менее, мы рассмотрим здесь, в разделе "Собственно перевод", любые преобразования деревьев, безотносительно к конкретному характеру их использования в той или иной системе АП.

Из работ, посвященных в 1970-1974 гг. проблематике преобразования синтаксических деревьев в рамках АОТ, мы очень кратко изложим следующие пять:

1) Q-системы А.Кольмероэра.
2) Δ-грамматика А.В.Гладкого - И.А.Мельчука.
3) R-грамматика О.С.Кулагиной.
4) Трансформационная система Т.Д.Корельской - Е.В.Падучевой.
5) Древесные преобразователи Ж.Шоше.

1) Под Q-системами [122] понимаются формальные грамматики, описывающие переработку деревьев и множеств деревьев. Точнее, Q-система есть совокупность правил вида $a_1 + a_2 + \ldots + a_p \Longleftrightarrow b_1 + b_2 + \ldots + b_q \mid$ УСЛОВИЯ, где a_i и b_j - деревья с помеченными узлами, а УСЛОВИЯ - булевы формулы, определяющие допустимые соотношения между свойствами деревьев в левой и правой час-

сти. На языке Q-систем удается записать все грамматики английского и французского языков, составляющие лингвистическую часть системы англо-французского АП Монреальского ун-та (см. выше, стр. 38). Применение любых Q-систем в процессе АОТ осуществляется универсальным Q-преобразователем, разработанным и поставленным на ЭВМ CDC-6400 в конце 1969 г. под руководством А.Кольмерсэра. Таким образом, Монреальская система АП четко распадается на "грамматику" и "механизм": грамматика здесь - это 15 Q-систем, написанных лингвистами и задающих все сведения, необходимые для АМАн и АСАн английского текста, собственно перевода, и АССинт и АМСинт французского текста; а механизм - Q-преобразователь.

2) В статье [18] вводятся формальные грамматики, ориентированные на переработку деревьев зависимостей, - так называемые Δ-грамматики. Они основаны на операции подстановки в деревьях (т.е. являются, в некотором смысле, обобщением грамматик Хомского на деревья); выделены, в частности, три типа специальных элементарных синтаксических операций: отбрасывание узла (= стяжение ветви в узел), добавление узла (= "растяжение" узла в ветвь) и переподчинение узла (= перевешивание ветви). Доказывается, что произвольное элементарное преобразование дерева моделируется той или иной последовательностью специальных операций трех указанных типов. Благодаря этому из специальных операций может быть построена универсальная синтаксическая Δ-грамматика. В [19] дана развернутая иллюстрация: Δ-грамматика, описывающая переход от ГСС к ПСС в русском языке. См. также [67], с. 162-166, 237-256.

3) Аналогичная задача решается О.С.Кулагиной в [42]: вводятся понятия, необходимые для точного формулирования правил преобразования деревьев зависимостей (правил ПДЗ), и определяется вид этих правил; в [42] рассматривается, в частности, и вопрос о представлении произвольных ПДЗ через специальные элементарные правила трех типов: замена ветви на узел, замена узла на ветвь и замена одной ветви на другую. О.С.Кулагиной определяется также трансформационная грамматика для деревьев ТГД (= совокупность правил ПДЗ), для множеств деревьев ("обобщенные λ-деревья") и R-грамматика - такая ТГД, в которой на порядок применения правил наложены определенные ограничения (путем объединения правил в "цепочки",

указания отсылок от одних правил или "цепочек" к другим и задания запрещенных последовательностей правил или цепочек; ср. матричные грамматики Абрахама, программированные грамматики Розенкранца и т.п.: [17], с. 111-115). R-грамматика, содержащая только элементарные правила, называется элементарной; доказывается, что для всякой R-грамматики можно построить полностью эквивалентную ей элементарную R-грамматику. Кроме того, в [42] определяются алгоритмы ПДЗ, конструируемые из правил предложенного вида, и освещаются пути их возможной машинной реализации. Теоретические положения данной работы практически использованы в системе АП "ФР-П", о которой шла речь выше.

4) Несколько иначе, чем в [18] и [42], трактуются преобразования деревьев Т.Д.Корельской и Е.В.Падучевой [38]. По существу, в [38] рассматриваются преобразования не просто деревьев, а более сложных объектов - синтаксических представлений, состоящих из трех компонентов (см. выше, стр.46). Синтаксическое преобразование задается парой ⟨условие применимости УП, собственно преобразование Т⟩; УП - это набор требований к преобразуемому СП (не только к дереву!), а Т - цепочка элементарных преобразований (среди которых различаются минимальные - задаваемые прямо указанием результата их применения, и производные - определяемые через минимальные). Конструкция, предлагаемая в [38], достаточно близка к трансформационной теории современной лингвистики, где трансформация определяется как пара

⟨structural description, structural change⟩.

5) В работах [118] и [119] исследуются различные математические аспекты преобразований синтаксических деревьев при АОТ: представимость подобных преобразований в терминах конечных преобразователей, композиция преобразователей и т.п.

В 1970-1974 гг. было опубликовано также большое количество работ по проблеме формального описания преобразований синтаксических структур вне прямой связи с АП или вообще АОТ; из таких работ мы отметим здесь только [103] и ограничимся отсылкой к [17], где на стр. 128-133 дан достаточно подробный обзор "древесных" грамматик.

Выбор лексических эквивалентов.

Среди многочисленных работ на эту тему мы выделим статью [33], где показано, как с помощью аппарата лексических функций и системы перифразирования [32] можно автоматически получать лексические эквиваленты в случаях существенной лексической неконгруэнтности двух языков (She nagged him into his grave ⇒ Она запилила его до смерти, We talked him out of going to Bermuda ⇒ Мы отговорили его от поездки на Бермуды).

За исключением названной статьи (а также статьи И.Беллерт [105], рассматривающей использование лексических функций во французско-английском словаре) остальные известные нам работы шли по традиционному пути: лексическая неоднозначность разрешается в основном за счет перечней "диагностирующих" контекстов, т.е. лексемных сочетаний. Так, в [85] для обработки 50 английских неоднозначных лексем предлагается 50 подпрограмм (по одной на слово) и 90 списков (конкретных слов или слов с определенным семантическим признаком; таких признаков - 66). Совершенно та же методика предложена в [14] и в [57]. С.Вейс [200], подробно рассмотрев известные ему способы разрешения лексической неоднозначности, все же останавливается на списках "диагностирующих лексем" (английский язык).

Статья Г.Циммерманна [215] посвящена разрешению лексической неоднозначности в немецком языке на основе различия морфологических и синтаксических свойств разных значений, а статья А.Роткегель [175] - на основе вхождения слова в том или ином значении в устойчивое словосочетание. Однако в рамках систем АП второго (а тем более - третьего) поколения отдельного этапа разрешения лексической неоднозначности за счет морфологии или синтаксиса (в широком смысле, включая в синтаксис и фраземы) быть не может: "глобальные" анализаторы на уровне морфологии и синтаксиса автоматически обеспечивают выбор всех тех и только тех значений (рабочих слов), словарная информация которых позволяет им выступать в данном месте словоформы или, соответственно, синтаксической структуры. Поэтому направление исследований типа [57] (НАИ, например, использование сильного управления специально и особо для разрешения лексической неоднознач-

ности) представляется бесперспективным. По мнению автора обзора, имеются три пути к решению проблемы лексической неоднозначности (мы имеем здесь в виду ту неоднозначность, которая не может быть разрешена за счет морфологии и синтаксиса):

- Широкое использование аппарата лексических функций (стандартных и нестандартных).

- Использование СемП лексем и проверка семантической "согласованности" лексем анализируемой фразы - на основе аксиом семантической правильности высказываний (ср. [49]), на основе взаимных семантических требований лексем ([51],[93]) и т.п.

- Использование энциклопедических сведений (например, банка данных ИПС, подключенной к системе АП).

Первый путь легко осуществим в настоящее время, второй - намного сложнее, но все же (по крайней мере - частично) тоже осуществим уже сейчас, а о третьем пока можно говорить лишь в аспекте первых и еще робких попыток. Наиболее практичным изо всех реальных в настоящий момент способов разрешения лексической неоднозначности (для систем АП третьего поколения) представляется - в тех случаях, когда лексические функции помочь не могут, - обращение ЭВМ за справкой к одноязычному предредактору (как это делает DISAMBIGUATOR в системе MIND) или одноязычному постредактору (как это предполагается в проекте системы АРАП).

Особый интерес в связи с рассматриваемым здесь этапом процесса АП представляют две следующие статьи, специально посвященные полному преобразованию результатов анализа в исходные данные синтеза (в составе систем АП).

В [50] Б.М.Лейкина рассматривает этап преобразования в экспериментальной системе англо-русского АП ВЦ ЛГУ (см. выше, стр. 15). Преобразование выполняется двумя механизмами: ПП и ПШ. ПП производит три операции: 1) браковку неправильных вариантов синтаксического анализа (на основе привлечения дополнительной семантической информации), 2) интерпретацию оставшихся вариантов (выявление "истинных" СС из "спрессованного" представления нескольких альтернатив; выявление семантики грамматических форм; выработка информации о типах синтаксической связи и об актуальном членении; разрешение лексической неоднозначности) и 3) нор-

мализацию оставшихся вариантов - "приведение к ЯП" (элиминация всех служебных слов, стандартизация специфически английских конструкций и т.п.). ПП также производит три операции: 1) выбор нужной основы для каждого узла выходного дерева, 2) грамматическую интерпретацию и преобразование СС в соответствии с требованиями русского языка и 3) браковку неправильных выходных вариантов. Преобразование в данной системе АП является множественным (как и все прочие этапы АП) и выполняется на основе формальных "преобразующих" грамматик, включающие правила (= трансформации) четырех основных типов (изменение информации об узле, переподчинение узлов, добавление узлов, элиминация различных объектов).

В [123] полностью приведена грамматика (в виде двух Q-систем), обеспечивающая переход от английской ГСС к французской ГСС. Само преобразование выполняется в три этапа: 1) стандартизация (= "деидиоматизация") английской ГСС; 2) лексический переход; 3) идиоматизация (= "дестандартизация") входной ГСС - превращение ее во французскую.

Укажем в заключение статьи [135] и [136] как образцы работ, посвященных семантике грамматических категорий и проблеме их перевода при переходе от одного языка к другому.

Г. Синтез.

1. Семантический синтез. Мы охарактеризуем здесь три действующие системы АСемСинт:

1) Система Р.Симмонса.
2) Система Ш.Клейна.
3) Система BABEL Н.Голдмана - К.Ризбека.

1) Система АСемСинт Р.Симмонса [185] преобразует задаваемую ей семантическую сеть (см. выше, стр. 47) в правильные английские высказывания. В словаре системы для каждой лексемы, наряду с грамматической информацией, указываются имплицируемые ею другие лексемы, что обеспечивает возможность синонимического перифразирования. Сведения о выходном (= английском) языке записаны в виде сетевой грамматики (см. выше, стр. 34), в которой узлы

помечены символами составляющих, лексем и грамматических морфем, а дуги - символами СемО. Каждому СемО сопоставлен стандартный оператор (записанный на LISP 1.5), который обеспечивает все допустимые в английском языке оформления этого СемО. Управляющий алгоритм перебирает произвольным образом узлы исходной сети и строит для выходящих из рабочего узла отношений все возможные выражения. Система реализована на ЭВМ CDC-6600 и опробована в ряде экспериментов.

2) Система АСемСинт Ш.Клейна [153] является центральным блоком в составе другой, более общей системы, моделирующей сочинение новелл (на английском языке). Системе задается исходный "мир", описанный набором семантических троек (см. стр.48) - перечень действующих лиц, отношений между ними в момент t_0 и т.п., т.е. СемП$_{t_0}$. Имеется набор аксиом действительности - правил, указывающих, что именно и с какой вероятностью может произойти в "мире" (например: "Если мужчине нравится девушка и они окажутся наедине, то он попытается поцеловать ее с вероятностью 1/3; если он тоже нравится ей, то вероятность поцелуя - 5/6"). На основе этих аксиом происходит преобразование исходного СемП$_{t_0}$ в ряд исследовательных СемП$_{t_1}$, СемП$_{t_2}$ и т.п. - моделируется развитие "мира". Блок АСемСинт переводит СемП$_{t_1}$ в английский текст и так автоматически получается новелла; этот блок включает в себя следующие компоненты:

- семантическо-английский словарь, где именам вершин и дуг, используемых в СемП, сопоставлены английские лексические единицы;

- порождающая грамматика, строящая ГСС английских фраз и преобразующая эти ГСС в реальные фразы;

- механизм управления лексическим заполнением порождаемых структур (в них вставляются лексические единицы, "считываемые" с СемП, так что ГСС и получаемая из нее фраза сохраняют смысл этого СемП).

Вся система целиком, включая блок АСемСинт, была реализована на ЭВМ Burroughs B5500 , а затем - на UNIVAC 1108 ; она опробована в ряде экспериментов. См. ее подробное описание в где приведен полностью детективный рассказ (2100 слов) с ревно-

стью и убийством, написанный ЭВМ за 19 сек.

3) Система BABEL (= Better Analytic Basis for Encoding in Language; [131] и [132]) входит в состав описывавшейся выше системы MARGIE . Основной компонент системы BABEL - набор семантических правил $П_{сем}$ (= discrimination nets) . Левая часть каждого $П_{сем}$ - булева формула, составленная из определенных условий, а правая - имя английской лексемы. Условия в левых частях $П_{сем}$ бывают трех типов:

- проверки наличия в СемП определенных подграфов, т.е. описаний тех или иных объектов, явлений и т.п.;

- проверки наличия у объектов и явлений, упомянутых в условиях предыдущего типа, определенных свойств (выполняются в энциклопедической зоне памяти);

- проверки наличия в памяти системы определенных сведений о рассматриваемых предметах и явлениях.

По существу, левая часть $П_{сем}$ - это толкование некоторого английского слова, включающее и пресуппозиции, а правая - заглавие словарной статьи, в которой дается основа, морфологические сведения о ней и ее модель управления (= frame) , где указаны СО, выражаемые при данной лексеме, оформляющие их предлоги и соответствующие СемО (см. [132], с. 42-43).

С помощью набора $П_{сем}$ BABEL строит для исходного квазиграфа синтаксическую структуру, которая затем превращается в реальную фразу дополнительным механизмом, основанным на грамматике и программе Р.Симмонса [185].

Главная особенность системы BABEL по сравнению с другими действующими системами АСемСинт, в частности - с системами Симмонса и Клейна, состоит в том, что BABEL начинает свою работу с такого СемП, в котором еще нет ни английских слов, ни английских СО (или хотя бы чего-то, достаточно близкого к английским СО). Таким образом, BABEL является самой семантичной из всех известных в настоящее время действующих систем АСемСинт.

2. Синтаксический синтез.

По проблемам АССинт в 1970-1974 гг. было опубликовано достаточно много работ; мы бегло охарактеризуем некоторые из них, по-

священные, в первую очередь, лингвистическим аспектам АССинт. (О т д е л ь н ы е описания действующих систем или блоков АССинт автору обзора неизвестны, хотя соответствующие компоненты имеются в составе большинства действующих систем АП и многих систем АОТ.)

Глубинному АССинт для русского языка посвящена уже упоминавшаяся статья [19] : в ней приводятся правила (в виде правил Δ -грамматики, см. стр.52), описывающие переход от ГСС русских фраз к соотьетствующим ПСС. В [72] подробно рассматриваются возможности построения сложных русских предложений из пары простых; выделяются, в частности, типы семантико-синтаксических отношений между компонентами сложного предложения.

Этап АССинт для немецкого языка описан Л.С.Годабрелидзе в [22]. В качестве исходных данных здесь предполагается предикатное представление фразы - предикаты с заполненными местами, связанные между собой элементарными смысловыми отношениями (этих отношений - всего 20). Предикатное представление перерабатывается в ходе синтеза в четыре последовательных (все более поверхностных) представления:

- ядерное, состоящее из ядерных (= неразложимых) предложений;

- сегментное, состоящее из неминимальных цепочек, не содержащих, однако, разного рода обороты (= satzwertige Wortgruppen) [эти цепочки - прообразы простых предложений];

- сентенционное, состоящее из цепочек, отвечающих целым предложениям; и

- морфологическое: цепочка лексем с полной морфологической информацией, соответствующая фразе в ее реальном виде.

Лингвистические сведения, необходимые для осуществления АССинт указанного типа, даны в статьях [20] и [21].

К.Дансеро [127],[128] предлагает достаточно полную систему АССинт для французского языка; среди рассматриваемых им преобразований, необходимых при переходе от ГСС (= "нормализованная структура") к ПСС французской фразы, укажем такие, как переход от признака "ВОПРОС" к французской вопросительной конструкции Est-ce que... , замена фиктивного субъекта лексема-

мы on или cela, выбор пассивной конструкции (там, где это необходимо в связи с порядком слов) и т.п. Полученная ПСС морфологизуется и линеаризуется посредством применения неупорядоченного набора правил подстановки (их - около двухсот).

Вопросы линеаризации синтаксических деревьев при АССинт исследованы Д.А.Грикуровой на материале грузинского языка [25]: предлагается система автоматического определения порядка слов в некоторых типах грузинских именных словосочетаний.

Наконец, частный, но интересный вопрос: выработка информации об артикле в процессе АССинт немецкого текста (при АП с русского языка, где артикля нет!) - рассмотрен в статье И.Брандт [116].

3. Морфологический синтез.

Здесь мы можем, по-видимому, ограничиться несколькими краткими ссылками.

Полная система АМСинт для русского языка была разработана в связи с системой АП "ФР-Д" (см. выше, стр. 11) и функционирует в ее составе [37].

Синтез форм русского глагола описан в [76] и [156]; в последней статье все формы образуются от одной основы - в соответствии с известной теорией Р.О.Якобсона.

Полная система АМСинт для французского языка [117] включена в состав Монреальской системы англо-французского АП (стр. 12).

В [197] дана автоматически полученная морфологическая классификация всех эстонских глаголов из большого одноязычного словаря ("Õigekeelsuse sõnaraamat") ; приводимые списки глаголов лягут в основу системы АМСинт (а затем - и АМАн) эстонских глагольных словоформ.

Наконец, АМСинт для чешского языка кратко охарактеризован в [157] , а для азербайджанского языка (глагольные словоформы) - в [75].

В настоящее время АМСинт и последующая печать выходного текста завершают процесс АП; однако, возможно, будет нелишним упомянуть - в качестве заманчивой перспективы - и работы по автоматическому фонетическому синтезу: этот этап станет необходимым,

когда дело дойдет до устного АП. Пока мы можем ограничиться ссылками на статьи [88] и [125].

Д. Автоматические словари.

В области разработки словарей, которые должны обеспечивать системы АОТ всей необходимой информацией о с л о в а х привлекаемого языка (сведения об их семантике, синтаксических свойствах, морфологии и т.п.), мы отметим четыре следующие начинания:

1) Толково-комбинаторный словарь (ТКС).

2) Машинный словарь Дж.Олни.

3) Саарбрюкенские автоматические словари.

4) Частотно-машинные словари группы "Статистика речи".

1) ТКС был предложен еще в 1967 г. [32]. Одна из его главных особенностей - это достаточно формализованная схема словарной статьи, опирающаяся на лингвистическую концепцию "Смысл ⇔ Текст". Этой схемой для каждой словарной единицы - лексемы - предусматривается описание следующих аспектов:

- Морфология (склонение/спряжение).

- Семантика (толкование в терминах более простых по смыслу лексем, исключающее порочные круги и достаточно стандартизованное; в принципе возможен переход к СемП данной лексемы в виде семантического графа).

- Синтаксическая сочетаемость (с одной стороны, модель управления, задающая все активные валентности лексемы, способы их оформления - падежами, предлогами и т.п., степень обязательности их насыщения и т.п.; с другой стороны, синтаксические признаки, характеризующие все пассивные валентности лексемы. Каждая активная синтаксическая валентность соотнесена с семантической валентностью данной лексемы; кроме того, для синтаксических валентностей указаны семантические ограничения на их заполнение).

- Лексическая сочетаемость (значения всевозможных лексических функций данной лексемы.

Словарная статья ТКС содержит очень большое количество информации о слове. В этой связи любопытно отметить, что попытка провести автоматический анализ русского текста на основе сведений

только о модели управления (с использованием более чем примитивной грамматики - менее двух десятков правил) дала неожиданно хорошие результаты ([3]).

В целях машинной реализации словарей типа ТКС был разработан специальный формальный язык для записи словарных статей соответствующего формата [5] .

В настоящее время опубликовано довольно большое количество русских словарных статей ТКС и некоторое количество английских и польских словарных статей: см. [81] , где на стр. 37-46 дан перечень всех выпусков серии "Предв. публикации Проблемн. группы по экспер. и прикл. лингвистике при Ин-те Рус. Яз. АН СССР" за 1970-1974 гг., т.е., в частности, и всех опубликованных словарных статей, а также [34] и целую серию работ в [60] , посвященных словарным статьям англо-русского многоаспектного автоматического словаря, который разрабатывается в Лаборатории маш. перевода Моск. Гос. Пед. ин-та иностр. языков под руководством З.М.Шаляпиной (этот словарь является развитием ТКС).

Отметим, кроме того, разработку ТКС для системы французско-русского АП в "Информэлектро" (Москва) [2] , а также для грузинского языка [87].

2) Группа во главе с Дж.Олни [168] подготовила машинный словарь английского языка, представляющий собой введенное в ЭВМ объединение двух лучших английских словарей: седьмого издания словаря Вебстера и его карманного варианта. При этом закодирована практически вся лингвистическая информация, содержавшаяся в этих словарях, в том числе - толкования. Создана система программ, позволяющая сопоставлять толкования различных слов, а также заменять слова в составе толкований их собственными толкованиями; это обеспечивает выявление семантически элементарных слов и совершенствование словарных толкований на основе перифразирования текстов. Словарь снабжен также программой, строящей все формы всех представленных в нем лексем. Кроме того, предложено устройство автоматической проверки словаря - **Lexicon Tester** , обеспечивающий, во-первых, внесение в словари охарактеризованного типа любых изменений и дополнений, а во-вторых, - выдачу любых сведений о внутреннем состоянии словаря. Словарь Олни используют, в той или иной степени, более двадцати различных организаций в США.

Ориентация на семантику и хорошая система обслуживания (без которой, как уже отмечалось, построение больших машинных словарей и успешная работа с ними невозможны) – вот главные достоинства данного лексикографического предприятия.

3) В рамках разработки систем русско-немецкого АП и автоматического анализа немецких текстов коллективом Саарского ун-та (Саарбрюккен, ФРГ) подготовлены большой машинный немецкий словарь и машинный русско-немецкий словарь – достаточно типичные словари второго поколения. Немецкому словарю посвящена целая серия статей в книге [174]; рассматривается состав словарной информации к основам всех частей речи, принципы занесения в словарь производных слов и т.д. Этот словарь будет включать в себя все слова из "Deutsches Wörterbuch" G.Wahrig'a плюс все слова, обнаруживаемые в обрабатываемых текстах, но отсутствующие в этом словаре.

Русско-немецкий словарь подробно описан в [160]; это словарь коротких (не более четырех букв) словоформ и основ. О современном состоянии словарной системы Саарского ун-та сообщается в [216].

4) Группа "Статистика речи" (Ленинград) опубликовала в 1970-1974 гг. очень большое количество сообщений о подготовке автоматических словарей АС для ряда языков и широкого спектра тематики текстов. Соответствующие ссылки до 1972 г. включительно можно найти в [9]; сверх того, укажем еще [13], [24] и [78].

Принципиальные установки, на которые опирается словарная работа данной группы, изложены в статье [13]; грубо говоря, они сводятся к разделению двуязычного АС на словарь общеупотребительной лексики и сменные отраслевые словари; особый упор делается на автоматический отбор лексики (из текстов нужной тематики) по статистическим критериям, т.е. по частоте. Существенным признается также использование АС "оборотов" – наиболее частотных триад и тетрад (= последовательности из трех/четырех словоформ); предполагается [10], что подобный словарь может позволить обходиться без синтаксического анализа. Создан пробный англо-русский АС "оборотов" в 565 единиц, обеспечивающий покрытие текста на 5%.

Насколько можно судить по имеющимся публикациям, словари группы "Статистика речи" типологически относятся к первому (иногда –

ко второму) поколению АС.

В заключение мы отметим работу [112], посвященную созданию формализма для реализации и обслуживания больших АС. Вводится понятие "астрабонда" (ASTRABONDE= Arborescence STRAtifiée PONDérée) – расслоенного дерева с условными весами на узлах и предлагается записывать словарные статьи в виде астрабондов (дерево образуется иерархией значений и подзначений; слои – это разные типы словарной информации; а веса – вероятности появления тех или иных значений и т.п.). Формулируется набор стандартных операций над словарем – множеством астрабондов (сокращение астрабонда, сращение астрабондов; переупорядочение узлов и слоев; объединение и пересечение словарей и т.п.).

x

x x

Рукопись настоящего обзора была прочитана Л.Н.Иорданской и О.С.Кулагиной; автор выражает им свою искреннюю признательность за деловую критику и полезные советы, которыми он воспользовался в меру своих возможностей.

БИБЛИОГРАФИЯ

1. Автоматический перевод. Сб. статей. Ред. и предисл. Кулагиной О.С., Мельчука И.А. М., 1971. 368 с. (РЖИнф, 1971, 71.10.88).

2. Апресян Ю.Д., Богуславский И.М., Волковыская Е.В., Гецелевич Е.В., Крысин Л.П. Фрагменты толково-комбинаторных словарей французского и русского языков для автоматического перевода. (Отд-ие ВНИИЭМ по науч.-техн. информ. в электротехнике. Рукопись деп. в ОНИИЭМ 3 апр. 1974 г., № 471-Д). М., 1973. 200 с. (РЖИнф, 1974, 74.8.133 ДЕП).

3. Арсентьева Н.Г. Об одном подходе к анализу предложений русского языка. – В сб. "Маш. пер. и прикл. лингвист." Вып. 14. М., 1971, 136-164 (РЖМат, 1972, 6В614; РЖИнф, 1972, 72.5.98).

4. —— , Кулагина О.С., Международная конференция по вычислительной лингвистике (Пиза, Италия, 27 августа – 1 сентября 1973 г.) В сб. "Маш. пер. и прикл. лингвистика". Вып. 17. М., 1974, 196-207.

5. —— , Шаляпина З.М., Язык для записи лингвистической информации в автоматическом словаре. В сб. "Маш. пер. и прикл. лингвист." Вып. 16. М., 1972, 113-152 (РЖИнф, 1973, 73.11.105; РЖМат, 1973, 11.В932).

6. Барбакадзе М.М. Интерпретирующая система для алгоритма русско-грузинского машинного перевода (БЭСМ-4). В сб. "Маш. перевод" (Ин-т систем управления АН Груз.ССР. Труды. 9, № 3), Тбилиси, "Мецниереба", 1971, 78-111 (РЖИнф, 1972, 72.1.87).

7. —— , Программа интерпретирующей системы для алгоритма русско-грузинского машинного перевода (БЭСМ-4). В сб. "Маш. перевод" (Ин-т систем управления АН ГрузССР. Труды. 10, № 3), Тбилиси, "Мецниереба", 1972, 51-76.

8. —— , Некоторые программы интерпретирующей системы для алгоритма русско-грузинского машинного перевода (БЭСМ-4). В сб. "Маш. перевод" (Ин-т систем управления АН ГрузССР. Труды. 11-12, № 3), Тбилиси, "Мецниереба", 1974, 135-140 (РЖИнф, 1974, 74.10.92).

9. Бектаев К.Б., Статистика речи (библиогр. указатель). Алма-Ата, 1972. 129 с. (РЖИнф, 1973, 73.9.109).

10. ——— , Белоцерковская Л.И., Борисевич А.Д., Букович В.А., Булашева Н.С., Гончаренко В.В., Данейко М.В., Зуева Т.Р., Исабекова Н.И., Нехай О.А., Пиотровский Р.Г., Сорокина В.А., Статистическое выделение русских машинных оборотов и построение двуязычных автоматических словарей. В сб. "Статистика речи и автомат. анализ текста 1972". Л., "Наука", 1973, 107-130 (РЖИнф, 1974, 74.5.142).

11. Василевский А.Л., Поволоцкая С.К., Самойлович М.В., Тарасова Г.А., Эмдина Ю.М., Экспериментальный алгоритм автоматического перевода патентной документации с английского языка на русский. "Научно-техн. информ. сб. Всес. ин-т научн. и техн. информ.", 1971, сер. 2, № 1, 29-34 (РЖИнф, 1971, 71.6.110).

12. ——— , Васильева Л.Я , Поволоцкая С.К., Самойлович М.В., Тарасова Г.А., Эмдина Ю.М., Алгоритм автоматического перевода описаний изобретений с английского языка на русский. (Тр. ЦНИИ патент. информ. и техн.-экон. исслед. Сер.3). М., 1972. 157 с. (РЖИнф, 1973, 73. 1.103).

13. Вертель В.А., Вертель Е.В., Крисевич В.С., Пиотровский Р.Г., Трибис Л.И., Автоматические словари в системе бинарного вероятностного МП. Уч. зап. Ленингр. гос. пед. ин-т им. А.И. Герцена, 1971, т. 458, ч. 1, 4-147 (РЖИнф, 1972, 72.2.96).

14. Вишнякова С.М., Устранение лексической многозначности при автоматическом переводе патентных текстов с английского языка на русский. В сб. "Автомат. анализ патентн. текстов", М., 1971, 181-190 (РЖИнф, 1971, 71.8.112).

15. Волосевич И.И., Машинная сегментация английского предложения. В сб. "Статистика речи и автомат. анализ текста 1972". Л., "Наука", 1973, 293-322 (РЖИнф, 1974, 74.4.171).

16. Гладкий А.В., Описание синтаксической структуры предложения с помощью систем синтаксических групп. I. Формальный аппарат. "Научно-техн. информ. сб. Всес. ин-т научн. и техн. информ.", 1971, сер. 2, № 9, 35-38 (РЖИнф, 1972, 72.3.116).

17. ——— , Диковский А.Я. Теория формальных грамматик. "Итоги науки и техники. Теория вероятн. Мат. статист. Теор. киберн. Том 10", М., 1972, 107-142 (РЖИнф, 1973, 73.3.99).

18. ——— , Мельчук И.А., Грамматики деревьев. I. Опыт формализации преобразований синтаксических структур естественного языка. В сб. "Информ. вопр. семиотики, лингвист. и автомат. перевода". Вып. 1, М., 1971, 16-41 (РЖИнф, 1971, 71.5.126).

19. ——— , Мельчук И.А., Грамматика деревьев. II. К построению -грамматики для русского языка. В сб. "Информ. вопр. семиотики, лингвист. и автомат. перевода". Вып. 4, М., 1974, 4-29 (РЖИнф, 1974, 74.9.124).

20. Годабрелидзе Л.С., Синтез простых предложений, содержащих обороты. В сб. "Машинный перевод" (Труды Ин-та электрон., автоматики и телемех. АН ГрузССР, 8, № 3). Тбилиси, 1970, 34-46 (РЖИнф, 1971, 71.2.105).

21. ——— , Некоторые вопросы трансформации простых предложений в сложное. В сб. "Маш. перевод" (Тр. Ин-та электрон., автоматики и телемех. АН ГрузССР, 8, № 3). Тбилиси, 1970, 63-75 (РЖИнф, 1971, 71.2.107).

22. ——— , Синтаксический этап немецкого синтеза. В сб. "Машинный перевод" (Сакартвелос ССР Мецниереба та Академиа. Мартвис системебис институти шромеби, Тр. Ин-т систем упр. АН ГрузССР, 11-12, № 3). Тбилиси, "Мецниереба", 1974, 100-112 (РЖИнф, 1974, 74.10.95).

23. ——— , Морфологический анализ немецких словоформ (формообразование). В сб. "Машинный перевод" (Сакартвелос ССР Мецниереба та Академиа. Мартвис системебис институти шромеби, Тр. Ин-т систем упр. АН ГрузССР, 11-12, № 3). Тбилиси, "Мецниереба", 1974, 113-134 (РЖИнф, 1974, 74.10.94).

24. Гончаренко В.В., Добрускина Э.М., Лексический перевод английских научно-технических текстов с помощью ЭВМ. В сб. "Автоматическая переработка текста", Кишинев, "Штиинца", 1972, 5-40 (РЖИнф, 1973, 73.4.100).

25. Грикурова Д.А., О порядке слов в некоторых словосочетаниях грузинского языка в процессе автоматического синтеза (именные, наречные, масдарные и причастные словосочетания). В сб. "Информ. вопр. семиотики, лингвист. и автомат. перевода". Вып. 5. М., 1974, 111-124 (РЖИнф, 1974, 74.12.124).

26. Грязнухина Т.А., Куземская Н.А., Скороходько Э.Ф., Семантический анализ в системе автоматического перевода на информационный язык ...ах-кодов. "Научно-техн. информ. Сб. Всес. ин-т научн. и техн. информ." 1974, сер. 2, № 4, 11-17 (РЖМат, 1974, 9В1077).

27. Гюльмисарян С.А., Дрейзин Ф.А., Тер-Мисакянц З.Т. Математические формулы в широком контексте. "Научно-техн. информ. Сб. Всес. ин-т научн. и техн. информ.", 1971, сер. 2, № 3, 33-38 (РЖИнф, 1971, 71.8.105).

28. Данейко И.В., Крисевич В.С., Нехай О.А., Пиотровский Р.Г., Приймов Ю.Г., Сорокина В.Л., Ястребова С.В., Синтаксический анализ предложения в системе англо-русского вероятностного машинного перевода. В сб.: "Частные вопросы автомат. анализа текстов", Минск, 1972, 279-289 (РЖИнф, 1973, 73.6.91).

29. Дрейзин Ф.А., О выборе правильного по смыслу синтаксического анализа из нескольких формально возможных при автоматическом анализе русских текстов. Маш. пер. и прикл. лингвист. Вып.10, 1967, 3-20.

30. Еськова Н.А., Мельчук И.А., Санников В.З. Формальная модель русской морфологии. I. Формообразование существительных и прилагательных. (Ин-т рус. яз. АН СССР. Проблемн. группа по эксперим. и прикл. лингвистике. Предвар. публикации). М., 1971. 71 с.

31. Колковский А.К., Леонтьева Н.Н., Мартемьянов Ю.С., О принципиальном использовании смысла при машинном переводе. В сб. "Машинный перевод" (Труды ИТМиВТ АН СССР, Вып. 2). М., 1961, 17-46.

32. – , Мельчук И.А., О семантическом синтезе. В сб. "Проблемы кибернетики". Вып. 19, М., "Наука", 1967, 177-238.

33. —— , Пивоварова Т.В., Into и out of . (К семантическому толкованию двух английских предлогов). В сб. "Информ. вопр. семиотики, лингвист. и автомат. перевода". Вып. 3, М., 1972, 86-96 (РЖИнф, 1972, 72.7.105).

34. —— , Разлогова Е.Э., О системе семантического синтеза. IУ. Образцы словарных статей. В сб. "Информ. вопр. семиотики, лингвистики и автомат. перевода". Вып. 4, М., 1974, 36-62 (РЖИнф, 1974, 74.8.132).

35. Зубов А.В., Возможная схема автоматического промышленного перевода. В сб. "Статистика текста. Т. 2", Минск, 1970, 154-172 (РЖИнф, 1971, 71.2.95).

36. Кибрик А.Е., Модель автоматического анализа письменного текста. (На материале ограниченного военного подъязыка). М., 1970. 365 стр. (РЖИнф, 1971, 71.3.88).

37. Козьмина Е.Л., Коровина Т.И., Морфологический синтез в системе французско-русского машинного перевода (ИПМ АН СССР). М., 1972, 52 с. (РЖИнф, 1972, 72.9.77).

38. Корельская Т.Д., Падучева Е.В., О формальном аппарате синтаксических преобразований. В сб.: "Исслед. по мат. лингвист., мат. логике и информ. языкам", М., 1972, 27-58 (РЖМат, 1972, 8В704).

39. Коровина Т.И., Морфологический анализ в системе французско-русского машинного перевода. (Ин-т прикл. мат. АН СССР. Препринт № 71). М., 1972. 33 с. (РЖИнф, 1973, 73.5.112).

40. —— , Синтез русских именных словоформ с помощью машины. "Научно-техн. информ. Сб. Всес. ин-т науч. и техн. информ.", 1975, серия 2, №

41. Кулагина О.С., Алгоритм синтаксического анализа в системе французско-русского машинного перевода. (ИПМ АН СССР. Препринт № 3). М., 1970. 47 с.

42. —— , Некоторые вопросы преобразования деревьев зависимостей при машинном переводе. В сб. "Пробл. кибернетики". Вып. 24, М., "Наука", 1971, 161-190 (РЖМат, 1972, 4В616).

43. ——— , О машинном переводе текстов на естественных языках. В сб. "Пробл. кибернетики". Вып. 27, М., "Наука", 1973, 33-45 (РЖИнф, 1974, 74.5.125; РЖМат, 1974, 5В1000).

44. ——— , О системе французско-русского машинного перевода ФР-П. В сб. "Пробл. кибернетики". Вып. 27, М., "Наука", 1973, 47-61 (РЖИнф, 1974, 74.5.126; РЖМат, 1974, 5В1001).

45. ——— , Трансформации деревьев зависимостей при французско-русском машинном переводе. (ИПМ АН СССР. Препринт № 3). М., 1973, 24 с. (РЖИнф, 1973, 73.10.102).

46. ——— , Мельчук И.А., Автоматический перевод: краткая история, современное состояние, возможные перспективы. В сб. "Автоматический перевод", М., 1971, 3-25.

47. ——— , Мельчук И.А., Эрастов К.О., Об одной возможной системе машинного перевода. (Ин-т рус. языка АН СССР. Пробл. группа по эксперим. и прикл. лингвист. Предварит. публикации, вып. 21). М., 1971, 76 с. (РЖИнф, 1972, 72.11.117).

48. Лейкина Б.М., Независимый морфологический анализ английского научно-технического текста. В сб. "Автомат. анализ патентн. текстов". М., 1971, 12-25 (РЖИнф, 1971, 71.9.72).

49. ——— , К проблеме семантической правильности. "Научно-техн. информ. Сб. Всес. ин-т научн. и техн. информ.", 1971, сер. 2, № 7, 24-28.

50. ——— , Задачи этапа преобразования в системе АП. В сб. "Лингвист. пробл. функц. моделир. речевой деят-сти". Вып. I. Л., 1973, 75-92 (РЖИнф, 1974, 74.2.152).

51. Леонтьева Н.Н., Создание информационного языка на базе семантического анализа текста. "Научно-техн. информ. Сб. Всес. ин-т научн. и техн. информ.", 1971, сер. 2, № 8, 8-15 (РЖИнф, 1972, 72.1.81).

52. ——— , Урысон Е.В., Алгоритм построения информационной записи текста. I этап. "Научно-техн. информ. Сб. Всес. ин-т научн. и техн. информ.", 1973, сер. 2, № 12, 3-13 (РЖМат, 1974, 5В1003).

53. Логачева Т.Д., Першикова М.С., Описание эксперимента по "упрощенному синтаксическому анализу русских научно-технических текстов. В сб. "Лингв. проблемы функцион. моделир. реч. деятельности". Вып. I., Л., 1973, 102-110 (РЖИнф, 1974, 74.2. 153).

54. —— , —— , Эксперимент по синтаксическому анализу русского простого предложения на основе упрощенной грамматики с позиционными условиями. В сб. "Лингвист. проблемы функц. моделир. речевой деят." Вып. II. Л., Изд-во ЛГУ, 1974, 110-117.

55. Малаховский Л.В., О способах обработки знаков препинания при автоматическом переводе с английского языка на русский. В сб. "Статистика речи и автомат. анализ текста", Л., 1971, 452-463 (РЖИнф, 1971, 71.12.119).

56. Мануков З.Г., Зубов А.В., Автоматический синтаксический анализ предложения при переводе немецких публицистических текстов. В сб. "Частные вопр. автомат. анализа текстов". Минск, 1972, 311-330 (РЖИнф, 1973, 73.9.131).

57. Марчук Ю.Н., Опыт машинной реализации дистрибутивной методики определения лексических значений. В сб. "Статистика речи и автомат. анализ текста 1972". Л., "Наука", 1973, 181-230 (РЖИнф, 1974, 74.6.179).

58. Материалы семантико-синтаксического анализа английского текста в системе англо-русского автоматического перевода. 1. Л.А.Афонасьева. Материалы анализа фразы **Projection indicators for each aircraft show height, airplane position and flight attitude of aircraft.** 2. М.Г.Шаталова. Материалы анализа фразы **This is f-m or swept-frequency jamming** (Ин-т рус. языка АН СССР. Проблемн. группа по эксперим. и прикл. лингвист. Предварит. публикации, вып. 48). М., 1974, 44 с. (РЖИнф, 1974, 74.12.132).

59. Машинный перевод и прикладная лингвистика. (Труды 1 Моск. Гос. пед. ин-та иностр. яз.). Вып. 8. М., 1964. 252 с.

60. Машинный перевод и прикладная лингвистика. (Труды Моск. гос. пед. ин-та иностр. яз. им. М.Тореза). Вып. 17. М., 1974, 209 с.

61. Мельчук И.А., Некоторые вопросы машинного перевода за рубежом. (АН СССР. Ин-т научн. информации. Доклады Конф. по обр. информ., маш. пер. и автом. чтению текста. Вып. 6). М., 1961. 44 с.

62. —— , Об одном эксперименте АП с русского языка на французский в рамках Гренобльской системы. "Научно-техн. информ. Сб. Всес. ин-т научн. и техн. информ.", 1969, сер. 2, № 12, 23-29 (РЖИнф, 1970, 70.7.88).

63. —— , Русский язык в исследованиях по автоматическому переводу, ч. 1. Русский язык в исследованиях по АП. М., 1973, 8-55.

64. —— , О книге "Автоматический перевод. 2. С русского и английского языков на немецкий". "Научно-техн. информ. Сб. Всес. ин-т научн. и техн. информ.", 1973, сер. 2, № 12, 27-31.

65. —— , Об одной модели понимания речи (семантическая теория Р.Шенка). I. Задачи лингвистической теории и семантическое представление высказываний. "Научно-техн. информ. Сб. Всес. ин-т научн. и техн. информ.", 1974, сер. 2, № 6, 35-46 (РЖИнф, 1974, 74.11.123).

66. —— , Об одной модели понимания текстов на естественном языке (семантическая теория Р.Шенка). II. Общее строение модели "Текст ⇒ Смысл" и используемые в ней семантические средства; системы SPINOZA и MARGIE . "Научно-техн. информ. Сб. Всес. ин-т научн. и техн. информ.", 1974, сер. 2, № 8, 33-44 (РЖИнф, 1974, 74.12.127; РЖМат, 1975, 1В1138).

67. —— , Опыт теории лингвистических моделей "Смысл ⇔ Текст". М., 1974. 314 стр. (РЖИнф, 1974, 74.7.137).

68. —— , Перцов Н.В., Фрагмент модели английского поверхностного синтаксиса. (Ин-т рус. яз. АН СССР. Пробл. мн. группа по эксперим. и прикл. лингвистике. Предвар. публикации. Вып. 35). М., 1973, 3-18.

69. —— , —— , Поверхностно-синтаксические отношения в английском языке. (Ин-т рус. яз. АН СССР. Проблемн. группа по эксперим. и прикл. лингвистике. Предвар. публикации. Вып. 43). М., 1973. 80 с.

70. —— , Равич Р.Д., Автоматический перевод: 1949-1963. М., 1967. 517 с.

71. —— , —— . Автоматический перевод: 1964-1970. М., 1975.

72. Падучева Е.В., О порождении сложного предложения из простых. В сб. "Исслед. по мат. лингвист., мат. логике и информ. языкам", М., 1972, 59-78 (РЖМат, 1972, 8В687; РЖИнф, 1972, 72. 7.102).

73. —— , О семантике синтаксиса. Материалы к трансформационной грамматике русского языка. М., "Наука", 1974. 292 с.

74. Перцов Н.В., Об исследованиях по автоматическому переводу в Монреальском университете (Канада). "Научно-техн. информ. Сб. Всес. ин-т научн. и техн. информ", 1973, сер. 2, № 1, 36-44 (РЖИнф, 1973, 73.7.85).

75. Пинес В.Я., Об одной модели автоматического синтеза (применительно к личным формам азербайджанского глагола). "Научно-техн. информ. Сб. Всес. ин-т научн. и техн. информ.", 1970, сер. 2, № 1, 42-47.

76. Пиотровская А.А., Машинная морфология русского глагола. В сб. "Статистика речи и автомат. анализ текста 1972". Л., "Наука", 1973, 260-277 (РЖИнф, 1974, 74.5.149).

77. Пиотровский Р.Г., Отраслевой вероятностный машинный перевод. В сб. "Статистика текста. Т. 2". Минск, 1970, 5-32 (РЖИнф, 1971, 71.2.94).

78. —— , Машинный перевод. (Некоторые итоги и перспективы). В сб. "Проблемы структурн. лингвистики 1971". М., 1972, 521-533 (РЖИнф, 1973, 73.1.95; РЖМат, 1972, 12В539).

79. Приймов Ю.Г., Сорокина В.А., Алгоритм автоматического выделения именной группы в техническом тексте на английском языке. В сб. "Статистика текста. Т. 2", Минск, 1970, 189-214 (РЖИнф, 1971, 71.2.103).

80. Розенцвейг В.Ю., Машинный перевод. Теорет. проблемы сов. языкознания, М., 1968, 171-192.

81. ——— , Опыт лингвистического описания лексико-семантических ошибок в речи на неродном языке. (Ин-т рус. яз. АН СССР. Проблемн. группа по экспер. и прикл. лингвистике. Предвар. публикации. Вып. 50). М., 1974. 46 с.

82. Самойлович М.В., Автоматический анализ предложных групп. В сб. "Автомат. анализ патентн. текстов", М., 1971, 45-56 (РЖИнф, 1971, 71.9.75).

83. Сгалл П., Паневова Я., Основные черты семантической записи предложения и ее формальная структура. "Научно-техн. информ. Сб. Всес. ин-т научн. и техн. информ.", 1972, сер. 2, 37-40 (РЖМат, 1973, 4В750).

84. Скороходько Э.Ф., Лингвистические проблемы обработки текстов в автоматизированных информационно-поисковых системах. В сб. "Вопросы информ. теории и практики", № 25, М., ВИНИТИ, 1974, 5-120.

85. Славина Л.А., Трибис Л.И., Об алгоритмическом устранении двуязычной лексической неоднозначности. В сб. "Частные вопр. автомат. анализа текстов", Минск, 1972, 290-310 (РЖИнф, 1973, 73.6.106; РЖМат, 1973, 5.В894).

86. Смиренский В.Б., Автоматический анализ глагольных конструкций. В сб. "Автомат. анализ патентн. текстов", М., 1971, 34-44 (РЖИнф, 1971, 71.9.74).

87. Тамарашвили Л.Р., Несколько словарных единиц для грузинского толково-комбинаторного языка. В сб. "Машинный перевод" (Сакартвелос ССР Мецниереба та Академиа. Мартвис системебис институти шромеби, Тр. Ин-т систем упр. АН ГрузССР, 11-12, № 3). Тбилиси, "Мецниереба", 1974, 30-49 (РЖИнф, 1974, 74.10.106).

88. Фролов Г.Д., Автоматическое преобразование с помощью ЭВМ русской письменной речи в устную. В сб. "Пробл. кибернетики". Вып. 28, М., "Наука", 1974, 245-268 (РЖИнф, 1974, 74.10.88).

www.ruslang.ru

89. Цейтин Г.С., Алгоритм для упрощенного синтаксического анализа. В сб. "Пробл. кибернетики". Вып. 24, М., "Наука", 1971, 227-242 (РЖИнф, 1972, 72.4.103).

90. Чехов А.С., Некоторые средства обработки местоимений-заместителей при синтаксическом анализе. В сб. "Маш. пер. и прикл. лингвист." Вып. 15, М., 1972, 35-53 (РЖИнф, 1973, 73.7.96).

91. Чикоидзе Г.Б., Кодирование грузинских корневых морф типа В сб. "Маш. перевод" (Ин-т систем управления АН ГрузССР. Труды. 11-12, № 3). Тбилиси, "Мецниереба", 1974, 147-154.

92. —— , Кодирование грузинских фонем артикуляционными признаками. В сб. "Маш. перевод" (Ин-т систем управления АН ГрузССР. Труды. 11-12, № 3). Тбилиси, "Мецниереба", 1974, 155-163.

93. Шаляпина З.М., Семантические элементы как основа лексикографического описания: общие принципы и формальный аппарат. В сб. "Маш. пер. и прикл. лингвист." Вып. 15, М., 1972, 54-98 (РЖМат, 1973, 6Б765).

94. —— , Семантико-синтаксический анализ в системе англо-русского автоматического перевода (АРАП). 1. Общие принципы анализа и описание ручного эксперимента. 2. Материалы анализа фразы **This is termed amplitude-modulated c - w jamming** (Ин-т рус. языка АН СССР. Проблемн. группа по эксперим. и прикл. лингвист. Предварит. публикации, вып. 47). М., 1974. 58 с. (РЖИнф, 1974. 74.12.131).

95. —— , Англо-русский многоаспектный автоматический словарь (АРМАС). "Маш. пер. и прикл. лингвистика". Вып. 17, М., 1974, 7-67.

96. Экспериментальная система англо-русского автоматического перевода патентной документации. (Центр. н.-и. ин-т патент. информ. и техн.-экон. исслед.). М., 1970. 133 стр.

97. Эндина Н.М., Автоматический анализ и перевод многозначных английских именных сочетаний. В сб. "Автомат. анализ патентн. текстов", М., 1971, 122-136 (РЖИнф, 1971, 71.8.110).

98. Anderson B.B., Transformationally based English strings and their word subclasses. (New York Univ. String program reports , №7). N.Y., 1970. 188 p. (РЖИнф, 1972, 72.4.100).

99. Andreewsky A., Just N., Formalisation des structures syntaxiques. Problèmes des chaînes longues. "Note. CEA. №1297". Saclay, 1970. 55 p. (РЖМат, 1974, 3Б900).

100. ——. , Colnat-Verdier N., Analyse locale et globale de la langue. Filtres et classes linguistiques. "Note. CEA. №1648". Saclay, 1973. 103 p. (РЖМат, 1974, 3Б901).

101. Automatische Sprachübersetzung. 2. Russisch-Deutsch. Englisch-Deutsch. Berlin, Akademie-Verlag. 1972. 237 S. [См. подробную рецензию на этот сборник: [64]].

102. Balogh Z., Programrendszer magyar nyelvű szövegek szókainak tövesítéséhez. Tud. és müsz. tájék., 1974, 21, №7, 487-494 (РЖИнф, 1974, 74.12.133).

103. Barbault M.C., Desclés J.P. Vers une formalisation des grammaires transformationnelles. Mathém. et sciences hum., 1971, №34, 27-41 (РЖМат, 1972, 6Б621).

104. Beiträge zur maschinellen Sprachdatenverarbeitung (Hrsg. Zenker L.). Berlin, Akad.-Verl., 1973. VI, 168 S. (РЖИнф, 1974, 74.2.107).

105. Bellert I., Lexical functions in a MT Dictionary. "Étud. linguist. appl. trad. autom. Univ. Mintréal". Montréal, 1971, 1-14 (РЖИнф, 1973, 73.12.110).

106. Bień J.S., An alphabetic code for the inflectional analysis of Polish texts. Algorytmy, 1971, 8, №14, 5-25 (РЖМат, 1972, 8В706).

107. —— , Towards computer systems for conversing in Polish. /Warszawa, 1973/. 27 p.

108. Bisbey R.L., A tree grapher for the linguist. VAIDE Proc. 10th Annu. Meet., Los Angeles, Calif., 1971. S.l. s.a, 4/28-4/47 (РЖМат, 1974, 3В895).

109. Bobrow D., Natural language-interaction systems. Piot. Lang. Mach., L.-N.Y., 1970, 31-65 (РЖИнф, 1971, 71.11. 107; РЖМат, 1971, 12B.1063).

110. Bocşa M., Codage de l'alphabet généralisé du roumain écrit pour l'ordinateur IRIS-50(FELIX C-256). I. Le substantif. Cahiers de linguist. théor. et appliquée, 1973, 10, fasc.2, 139-151.

111. Boitet Ch., Toward an adaptive and interaction system in automatique translation. Grenoble, 1973. 28 p. (Intern. Conference on Comput. Linguistics, Pisa, Italy, 1973).

112. —— , Astrabondes et dictionnares. (Univ. Scient. et Médic. de Grenoble). Grenoble, 1973. 34 p.

113. —— , Heuristique et analyse syntaxique predictive. (Univ. de Grenoble. Groupe d'études pour la TA. G - 2900 - A.) Grenoble, 1974. 60 p.

114. Booth K.V.H., Brown C., Stock C., English-French translation on a computer. Pensiero e linguag. oper., 1971, 2, №7-8, 231-238 (РЖИнф, 1972, 72.5.99).

115. Borillo A., Problèmes d'analyse syntactico-sémantique de constructions interrogatives en français dans le cadre du traitement automatique de questions. (CNRS. Unité de recherche "Analyse docum. et calcul en archéologie"). Marseille, 1973. 16 p.

116. Brandt I., Der Artikel in deutschen Präpositionalgruppen. "Autom. Sprachübersetzung. 2. Russisch —→ Deutsch. Englisch —→ Deutsch". B., 1972, 11-27(РЖИнф, 1973, 73.4.94).

117. Caneghem M., van, Morphographie française. "TAUM 71", Montréal, 1971, 205-223 (РЖИнф, 1972, 72.9.93).

118. Chauché J., Transduction d'arborescences. Application aux manipulations de formules sur ordinateur. (Thèse doct. Univ. de Grenoble). Grenoble, 1971, 84 p. (РЖИнф, 1972, 72.4.92).

119. —— , Arborescences et transformations. (Univ. scient. et méd. de Grenoble. Groupe d'études pour la T.A. G-2700-A). Grenoble, 1972. 52 p.(РЖМат,1973,8В747 К).

120. —— , Guillaume P., Quezel-Ambrunaz M., Le système A.T.E.F. (analyse de textes en états finis). (G-2600-A. Groupe étud. trad. automat. Univ. sci. et méd. Grenoble). Grenoble, 1972. 72 p. (РЖИнф, 1973, 73.4.90).

121. —— , —— , —— , Système C.E.T.A.I. Le langage. Grenoble, 1973.

122. Colmerauer A., Les systèmes Q, ou un formalisme pour analyser et synthétiser des phrases sur ordinateur. Univ. Montréal. Dep. Inform. Publ. interne №43. Montréal, 1970, 45 p. (РЖИнф, 1971, 71.5.147; РЖИнф, 1972, 72.9.88)

123. —— , Phase de transfert. "TAUM 71". Montréal, 1971, 134-171.

124. Computer models of thought and language (Eds. R.Schank and K.Colby). San Francisco, 1973.

125. Cooper F., Machines and speech. "Res. Trends Comput. Linguist." Arlington (Va), 1972, 38-66(РЖИнф,1974,74.6.157).

126. Crawford T.D., The Cardiff machine translation project. (Pap. Conf. Assoc. Teachers Ital. Penarth, March 25th 1972). Cardiff, 1972. 10 pp. (РЖИнф, 1974, 74.1.89).

127. Dansereau J., Grammaire du francais: notes explicatives. "TAUM 71", Montréal, 1971, 172-204(РЖИнф, 1972, 72.9.92).

128. —— , Modèle génératif de syntaxe française (Thèse maîtrise (linguist). Fac. étud. supér. Univ. Montréal). Montréal, 1972. 149 p. (РЖИнф, 1973, 73.5.94).

129. Del Campo I., González I., Molina T., Marcos F., The automatic syntactic analysis as an aid in dictionary making. /Madrid, 1973/. 16 p.

130. Friedman J. (with Th.H.Bredt, R.W.Doran, B.W.Pollack, Th.S. Martner), A computer model of transformational grammar. N.Y., "Amer. Elsevier Publ. Comp.", 1971. 166 p.

131. Goldman N.M., Computer generation of natural language f.om a deep conceptual base. (Istituto per gli studi semant. e cognit.). Castagnola (Switzerland), 1974. 309 p.

132. —— , Riesbeck Ch.K., A conceptually based sentence paraphraser. (Stanford Artif. Intell. Lab. Memo AIM-196). Stanford (Calif.), 1973. 88 pp.(РЖИнф, 1974, 74.1.116).

133. Grishman R., Sager N., Raze C., Bookchin B., The linguistic string parser. AFIPS Conf. Proc. Vol. 42. Nat. Comput. Conf. and Expos., New York, N.Y., 1973. Montvale (N.J.), 1973, 427-434 (РЖМат, 1974, 6В 1232).

134. Gross L.N., A computer program for testing grammars on-line. Pap. 6th Annu. Meet. Assoc. Mach. Transl. and Comput. Ling. Urbana, Ill. (Inform. Syst. Lang. Stud. №19. MITRE Corp.). Bedford (Mass.), 1968. (РЖИнф, 1971, 71.7.100).

135. Hajičová E., Panevová J., Sgall P., Recursive properties of tense in Czech and English. Prague Bull. Math. Linguist., №13, 1970, 9-43 (РЖИнф, 1971, 71.4.99).

136. Haltof B., Steube A., Zur semantischen charakterisierung der Tempora im Russischen und Deutschen. Ling. Arbeitsber. Sekt. Theor. und angew. Spachwiss., Leipzig, 1970, 37-52 (РЖИнф, 1971, 71.7.96).

137. Harris B., "Polygram": grapho-morphology analyzer for English. "TAUM 71". Montréal, 1971, 46-105 (РЖИнф, 1972, 72.9. 89).

138. Harris Z., String analysis. The Hague, 1962.

139. Hays D.G., Types of processes on cognitive networks. /Buffalo (N.Y.), 1973/. 16 p. (Препринт доклада на Inter. conference on Comput. Linguistics, Pisa, Italy, 1973).

140. Henisz-Dostert B., Thompson F.B., Syntactic analysis in REL English. Statist. Meth. Linguist., 8, 1972, 5-38 (РЖИнф, 1973, 73.4.111).

141. Herskovits A., The generation of French from a semantic representation. (Stanford Artif. Intell. Lab. Memo AIM-212. STAN-CS-73-384). Stanford, 1973. 20 p.

142. Hofmann Th.R., Utilization of C-networks in translation of natural language. Étud. linguist. appl. trad. autom. Univ. Montréal. Montréal, 1971, 23-28(РЖИнф, 1973, 73.12.112).

143. —— , Interpretation and integration of sentences into a C-network. Étud. linguist. appl. trad. automat. Univ. Montréal. Montréal, 1971, 50-62 (РЖИнф, 1973, 73. 12.113).

144. Ichbiah M., Valabrègue H., L'analyse syntaxique automatique de phrases françaises en vue de la constitution de fichiers. TA informs, 1972, 13, №2, 25-37(РЖМат, 1973, 7В735).

145. Josselson H.H., A linguistic interpretation of MT in the sixties. Res. Comput. -aid. Transl. Russ. —→ Engl. 11th Annu. Progr. Rept. Wayne State Univ., 1969, Detroit (Mich.), 1969, I/1 - I/72 (РЖИнф, 1971, 71.1.148).

146. —— , Automatic translation of languages since 1960. A linguist's view. Adv. comput. Vol. II. New York - London, 1971, 1-58 (РЖМат, 1972, 1В1145).

147. Kaplan R.M., A multi-processing approach to natural language. AFIPS Conf. Proc. Vol. 42. Nat. Comput. Conf. and Expos., New York, N.Y., 1973. Montvale (N.J.), 1973, 435-440 (РЖМат, 1974, 6В 1237).

148. Kay M., Automatic translation of natural languages. Daedalus, 1973, 217-230. (Proceedings of the American Academy of Arts and Sciences, 102, №3).

149. —— , Sparck-Jones K., Automated language processing. Ann. Rev. Inf. Sci. and Technol. Vol. 6, Chicago, 1971, 146-166 (РЖИнф, 1972, 72.12.163).

150. Kittredge R., A grammar for normalizing English sentences. "TAUM 71". Montréal, 1971, 106-133 (РЖИнф, 1972, 72.9.90).

151. Klein Sh., Computer simulation of language contact models. (Univ. Wisc. Comput. Sci. Dept Tech. Rept №163). Madison, 1972. 28 pp. (РЖИнф, 1973, 73.12.104).

152. —— , Automatic inference of semantic deep structure rules in generative semantic grammars. (Univ. Wisc. Comput. Sci. Dep. Tech. Rept №180). Madison, 1973. 22 p. (РЖИнф, 1973, 73.12.105).

153. —— , Oakley J.D., Suurballe D.J., Ziesemer R.A., A program for generating reports on the status and history of stochastically modifiable semantic models of arbitrary universes. Statist. Meth. Linguist., 8, 1972, 64-93 (РЖИнф, 1973, 73.5.116).

154. —— , Aeschlimann J.F., Balsiger D.F., Converse S.L., Court C., Foster M., Lao R., Oakley J.D., Smith J., Automatic novel writing: a status report. (Comput. Sci. Dept. Univ. Wiscon. Techn. Rept №186). Madison (Wisc.), 1973. 109 p.

155. Klimonow G., Die Syntax der russischen Adverbien. Eine Studie zur automatischen syntaktischen Analyse des Russischen. Autom. Sprachübersetz. 2. Russisch —→ Deutsch. Englisch —→ Deutsch. B., 1972, 29-127(РЖИнф, 1973, 73.4.92).

156. Kortlandt F.H.H., Russian conjugation: computer synthesis of Russian verb forms. Tijdschrift voor Slavische Talen Letterkunde, 1972, 1, 51-80.

157. Krejčová I., Notes on morphological synthesis for mechanical translation of technical texts on computer EPOS-2. Inform. Process. Machines, 1971, №15, 91-96 (РЖМат., 1973, 1B907).

158. Kunze J., Die Komponenten der Darstellung syntaktischer Strukturen in einer Abhängigkeitsgrammatik. Prague Bull. Math. Linguist., 18, 1972, 15-27(РЖИнф, 1973, 73.3.107).

159. Langmaack H., Über ein Verfahren zur systematischen Auffindung aller Strukturbäume eines Satzes in einer kontextfreien Grammatik. Referate des 1. SFB - Kolloq. von 20.-22. Mai 1971. Saarbrücken, 1971, 63-66 (РЖИнф, 1973, 73.2.119).

160. Luckhardt H.-D., Beschreibung eines Computer-Lexicons zur automatischen russisch-deutschen Übersetzung. Ber. Sonderforschungsbereich Elektron. Sprachforsch. Univ. Saarland., 1972, N° 9/LA 11, 1-24 (РЖИнф, 1973, 73.6.94; РЖМат, 1973, 7В730).

161. Maas H.D., Ein Programm zur russischen syntaktischen Analyse. Ber. Sonderforschungsbereich Elektron. Sprachforsch. Univ. Saarland., 1972, N G9/LA 11, 43-102 (РЖМат, 1973, 7В730).

162. Mambrú Laporta J., Documentación y traducción automática en el campo textil. Cienc. y técn. en el mundo, 1974, 26, №424, 129-162 (РЖИнф, 1974, 74.9.135).

163. Melby A.K., Junction grammar and machine assisted translation. S.P., s.a. 13 pp., ill. (Доклад на Intern. conference on Comput. linguistics, Pisa, Italy, 1973).

164. Mishelevich D.J., Computer-based semantic analyzers. Int. J. Comput. and Inform. Sci., 1972, 1, N 3, 267-286 (РЖМат, 1973, 2В701).

165. Montgomery Ch.A., Automated language processing. Annu. Rev. Inform. Sci. and Technol. Vol. 4. Chicago, 1969, 145-174 (РЖИнф, 1970, 70.12.79).

166. Natural language processing (Ed. Rustin R.). N.Y., 1973.

167. Newell A., et al. Speech understanding systems. Amsterdam, 1973.

168. Olney J., Ramsey D., From machine-readable dictionaries to a lexicon tester: progress, plans, and an offer. Comput. Stud. Hum. and Verb. Behav., 1972, 3, №4, 213-220 (РЖИнф, 1973, 73.5.103).

169. Pala K., Об автоматическом синтаксическом анализе чешских предложений . Prague Stud. Math. Linguist., 1972, 3, 183-192 (РЖИнф, 1972, 72.8.109; РЖМат, 1972, 8В 705).

170. Perschke S., Possibilities of further development of automatic language translation. Pensiero e linguag. oper., 1970, 1, №1, 83-93 (англ.) , 94-105 (итал.) (РЖИнф, 1970, 70.7.74).

171. Peters P., Stanley R., Ritchie R.W., Context-sensitive immediate constituent analysis: context-free languages revisited. Math. Syst. Theory, 1973, 6, №4, 324-333 (РЖИнф, 1973, 73.11.93; РЖМат, 1973, 12 В802).

172. Petrick S.R., Computer-oriented grammars and parsing. Res. Trends Comput. linguist., Arlington (Va), 1972, 22-37 (РЖИнф, 1974, 74.6.184).

173. Rainer D., Die Beschreibung adverbialer Gruppen. I. Eine Erweiterung der "Skizze einer deutschen Grammatik". Ber. Sonderforschungsbereich. Elektron. Sprachforsch. Univ. Saarlandes, 1973, 27-48 (РЖИнф, 1974, 74.8.122).

174. Referate des 1. SFB-Kolloquiums. (Ber. Sonderforschungsbereich "Elektron.Sprach-forschung". Univ. des Saarlandes). Saarbrücken, 1971 (РЖИнф, 1973, 73.2.99).

175. Rothkegel A., Feste Syntagmen. Ber. Sonderforschungsbereich Elektron. Sprachforsch. Univ. Saarlandes, №10-72: G, Teil 2. /Saarbrücken, 1972/, 95-137 (РЖИнф, 1974, 74.5.144).

176. Rüdiger B., Einige Lösungsmöglichkeiten für syntaktischen Mehrdeutigkeiten im Englischen und Deutschen. Automat. Sprachübersetz. 2. Russisch —► Deutsch. Englisch —► Deutsch. Berlin, 1972, 159-199 (РЖИнф, 1973, 73.4.107).

177. Sågvall A.-L., A system for automatic inflectional analysis (implemented for Russian). Stockholm, 1973. 142 pp. (РЖМат, 1975,

178. Salkoff M., Analyse syntaxique automatique utilisant une grammaire en chaine (string grammar). Math. et sci. hum., 1971, №35, 19-30 (РЖМат, 1972, 6В623).

179. —— , Une grammaire en chaîne du français (analyse distributionnelle). Paris, Dunod, 1973. 199 p. (Monographies de linguistique mathématique. 6).

180. Schank R.C., Conceptual dependency: a theory of natural language understanding. Cognit. Psychol., 1972, 3, №4, 552-631 (РЖМат, 1973, 2В948).

181. Schmid E.L., Thiel M., Zur Behandlung nicht inventarisierter Komposita. Ber. Sonderforschungsbereich Elektron. Sprachforsch. Univ. Saarlandes, 1972, NG10/6A12, 82-94 (РЖМат, 1973, 6Б763).

182. Schott G., Linguistische Algorithmen zur Deflexion als Mittel zum automatischen Indexieren im Deutschen. Lect. Notes Comput. Sci., 1973, 1, 421-430 (РЖМат, 1974, 7Б1182).

183. Simmons R.F., Natural language question-answering systems: 1969. Communs ACM, 1970, 13, №1, 15-30 (РЖИнф, 1970, 70.10.58).

184. ——, Integrated computer systems for language. Res. Trends Comput. Linguist., Arlington (Virginia), 1972, 7-21 (РЖИнф, 1974, 74.6.156).

185. ——, Slocum J., Generating English discourse from semantic networks. Communs ACM, 1972, 15, №10, 891-905.

186. Sinaiko H.W., Klare G.R., Further experiments in language translation: a second evaluation of the readability of computer translations. ITL, 1973, 19, 29-52.

187. Starke I., Die hauptsachlichen Typen syntaktischer Mehrdeutigkeit im Russischen und einige Möglichkeiten ihrer Lösung mit syntaktischen Mitteln. Automat. Sprachübersetz. 2. Russisch → Deutsch. Englisch → Deutsch. Berlin, 1972, 201-221 (РЖИнф, 1973, 73.4.109).

188. TAUM 71. Rapp. janvier 1971. (Proj. Trad. Automat. Univ. Montréal). Montréal, 1971. 239 p. (РЖИнф, 1972, 72.9.87).

189. Thiel M., Satzadverbien. Ber. Sonderforschungsbereich. Elektron. Sprachforsch. Univ. Saarlandes, 1973, 49-72 (РЖИнф, 1974, 74.8.123).

190. Uhlířová L'., On the non-projective constructions in Czech. Prague Stud. Math. Linguist., 1972, 3, 171-181 (РЖИнф, 1972, 72.8.108).

191. Valach M., Hardware processor for English. Res. 1971-1972: Annu. Progr. Rep. Sch. Inform. and Comput. Sci. Ga Inst. Technol. Atlanta (Ga), 1972, 93-99 (РЖИнф, 1974, 74.6.186).

192. ——, Hardware processor for English. Res. 1972-1973: Annu. Progr. Rept. Sch. Inform. and Comput. Sci. Ga Inst. Technol. Atlanta (Ga), 1973, 74-80 (РЖИнф, 1974, 74.6.185).

193. ——, Deimel L.E., Structural representation of English. Res. 1971-1972: Annu. Progr. Rept. Sch. Inform. and Comput. Sci. Ga Inst. Technol. Atlanta (Ga), 1972, 12-26 (РЖИнф, 1974, 74.6.166).

194. ——, Deimel L.E., Structural representation of English. Res. 1972-1973: Annu. Progr. Rept. Sch. Inform. and Comput. Sci. Ga Inst. Technol. Atlanta (Ga), 1973, 28-33 (РЖИнф, 1974, 74.6.167).

195. Vauquois B., Modèles pour la traduction automatique. Math. et sci. hum., N 34, 1971, 61-70 (РЖИнф, 1972, 72.5.84; РЖМат, 1972, 6В637).

196. Verloren van Themaat W.A., Automatic analysis of Dutch compound words. (Mathem. Center Tracts. N 38). Amsterdam, 1972. 174 p. (РЖМат, 1972, 7В675).

197. Viks Ü., Remmel M., Eesti verbid - Estonian Verbs. (Ин-т Языка и Лит. АН ЭССР . Preprint 3). Tallinn, 1974. 60 pp.

198. Walker D.E., Automated language processing. (Stanford Res. Inst. Art. Intell. Cent. Techn. Note №77). Menlo Park (Calif.) 1973, 85 pp. (РЖИнф, 1974, 74.6.155).

199. ——, Speech understanding research. (Stanford Res. Inst. Ann. Techn. Rept, 4 Oct. 1971 - 2 Oct. 1972). Menlo Park (Calif.), 1973. 62 pp. (РЖИнф, 1973, 73.12.103).

200. Weiss S.F., Automatic resolution of ambiguities from natural language text. Inform. Storage and Retriev. Dep. Comput. Sci. Cornell Univ. Sci. Rept N 18. Ithaca (N.Y.), 1970, IV/1-IV/51 (РЖИнф, 1971, 71.11.105).

201. Weissgerber M., Zur Behandlung der Negation in der maschinellen Satzanalyse. Ber. Sonderforschungsbereich. Elektron. Sprachforschung. Univ. Saarlandes, 1973, 73-91 (РЖИнф, 1974, 74.8.120).

202. Wierzbicka A., Semantic primitives. Frankfurt/Main, "Athenäum", 1972. 235 p.

203. Wilks Y., An artificial intelligence approach to machine translation. (Stanford Artificial Intell. Project. Memo AIM-161. STAN-CS-264-72). Stanford, 1972. 44 p. (РЖИнф, 1973, 73.12.103).

204. —— , Grammar, meaning and the machine analysis of language. London, 1972. 198 p.

205. —— , Preference semantics. (Stanford Artif. Intell. Lab. Memo AIM-206). Stanford, 1973. 21 pp. (РЖИнф, 1974, 74.9.133).

206. —— , Natural language inference. (Stanford Artificial Intelligence Laboratory. Memo AIM-211. Computer Science Dept. Report CS-383). Stanford, 1973. 24 pp.

207. Winograd T., Procedures as a representation for data in a computer program for understanding natural language. (MAC TR-84. Artif. Intell. Lab. Mass. Inst. Techol.). Cambridge (Mass.), 1971. 462 pp. (РЖИнф, 1974, 74.10.89).

208. —— , Understanding natural language. Edinburgh, 1972. 195 pp. (РЖИнф, 1973, 73.6.100).

209. —— , Five lectures on artificial intelligence. (Stanford Artif. Intell. Lab. Memo AIM-246. STAN-CS-74-459). Stanford (Calif.), 1974. 93 p.

210. Woods W.A., Transition network grammars for natural language analysis. Communs ACM, 1970, 13, №10, 591-606 (РЖИнф, 1971, 71.4.89).

211. —— , Progress in natural language understanding: an application to lunar geology. AFIPS Conf. Proc. Vol. 42. Nat. Comput. Conf. and Expos., New York, N.Y., 1973. Montvale (N.J.), 1973, 441-450 (РЖМат, 1974, 6В1234).

212. —— , Kaplan R.M., Nash-Webber B., The lunar sciences natural language information system: final report. (BBN Report 2378). Cambridge (Mass.), 1972.

213. —— , Makhoul J., Mechanical inference problems in continuous speech understanding. Artif. Intelligence, 1974, vol. 5, №1, 73 - 91 (РЖМат, 1975,

214. Wyatt J.L., A computer-validated Portuguese to English transformational grammar. The Hague - Paris, 1972. 437 pp. (РЖИнф, 1973, 73.6.85).

215. Zimmermann H.H., Zur Auflösung von Mehrdeutigkeiten bei einer maschinellen Analyse des Deutschen. Beitr. Linguist. und Informationsverarb., № 12, 1971, 36-49 (РЖМат, 1971, 2В771).

216. —— , Das Lexikonsystem zur maschinellen Sprachbearbeitung. Ber. Sonderforschungsbereich. Elektron. Sprachforschung. Univ. Saarlandes, №10-72: C, Teil 1 /Saarbrücken, 1972/, 62-80 (РЖИнф, 1974, 74.5.140).

ИНСТИТУТ РУССКОГО ЯЗЫКА им. В. В. ВИНОГРАДОВА

БИБЛИОТЕКА ИНСТИТУТА РУССКОГО ЯЗЫКА АН СССР

www.ruslang.ru

А05169 от 13.06.75 г. 5,5 п.л. 190 экз. Зак. 686

Тип. МГПИИЯ

. 6.86

30 г

Бесплатно