

С
ТАТИСТИКА
РЕЧИ
И АВТОМАТИЧЕСКИЙ
АНАЛИЗ
ТЕКСТА

1972

KB

АКАДЕМИЯ НАУК СССР
ИНСТИТУТ ЯЗЫКОЗНАНИЯ

СТАТИСТИКА
РЕЧИ
И АВТОМАТИЧЕСКИЙ
АНАЛИЗ
ТЕКСТА

1972

ИЗДАТЕЛЬСТВО «НАУКА»
ЛЕНИНГРАДСКОЕ ОТДЕЛЕНИЕ
ЛЕНИНГРАД • 1973



Сборник представляет собой коллективную монографию, в которой исследуется статистическая структура текстов на основных европейских языках, а также рассматриваются вопросы лексико-семантического, морфологического и синтаксического вероятностного машинного перевода. Реализованные на ЭВМ алгоритмы сборника представляют интерес с точки зрения развития эвристического программирования и формализации лингвистической интуиции человека.

Сборник адресован лингвистам, математикам и инженерам, занимающимся вопросами информатики и автоматизации.

СОДЕРЖАНИЕ

Стр.

Редакционная коллегия:

П. М. АЛЕКСЕЕВ, А. В. ГРОШЕВА,
Р. Г. ПИОТРОВСКИЙ (ответственный редактор),
Л. З. СОВА, С. А. ШУБИК.

Предисловие	5
Сокращения и условные обозначения	7
Часть I. Статистическая структура текста	
В. В. Богданов. Статистические концепции языка и речи	9
П. М. Алексеев. Семантические частотные словари	20
М. Е. Каширина. Частотный словарь английского подъязыка физики твердого тела	37
Е. М. Тарасова. Частотный словарь английского подъязыка физики элементарных частиц	47
Л. В. Малаховский. О частотном распределении английских количественных числительных первой сотни	56
Д. С. Демьянова. Частотный словарь немецкого подъязыка автоматизации	67
Э. Я. Рудник, А. И. Галкин. Частотный словарь немецкого подъязыка почвоведения	74
С. Е. Мазарская, Л. П. Просветова. Частотный словарь существительных немецкого подъязыка неорганической химии	79
С. А. Шубик. О размерах предложения в современном немецком языке	86
А. В. Грошева. Частотность классов латинских фонем	98

Часть II. Лексико-семантический машинный перевод

К. Б. Бектаев, Л. И. Белоцерковская, А. Д. Борисевич, <u>В. А. Букович</u> , Н. С. Булашева, В. В. Гончаренко, М. В. Данейко, Т. Р. Зуева, Н. И. Исабекова, О. А. Нехай, Р. Г. Пиотровский, В. А. Соркина. Статистическое выделение русских машинных оборотов и построение двуязычных автоматических словарей	107
Л. И. Тривис. Об одной модели распознавания лексических значений неоднозначных слов	131

	Стр.
Ю. Н. Марчук. Опыт машинной реализации дистрибутивной методики определения лексических значений	181
В. Я. Шабес, В. В. Рожков. Устранение предположно-падежной многозначности на базе тезауруса	231

Часть III. Машинная морфология

М. Г. Зорев. Машинная морфология и ее применение в немецко-русском автоматическом словаре	245
А. А. Пиотровская. Машинная морфология русского глагола	260

Часть IV. Синтаксический машинный перевод

Л. Э. Сова. Инвариант структурно-синтагматических отношений	278
И. И. Волосевич. Машинная сегментация английского предложения	293
О. А. Афаунова. Автоматическое устранение грамматической многозначности английских ing-овых форм	323

ПРЕДИСЛОВИЕ

Предлагаемый вниманию читателя сборник охватывает работы, выполненные за три года (1968—1971) в группе «Статистика речи» и в сотрудничающих с ней коллективах.

Научная стратегия группы в этот период не изменилась в своей основе. В группе продолжались исследования статистической структуры текста, результаты которых использовались для построения вероятностного машинного перевода и реферирования текстов на основных европейских языках.

К началу 70-х гг. эта научная стратегия начала давать свои первые плоды. В 1970—1971 гг. минский, кишиневский и ленинградский коллективы построили первую экспериментальную автоматическую систему лингвистического обслуживания, включающую пословный перевод, перевод-аннотацию и реферат научно-технических текстов на английском языке.

Устойчивая работа системы на отечественных ЭВМ демонстрировалась общественности, освещалась в центральной и республиканской печати («Известия», от 30 мая 1970 г.; «Советская индустрия», от 26 декабря 1971 г.; «Вечерняя Алма-Ата», от 22 мая 1970 г.; «Молдова социалистэ», от 23 октября 1971 г. и др.).

Большой автоматический англо-русский словарь, составлявший основу системы, построен на аксиоматико-иконическом принципе, который заключается в прямом соотношении входных (английских) словоформ и словосочетаний выходным (русским) лексическим единицам (ИЛ, стр. 4—147, СТ II, стр. 331—614), при этом система включает также и построенный на вероятностной основе алгоритм распознавания смыслового образа текста (УЗК, стр. 66—74). Испытание, а затем и внедрение указанной системы показало, что автоматический анализ текста должен и в дальнейшем опираться на сочетание аксиоматико-иконического подхода, отражающего норму языка, и вероятностного «эвристического» алгоритмизирования, отражающего порождение правильных речевых цепочек системой языка.

Оба эти подхода представлены в статьях II—IV частей сборника. С одной стороны, в работе К. Б. Бектаева, Л. И. Белоцерковской и других развивается методика построения двуязычных

словарей машинных оборотов, словарей, позволяющих получать вполне нормативный со стилистической точки зрения перевод целых сегментов входного текста. Эта статья интересна также и тем, что она продолжает разработку проблемы марковских связей внутри лингвистического текста, начатую в предшествующих коллективных монографиях группы «Статистика речи» (см. СтР, стр. 4, 200—224; СТ I, стр. 376—394, 434—523; ИЛ, стр. 20—33; СтРААТ, стр. 47—112, 262—285).

С другой стороны, Л. И. Трибис, Ю. Н. Марчук, В. Я. Шабес, А. А. Пиотровская, И. И. Волосевич, О. А. Афаунова описывают в своих статьях реализованные на ЭВМ «эвристические» алгоритмы смыслового анализа текста и снятия синтаксической неоднозначности. Построение «эвристических» машинных алгоритмов вызывает к жизни большое число теоретических вопросов, связанных с соотношением означаемого и означающего и преодолением асимметрии лингвистического знака. Эти вопросы затронуты в работе Л. И. Трибис, им посвящена также теоретическая статья Л. Э. Совы.

В первой части сборника рассматриваются общие вопросы лингвостатистики (В. В. Богданов, П. М. Алексеев), а также публикуются результаты лексико-статистических (М. Е. Каширина, Е. М. Тарасова и другие), синтактико-стилистических (С. А. Шубик) и фонетико-статистических (А. В. Грошева) исследований. Эти результаты могут представлять интерес для будущих работ по автоматическому анализу текста.

Сокращения и условные обозначения

АКД	— автореферат кандидатской диссертации.
АН	— Академия наук.
АП	— сб.: Автоматический перевод. Москва, 1971.
АПТМА	— сб.: Автоматическая переработка текста методами прикладной лингвистики. Материалы Всесоюзной конференции 6—8 октября 1971 г. Кипшиев, 1971.
АС	— автоматический словарь.
буд.	— будущее (время).
вин.	— винительный (падеж).
ВИНТИ	— Всесоюзный институт научной и технической информации.
ВП	— Вопросы психологии. Москва.
вр.	— время (глагола).
ВФ	— Вопросы философии. Москва.
ВЯ	— Вопросы языкознания. Москва.
ГПИ	— государственный педагогический институт.
ГПИИЯ	— государственный педагогический институт иностранных языков.
дат.	— дательный (падеж).
дв. ед.	— двоячая (-ные) единица (-цы).
деепр.	— деепричастие.
ед. ч.	— единственное число.
ж. р.	— женский род.
изъяв.	— изъявительное (наклонение).
ИЛ	— сб.: Инженерная лингвистика. Ученые записки Ленинградского государственного педагогического института им. А. И. Герцена, том 458, части I—II, 1971.
им.	— именительный (падеж).
им. ф.	— именная форма (существительное, прилагательное, причастие, местоимение).
инф.	— инфинитив.
л.	— лицо (грамматическое).
ЛГУ	— Ленинградский государственный университет.
МГУ	— Московский государственный университет.
м. р.	— мужской род.
мн. ч.	— множественное число.
МНК	— Материалы научной конференции профессорско-преподавательского состава Чимкентского государственного педагогического института, посвященной 100-летию со дня рождения В. И. Ленина и 50-летию Казахстана. 1970.
МОЗУ	— магнитное оперативное запоминающее устройство.
МП	— машинный перевод.
накл.	— наклонение (глагола).
наст.	— настоящее (время).
несов.	— несовершенный (вид).

НЛ	— сб.: Новое в лингвистике. Москва.
НТИ	— сб.: Научно-техническая информация. Москва.
ОЗУ	— оперативное запоминающее устройство.
п.	— пункт.
пад.	— падеж.
ПК	— сб.: Проблемы кибернетики. Москва.
предл.	— предложный (падеж).
прил.	— прилагательное.
прич.	— причастие.
прош.	— прошедшее (время).
род.	— родительный (падеж).
сов.	— совершенный (вид).
сосл.	— сослагательное (наклонение).
сп.	— список.
спр. ф.	— спрягаемая форма.
ср. р.	— средний род.
СТ I	— сб.: Статистика текста, том I. Лингвостатистические исследования. Минск, 1969.
СТ II	— сб.: Статистика текста, том II. Автоматическая переработка текста. Минск, 1970.
СтР	— сб.: Статистика речи. Ленинград, 1968.
СтРААТ	— сб.: Статистика речи и автоматический анализ текста. Ленинград, 1971.
сущ.	— существительное.
твор.	— творительный (падеж).
УЗК	— Ученые записки Калининского государственного педагогического института им. М. И. Калинина.
ЦНИИПИ	— Центральный научно-исследовательский институт патентной информации.
ЭВМ	— электронно-вычислительная машина.
ЭЯСР	— сб.: Энтропия языка и статистика речи. Минск, 1966.

<i>ab</i>	— abbreviation.	MT	— Mechanical Translation and Computational Linguistics — An International Journal. Chicago.
<i>adj</i>	— adjective.	<i>n</i>	— noun.
<i>adv</i>	— adverb.	<i>not</i>	— notional (verb).
<i>art</i>	— article.	<i>num</i>	— numeral.
<i>attr</i>	— attributive, Adjektiv (нем.).	<i>p I</i>	— participle I.
<i>aux</i>	— auxiliary (verb).	<i>p II</i>	— participle II.
<i>C</i>	— цифра.	<i>part</i>	— particle.
<i>ej</i>	— conjunction.	<i>past indef</i>	— past indefinite.
<i>dem</i>	— demonstrative.	<i>pp</i>	— participle past.
<i>F</i>	— частота.	<i>pref</i>	— prefix, Präfix (нем.).
<i>ger</i>	— gerund.	<i>pres indef</i>	— present indefinite.
<i>i</i>	— ранг словарной единицы.	<i>pron</i>	— pronoun.
<i>imp</i>	— imperative.	<i>prp</i>	— preposition.
<i>inf</i>	— infinitive.	<i>rel</i>	— relative.
<i>interrog</i>	— interrogative.	<i>sb</i>	— subject.
<i>introd</i>	— introductory.	<i>sing</i>	— singular.
<i>l</i>	— link (verb).	<i>v</i>	— verb.
<i>m</i>	— количество единиц с одинаковой частотой.	<i>vbl n</i>	— verbal noun.
<i>mod</i>	— modal (verb).	<i>Z</i>	— цифра.

Часть I. СТАТИСТИЧЕСКАЯ СТРУКТУРА ТЕКСТА

В. В. Богданов

СТАТИСТИЧЕСКИЕ КОНЦЕПЦИИ ЯЗЫКА И РЕЧИ

Общеизвестен знаменитый афоризм Н. С. Трубецкого: «Язык лежит вне „меры и числа“».¹ Как понимать это высказывание? Как утверждение о том, что к языку неприменимы те методы количественного анализа, которые пригодны для исследования речи, или же как констатацию принципиальной невозможности количественного изучения языка? С первой интерпретацией этого тезиса еще можно было бы, с известными оговорками, согласиться. Однако имеются все основания полагать, что великий фонолог вкладывал в свою формулу именно второй смысл, который следует признать неприемлемым, хотя в лингвостатистической литературе иногда и можно встретить суждения, отражающие согласие их авторов с формулировкой Н. С. Трубецкого. Так, в одной из работ по этому поводу сказано: «Следует признать, что до сих пор развитие лингвистической статистики не сняло положения Н. С. Трубецкого, что „язык лежит вне меры и числа“».²

Исследования последних полутора десятков лет свидетельствуют, пожалуй, об обратном, и лучшим опровержением указанного тезиса Н. С. Трубецкого является тот факт, что лингвостатистика наряду с продолжающимся расширением и углублением специальных исследований обратилась к разрешению некоторых фундаментальных вопросов лингвистической теории. Это нашло выражение в попытке дать статистическую интерпретацию понятий «языка» и «речи».

Первым, кто обратился к этой проблеме, был Г. Хердан. Уже в 1956 г. он писал: «Стабильность относительных частот, которую мы обнаруживаем как составную часть различных элемен-

¹ Н. С. Трубецкой. Основы фонологии. Русск. пер. М., 1960, стр. 15.

² Г. С. Клычков. Некоторые замечания о соотношении статистики речи и структуры языка. В сб.: Иностранные языки в высшей школе, вып. II, Московский ГПИИЯ, 1963, стр. 12.

тов данного ряда языковых форм, неизбежно приводит к выводу, что „la langue“ включает в себя не только энграммы в виде лексических форм, но эти энграммы плюс соответствующие вероятности их встречаемости. Это как раз то, что я назвал статистической проекцией (view) с осью ровской дихотомии»³ (выделено Г. Херданом, — В. Б.). Характерно, что от этой мысли Г. Хердан не отказался и в своих последующих работах,⁴ хотя иногда он склонен был свойство стабильности относительных частот приписывать не языку (langue), а речевой деятельности (langage), не придавая, впрочем, этому уточнению слишком большого значения и считая его в значительной мере «делом вкуса» (a matter of taste).⁵ Это значит, что взгляды Г. Хердана по данному вопросу остались в целом неизменными, несмотря на ряд критических замечаний, которым подвергся его основной тезис.⁶ В соответствии с этими взглядами язык получает характеристики генеральной совокупности, а речь — свойства выборки.

К концепции Г. Хердана близки взгляды П. Гиро. Следует отметить, что П. Гиро не выразил свою точку зрения по этому вопросу с такой определенностью, как Г. Хердан, однако содержащиеся в работе французского лингвостатистика высказывания не оставляют сомнения в том, что он либо разделял точку зрения Г. Хердана, либо, во всяком случае, был достаточно близок к ней. В этом легко убедиться, если обратиться к его известной книге. П. Гиро пишет: «Частота знака не может рассматриваться как характеристика речи (parole); она является объективной характеристикой языка (langue), столь же важной, как формы или значения данного языка».⁷ Хотя П. Гиро пишет здесь о частоте, на самом деле он имеет в виду вероятность, поскольку на протяжении всей книги говорит о стабильности относительных частот как показателе языка. Что же касается речи, то, по его словам, «любая

³ G. H e r d a n. Language as Choice and Chance. Groningen, 1956, p. 79.

⁴ G. H e r d a n. 1) Type-Token Mathematics. 's-Gravenhage, 1960, p. 39; 2) The Calculus of Linguistic Observations. 's-Gravenhage, 1962, p. 19; 3) Quantitative Linguistics. London, 1964, p. 3; 4) The Advanced Theory of Language as Choice and Chance. New York, 1966, p. 27; см. также одну из последних его статей: L'elemento formale matematico nelle lingue naturali. Lingua e stile, № 3, anno 2, 1967, где приведено высказывание (стр. 286), полностью совпадающее с формулировкой, содержащейся в его первой книге.

⁵ G. H e r d a n. Quantitative Linguistics, p. 4.

⁶ Сам Г. Хердан выражает сожаление в связи с тем, что его основное положение неправильно было понято М. Халле. См.: G. H e r d a n. Quantitative Linguistics, pp. 4—5; см. также: P. M. Ф р у м к и н а. Рец. на: G. H e r d a n. Type-Token Mathematics. The Hague, 1960. — ВЯ, 1963, № 3, стр. 127—129; H. S p r a n g - H a n s s e n. Рец. на: G. H e r d a n. The Calculus of Linguistic Observations. The Hague, 1962. — Lingua, № 2, v. 19, 1967, p. 221.

⁷ P. G u i r a u d. Problèmes et méthodes de la statistique linguistique. Paris, 1960, pp. 17—18.

речь, любой текст представляет собой образец (выборку, фр. échantillon, — В. Б.) некоторого состояния языка, численную структуру которого, а также возможности его семантических реализаций он отражает».⁸

Плая точка зрения была выдвинута в совместной работе Н. Д. Андреева и Л. Р. Зиндера.⁹ Авторы вводят понятие «речевой вероятности» в качестве четвертого компонента к трем основным понятиям Ф. де Соссюра. «Система речевой вероятности... — пишут они, — есть совокупность относительных количественных характеристик, описывающих численные отношения между элементами (или группами элементов) в некотором массиве текстов».¹⁰ Эта вероятность, как показывает ее название, характеризует речь. С другой стороны, она как «отвлечение от речи» начинает соотноситься с языком, ибо представляет собой, как пишут авторы, «многомерную структуру».¹¹

Вопрос о вероятностной характеристике языка и речи рассматривается также в книге Н. Д. Андреева,¹² где вводятся понятия парадигматической и синтагматической вероятностей. Первая определяется удельным весом языковой единицы в инвентаре соответствующих единиц, а вторая — ее встречаемостью в речевой цепи.

«Парадигматические вероятности, — пишет Н. Д. Андреев, — принадлежат языку и всегда равны у разных единиц одной группы; синтагматические вероятности принадлежат речи и, за редкими исключениями, не равны у различных единиц одной и той же группы»¹³ (разрядка Н. Д. Андреева, — В. Б.).

Таким образом, отличие этой концепции от предыдущей заключается в том, что там одна «речевая вероятность» характеризует язык и речь, тогда как здесь свои количественные показатели получают оба члена соссюрсовской антиномии.

Аналогичные проблемы привлекают и внимание Б. Н. Головина в его недавно опубликованной книге.¹⁴ Б. Н. Головин признаёт допустимость квантитативного изучения языка и речи. «Одним из реальных оснований применения статистики в изучении языка и речи, — пишет он, — нужно признать об объективную присущность языку количественных признаков, количественных характеристик»¹⁵ (раз-

⁸ Ibid., p. 18.

⁹ Н. Д. Андреев, Л. Р. Зиндер. О понятиях речевого акта, речи, речевой вероятности языка. ВЯ, 1963, № 3, стр. 15—21.

¹⁰ Там же, стр. 19.

¹¹ Там же, стр. 19.

¹² Н. Д. Андреев. Статистико-комбинаторные методы в теоретическом и прикладном языковедении. Л., 1967.

¹³ Там же, стр. 17.

¹⁴ Б. Н. Головин. Язык и статистика. М., 1971.

¹⁵ Там же, стр. 11.

рядка Б. Н. Головина, — В. Б.). Однако он считает неприемлемым понятие «парадигматической вероятности»,¹⁶ выдвинутое Н. Д. Андреевым. В целом взгляды Б. Н. Головина весьма близки к воззрениям Г. Хердана и П. Гиро.

Своеобразное положение по рассматриваемому здесь вопросу выдвинул Г. С. Клычков.¹⁷ Он считает, что статистическая информация о языке заключена не в самих вероятностях, а в отклонениях от них. «Если мы хотим приблизиться к познанию системы языка через статистическое изучение речи, — пишет он, — мы должны искать в речевом потоке такие отклонения от вероятностей, которые взаимосвязаны и образуют систему».¹⁸

Не совсем обычна и точка зрения Б. Ульвестада,¹⁹ который предлагает границу между языком и речью проводить на уровне 80—90% регулярной встречаемости высказываний и форм слов в речи. Все, что лежит выше этого уровня, относится к языку, все остальное принадлежит речи.

Наконец в самое последнее время в диссертации Л. А. Турыгиной,²⁰ а также в совместной статье Р. Г. Пиотровского и Л. А. Турыгиной²¹ предложена новая статистическая концепция языка и речи. Авторы поставили перед собой задачу показать, какая из двух схем — двучленная сосюрская «язык — речь» или трехчленная схема Э. Косериу «язык — норма — речь» — более отвечает реальным фактам. С этой целью они провели обширный статистический эксперимент с использованием различных статистико-вероятностных критериев оценки полученных результатов. «Если наш эксперимент, — пишут Р. Г. Пиотровский и Л. А. Турыгина, — обнаружит в тексте такую статистическую упорядоченность, которая отражает лингвистическую стабильность нормы, мы сможем более уверенно говорить о целесообразности введения третьего компонента в антиномию „язык — речь“. Если же наш эксперимент не обнаружит статистической упорядоченности и нормированности в тексте, то целесообразность самого понятия нормы будет поставлена под вопрос».²²

Проведенный авторами эксперимент подтвердил справедливость гипотезы о нормированности текста. Отсюда был сделан логический вывод, что в языке существует некоторый эталон «статис-

тического построения и упорядочения текста».²³ Таким эталоном является норма, занимающая как бы промежуточное положение между нестатистической системой языка и порождаемым ею текстом, или речью.

Таковы существующие в настоящее время статистические концепции языка и речи. Для всех этих концепций характерно, что их авторы (даже Г. С. Клычков, в известной степени противоречащий себе) считают возможным применять количественную меру не только по отношению к речи, но и по отношению к языку, вопреки приведенному в начале настоящей статьи высказыванию Н. С. Трубедко. И это вполне естественно. Как пишут Р. Г. Пиотровский и Л. А. Турыгина, «сама логика разрешения антиномии „язык — речь“ приводит языковедение к использованию статистического аппарата и вероятностной меры».²⁴

Теперь необходимо перейти к оценке рассмотренных концепций с точки зрения того вклада, который они вносят в наше представление о природе языка и речи. Следует сразу же сделать оговорку, что здесь не будет анализироваться концепция Б. Ульвестада. Она в достаточной мере произвольна, поскольку выбор порога по уровню 80—90% ничем не мотивирован. С таким же успехом в качестве этого уровня можно было бы взять 70 или, скажем, 95%.

Все остальные концепции можно разделить на три группы (из дальнейших рассуждений станет очевидным смысл такого разделения):

- 1) концепции Л. Р. Зиндера—Н. Д. Андреева и Н. Д. Андреева;
- 2) концепция Г. С. Клычкова;
- 3) концепции Г. Хердана, П. Гиро, Б. Н. Головина и Р. Г. Пиотровского—Л. А. Турыгиной.

Рассмотрим особенности каждой из групп.

В концепции Л. Р. Зиндера—Н. Д. Андреева, авторы которой оперируют понятием «речевой вероятности», не уточнено, какая количественная мера характеризует язык и какая — речь и как соотносятся эти меры. Иначе говоря, неясно, куда поместить «речевую вероятность» в схеме Ф. де Соссюра



Если, как пишут авторы, «речевая вероятность» характеризует и язык, и речь, т. е. если получаем схему типа:

¹⁶ Там же, стр. 46.

²⁴ Там же, стр. 14.

¹⁶ Там же, стр. 115—117.

¹⁷ Г. С. Клычков, ук. соч., стр. 10—13.

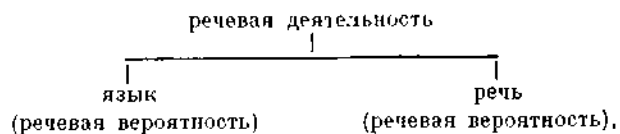
¹⁸ Там же, стр. 13.

¹⁹ B. Ulvestad. Statistik und Sprachbeschreibung. In: Das Ringen um eine neue deutsche Grammatik, Darmstadt, 1962, S. 61—73.

²⁰ Л. А. Турыгина. Статистическая интерпретация антиномии язык и речь. АКД, Л., 1970.

²¹ Р. Г. Пиотровский и Л. А. Турыгина. Антиномия «язык — речь» и статистическая интерпретация нормы языка. СтРААТ, стр. 5—46.

²² Там же, стр. 14.



то введение четвертого компонента в триаду Ф. де Соссюра позволяет опустить два других компонента, а именно: язык и речь, в результате чего получим двучленное противопоставление: «речевая деятельность — речевая вероятность».

Концепция Н. Д. Андреева может быть представлена в виде следующей схемы:



Соссюровская триада при такой интерпретации оказывается вполне оправданной. Интересно, что та вероятность, которую Н. Д. Андреев приписывает речи, у Г. Хердана, П. Гиро и Б. Н. Головина выступает как показатель языка, хотя они и не называют ее «синтагматической». Н. Д. Андреев, как отмечалось выше, ввел еще понятие «парадигматической вероятности», которое не используется больше ни в одной статистической концепции языка и речи. Благодаря этому одноуровневым единицам языка он приписал свойство равновозможности. Но в этом случае получается, что, подвергая речь статистическому анализу, мы ровным счетом ничего не узнаем о языке, поскольку постулируемая Н. Д. Андреевым равновозможность заложена в языке априори, между тем как определяемая по данным речи синтагматическая вероятность не обладает, как указывает сам автор, свойством равновозможности. Возможно, что именно эти соображения послужили Б. Н. Головину основанием для утверждения о том, что «„парадигматическая“ вероятность просто не существует».²⁵ Однако понятие «парадигматической вероятности» в принципе не противоречит требованиям теории вероятностей, поскольку в этой теории существует несколько определений вероятности, а именно: 1) классическое, или лапласовское, сводящее понятие вероятности к понятию равновозможности; 2) статистическое, или мизесовское,²⁶ определяющее вероятность как предел относительной частоты;

²⁵ Б. Н. Головин. Рец. на: Н. Д. Андреев. Статистико-комбинаторные методы в теоретическом и прикладном языковедении. Л., 1967. — ВЯ, 1969, № 1, стр. 113; см. также: Б. Н. Головин. Язык и статистика, стр. 116—117.

²⁶ R. von Mises. 1) Wahrscheinlichkeit, Statistik und Wahrheit. Wien, 1951; 2) Mathematical Theory of Probability and Statistics. New York—London, 1964.

3) психологическое, или логическое;²⁷ 4) аксиоматическое;²⁸ 5) геометрическое. Сопоставляя с этими математическими определениями вероятности понятия парадигматической и синтагматической вероятностей, нетрудно усмотреть, что первая есть всего лишь лингвистическая переформулировка вероятности в классическом варианте ее определения, а вторая представляет собой переформулировку в статистическом варианте определения. Другое дело, что парадигматическая вероятность оказывается лингвистически бессодержательной.²⁹ Кстати, именно такую оценку дают многие математики и ее математическому аналогу — лапласовской вероятности, о которой обычно говорят, что она оказывается недостаточной,³⁰ «представляет собой до известной степени бессодержательную тавтологию»,³¹ а ее приложение «не выходит за сферу азартных игр и исчисления вероятностей исхода таких игр».³² В лингвистике парадигматическая вероятность в том определении, которое дал ей Н. Д. Андреев, может найти единственное применение в типологии при сопоставлении инвентарей сходных единиц.³³ Строго говоря, эту вероятность правильнее было бы назвать удельным весом единицы в инвентаре.

²⁷ Р. Карнап. Философские основания физики. Русск. пер. М., 1971, стр. 59—85.

²⁸ А. Н. Колмогоров. Основные понятия теории вероятностей. М.—Л., 1936.

²⁹ В беседе с нами Н. Д. Андреев высказал соображение о том, что парадигматическую вероятность нельзя считать лингвистически бессодержательной, поскольку она может быть интерпретирована в терминах теории информации. Однако информационное измерение, которое может получить парадигматическая вероятность, не добавляет ничего нового к нашему представлению о ней, поскольку, при условии равновозможности, и вероятность каждой единицы в инвентаре, и количество информации, приходящееся на эту единицу, всецело определены объемом инвентаря и в этом смысле априорны. Действительно, если нам известно, что число фонем в инвентаре $n=32$, то тем самым мы уже знаем, что парадигматическая вероятность одной фонемы $P = \frac{1}{n} = \frac{1}{32}$ и что, следовательно, количество информации, приходящееся

на одну фонему, $H = \log_2 n = \log_2 32 = 5$ дв. ед. Если же объем инвентаря неизвестен, то теряет смысл сама постановка вопроса о парадигматической вероятности и ее информационном измерении. Именно поэтому К. Шеннон постоянно подчеркивает в своих работах, что количество информации не имеет ничего общего с подлинным содержанием. См., например, его статьи: 1) Математическая теория связи, стр. 243; 2) Современные достижения теории связи, стр. 405. В кн.: К. Шеннон. Работы по теории информации и кибернетике. Русск. пер. М., 1963. — Иная картина будет, по-видимому, иметь место при отсутствии условия равновозможности, о чем речь пойдет ниже.

³⁰ Б. Г. Гнеденко. Курс теории вероятностей. М., 1961, стр. 34.

³¹ А. Я. Хичин. Частотная теория Р. Мизеса и современные идеи теории вероятностей. ВФ, 1961, № 1, стр. 95.

³² З. Павловский. Введение в математическую статистику. М., 1967, стр. 10.

³³ В лингвистике известны работы, в которых производилось типологическое сопоставление языковых инвентарей под количественным углом зрения, хотя и без применения теории вероятностей. Например, такой анализ

Что касается синтагматической вероятности, то, отнеся ее всецело к речи, Н. Д. Андреев исключил синтагматическую ось из сферы языка. В этом случае становится непонятным, под действием каких причин инвентарь языка, единицы которого обладают равными парадигматическими вероятностями, способен вплотную встать с иным, неравным характером вероятностей. Поэтому не остается ничего иного, как признать, что в языке имеются не только инвентарь единиц, но и заложенные в этих единицах потенциальные возможности их взаимной сочетаемости или несочетаемости. Такие возможности реализуются в речи с различной мерой повторяемости для разных комбинаций единиц. Важно подчеркнуть, что потенциальная сочетаемость языковых единиц (т. е. их языковая синтагматика) и конкретная реализация этой сочетаемости с той или иной степенью повторяемости комбинаций (т. е. речь) представляют собой разные явления, могущие получить различную количественную оценку.

Если обратиться теперь к оставшимся двум группам концепций, то о них можно сказать следующее. Точка зрения Г. С. Клычкова, с одной стороны, и Г. Хердана, П. Гиро, Б. Н. Головина и Р. Г. Пиотровского—Л. А. Турыгиной—с другой, в определенном смысле противоположны, так как первый усматривает свойство языка в отклонениях от вероятностей, тогда как остальные видят его в устойчивом характере вероятностей. Положение Г. С. Клычкова кажется малоубедительным потому, что он, вместо того чтобы приписывать языку инвариантные свойства, проявляющиеся в фактах постоянства (в том числе и количественного), характеризует язык только вариантными свойствами, которые обнаруживаются в фактах количественной неустойчивости, или колеблемости. Данные об отклонениях от общезыковых вероятностей будут раскрывать только одну сторону языка, а именно: факт расслоения его на подязыки, причем в пределах каждого подязыка чисто «языковая» сущность опять же будет проявляться в стабильности присущих ему вероятностей. За основную же массу отклонений от вероятностей будет ответственна речь.

Все предшественники третьей группы концепций характеризуют язык свойством стабильности относительных частот, т. е. постоянными вероятностями. Однако эту стабильность отдельные ученые склонны приписывать разным факторам. Г. Хердан, П. Гиро и Б. Н. Головин считает ее неотъемлемой характеристикой языка, Р. Г. Пиотровский и Л. А. Турыгина полагают, что эта стабильность обусловлена влиянием нормы. Однако объяснение указанного свойства влиянием нормы не противоречит точке зрения Г. Хердана, П. Гиро и Б. Н. Головина, а только дополняет и углуб-

применительно к фонологическому уровню осуществлен в работах: Т. Милевский. Предпосылки типологического языкознания. Русск. пер. В сб.: Исследования по структурной типологии, М., 1963, стр. 3—31; Т. Милевский. Językoznawstwo. Warszawa, 1965, ss. 203—253.

ляет ее. Действительно, как указывал Э. Косериу, норма является компонентом языка, а не речи. «Функциональный язык (язык, на котором можно говорить), — писал Э. Косериу, — это «система функциональных противопоставлений и нормальных реализаций», или, точнее, это система и норма» (подчеркнуто Э. Косериу, — В. Б.).³⁴

При этом добавление к соскоровской схеме «нормы»³⁵ или «нормы» и «системы» не меняет ее принципиального характера, не превращает ее из двучленной в трехчленную. Ведь Ф. де Соссюр в действительности тоже пользуется трехчленной схемой: речевая деятельность (langage), язык (langue) и речь (parole). Однако эта схема своеобразна в том смысле, что логические отношения между ее членами неодинаковы. Фактически здесь наблюдается раздвоение целого, т. е. «речевой деятельности», на две противоположности: «язык» и «речь». Между «речевой деятельностью», с одной стороны, и «языком» и «речью», с другой, существует отношение включения («речевая деятельность» включает в себя «язык» и «речь»), а между «языком» и «речью» — отношение исключения («язык» исключает «речь», и наоборот). Это свойство «языка» и «речи» уже отмечалось некоторыми исследователями. Так, в одной работе по этому поводу сказано следующее: «Если мы обратимся к правилам деления понятия, то отметим, во-первых, что langue и parole в сумме должны давать langage, во-вторых, langue и parole как члены деления (подчеркнуто авторами, — В. Б.) должны исключать друг друга (по правилу дихотомии), следовательно, понятия langue и parole должны быть понятиями несовместимыми, или, точнее, противоречащими».³⁶

Если, с соблюдением указанных уточнений, схематически представить триаду Ф. де Соссюра, то ее члены должны находиться на разных уровнях иерархии, как показано на самой первой схеме (см. стр. 13). Отсюда следует, что если добавить к антиномии «язык — речь» третий член — «норму», или, по Э. Косериу, «норму» и «систему», то в этом случае будет иметь место включение в одно двучленное противопоставление другого, или вложение одной триады в другую, или, точнее, перевод одной из антиномий на одну ступень иерархии ниже, т. е. схематически

³⁴ Э. Косериу. Синхрония, диахрония и история. Русск. пер. НЛ, вып. III, М., 1963, стр. 174.

³⁵ Под «нормой» в настоящей статье мы, вслед за Э. Косериу, понимаем действующее в языке «синхронное... равновесие системы», поэтому здесь не затрагиваются вопросы, связанные с кодификацией, т. е. регистрацией нормы в сводах правил. О норме и кодификации см.: В. А. Ицкович. 1) Языковая норма. М., 1968, стр. 4—10; 2) Норма и ее кодификация. В сб.: Актуальные проблемы культуры речи, М., 1970, стр. 9—39; Л. К. Граудина. Норма и статистика. Там же, стр. 326—368.

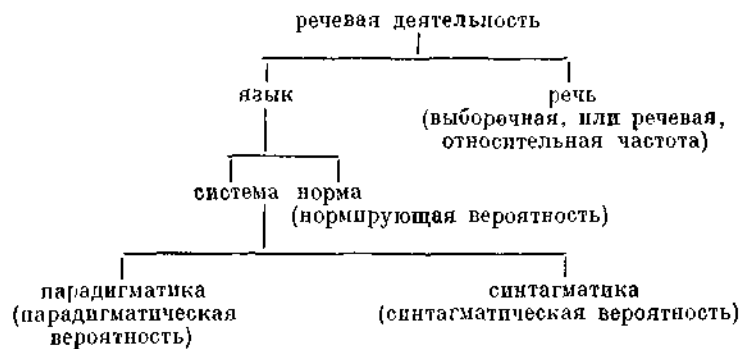
³⁶ Ю. А. Левицкий, Г. В. Петрова. Термины «язык» и «речь» в современном языкознании. Вестник МГУ, филология, № 6, 1968, стр. 17.



Итак, норму характеризует стабильность относительных частот, иначе говоря, нормирующие вероятности, определяемые на генеральной совокупности, в качестве которой может выступать язык или подязык. Что касается системы, то она, как указывают Р. Г. Пиотровский и Л. А. Турыгина,³⁷ является действительно нестатистической, но только в том смысле, что не описывается статистической вероятностью. Однако система также может получить количественную оценку на базе введенных Н. Д. Андреевым понятий парадигматической и синтагматической вероятностей, с той только разницей, что синтагматические вероятности будут относиться не к речи, а к языку (к его синтагматической оси), и будут вычисляться не по речевым выборкам, а по матрице сочетаемости единиц определенного уровня языка.

В отношении же речи можно сказать, что она будет описываться конкретными выборочными относительными частотами (их можно назвать речевыми относительными частотами). Эти частоты постоянно колеблются в больших или меньших пределах. В самом факте их колеблемости отражается их чисто речевая природа. Однако эти колебания всегда совершаются в определенных пределах, т. е. с определенной степенью стабильности, которая представляет собой не что иное, как отражение упорядочивающего характера нормы, т. е. в конечном счете языка.

Этот новый вариант статистической интерпретации языка и речи, учитывающий рассмотренные выше концепции, схематически можно представить следующим образом:



³⁷ Р. Г. Пиотровский и Л. А. Турыгина, ук. соч., стр. 46.

Парадигматические вероятности обладают свойством равно-возможности на уровне единиц языкового инвентаря и неравно-возможности на уровне их классов. Синтагматические вероятности не равновозможны как на уровне единиц, так и на уровне классов. Если парадигматические вероятности единиц могут быть описаны в рамках классического варианта теории вероятностей, то для интерпретации парадигматических вероятностей классов и синтагматических вероятностей единиц и классов в математической теории вероятностей вообще не предусмотрено специального аппарата. Классическое определение вероятности к ним неприменимо,³⁸ а эффективность применения понятия вероятности в статистическом варианте ее определения достаточно сомнительна (по крайней мере по отношению к явлениям нижних уровней языка — фонемному и морфемному). Что касается нормирующей вероятности, то ее математическим аналогом является статистическая вероятность.³⁹ Подведем итоги:

1) язык представляет собой систему с количественными ограничениями, которые проявляются как в устройстве самой системы, так и в процессе ее функционирования;

2) количественные ограничения выражаются в том, что при переходе от речи к языку в пределах речевой деятельности, от нормы к системе в пределах языка и от синтагматики к парадигматике в рамках системы снижается вариативность количественных характеристик и возрастает их жесткость; на одном конце этой цепи переходов находится речь, которая характеризуется постоянно колеблющимися относительными частотами, на другой — парадигматика, для которой, по крайней мере на уровне единиц, характерно полное отсутствие какой бы то ни было вариативности ввиду равенства парадигматических вероятностей единиц; в пределах же указанного интервала наблюдается целая гамма постепенных переходов от большей вариативности к меньшей;

3) тезис Н. С. Трубецкого, будто «язык лежит вне меры и числа», следует признать ошибочным, поскольку язык поддается измерению по трем координатам: парадигматика, синтагматика и норма.

Рассмотренные факты свидетельствуют, что лингвостатистика позволяет приоткрыть еще одну тайну человеческого языка, которая могла бы остаться не обнаруженной, если бы эта наука не получила такого развития, какое сейчас наблюдается.

³⁸ Моделью, с помощью которой обычно иллюстрируется понятие классической вероятности, является правильный многогранник (например, шестигранная игральная кость). Эта модель объясняет и парадигматическую вероятность на уровне единиц, однако не способна объяснить остальные виды вероятностей, с которыми приходится иметь дело, рассматривая систему языка. Приближенной моделью этих вероятностей может служить неправильный многогранник с различными по величине и форме гранями.

³⁹ Добавление условных вероятностей, которые, с учетом тех же ограничений, могут характеризовать систему и норму, не меняет представленной в этой схеме общей картины.

И. М. Алексеев

СЕМАНТИЧЕСКИЕ ЧАСТОТНЫЕ СЛОВАРИ

§ 1. НАЗНАЧЕНИЕ, РАЗНОВИДНОСТИ И МЕТОДЫ СОСТАВЛЕНИЯ

Частотные словари и статистическая лексикография, охватывающая работы по их изготовлению, становятся для отечественной лингвистики вполне привычным явлением. В нашей стране издано несколько десятков частотных словарей и списков; более сорока из них представлены в публикациях общесоюзной группы «Статистика речи».

Словарям этого типа начинают отводить место в лексикографических классификациях;¹ обсуждаются теоретические и методические проблемы их составления.² Лексической статистике и вообще квантитативному анализу явлений языка и речи предлагается уделять должное внимание в вузовских курсах языкознания.³

Среди частотных словарей выделяются те, в которых регистрируется семантика их входных единиц. Такие словари целесообразно называть семантическими частотными словарями. Можно ожидать, что усилия лексикографов-статистиков в будущем сконцентрируются на создании именно этой разновидности частотных словарей. «Обычный», несемантический частотный словарь предполагает дальнейшую его обработку — в зависимости от интересов того, кому он адресован. Он является, таким образом, как бы полуфабрикатом для других словарей, например для учебного

¹ Н. М. Шанский. Лексикология современного русского языка. М., 1964, стр. 309—311; А. В. Калинин. Лексика русского языка. М., 1971, стр. 227; О. С. Ахманова. Словарь лингвистических терминов. М., 1966, стр. 421; Энциклопедия кибернетики. Киев (в печати).

² См.: ИЛ, СТИ, СТИ, СтР, СтРААТ, ИЛ; А. Н. Колгушкин. Лингвистика в военном деле. М., 1970. В. М. Андрющенко. Некоторые вопросы теории и техники составления частотных словарей. Иностраные языки в школе, 1969, № 1.

³ Б. Н. Головин. Язык и статистика. М., 1971, стр. 3—4.

словаря-минимума или лексического справочника, для машинного словаря в системе автоматического перевода и автоматического поиска, а также материалом для дальнейших операций «качественного» и количественного лингвистического анализа. Семантический частотный словарь может служить учебным пособием (или почти готовым машинным словарем), сохраняя при этом все достоинства обычного частотного словаря, но зато существенно расширяя сферу своего применения и круг тех, кто им будет пользоваться.

Различия между такими словарями будут вызваны прежде всего их направленностью, однако и существующие в настоящее время немногочисленные семантические частотные словари имеют смысл рассмотреть также с точки зрения более формальных их особенностей.

Естественно различать одноязычные и неодноразличные словари; эти последние делятся на двуязычные и многоязычные. К одноязычным семантическим частотным словарям можно отнести словари языка писателя,⁴ если в них толкуются значения и указываются частоты слов, хотя эти словари не являются собственно частотными. Многоязычные представлены пока лишь одним частотным словарем Э. Итон (подробнее о конкретных словарях см. ниже).

Семантические частотные словари могут претендовать на охват лексики «языка в целом» — общие словари — или какой-то его разновидности (стиля, подязыка) — отраслевые словари. Более перспективными являются вторые, хотя бы потому, что для их изготовления требуется неизмеримо меньше усилий, чем для надежного общего словаря, и, следовательно, можно ожидать их появления в обозримые сроки. К общим относятся словари Э. Итон, Э. Торндайка, И. Лорджа, И. Лорджа—Э. Торндайка, частотные фразеологические словари, изданные в США. Все остальные семантические частотные словари являются отраслевыми.

Лингвистический уровень, представленный входными единицами словаря, также определяет различия между семантическими частотными словарями. Известны словари и списки слов (их большинство), словоформ,⁵ словосочетаний, морфем (суффиксов).

Существенная особенность того или иного семантического частотного словаря состоит в том, регистрирует ли он статистику значений, и если регистрирует, то как именно. Списки И. Лорджа и И. Лорджа—Э. Торндайка указывают номера значений слова по Большому Оксфордскому словарю и доли этих значений в сумме общей частоты употребления слова. Э. Итон приводит

⁴ Словарь языка Пушкина, тт. 1—4. М., 1956—1961; Словарь языка Абая. Алма-Ата, 1968.

⁵ Англо-русский частотный словарь словоформ (подязык электроники) составляется, например, Л. А. Пустовойтовой (группа «Статистика речи»).

показатель вхождения слова и каждого его эквивалента в других языках в одну из шести тысяч самых частых слов (цифровой индекс) и в одну из двух полутысяч этой тысячи (буквенный индекс). В опубликованных переводных частотных словарях статистика значений (в принципе⁶) отражена в произвольном выборе наиболее вероятных, по мнению составителя, выходных эквивалентов слова.

Регистрация статистики значений связана со способом составления семантического частотного словаря, и в этом способе заключается следующее различие между такими словарями. Можно назвать, например, такие способы.

1) Метод компиляции, объединения одноязычных частотных словарей. Таким способом составлен словарь Э. Итон. Возможно объединение частотных словарей, полученных в результате анализа текстов, близких по содержанию, и объединение разноязычных частотных словарей параллельных текстов. Этот последний способ связан с затруднениями чисто «технического» характера — не всегда удается найти параллельные тексты достаточного объема, разнообразия и качества. Большинство компилятивных словарей, над которыми работают в группе «Статистика речи», базируется на текстах первого типа.

2) Следующий способ можно было бы назвать «методом оценки толкового словаря по тексту». Входные единицы толкового словаря сверяются с соответствующими словоупотреблениями в репрезентативной выборке из текста. Для словарей И. Лорджа — Э. Торндайка и М. Уэста исследована выборка объемом 5 млн словоупотреблений. Сопоставлением Большого Оксфордского словаря с этими текстами многие исполнители занимались в течение десяти лет, что свидетельствует о необычной трудоемкости такого предприятия.

3) Метод анализа параллельных текстов. Выписывая очередное слово (словоформу, словосочетание) из «входного» текста, составитель регистрирует и его эквивалент так, как он представлен на языке «выходного» текста. Составлены такие англо-русские словари для подъязыков двигателей внутреннего сгорания и электроники.⁷

⁶ Именно так, поскольку в некоторых случаях словарная статья включает в себя нецифичные для данного подъязыка значения. См.: L. Hoffmann a n. 1) Fachwortschatz. Medizin. Häufigkeitwörterbuch. Leipzig, 1970, S. 71 (где английскому слову will присваивается значение wollen); 2) Fachwortschatz. Physik. Häufigkeitwörterbuch. Leipzig, 1970, S. 72 (где значение английского will объясняется через английский и немецкий примеры: I will и ich will);

⁷ В. А. Москвич. Статистика и семантика. М., 1969, стр. 159—162, где сообщается о картотеке по выборке 18 000 словоупотреблений (английский текст) и 14 000 словоупотреблений (русский текст) из подъязыка двигателей внутреннего сгорания; П. М. Алексеев. Частотный англо-русский словарь-минимум по электронике. М., 1971, раздел: частотный англо-русский словарь-минимум терминологических словосочетаний по электронике.

4) Метод регистрации значений по одноязычному тексту. Каждая учитываемая единица переводится на выходной язык в процессе расписывания одноязычного текста. Переводной эквивалент подсказывается окружением учитываемой единицы и содержанием текста.⁸

5) Метод перевода частотного словаря. Этот способ является самым простым. Одноязычный частотный словарь переводится на выходной язык без непосредственного обращения к тексту; в случае необходимости используют обычные словари, например двуязычные терминологические. Большинство существующих семантических частотных словарей составлено именно таким образом.

Наиболее аккуратным образом учитывается семантическая статистика лингвистического элемента в двуязычной ситуации двумя из перечисленных способов: при анализе параллельных текстов и при переводе единиц текста в процессе расписывания. Оба эти приема крайне трудоемки; с этой точки зрения перевод готового частотного словаря доступнее других способов. У метода компиляции имеется важное преимущество, так как он обещает сравнительно недорогое и быстрое изготовление двуязычных словарей за счет частичной автоматизации работ, т. е. независимого составления с помощью ЭВМ «левой» и «правой» половин такого словаря. Переводной словарь, особенно отраслевой и тем более терминологический, получить легче других, так как из всех семантических возможностей слова в словаре отмечаются только те значения, которые вероятнее всего связаны с употреблением слова в данном подъязыке. Выбор переводного эквивалента, следовательно, целиком зависит от осведомленности составителя в той области функционирования языка, к которой относится подъязык или совокупность текстов.

Словарь компилятивный позволяет использовать более строгие процедуры, основанные уже не только на «интуитивной» статистике, однако сейчас еще трудно рассчитывать на полностью формализованную методику объединения одноязычных частотных словарей. Представляются реалистичными несколько возможностей, первая из которых состоит в переводе одноязычного частотного словаря, а еще одна включает в себя первую как необходимый этап, после которого статистика переводов словарных единиц проверяется или, точнее, определяется с помощью частотного словаря выходного языка. При этом, разумеется, чем более ограничен тематикой текстов подъязык, тем надежнее статистика значений.

Участие ЭВМ в такой работе будет иметь большое значение, но не является принципиально решающим, поскольку самый ответственный этап — подбор выходных эквивалентов — остается за человеком. Можно надеяться, что в дальнейшем ЭВМ сумеет

⁸ По этой методике составляется, например, немецко-русский частотный словарь подъязыка хирургии (Н. Н. Яблонская, группа «Статистика речи»).

сама переводить частотные словари, но при создании машинного словаря достаточно большого объема для такого перевода, видимо, опять не обойтись без человека.

Существенную помощь при переводе частотного словаря может оказать частотный же список микроконтекстов для входного слова, анализируя который составитель и подбирает соответствующие выходные эквиваленты. Статистика микроконтекстов может быть получена как вручную, так и с помощью ЭВМ, относительно всех слов или только многозначных, выделенных составителем для этой цели при переводе словаря.

Очевидно, что еще не скоро удастся полностью автоматизировать лексикографические работы. Только обширные эксперименты с применением самых различных методов могут приблизить то время, когда на электронно-вычислительной машине станет так же легко получить неодноязычный словарь, как сегодня «простой» частотный.

§ 2. ОБЗОР СУЩЕСТВУЮЩИХ СЕМАНТИЧЕСКИХ ЧАСТОТНЫХ СЛОВАРЕЙ

I. Общие словари

1) I. L o r g e, E. L. T h o r n d i k e. A Semantic Count of English Words. New York, 1938.

2) I. L o r g e. The Semantic Count of 570 Commonest English Words. New York, 1949.

Вторым списком дополняется первый, в который не вошли 570 наиболее частых слов, исключенных первоначально из подсчетов. Объем выборки для всего словаря равен 5 000 000 словоупотреблений, значения которых определялись по Большому Оксфордскому словарю.

П р и м е р:

	OED (Th. 1b)	Per Mille (638)	No. of Units (29)
game			
<i>sb</i>	1	005	3
	3	092	18
	4	271	27
	4b	036	10
Modern Olympics		044	8
4c		009	4
	.	.	.
	.	.	.
	16	044	8
	17	005	2
<i>adj</i>	1	006	4
<i>v</i>	1	003	2
<i>vbl sb</i>	1	009	6

Расшифровка индексов такова: OED — Большой Оксфордский словарь (БОС); Th. 1b — слово входит во вторую половину первой тысячи самых частых слов в словаре Э. Торндайка. Цифры в этой же колонке соответствуют номерам значений слова по БОС, *sb* — существительное; *adj.* — прилагательное; *vbl. sb.* — отглагольное существительное. В колонке «Per Mille» число в скобках показывает частоту слова в исследованной выборке; цифры под ним означают отношение частоты каждого значения к частоте слова (в десятичной дроби опущены нуль и запятая после него). В последней колонке указано общее число источников, в которых встретилось слово (в скобках), и под ним число источников для каждого значения.

3) M. W e s t. A General Service List of English Words, London, 1953 (11th impression — 1959).

В этом словаре приводятся 2000 самых «важных» слов английского языка, опубликованные ранее (Interim Report on Vocabulary Selection. London, 1936) группой ведущих английских и американских методистов. После того как был составлен семантический словарь И. Лорджа—Э. Торндайка, М. Уэст использовал его данные для определения статистики значений каждого из этих 2000 слов.

П р и м е р:

G A M E	638	(1)	(amusement, Children's play)	
			Fun and games	
			It's not serious; it's just a game	9%
		(2)	(with the idea of competition, e. g. cards, football, etc.)	
			A game of football	
			Indoor games; outdoor games	38%
		(3)	(a particular contest)	
			We won, six games to three	
			I played a poor game	
			Play a losing game (10.5%)	23%
		(4)	(games = athletic contest)	
			Olympic Games	8%
			? [=animals, 11%; game-/, game-birds, etc., 5%]	
			[=fun Make game of, 0.5%]	

4) H. S. E a t o n. Semantic Frequency List for English, French, German and Spanish. Chicago, 1940;

H. S. E a t o n. An English—French—German—Spanish Word Frequency Dictionary. New York, 1961. Оба издания идентичны, различаясь только в названиях.

Словарь Э. Итон заслуживает особого внимания как первая и, кажется, единственная попытка создания многоязычного час-

тотного словаря на базе одноязычных. Используемые ею словари французского, немецкого и испанского языков⁹ дают частоты употребления зарегистрированных в них слов, а словарь английского языка указывает лишь, в какую полутысячу и тысячу самых частых по порядку слов входит данное слово. Поэтому Э. Торндайка, так как входным языком в ее словаре является английский. Таким образом, она, вслед за Э. Торндайком, вынуждена оперировать не частотами, а рангами (точнее, ранговыми интервалами по 500 слов в интервале).¹⁰ Это, разумеется, значительно облегчило работу, но точность данных сводного словаря сильно понизилась. Существенно пострадал словарь Э. Итон с точки зрения статистической корректности еще и в связи с различиями в объеме выборки, на которых базировались все четыре словаря. Немецкий частотный словарь составлен по выборке 11 млн словоупотреблений, английский — 9,6, испанский — 1,2 и французский — 1 млн словоупотреблений соответственно.

Чтобы как-то нейтрализовать столь большие различия в выборках, каждому языку присвоен коэффициент: немецкому — 4, английскому — 4, испанскому — 1 и французскому — 1 (Э. Итон не объясняет, почему были взяты именно такие соотношения, ведь английская выборка в 8 раз и немецкая в 9 раз превосходят испанскую, в 9,6 и в 11 раз — французскую).

Словник объемом 6,5 тыс. слов-понятий (эти цифры объясняются легко: во французский и испанский словари помещены по 6 тыс. самых частых слов) разбит на 7 частей. В каждую из шести первых входит по тысяче понятий и в седьмую около пятисот. Имеется и более дробное деление на разделы; их всего 115. Внутри раздела понятия расположены по алфавиту входных английских слов. Группировка понятий по разделам проста и остроумна.

При рассмотрении очередного понятия складываются коэффициенты всех языков, однако величина коэффициента может меняться в зависимости от того, в какую тысячу самых частых слов данного языка входит рассматриваемое слово-понятие. Например, если слово-понятие фиксируется в числе первой тысячи слов всеми четырьмя словарями, то коэффициенты для каждого языка остаются неизменными: 4, 4, 1 и 1. Сумма этих коэффициентов 10 делится на 10 и результат 1 дает основание помещать данное понятие в раздел 1. Если, далее, слово-понятие отмечено в числе первой тысячи английских, французских и испанских слов,

⁹ G. E. Vander Beke. French Word Book. New York, 1929; F. W. Kaeding. Häufigkeitwörterbuch der deutschen Sprache. Steglitz bei Berlin, 1897—1898; M. A. Buchanan. Graded Spanish Word Book. Toronto, 1927.

¹⁰ E. L. Thorndike. Teacher's Word Book of 20 000 Words. New York, 1932.

а в немецком словаре оно входит во вторую тысячу, то коэффициент для немецкого языка удваивается. Сумма $4+1+1+8=14$ делится на 10, и результат 1,4 обязывает поместить данное понятие в раздел 1,4.

Каждое из четырех слов, выражающих понятие, сопровождается индексом рангового интервала, в пределах которого это слово регистрируется соответствующим частотным словарем. Индекс 1b означает, что слово находится в пределах рангов 501 и 1000 соответствующего частотного словаря, но не определяет места понятия в общем списке Э. Итон. Место понятия устанавливается с помощью коэффициентов. Например, для английского слова game с ранговым индексом 1b в значении play найдены французское соответствие jeu^{1b}, немецкое Spiel^{2a} и испанское juego^{2a}. В результате операций с коэффициентами game (play) попало в раздел 1,5 (вторая тысяча понятий): $4+1+8+2$, деленное на 10, дает 1,5. Слово game с тем же общим для этого слова индексом 1b, но ассоциирующееся с to shoot, помещено в раздел 2,7 (третья тысяча понятий), так как имеет соответствия во французском gibier^{5a}, немецком Wild^{4a} и испанском языке caza^{2b}: $4+5+16+2$, деленное на 10, дает 2,7.

Для каждого языка в словаре имеется алфавитный указатель слов, где слово снабжено ранговым индексом и ссылкой на раздел, в котором находится соответствующее слово-понятие.

П р и м е р ы.

Section 1.5. The Second Thousand Concepts

English	French	German	Spanish
1. game ^{1b} (play)	jeu ^{1b}	Spiel ^{2a}	juego ^{2a}

Section 2.7. The Third Thousand Concepts

2. game ^{1b} (to shoot)	gibier ^{5a}	Wild ^{4a}	caza ^{2b}
----------------------------------	----------------------	--------------------	--------------------

Index to English Words in the List

3. game ^{1b}	
(play) 1.5	
(to shoot) 2.7	

Целая серия семантических частотных словарей составлена и опубликована Комитетом по обучению иностранным языкам (США). Этот комитет по инициативе Э. Торндайка с середины 20-х гг. осуществлял работы по описанию базовых языков на уровнях лексики, фразеологии и синтаксиса. Результатом явились, в частности, следующие фразеологические словари:

5) H. Keniston. Spanish Idiom List. New York, 1929;

- 6) E. Hauch. A German Idiom List. New York, 1931;
 7) F. D. Cheydler. French Idiom List. New York, 1929 (1930; 1942);
 8) Ch. B. Brown, M. L. Shane. Brazilian Portuguese Idiom List. New York, 1951.

Для каждого словаря исследовалась выборка объемом 1.2 миллиона словоупотреблений из 120 источников (по 10 000 словоупотреблений каждый), представляющих современный общелитературный язык.

Идиома определялась как словосочетание, которое при словном переводе на английский язык не сохраняет адекватного значения. Из существовавших к началу исследования фразеологических словарей извлекался канонический список идиом (французский список включал, например, 2700 идиом; бразильско-португальский — более 3500). Исполнители, распределив между собой тексты-источники, регистрировали по ним частоты употребления идиом канонического списка, одновременно выписывая из текста идиомы, отсутствующие в каноническом списке.

Словарь состоит из разделов: идиом, составных союзов, составных предлогов, возвратных глаголов, глаголов, требующих предлога, алфавитного указателя ядерных слов сочетания. Во французском словаре имеется также список отрицаний, а в испанском и бразильско-португальском — список глаголов, не требующих предлога или наречия.

В списке приводится ядерное слово в исходной форме, словосочетание и его перевод на английский язык. Упорядочен каждый список (кроме алфавитного) по числу источников, в которых встретилось словосочетание, т. е. по показателю распространенности range. Указывается также частота употребления сочетания в обследованной выборке. Количественный критерий для отбора словосочетаний в публикуемую часть словаря для разных языков различен. Во французском, например, помещено 1724 сочетания с показателем распространенности не менее 3, в бразильско-португальском — 1644 и 5 соответственно.

Пример.

Бразильско-португальский словарь

Part A

	Idioms	Range	Frequency
haver	há (muita gente aqui), «there are (many people here)»	120	2158
outro	outro (livro), «another (book)»	118	769
.	.	.	.
.	.	.	.
vez	cada vez mais (lento), «(slower) and (slower)»	69	158

Пример.

Французский словарь

List A

	Idioms and Locutions	Range	Frequency
faire	il fait (bâtir une maison), «he has (a house built)»	84	1140
avoir	il y a (des plumes sur la table), «there are (pens on the table)»	76	1638
.	.	.	.
.	.	.	.
avoir	il a peur de (tomber), «he is afraid of (falling)»	47	173

9) E. L. Thorndike. The Teaching of English Suffixes. New York, 1941.

В книге Э. Торндайка приводится список 90 английских суффиксов, наиболее важных с точки зрения обучения языку (суффикс наречия -ly, например, не рассматривается). Материалом послужили 50 000 самых употребительных слов из семантического словаря, частично опубликованного И. Лорджем и Э. Торндайком (см. №№ 1—2). В объяснительном варианте списка приводятся следующие данные: суффикс, количество разных слов, образованных суффиксом, независимо от их частоты, распределение этого количества по частотным зонам словаря, коэффициент анализа (равный проценту от числа 16-летних информантов, сумевших определить, что представляющие слова с этим суффиксом производны), коэффициент опознания (процент тех же информантов, указавших значение слова по известной основе) и перечня значений, которые придает этот суффикс словам. При каждом значении указывается количество разных слов, образованных этим суффиксом.

Эти же данные, кроме перечня и объяснения значений, представлены в более сжатом, табличном виде во втором списке суффиксов.

10) Р. И. Бигаев, Э. Р. Гукасянц, Г. Н. Михайлов, Ф. Н. Нигматова, И. Н. Соловьева, Г. Ш. Шарипов. Частотный русско-узбекский словарь-минимум. Ташкент, 1967.

Словарь предназначен в качестве пособия для студентов и учащихся средних специальных учебных заведений Узбекистана, изучающих русский язык как неродной. Объем выборки 360 тыс. словоупотреблений из общественно-политической и художествен-

ной литературы, учебников и хрестоматий, газет и журналов (всего 102 источника). 3500 слов с частотами не ниже 3 приводятся в виде алфавитно-частотного русско-узбекского словаря и частотного русского списка. Ряд производных слов, например, *колхозник*, *домик*, исключены из словника и объединены в статье со словами *колхоз* и *дом*. Некоторое количество слов, по мнению составителей, достаточно обиходных, но не обнаруженных в выборке, было взято из словаря Э. А. Штейнфельдт; каждое из них имеет помету «Ш». Необходимые для обучения слова, не отмеченные ни в выборке, ни в словаре Э. А. Штейнфельдт, включены в словарь без указания частот. По способу составления словарь является переводным. Хронологически это первый двуязычный частотный словарь, изданный в СССР.

Пример.

Делать 290. 1) қилмоқ, этмоқ, иш қилмоқ
2) ишлаб чиқармоқ, ясамоқ

Плохой *қисқа ф.* плох, плох^а, плох^о 49 ёмон, ярамас.

II. Отраслевые словари

1) С. Д. Береснев, А. И. Соловьева. Зоотехнический частотный словарь немецкого языка. Свердловск, 1968 (ротапринт).

Частотный немецко-русский словарь составлен вручную по выборке объемом около 151 000 словоупотреблений. Общее число разных слов, обнаруженных в выборке, не указывается.

Список переведенных на русский язык 2050 слов упорядочен по убыванию относительных частот, наименьшая из которых равна 0.00006 и соответствует при данной выборке абсолютной величине 9.

Пример.

№№	Слово и его значение	Относит. частота	Накопл. относит. частота	Ранг слова
1	die (f) артикль ж. р.	0.02765	0.02765	1
2	und и, а	0.02751	0.05516	2
.
.
148	geben давать	0.00095	0.54646	131
149	jeder, jede, jedes каждый	0.00095	0.54741	
150	Schwein, и свинья	0.00094	0.54835	132
151	verschieden разный, различный	0.00094	0.54929	

2) Г. Ф. Шанаурова. Медицинский частотный словарь немецкого языка. Свердловск, 1969 (ротапринт).

Словарь базируется на расписанной вручную выборке объемом 172 000 словоупотреблений. Общее число разных слов не указывается. 1998 из них расположены в частотном порядке; последнее слово имеет относительную частоту 0.00005, т. е. абсолютную — 8 или 9.

Построение обоих словарей одинаково; второй имеет в виде приложения алфавитный список немецких слов, при которых указаны их номера из первой колонки слева (см. пример).

3) С. Д. Береснев, М. Т. Есаулова. Ветеринарно-зоотехнический частотный словарь немецкого языка. Свердловск, 1969 (ротапринт).

Этот словарь объединяет результаты обследования двух выборок: из текстов по зоотехнике (см. выше, разд. II, № 1) и ветеринарии, общим объемом около 300 тыс. словоупотреблений. В числе первых по частоте двух тысяч слов из каждой выборки оказались общими для обоих подязыков 1035 слов. Из них в немецко-русский список вошли 870 знаменательных слов. Приводятся также списки «по убыванию частот» (частоты, впрочем, не указаны; порядок не алфавитный) 600 слов, характерных для зоотехнических текстов, и 800 слов, встретившихся только в текстах по ветеринарии. Слова во всех трех списках сопровождаются русскими переводами. Прилагается список служебных слов без указания частот.

В отличие от словарей №№ 1 и 2 данного раздела изъята колонка «ранг слова» и введена колонка «абсолютная частота».

Оформление всех трех словарей нельзя признать удачным, хотя сам факт их появления весьма отраден — это первые отечественные неодноязычные отраслевые частотные словари.

Если такой словарь рассчитан на массового потребителя, особенно учащегося (а так оно и есть), то обилие цифровой информации только мешает ему пользоваться словарем. Излишни, например, относительные частоты и то, что составители называют «рангом слова» (об этом ниже). Накопленные относительные частоты окажутся доступнее и нагляднее, если они даются в отдельной таблице и не для каждого слова, а через интервалы.

Если такой словарь рассчитан на специалиста в области лингвостатистики, то прежде всего каждое слово должно сопровождаться абсолютной частотой, так как относительные частоты во многих случаях определены с округлением, из-за чего при обратном переходе к абсолютной частоте она может оказаться дробной (например, $0.00005 \cdot 172000 = 8.6$).

Досадное упущение составителей первых двух словарей заключается в ошибочном понимании термина «ранг слова». В действительности они имеют в виду ранг частоты, а не слова (см.

пример). Последнее из 2050 слов (словарь № 1, разд. II) получает, таким образом, ранг не 2050, а 256, так как именно 256 разных частот оказалось на 2050 самых частых слов.

Этой ошибки не избежал и проф. Л. Гофман из ГДР (№№ 4—9), который также называет порядковый номер (ранг) частоты рангом слова.¹¹

L. H o f f m a n n. Fachwortschatz Medizin Häufigkeitswörterbuch. Leipzig, 1970:

4. Russisch;
5. Englisch;
6. Französisch.

L. H o f f m a n n. Fachwortschatz Physik Häufigkeitswörterbuch. Leipzig, 1970:

7. Russisch;
8. English;
9. Französisch.

Каждый из шести словарей включает в себя 1100—1200 русских, английских и французских слов (кроме артиклей), представленных в двух списках — частотном (без перевода на немецкий язык) и алфавитно-частотном (с переводом). При входном слове указан ранг слова (в действительности это ранг частоты) и относительная частота с точностью до шестого знака после запятой, что не представляется оправданным, поскольку не известны ни объем выборки, ни абсолютные частоты. В связи с этим цифровые данные в значительной мере теряют свое значение как лингвостатистический материал.

¹¹ Можно еще и сегодня сетовать по поводу неустоявшейся в лингвостатистике терминологии, однако понятие «ранг слова» настолько прочно вошло в обиход, особенно после ставших классическими работ Дж. Циффа и его последователей, что лишь недоумение вызывает путаница в использовании этого термина. Ср., например, два разных его употребления в одной работе: Р. М. Ф р у м к и н а. Вероятность элементов текста и речевое поведение. М., 1971; на стр. 14 рангом называется порядковый номер слова в списке по убыванию частот слов, а на стр. 56 (сноска 3) утверждается, что в лексической статистике рангом «обычно» называют число разных текстов, содержащих данное слово. Другой пример еще более удивительной непоследовательности см. в работе: Л. Н. З а с о р и н а. Статистика и автоматизация в лексикографии. Л., 1966, где на стр. 61, 63, 82 рангом называется номер слова, а на стр. 98 — номер частоты. Этот же автор (рец. на: N. P. V a k a r. A Word Count of Spoken Russian. — Филологические науки, 1968, № 5, стр. 107) упрекает составителя частотного словаря (пищущего на английском языке) в неправильном (!) употреблении термина range ('область', 'диапазон', — П. А.) в значении 'ранг', спутав его со словом rank ('ранг', — П. А.). . .

Пример (№ 4):

Häufigkeitsliste Russisch

Rang	Wort	relative Häufigkeit
1	в	0.041714
2	и; и . . .	0.032685
.	.	.
.	.	.
.	.	.
26	являться	0.002971
27	заболевание	0.002942
	случай	
28	весь	0.002800

Alphabetisches Verzeichnis Russisch-Deutsch

в	1/0.041714	in; nach; um, am;
.	.	von, zu, bei
.	.	.
.	.	.
.	.	.
и; и . . . и	2/0.032685	und; sowohl . . . als auch

В случаях омонимии частей речи при входном слове в алфавитно-частотном списке указывается его общая частота и отдельно частоты для каждого омонима:

base 86/0.000400. Basis, Grundlage, Base 0.000 200; basieren, beruhen 0.000200

Это позволяет предполагать, что какая-то семантика, хотя бы на уровне частей речи, учитывалась в процессе анализа текста.

П. М. А л е к с е е в. Частотный англо-русский словарь-минимум по электронике. М., 1971.

Словарь состоит из двух частей.

10) Частотный англо-русский словарь-минимум однословных терминов по электронике.

Выборка объемом 200 000 словоупотреблений из журнальных и книжных текстов США — Великобритании распечатана вручную. Обнаружено более 10 тыс. разных словоформ, которые сведены к 7 тыс. исходных форм слов. Из этого количества в словарь попали 2867 однословных терминов с частотами не менее 2. Переводы к ним даны с учетом специфики подязыка электроники. При каждом входном слове указаны частота его употребления в английском тексте и количество текстов, в которых зарегистрировано слово (из 200).

Словарь представлен в двух видах — алфавитно-частотном и частотном. В последнем отсутствуют переводы и показатель распространенности, зато в отличие от первого указаны ранги слов.

Пример:

Алфавитно-частотный вариант				
addition	n	71	49	добавление, сложение
additive	a	3	2	аддитивный
address	n	6	4	адрес

Частотный вариант			
21	junction	n	337
22—23	material	n	331
	output	n	
24	crystal	n	324

11) Частотный англо-русский словарь-минимум терминологических сочетаний по электронике.

Выборка объемом 200 000 словоупотреблений (около 10 тыс. употреблений терминологических словосочетаний одиннадцати самых частых типов) из английского текста и его русского перевода расписывалась вручную, причем для каждого словосочетания в английском тексте сразу же отыскивался его эквивалент в русском тексте.

В словарь включены 2443 словосочетания с частотами не менее 2, представленные в двух видах — алфавитно-частотном (с переводами) и частотном (без перевода). Как и в словаре № 10, в алфавитно-частотном варианте не указаны ранги, но частота сопровождается показателем распространенности (из пяти, по числу основных групп текстов). В частотном этого показателя нет, но введены ранги словосочетаний.

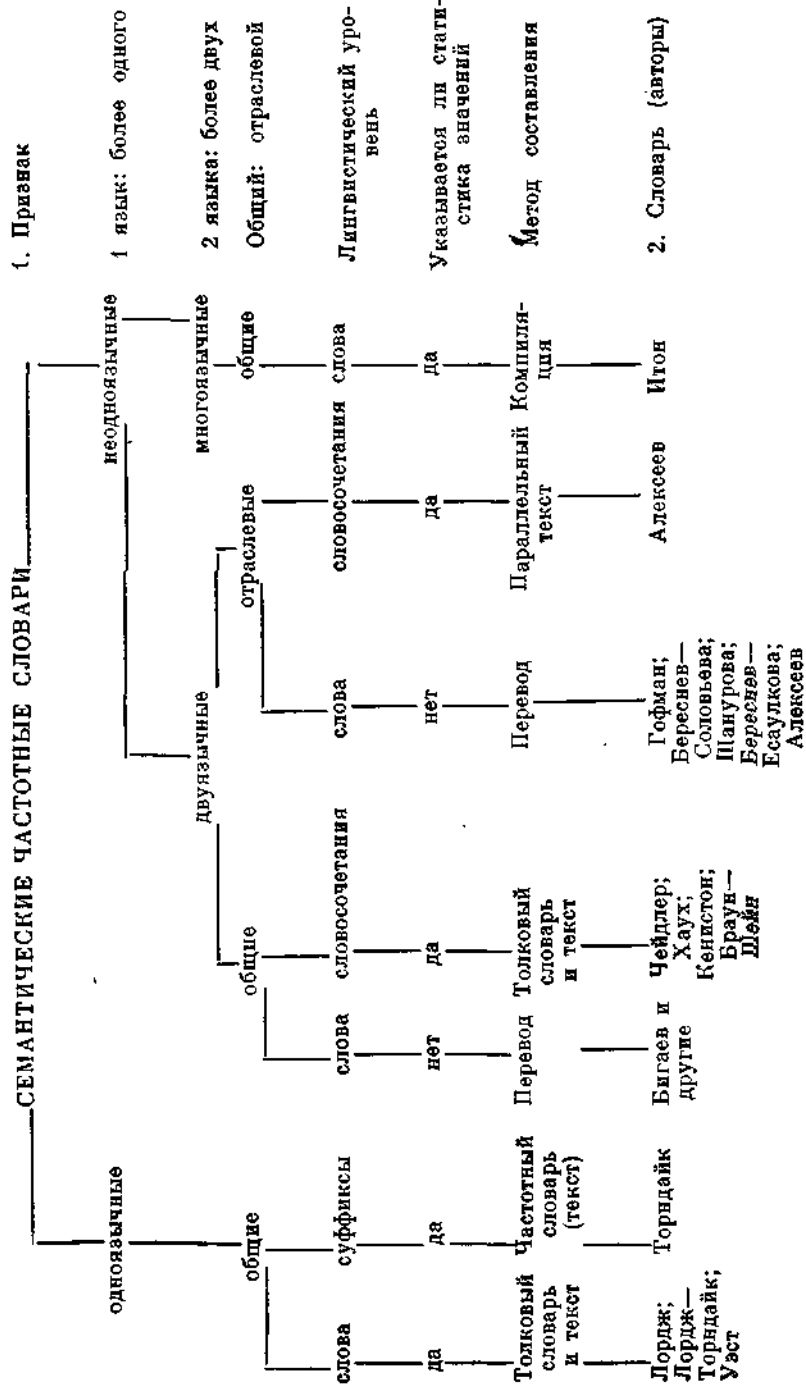
Пример:

Алфавитно-частотный вариант				
account				
a. for	5	2		учитывать, объяснять
accuracy				
angular a.	3	1		угловая точность
time a.	3	1		временная точность

Частотный вариант		
41—43	coaxial cable equivalent circuit similar to	21

Некоторое количество словосочетаний, не являясь собственно терминологическими, например similar to, вошло в словарь благо-

СЕМАНТИЧЕСКИЕ ЧАСТОТНЫЕ СЛОВАРИ



Классификация опубликованных семантических частотных словарей.

даря высокой степени их употребительности в научно-техническом тексте.

Выше были рассмотрены только те семантические частотные словари, с которыми автор сумел ознакомиться лично (кроме общих словарей №№ 1, 5 и 6, разд. I относительно содержания и методики которых, тем не менее, удалось получить достаточные сведения); их классификация представлена на рисунке. По-видимому, за пределами обзора оказалось лишь небольшое число опубликованных словарей подобного рода. Известно, что некоторое количество двуязычных частотных словарей существует в неизданном виде, готовится к печати и составляется. В одной группе «Статистика речи» ведутся работы по получению не менее двух десятков двуязычных отраслевых частотных словарей, в основном переводных и компилятивных, причем почти все новые исследования в этой группе ориентированы на использование ЭВМ. Более того, очевидно, что каждый отраслевой частотный словарь, составленный в коллективе «Статистики речи», будет в свое время преобразован в двуязычный и тем самым послужит дальнейшему развитию статистической лексикографии, еще более приблизив ее к запросам сегодняшнего и завтрашнего дня.

М. Е. Каширина

ЧАСТОТНЫЙ СЛОВАРЬ АНГЛИЙСКОГО ПОДЪЯЗЫКА ФИЗИКИ ТВЕРДОГО ТЕЛА

Частотный словарь (см. табл. 1) составлен в результате статистического исследования текстов по физике твердого тела длиной 200 тыс. словоупотреблений. Использовались статьи из английских, американских, канадских, индийских и японских журналов за 1965—1968 гг., материалы конференций, монографии и учебники. Словарь составлен вручную по методике, принятой в группе «Статистика речи».

Для обеспечения репрезентативности выборки предварительному анализу были подвергнуты двенадцать номеров реферативного журнала АН СССР «Физика», в результате чего получена следующая схема распределения минимальных выборок по разделам физики твердого тела:

- | | |
|---|--------|
| 1) твердые металлы | — 50; |
| 2) твердые тела | — 50; |
| 3) твердые диэлектрики и полупроводники | — 100. |

Ввиду того что количество публикаций в жанре статьи занимало около 3/4 общего числа названий, было решено отразить это соотношение в схеме выборки. С этой целью 100 минимальных выборок (текстов) по твердым полупроводникам и диэлектрикам разделены на две группы по 50 минимальных выборок в каждой (50 из журнальных статей, 50 из книг).

Приводимая схема распределения текстов по разделам и подразделам физики твердого тела выглядела нижеследующим образом (в количествах минимальных выборок).

1. Твердые металлы

- | | |
|--|------|
| 1) общие вопросы | — 1; |
| 2) теория твердых металлов | — 2; |
| 3) химическая связь и кристаллические поля | — 1; |
| 4) упругие и тепловые свойства; фононные спектры | — 3; |
| 5) диффузия и электроперенос | — 9; |

6) электрические и оптические свойства	— 19;
7) магнитные свойства	— 11;
8) взаимодействие излучений с твердыми металлами	— 4;
итоги	— 50.

2. Твердые тела

1) общие вопросы	— 4;
2) теория твердого тела	— 4;
3) теория и методы структурного анализа	— 4;
4) упругие и тепловые свойства; релаксация	— 2;
5) диффузия	— 2;
6) фазовые равновесия и фазовые превращения	— 4;
7) кристаллизация и плавление	— 12;
8) структура реального кристалла	— 10;
9) механические свойства	— 6;
10) взаимодействие излучений с твердыми телами	— 2;
итоги	— 50.

3. Твердые диэлектрики и полупроводники
(выборки отдельно по книгам и по статьям)

1) общие вопросы	— 2, 2;
2) теория твердых диэлектриков и полупроводников	— 7, 7;
3) химическая связь и кристаллические поля	— 5, 5;
4) упругие и тепловые свойства; фононные спектры	— 5, 5;
5) диффузия	— 1, 1;
6) электрические и оптические свойства	— 23, 23;
7) магнитные свойства	— 5, 5;
8) взаимодействие излучений с диэлектриками и полупроводниками	— 2, 2;
итоги	— 50, 50.

Обследованию не подвергались заглавия статей, аннотации, подписи под рисунками и раздел «Acknowledgements». Уравнения и формулы, занимающие отдельную строку, при подсчете не учитывались. Формулы, входящие в строку, считались одним словом-употреблением. Все математические, химические и другие символы учитывались под знаком «у», все слова с заглавной буквы — под знаком «N».

При подсчете лексико-грамматическая омонимия по возможности различалась. Части речи и грамматические категории обозначены при помощи общепринятой индексации.

Данный словарь отличается от частотного словаря английских текстов по физике твердого тела Р. С. Мелик-Гусейновой объемом исследованного материала (словарь Мелик-Гусейновой —

Частотный список словоформ

i	Словоформы	F
1	the a	20244
2	«у»	16459
3	of prp	9392
4	in prp	5402
5	and cj	4831
6	a a	3610
7	«N»	3108
8	is v/l	2495
9	to prp	2369
10	for prp	2336
11	by prp	2144
12	to part	1678
13	with prp	1557
14	that cj	1550
15	is v/aux	1484
16	as cj	1420
17	this pron	1397
18	from prp	1182
19	be v/aux	1134
20	which pron	1063
21	on prp	1017
22	it pron	982
23	at prp	928
24	we pron	878
25	an a	847
26	are v/aux	765
27	are v/l	727
28	was v/aux	707
29	energy n	630
30	temperature n	593
31	not part	587
32	can v	557
33	these pron	545
34	where adv/cj	539
35	were v	518
36	crystal n	505
37	or cj	486
38	field n	470
39	be v/l	467
40	fig. ab	460
41	between prp	452
42	but cj	438
43	been v/aux	436
44	two num	427
45—46	have v/aux, if cj	409
47	than cj	403
48	may v	397
49	given pp	354
50	only adv	349
51	has v/aux	347

Таблица 1 (продолжение)

i	Словоформа	F
52	also <i>adv</i>	344
53	values <i>n</i>	324
54	shown <i>pp</i>	323
55	value <i>n</i>	322
56	surface <i>n</i>	314
57	all <i>pron</i>	311
58	obtained <i>pp</i>	295
59	equation <i>n</i>	293
60	band <i>n</i>	290
61	case <i>n</i>	288
62	then <i>adv/cj</i>	285
63	used <i>pp</i>	281
64	will <i>v/aux</i>	275
65—86	crystals <i>n</i> , one <i>num</i>	269
67	electron <i>n</i>	262
68	such <i>pron</i>	260
69	lattice <i>n</i>	257
70	more <i>adv</i>	255
71	when <i>adv/cj</i>	247
72	effect <i>n</i>	245
73	however <i>cj</i>	244
74	same <i>pron</i>	242
75	function <i>n</i>	239
76	measurements <i>n</i>	237
77	there <i>part</i>	236
78	observed <i>pp</i>	235
79	thus <i>adv</i>	230
80	so <i>adv/cj</i>	229
81—83	model <i>n</i> , since <i>cj</i> , some <i>pron</i>	227
84—85	density <i>n</i> , one <i>pron</i>	223
86	data <i>n</i>	222
87—88	magnetic <i>adj</i> , that <i>pron</i>	219
89—90	electrons <i>n</i> , very <i>adv</i>	218
91—92	into <i>prp</i> , other <i>pron</i>	216
93	have <i>v</i>	211
94—95	form <i>n</i> , results <i>n</i>	207
96	work <i>n</i>	206
97	small <i>adj</i>	204
98	present <i>adj</i>	199
99	found <i>pp</i>	197
100	direction <i>n</i>	195
101—102	both <i>pron</i> , different <i>adj</i>	194
103	high <i>adj</i>	193
104	conductivity <i>n</i>	192
105—107	atoms <i>n</i> , made <i>pp</i> , method <i>n</i>	191
108	impurity <i>n</i>	190
109	no <i>pron</i>	189
110	number <i>n</i>	188
111	was <i>v/l</i>	187
112—113	eq. <i>ab</i> , temperatures <i>n</i>	183
114	free <i>adj</i>	181
115—118	absorbtion <i>n</i> , range <i>n</i> , terms <i>n</i> , those <i>pron</i>	179
119	about <i>prp</i>	178

Таблица 1 (продолжение)

i	Словоформы	F
120	each <i>pron</i>	176
121	state <i>n</i>	175
122—123	its <i>pron</i> , order <i>n</i>	173
124—126	equations <i>n</i> , experimental <i>adj</i> , sample <i>n</i>	172
127—128	they <i>pron</i> , their <i>pron</i>	171
129	measured <i>pp</i>	169
130	is <i>v</i>	168
131	point <i>n</i>	167
132	out <i>prp</i>	166
133—134	plane <i>n</i> , single <i>adj</i>	165
135—136	low <i>adj</i> , using <i>p</i>	163
137	using <i>n</i>	162
138—139	first <i>num</i> , structure <i>n</i>	161
140—142	due <i>adj</i> , general <i>adj</i> , time <i>n</i>	159
143	concentration <i>n</i>	158
144—145	current <i>n</i> , would <i>v</i>	157
146—148	calculated <i>pp</i> , thermal <i>adj</i> , type <i>n</i>	155
149	region <i>n</i>	154
150	constant <i>n</i>	153
151—153	most <i>adv</i> , optical <i>adj</i> , pressure <i>n</i>	151
154	large <i>adj</i>	150
155—156	here <i>adv</i> , therefore <i>adv</i>	149
157	now <i>adv</i>	148
158—160	figure <i>n</i> , must <i>v</i> , similar <i>adj</i>	147
161	axis <i>n</i>	144
162	should <i>v/mod</i>	142
163—165	applied <i>pp</i> , dependence <i>n</i> , material <i>adj</i> .	141
166—168	atom <i>n</i> , because <i>cj</i> , within <i>prp</i>	139
169	possible <i>adj</i>	138
170—171	are <i>v</i> , see <i>v</i>	136
172—175	curve <i>n</i> , diffusion <i>n</i> , symmetry <i>n</i> , well <i>adv</i>	135
176	our <i>pron</i>	134
177—179	determined <i>pp</i> , over <i>prp</i> , transition <i>n</i>	133
180—183	any <i>pron</i> , states <i>n</i> , stress <i>n</i> , three <i>num</i>	131
184	table <i>n</i>	130
185—188	parallel <i>adj</i> , properties <i>n</i> , specimens <i>n</i> , theory <i>n</i>	129
189	has <i>v</i>	128
190—191	approximation <i>n</i> , wave <i>n</i>	127
192—193	change <i>n</i> , ration <i>n</i>	125
194	corresponding <i>p</i>	124
195—196	distribution <i>n</i> , samples <i>n</i>	123
197—198	et (в сочетании: et al), light <i>n</i>	122
199—201	per <i>prp</i> , section <i>n</i> , term <i>n</i>	121
202—203	expression <i>n</i> , « <i>γ</i> » — type <i>n</i>	120
204	scattering <i>n</i>	119
205—208	al. <i>ab</i> , emission <i>n</i> , heat <i>n</i> , shows <i>n</i>	118
209—210	conditions <i>n</i> , specimen <i>n</i>	117
211	level <i>n</i>	116
212—213	condition <i>n</i> , constants <i>n</i>	115
214	after <i>prp</i>	114
215—216	curves <i>n</i> , under <i>n</i>	113
217	energies <i>n</i>	112
218	could <i>v</i>	110

Таблица 1 (продолжение)

i	Словоформы	F
219—220	factor <i>n</i> , less <i>adv</i>	109
221—222	rate <i>n</i> , solution <i>n</i>	108
223—224	associated <i>pp</i> , dislocation <i>n</i>	107
225—228	described <i>pp</i> , effects <i>n</i> , i. e. <i>ab</i> taken <i>pp</i>	106
229—230	along <i>prp</i> , system <i>n</i>	105
231	were <i>v/l</i>	104
232—234	discussed <i>pp</i> , respectively <i>adv</i> , strain <i>n</i>	103
235—237	higher <i>a</i> , pure <i>a</i> , variation <i>n</i>	102
238	analysis <i>n</i>	101
239—240	components <i>n</i> , through <i>adv/prp</i>	100
241—245	effective <i>adj</i> , excitation <i>adj</i> , experiments <i>s</i> , way <i>n</i> , while <i>cj</i>	99
246—247	equilibrium <i>n</i> , much <i>adv</i>	98
248—249	obtain <i>v</i> , simple <i>adj</i>	97
250—252	dielectric <i>adj</i> , mass <i>n</i> , unit <i>n</i>	96
253—258	aluminium <i>n</i> , behavior <i>n</i> , difference <i>n</i> , electrical <i>a</i> , electric <i>a</i> , show <i>v</i>	95
259—264	carriers <i>n</i> , elements <i>n</i> , <i>ev ab</i> , give <i>v</i> , part <i>n</i> , volume <i>n</i>	94
265—266	potential <i>n</i> , problem <i>n</i>	93
267—270	axes <i>n</i> , copper <i>n</i> , frequency <i>n</i> , paper <i>n</i>	92
271—273	films <i>n</i> , hence <i>adv</i> , interaction <i>n</i>	91
274—275	phase <i>n</i> , points <i>n</i>	90
276—281	dislocations <i>n</i> , various <i>adj</i> , known <i>pp</i> , line <i>n</i> , lines <i>n</i> , several <i>pron/n</i>	89
282—287	atomic <i>adj</i> , bands <i>n</i> , <i>cm ab</i> , lower <i>adj</i> , particular <i>adj</i> , thickness <i>n</i>	88
288—291	condition <i>n</i> , consider <i>v</i> , elastic <i>adj</i> , parameters <i>n</i>	87
292—294	important <i>adj</i> , matrix <i>n</i> , planes <i>n</i>	85
295	up <i>adv/prp</i>	84
296—306	being <i>p/aux</i> , being <i>p/l</i> , component <i>n</i> , considered <i>pp</i> , example <i>n</i> , impurities <i>n</i> , magnitude <i>n</i> , materials <i>n</i> , presence <i>n</i> , shape <i>n</i> , velocity <i>n</i>	83
307—309	boundary <i>n</i> , cases <i>n</i> , expected <i>n</i>	82
310—313	changes <i>n</i> , relative <i>adj</i> , relation <i>n</i> , second <i>num</i>	81
314	polarization <i>n</i>	80
315—320	approximately <i>adv</i> , calculations <i>n</i> , fact <i>n</i> , functions <i>n</i> , levels <i>n</i> , procedure <i>n</i>	79
321—322	many <i>pron</i> , metals <i>n</i>	78
323—327	coefficients <i>n</i> , defined <i>pp</i> , force <i>n</i> , independent <i>adj</i> , orientation <i>n</i>	77
328—330	does <i>v/aux</i> , either <i>cj</i> , interface <i>n</i>	76
331—336	according <i>prp</i> , alloys <i>n</i> , angle <i>n</i> , gives <i>v</i> , junctions <i>n</i> , process <i>n</i>	75
337—339	below <i>prp</i> , means <i>n</i> , shall <i>v/aux</i>	74
340—344	above <i>prp/adv</i> , agreement <i>n</i> , basis <i>n</i> , becomes <i>v</i> , result <i>n</i>	73
345—352	film <i>n</i> , length <i>n</i> , linear <i>adj</i> , normal <i>adj</i> , occur <i>v</i> , photoconductivity <i>n</i> , position <i>n</i> , radiation <i>n</i>	72
353—358	assumed <i>pp</i> , necessary <i>adj</i> , rather <i>adv</i> , room <i>n</i> , them <i>pron</i> , zero <i>n</i>	71
359—362	even <i>adv</i> , holes <i>n</i> , <i>mm ab</i> , principal <i>adj</i>	70
363—371	activation <i>n</i> , calculation <i>n</i> , charge <i>n</i> , derived <i>pp</i> , edge <i>n</i> , quite <i>adv</i> , semiconductors <i>n</i> , will <i>v</i> , valence <i>n</i>	69
372—378	coefficient <i>n</i> , constant <i>adj</i> , directions <i>n</i> , follows <i>v</i> , ionizations <i>n</i> , metal <i>n</i> , spectra <i>n</i>	68

Таблица 1 (продолжение)

i	Словоформы	F
379—385	equal <i>adj</i> , features <i>n</i> , following <i>adj</i> , junction <i>n</i> , moment <i>n</i> , reported <i>pp</i> , vector <i>n</i>	67
386—391	basal <i>adj</i> , during <i>prp</i> , increase <i>n</i> , parameter <i>n</i> , voltage <i>n</i> , written <i>pp</i>	66
392—397	although <i>cj</i> , motion <i>n</i> , relaxation <i>n</i> , resonance <i>n</i> , thin <i>adj</i> , waves <i>n</i>	65
398—404	comparison <i>n</i> , experiment <i>n</i> , growth <i>n</i> , larger <i>adj</i> , measurement <i>n</i> , related <i>pp</i> , seen <i>pp</i>	64
405—411	cubic <i>adj</i> , peak <i>n</i> , prepared <i>pp</i> , proportional <i>adj</i> , set <i>n</i> , spectrum <i>n</i> , tensor <i>n</i>	63
412—419	alloy <i>n</i> , below <i>adv</i> , carried <i>pp</i> , distance <i>n</i> , fields <i>n</i> , make <i>v</i> , produced <i>pp</i> , spin <i>n</i>	62
420—427	carrier <i>n</i> , compared <i>pp</i> , germanium <i>n</i> , maximum <i>n</i> , mechanism <i>n</i> , methods <i>n</i> , positive <i>adj</i> , recombination <i>n</i>	61
428—435	completely <i>adv</i> , layer <i>n</i> , quantities <i>n</i> , resistivity <i>n</i> , solutions <i>n</i> , susceptibility <i>n</i> , theoretical <i>adj</i> , technique <i>n</i>	60
436—441	average <i>adj</i> , greater <i>adj</i> , intensity <i>n</i> , <i>sec ab</i> , step <i>n</i> , surfaces <i>n</i>	59
442—447	deformation <i>n</i> , directly <i>adv</i> , ion <i>n</i> , let <i>v</i> , previously <i>adv</i> , wavelength <i>n</i>	58
448—454	acceptor <i>n</i> , another <i>n</i> , area <i>n</i> , course <i>n</i> , hole <i>n</i> , magnetization <i>n</i> , silicon <i>n</i>	57
455—456	increasing <i>p</i> , sum <i>n</i>	56
457—468	always <i>adv</i> , based <i>pp</i> , contribution <i>n</i> , containing <i>p</i> , direct <i>adj</i> , electronic <i>adj</i> , formation <i>n</i> , investigated <i>pp</i> , studied <i>pp</i> , total <i>adj</i> , treatment <i>n</i>	55
469—476	cell <i>n</i> , compounds <i>n</i> , few <i>pron</i> , transitions <i>n</i> , us <i>pron</i> , use <i>n</i> , usually <i>adv</i> , width <i>n</i>	54
478—481	certain <i>adj</i> , discussion <i>n</i> , might <i>v</i> , negative <i>adj</i>	53
482—488	across <i>prp</i> , contact <i>n</i> , detailed <i>pp</i> , factors <i>n</i> , group <i>n</i> , resistance <i>n</i> , solid <i>n</i>	52
489—497	addition <i>n</i> , again <i>adv</i> , bulk <i>n</i> , four <i>num</i> , good <i>adj</i> , index <i>n</i> , nature <i>n</i> , still <i>adv</i> , too <i>adv</i>	51
498—506	account <i>n</i> , appear <i>v</i> , be <i>v</i> , following <i>p</i> , in <i>ab</i> , physical <i>adj</i> , plotted <i>pp</i> , solid <i>adj</i> , «x»-ray <i>n</i>	50
507—516	appears <i>v</i> , further <i>a</i> , increases <i>v</i> , ions <i>n</i> , platinum <i>n</i> , reduced <i>pp</i> , reflectance <i>n</i> , sites <i>n</i> , substrate <i>n</i> , take <i>v</i>	49
517—523	determine <i>v</i> , layers <i>n</i> , pits <i>n</i> , shear <i>n</i> , slip <i>n</i> , usual <i>adj</i> , without <i>prp</i>	48
524—535	assumption <i>n</i> , beam <i>n</i> , before <i>adv</i> , diffraction <i>n</i> , dispersion <i>n</i> , first <i>num</i> , information <i>n</i> , perfect <i>adj</i> , respect <i>n</i> , study <i>n</i> , somewhat <i>adv</i> , spectral <i>adj</i>	47
536—544	above <i>adv/adj</i> , appropriate <i>adj</i> , exited <i>pp</i> , flow <i>n</i> , forces <i>n</i> , generally <i>adv</i> , grown <i>pp</i> , new <i>adj</i> , recently <i>adv</i>	46
545—549	above <i>prp</i> , expressions <i>n</i> , neighbours <i>n</i> , separation <i>n</i> , vacuum <i>n</i>	45
550—555	available <i>adj</i> , expansion <i>n</i> , having <i>p/aux</i> , how <i>adv</i> , incident <i>adj</i> , size <i>n</i>	44
556—567	already <i>adv</i> , anisotropy <i>n</i> , degree <i>n</i> , find <i>v</i> , power <i>n</i> , real <i>adj</i> , series <i>n</i> , structures <i>n</i> , source <i>n</i> , use <i>v</i> , zink <i>n</i>	43
568—580	assumed <i>v</i> , concentrations <i>n</i> , cross <i>adj</i> , essentially <i>adv</i> , figs. <i>ab</i> , indicate <i>v</i> , long <i>adj</i> , rear <i>prp</i> , property <i>n</i> , pulse <i>n</i> , semiconductor <i>n</i> , simply <i>adv</i> , until <i>adv</i>	42

Таблица 1 (продолжение)

i	Словоформы	F
581—591	called <i>pp</i> , centres <i>v</i> , deviation <i>n</i> , difficult <i>adj</i> , expressed <i>pp</i> , formed <i>pp</i> , pair <i>n</i> , regions <i>n</i> , smaller <i>adj</i> , situation <i>n</i> , varies <i>v</i>	41
592—606	barrier <i>n</i> , depends <i>v</i> , except <i>pp</i> , introduced <i>pp</i> , iron <i>n</i> , melting <i>ger</i> , negligible <i>adj</i> , nuclei <i>n</i> , perpendicular <i>adj</i> , represents <i>v</i> , result <i>v</i> , space <i>n</i> , transformation <i>n</i> , upon <i>pp</i> , vacancies <i>n</i>	40
607—622	calculate <i>v</i> , chosen <i>pp</i> , clearly <i>adv</i> , differences <i>n</i> , evidence <i>n</i> , excess <i>n</i> , experimentally <i>adv</i> , further <i>adv</i> , interest <i>n</i> , just <i>adv</i> , minimum <i>n</i> , primary <i>a</i> , previous <i>a</i> , results <i>v</i> , specific <i>a</i> , zone <i>n</i>	39
623—646	around <i>pp/adv</i> , correction <i>n</i> , cyclotron <i>n</i> , complete <i>a</i> , displacement <i>n</i> , done <i>pp</i> , eqs. <i>ab</i> , exact <i>a</i> , frequencies <i>n</i> , figures <i>n</i> , limit <i>n</i> , manner <i>n</i> , note <i>n</i> , neglected <i>pp</i> , near <i>adv</i> , presented <i>pp</i> , polycrystalline <i>a</i> , positions <i>n</i> , relationship <i>n</i> , response <i>n</i> , stage <i>n</i> , there <i>adv</i> , transport <i>n</i> , uniform <i>a</i>	38
647—666	against <i>pp</i> , additional <i>a</i> , actual <i>a</i> , contrast <i>n</i> , center <i>n</i> , carbon <i>n</i> , double <i>a</i> , end <i>n</i> , grain <i>n</i> , injection <i>n</i> , molecules <i>n</i> , refractive <i>a</i> , relations <i>n</i> , sections <i>n</i> , showed <i>v</i> , shift <i>n</i> , typical <i>p</i> , true <i>a</i> , total <i>n</i> , times <i>n</i>	37
667—682	acceptors <i>n</i> , arrangement <i>n</i> , depend <i>v</i> , do <i>v/aux</i> , determination <i>n</i> , error <i>n</i> , involved <i>pp</i> , isotropic <i>a</i> , itself <i>pron</i> , mean <i>a</i> , molecule <i>n</i> , reflection <i>n</i> , relatively <i>adv</i> , side <i>n</i> , studies <i>n</i> , vacancy <i>n</i>	36
683—695	angular <i>a</i> , corresponds <i>v</i> , deposition <i>n</i> , ionic <i>a</i> , mentioned <i>pp</i> , mobility <i>n</i> , noted <i>pp</i> , occurs <i>v</i> , purpose <i>n</i> , required <i>pp</i> , superconducting <i>p l</i> , techniques <i>n</i> , whether <i>cj</i>	35
696—712	attenuation <i>n</i> , assuming <i>p</i> , become <i>v</i> , changed <i>pp</i> , initial <i>a</i> , least <i>a</i> , origin <i>n</i> , observations <i>n</i> , provide <i>n</i> , pattern <i>n</i> , phonon <i>n</i> , principle <i>n</i> , sign <i>n</i> , silver <i>n</i> , types <i>n</i> , whose <i>pron</i> , whereas <i>cj</i>	34
713—741	almost <i>adv</i> , amount <i>n</i> , basic <i>a</i> , background <i>n</i> , constituent <i>a</i> , contains <i>v</i> , close <i>a</i> , donors <i>n</i> , decrease <i>n</i> , e. g. <i>ab</i> , ferromagnetic <i>a</i> , height <i>n</i> , immediately <i>adv</i> , identical <i>a</i> , indicatrix <i>n</i> , indicated <i>pp</i> , intrinsic <i>a</i> , like <i>adv</i> , lead <i>n</i> , microscope <i>n</i> , parts <i>n</i> , placed <i>pp</i> , plate <i>n</i> , produce <i>v</i> , quartz <i>n</i> , seems <i>v</i> , slope <i>n</i> , though <i>cj</i> , taking <i>p</i>	33
742—775	average <i>n</i> , apply <i>v</i> , approximate <i>a</i> , breakdown <i>n</i> , boundaries <i>n</i> , centre <i>n</i> , clear <i>a</i> , clusters <i>n</i> , cleavage <i>n</i> , chapter <i>n</i> , connected <i>pp</i> , define <i>v</i> , detector <i>n</i> , depth <i>n</i> , exchange <i>n</i> , exist <i>v</i> , external <i>a</i> , gradient <i>n</i> , hand <i>n</i> , instead <i>pp</i> , integral <i>n</i> , interesting <i>a</i> , law <i>n</i> , mode <i>n</i> , observation <i>n</i> , path <i>n</i> , resulting <i>a</i> , slightly <i>adv</i> , static <i>p</i> , standard <i>n</i> , strength <i>n</i> , view <i>n</i> , water <i>n</i> , yet <i>adv/cj</i>	32
776—799	composition <i>n</i> , conventional <i>a</i> , convention <i>a</i> , decreases <i>v</i> , differential <i>a</i> , description <i>n</i> , donor <i>n</i> , element <i>n</i> , electrodes <i>n</i> , far <i>adv</i> , form <i>v</i> , flat <i>a</i> , gas <i>n</i> , interpretation <i>n</i> , infrared <i>n</i> , include <i>v</i> , p. <i>ab</i> , permittivity <i>n</i> , question <i>n</i> , resolution <i>n</i> , readily <i>adv</i> , rapidly <i>adv</i> , significant <i>a</i> , spherical <i>a</i>	31
800—823	absolute <i>a</i> , caused <i>pp</i> , cannot <i>v</i> , critical <i>a</i> , defects <i>n</i> , detail <i>n</i> , efficiency <i>n</i> , easily <i>adv</i> , imperfection <i>n</i> , indica-	30

Таблица 1 (продолжение)

i	Словоформы	F
	tes <i>v</i> , investigation <i>n</i> , mole <i>n</i> , place <i>n</i> , processes <i>n</i> , problems <i>n</i> , quantum <i>n</i> , reciprocal <i>a</i> , shapes <i>n</i> , similarly <i>adv</i> , splitting <i>n</i> , study <i>n</i> , sufficient <i>a</i> , systems <i>n</i> , target <i>n</i>	
824—856	accuracy <i>n</i> , arbitrary <i>a</i> , analogous <i>a</i> , confirmed <i>pp</i> , diameter <i>n</i> , diamond <i>n</i> , domain <i>n</i> , extent <i>n</i> , ellipsoid <i>n</i> , edges <i>n</i> , filled <i>pp</i> , finally <i>adv</i> , faces <i>n</i> , his <i>pron</i> , leads <i>v</i> , local <i>a</i> , limited <i>pp</i> , measure <i>v</i> , modes <i>n</i> , next <i>a</i> , often <i>adv</i> , purity <i>n</i> , pressures <i>n</i> , referred <i>pp</i> , ref. <i>ab</i> , steps <i>n</i> , straight <i>a</i> , strong <i>a</i> , suppose <i>v</i> , together <i>adv</i> , transmission <i>n</i> , takes <i>v</i> , useful <i>a</i>	29
857—883	classes <i>n</i> , demagnetizing <i>p</i> , enough <i>adv/a</i> , final <i>a</i> , halides <i>n</i> , influence <i>n</i> , lie <i>v</i> , luminescence <i>n</i> , last <i>a</i> , modified <i>pp</i> , momentum <i>n</i> , nearly <i>adv</i> , need <i>v</i> , notation <i>n</i> , ordered <i>pp</i> , probably <i>adv</i> , reasonable <i>a</i> , replaced <i>pp</i> , secondary <i>a</i> , sets <i>n</i> , solids <i>n</i> , special <i>a</i> , spin-orbit <i>n</i> , stresses <i>n</i> , sufficiently <i>adv</i> , two-dimensional <i>a</i>	28
884—915	alkali <i>n</i> , apparent <i>a</i> , application <i>n</i> , binding <i>n</i> , characteristics <i>n</i> , computed <i>pp</i> , consists <i>v</i> , deposited <i>pp</i> , examined <i>pp</i> , found <i>v</i> , fairly <i>adv</i> , gold <i>n</i> , helium <i>n</i> , homogenous <i>a</i> , infinite <i>a</i> , imperfections <i>n</i> , latter <i>a</i> , lifetime <i>n</i> , medium <i>n</i> , mixed <i>pp</i> , moments <i>n</i> , narrow <i>a</i> , numerical <i>a</i> , oxide <i>n</i> , particularly <i>adv</i> , precision <i>n</i> , quadric <i>n</i> , represent <i>a</i> , rise <i>n</i> , separated <i>pp</i> , tube <i>n</i> , varying <i>p</i>	27
916—954	angles <i>n</i> , approach <i>n</i> , areas <i>n</i> , characteristic <i>a</i> , central <i>a</i> , complicated <i>pp</i> , considering <i>p</i> , definition <i>n</i> , dependent <i>a</i> , describe <i>v</i> , did <i>v/aux</i> , employed <i>pp</i> , equivalent <i>a</i> , explain <i>v</i> , existence <i>n</i> , finite <i>a</i> , for <i>cj</i> , giving <i>p</i> , illustrated <i>pp</i> , internal <i>a</i> , interstitial <i>a</i> , listed <i>p</i> , makes <i>v</i> , opposite <i>a</i> , oxygen <i>n</i> , pairs <i>n</i> , peaks <i>n</i> , performed <i>pp</i> , phonons <i>n</i> , piezoelectric <i>a</i> , quantity <i>n</i> , reflected <i>pp</i> , reflections <i>n</i> , representation <i>n</i> , screw <i>n</i> , unity <i>n</i> , vectors <i>n</i> , weak <i>a</i> , wavelength <i>n</i>	26
955—992	anomaly <i>n</i> , appearance <i>n</i> , before <i>pp</i> , calibrated <i>pp</i> , change <i>v</i> , class <i>n</i> , common <i>a</i> , configuration <i>n</i> , considerably <i>adv</i> , contacts <i>n</i> , crystallographic <i>a</i> , equal <i>v</i> , explained <i>pp</i> , geometry <i>n</i> , imaginary <i>a</i> , increased <i>pp</i> , investigations <i>n</i> , kbar <i>n</i> , lowest <i>a</i> , making <i>p</i> , migration <i>n</i> , original <i>a</i> , patterns <i>n</i> , predictions <i>n</i> , radius <i>n</i> , recent <i>a</i> , reduces <i>v</i> , remains <i>v</i> , rotation <i>n</i> , scattered <i>pp</i> , spin-wave <i>n</i> , treated <i>pp</i> , thermocouple <i>n</i> , tin <i>n</i> , valid <i>a</i> , vapor <i>n</i> , write <i>v</i>	25
993—1038	absence <i>n</i> , applying <i>p</i> , authors <i>n</i> , bond <i>n</i> , capture <i>n</i> , changes <i>v</i> , considerable <i>a</i> , correlation <i>n</i> , covalent <i>a</i> , crystalline <i>n</i> , densities <i>n</i> , discuss <i>v</i> , disordered <i>pp</i> , doped <i>pp</i> , evaporation <i>n</i> , earlier <i>a</i> , generated <i>pp</i> , ground <i>n</i> , had <i>v</i> , had <i>v/aux</i> , is <i>v/mod</i> , know <i>v</i> , lead <i>v</i> , mathematical <i>a</i> , needed <i>pp</i> , nucleus <i>n</i> , only <i>a</i> , phenomena <i>n</i> , polarizability <i>n</i> , probability <i>n</i> , random <i>n</i> , removed <i>pp</i> , represented <i>pp</i> , rod <i>n</i> , scheme <i>n</i> , second-order <i>n</i> , self-diffusion <i>n</i> , sodium <i>n</i> , strains <i>n</i> , subsequent <i>a</i> , suggested <i>pp</i> , transmittance <i>n</i> , versus <i>pp</i> , whole <i>a</i> , wide <i>a</i> , «y»-band <i>n</i>	24
1039—1070	A <i>ab</i> , actually <i>adv</i> , abate <i>n</i> , approximations <i>n</i> , arsenide <i>n</i> , cavity <i>n</i> , chamber <i>n</i> , consideration <i>n</i> , defect <i>n</i> , dipole	23

Таблица 1 (продолжение)

i	Словоформы	F
1071—1105	<i>n</i> , estimated <i>pp</i> , flux <i>n</i> , furnace <i>n</i> , great <i>a</i> , he <i>pron</i> , highly <i>adv</i> , included <i>pp</i> , indeed <i>adv</i> , linearly <i>adv</i> , main <i>a</i> , obtains <i>v</i> , photon <i>n</i> , picture <i>n</i> , reference <i>s</i> , seem <i>v</i> , signal <i>n</i> , site <i>n</i> , spacing <i>n</i> , strongly <i>adv</i> , sulphide <i>n</i> , vary <i>v</i> , «у»-th <i>num</i>	22
1106—1148	cation <i>n</i> , chemical <i>a</i> , contained <i>v</i> , comparable <i>a</i> , correspond <i>v</i> , electrode <i>n</i> , essential <i>a</i> , fixed <i>pp</i> , followed <i>pp</i> , individual <i>a</i> , interpreted <i>pp</i> , importance <i>n</i> , lecturers <i>n</i> , liquid <i>a</i> , mechanisms <i>n</i> , molecular <i>a</i> , moreover <i>adv</i> , numbers <i>n</i> , paramagnetic <i>a</i> , pointed <i>pp</i> , propagation <i>n</i> , reflectivity <i>n</i> , responsible <i>a</i> , say <i>v</i> , simplified <i>pp</i> , stable <i>a</i> , steady <i>a</i> , symmetrical <i>a</i> , tip <i>n</i> , top <i>n</i> , uniformity <i>n</i> , upper <i>a</i> , variables <i>n</i> , was <i>v</i> , wire <i>n</i> annealing <i>n</i> , assumptions <i>n</i> , at. <i>ab</i> , beyond <i>prp/adv</i> , bottom <i>n</i> , contain <i>v</i> , contributions <i>n</i> , cut <i>pp</i> , coupling <i>n</i> , D. C. <i>ab</i> , diffusions <i>n</i> , earlier <i>adv</i> , face <i>n</i> , first-order <i>n</i> , fraction <i>n</i> , fringes <i>n</i> , gallium <i>n</i> , get <i>v</i> , group-«у» <i>n</i> , has <i>v/mod</i> , heated <i>p/p</i> , increase <i>v</i> , liquid <i>n</i> , loss <i>n</i> , microcopy <i>n</i> , off <i>adv/prep</i> , ones <i>pron</i> , penetration <i>n</i> , phenomenon <i>n</i> , proved <i>pp</i> , rates <i>n</i> , reason <i>n</i> , refer <i>v</i> , schocks <i>n</i> , sides <i>n</i> , substituting <i>pI</i> , throughout <i>adv</i> , ultraviolet <i>n</i> , uniaxial <i>a</i> , unique <i>a</i> , units <i>n</i> , what <i>pron</i> , «у»-axis <i>n</i>	21

Таблица 2

Распределение количества словоформ при $F \leq 20$

i	m	F	i	m	F
1149—1194	46	20	1846—1973	128	10
1195—1227	33	19	1974—2126	153	9
1228—1276	49	18	2127—2307	181	8
1277—1335	59	17	2308—2528	221	7
1336—1396	61	16	2529—2824	296	6
1397—1474	78	15	2825—3158	334	5
1475—1550	76	14	3159—3619	461	4
1551—1630	80	13	3620—4266	647	3
1631—1719	89	12	4267—5397	1131	2
1720—1845	126	11	5398—8214	2817	1

100 тыс. словоупотреблений, у нас — 200 тыс.), тематикой (словарь Мелик-Гусейновой составлен на основании одного раздела физики твердого тела из трех возможных) и процентным распределением количества минимальных выборок внутри раздела.

Выше (табл. 1) приводится частотный список словоформ при частоте $F > 20$. Распределение количества словоформ с частотой $F \leq 20$ представлено в табл. 2.

Е. М. Тарасова

ЧАСТОТНЫЙ СЛОВАРЬ АНГЛИЙСКОГО ПОДЪЯЗЫКА¹ ФИЗИКИ ЭЛЕМЕНТАРНЫХ ЧАСТИЦ

Частотный словарь (табл. 2) составлен вручную по выборке 200 тыс. словоупотреблений из периодических изданий США последних десяти лет.¹ Сплошному расписыванию подверглись 200 текстов длиной 1000 словоупотреблений каждый.

Общее число обнаруженных в выборке разных словоформ, включая лексико-грамматические омографы, равно 9051; из этого количества ниже (табл. 2) приводятся 1132 словоформы с частотой не менее 20. Распределение количества словоформ при $F < 20$ приведено в табл. 3.

Таблица 1

Схема выборки		Количество	
№№	Раздел	текстов	словоупотреблений
		Экспериментальные исследования	
1	Описание ускорителей, мишеней, пучковой камеры и т. д.	40	40 000
Теория			
2	Аксиоматические методы	40	40 000
3	Сильные взаимодействия	40	40 000
4	Слабые взаимодействия	40	40 000
5	Разное	40	40 000
Итого		200	200 000

¹ При определении подъязыка физики элементарных частиц и подборе текстов большую помощь оказал канд. физико-математических наук В. С. Рудаков.

Таблица 2

Частотный список словоформ

t	Словоформы	F
1	the	20517
2	of	9144
3	in	5505
4	and	5315
5	a	3817
6	we	3751
7	for <i>prp</i>	2971
8	is <i>l</i>	2854
9	to <i>prp</i>	2599
10	that <i>cj</i>	2457
11	to <i>part</i>	1991
12	by <i>prp</i>	1944
13	be <i>inf</i>	1826
14	this	1779
15	with	1748
16	as <i>cj</i>	1540
17	is <i>aux</i>	1381
18	which <i>cj</i>	1288
19	from	1255
20	at	1114
21	it	1032
22	are <i>l</i>	962
23	an	931
24	can	898
25	where <i>cj</i>	815
26	are <i>aux</i>	814
27	on	798
28	not	751
29	one <i>num</i>	692
30	these	651
31	eq.	606
32	only	589
33	energy	587
34	if <i>cj</i>	571
35	then	522
36	was <i>aux</i>	515
37	two <i>num</i>	513
38	all	462
39	terms	447
40	or	433
41	states <i>n</i>	431
42	state <i>n</i>	426
43	from	424
44—45	were <i>aux</i> , given	406
46	have <i>aux</i>	398
47	amplitude	397
48—49	between, case	392
50	cross <i>n</i>	372
51	mass	366
52	may	362

Таблица 2 (продолжение)

t	Словоформы	F
53	so <i>adv</i>	354
54	our	352
55	there <i>adv</i>	350
56	function	349
57	since <i>cj</i>	347
58	particles	346
59	fig.	344
60	same	336
61	value <i>n</i>	334
62	matrix	329
63	momentum	328
64—65	also, section	327
66	but <i>prp</i>	321
67—68	data, will <i>aux</i>	314
69	have <i>inf</i>	309
70	would <i>aux</i>	306
71	than	303
72—73	scattering <i>pI</i> , when	301
74	thus	297
75—76	no, obtained <i>pII</i>	296
77	values <i>n</i>	290
78	results <i>n</i>	286
79—81	first, has <i>aux</i> , other <i>a</i>	281
82—83	order <i>n</i> , such	278
84	been <i>pII</i>	277
85	used <i>pII</i>	276
86	amplitudes	273
87	about	271
88	field	266
89	any	265
90	factor	263
91—93	however, into, scattering <i>n</i>	261
94	now	260
95	equation	259
96	has <i>pres indef</i>	249
97	each	254
98	shown	250
99	experimental	245
100	particle	244
101	beam <i>n</i>	243
102—103	spin, above	240
104—105	both, system	239
106	events	238
107	model	234
108—109	decay, <i>n</i> , was <i>past indef</i>	229
110	because	227
111	constant <i>n</i>	223
112—113	region <i>s</i> , using <i>pI</i>	222
114	theory	219
115	interaction	209
116	large	208
117—120	over, small, term, wave <i>n</i>	207
121	approximation	206

Таблица 2 (продолжение)

i	Словоформы	F
122—123	angular, therefore	204
124—126	that, following <i>a</i> , some	203
127	more	202
128	function	200
129	relation	199
130—131	here, total	196
132—133	possible, shall <i>aux</i>	187
134	part <i>n</i>	186
135—138	number <i>n</i> , polarization, symmetry, target <i>n</i>	185
139	very	182
140	potential <i>a</i>	181
141	equations	175
142—143	must, solution	174
144	pole <i>n</i>	173
145—146	range <i>n</i> , reaction	172
147	effect <i>n</i>	169
148—150	angle, point <i>n</i> , they	168
151	proton	167
152—153	limit <i>n</i> , operator	165
154—155	physical, three	164
156—159	different, Eqs., factors, out	163
160	their	162
161	observed <i>pII</i>	159
162—163	exchange <i>n</i> , zero	157
164—168	corresponding <i>pI</i> , gives, its, made <i>pII</i> , under	156
169—170	meson, should <i>aux</i>	155
171	group <i>n</i>	154
172—174	defined <i>pII</i> , is (there), photon	153
175—176	due, energies	152
177—178	real, general <i>a</i>	151
179—180	phase, respectively	150
181—182	see <i>pres indef</i> , seem <i>pres indef</i>	149
183—184	have <i>inf</i> , vector	148
185	integral	147
186—188	channel, charge <i>n</i> , second	146
189—191	determined <i>pII</i> , production, resonance	145
192—194	contribution, distribution, those	144
195	time <i>n</i>	142
196	table <i>n</i>	141
197—200	integration, result <i>n</i> , sections, well <i>adv</i>	140
201—202	parameters, taken	139
203—204	let <i>inf</i> , relation	137
205	calculation	136
206—207	similar, space <i>n</i>	135
208	nucleon	134
209—210	protons, representation	133
211—212	coupling <i>pI</i> , known	132
213—214	present <i>a</i> , final	130
215	spectrum	129
216—218	current, us, does	128
219—221	find <i>pres indef</i> , found <i>pII</i> , were <i>past indef</i>	125
222	way <i>n</i>	124
223—225	coupling <i>n</i> , independent, through	123

Таблица 2 (продолжение)

i	Словоформы	F
226—227	fact, process <i>n</i>	122
228—231	bound <i>pII</i> , channels, consider <i>inf</i> , problem	121
232	magnetic	120
233—236	constants, photons, ratio, while	119
237	elements	118
238	incident	117
239	simple	115
240—244	baryon, calculations, operators, set <i>n</i> , threshold	114
245—249	dependence, effects <i>n</i> , electron, obtain <i>inf</i> , source	113
250—251	are (there), elastic	112
252—256	assumption, described <i>pII</i> , higher, isospin, method	111
257—258	experiment, shows <i>pres indef</i>	110
259—261	contributions, high, particular	109
262—265	angles, assume <i>inf</i> , counter <i>n</i> , unitary	108
266	being <i>pI</i>	106
267—268	points <i>n</i> , rate <i>n</i>	105
269—270	element, pion	104
271—275	calculated <i>pII</i> , example, near, strong, could	102
276—278	expression, most, within	101
279—280	positive, various	100
281—285	condition, follows, i. e., mesons, necessary	99
286	even <i>a</i>	98
287—290	agreement, important, masses, neutrons	97
291	potentials	96
292—295	distributions, measured <i>pII</i> , relative <i>a</i> , vertex	95
296—297	becomes, counters <i>n</i>	94
298—305	assumed <i>pII</i> , differential, equal, finite, Hamiltonian,	93
	hence, poles <i>n</i> , transition	
306—307	obtain <i>inf</i> , type	92
308—309	behavior, low <i>a</i>	91
310—312	dispersion, spectra, unitary	90
313—317	direction, discussed <i>pII</i> , lower, plane, processes <i>n</i>	89
318—321	after, further, interactions, times <i>n</i>	88
322	below	87
323—327	chamber, consider <i>inf</i> , discussion, good, several	86
328	octet	85
329—330	scattered <i>pII</i> , written	84
331—335	approximately, conditions, considered <i>pII</i> , neutron, pro-	83
	duct <i>n</i>	
336—338	cannot, direction, peak <i>n</i>	82
339—340	procedure, seen	81
341—343	Ref., related <i>pII</i> , singularities	80
344—348	analysis, fit <i>n</i> , initial, measurements, moments	79
349—352	cases, cut <i>n</i> , do <i>aux</i> , inelastic	78
353—359	fields, level <i>n</i> , matrices, structure, up, using <i>ger</i> , usual	77
360—365	axis, density, line, rather, single, width	76
366—368	absorption, just <i>adv</i> , solutions	75
369—373	corrections, emission, error <i>n</i> , sign <i>n</i> , use <i>n</i>	74
374—376	addition, correction, shift <i>n</i>	73
377—379	course <i>n</i> , representations, see <i>inf</i>	72
380—381	formula, much	71
382—387	compared <i>pII</i> , decays <i>n</i> , difference, linear, note <i>inf</i> , para-	70
	meter	

Таблица 2 (продолжение)

<i>i</i>	Словоформы	<i>F</i>
388—393	curve <i>n</i> , diagram, four, scalar, vacuum, work <i>n</i>	69
394—400	direct <i>a</i> , for <i>c_j</i> , imaginary, invariant, position, rates <i>n</i> , use <i>inf</i>	68
401—407	comparison, currents, either, errors <i>n</i> , expected <i>pII</i> , negative, variables	67
408—415	although, associated <i>pII</i> , electromagnetic, invariance, leads <i>pres indef</i> , less, sec., side	66
416—424	basis, excitation, intermediate internal, magnitude, paper <i>n</i> , take <i>pres indef</i> , waves, without	65
425—427	expressions, parity, <i>s</i> -wave	64
428—431	ions, light <i>n</i> , reduced <i>pII</i> , resonances	63
432	quantum	62
433—439	account, appendix, baryons, expressed <i>pII</i> , groups <i>n</i> , nuclear, numbers <i>n</i>	61
440—447	additional, consistent, deuteron, properties, show <i>pres indef</i> , transfer, two-body, whether	60
448—452	chosen, figure, fixed <i>pII</i> , namely, proportional	59
453—458	along, efficiency, implies <i>pres indef</i> , like <i>adj</i> , note <i>pres indef</i> , use <i>inf</i>	58
459—467	complex, corresponds, experiments <i>n</i> , included <i>pII</i> , length, nucleons, predictions, symmetric, them	57
468—474	complete, components, detector, effective, hand <i>n</i> , partial-wave, reactions, approach <i>n</i>	56
475—483	approach <i>n</i> appropriate, background, event, get <i>pres indef</i> , produced <i>pII</i> , quite, smaller, write <i>inf</i>	55
484—489	depends, give <i>inf</i> , quantities, reasonable, required <i>pII</i> , together	54
490—502	again, algebra, asymmetry, choice, experimentally, latter, give <i>pres indef</i> , make <i>inf</i> , many, normalization, numerical, scattering <i>ger</i> , solid	53
503—511	before, component, exactly, forward, infinite, information, next, taking <i>pI</i> , what	52
512—514	definition, levels, possibility	51
515—528	another, applied <i>pII</i> , axial-vector, contains, expansion, long, maximum, measurement, multiplet, non-relativistic, right-hand, rule <i>n</i> , series, tracks <i>n</i>	50
529—540	approximate, available, derived <i>pII</i> , except <i>pp</i> , explicitly, free <i>a</i> , minimum, perturbation, probability, third, true, whose	49
541—557	assuming <i>pI</i> , curves <i>n</i> , electric, electrons, excited <i>pII</i> , external, finally, having <i>pI</i> , hypothesis, lines <i>n</i> , nuclei, peaks <i>n</i> , predicted <i>pII</i> , situation, statistical, tensor, theorem	48
558—562	directly, had <i>aux</i> , momenta, shell, transformation	47
563—569	arbitrary, furthermore, how, larger, new, radiation, weak	46
570—578	assume <i>inf</i> , change <i>n</i> , helicity, identical, relativistic, square, still, temperature, track <i>n</i>	45
579—591	always, coefficients, coupled, cuts, diagrams, figs., force, means <i>pres indef</i> , partial, property, respect, seems, singlet	44
592—607	appears, asymptotic, average, based <i>pII</i> charged <i>pII</i> , clear, exact, far, frame <i>n</i> , ground <i>n</i> , high-energy, indices, make <i>inf</i> , pseudoscalar, satisfied <i>pII</i> , singularity	43

Таблица 2 (продолжение)

<i>i</i>	Словоформы	<i>F</i>
608—622	completely, considerations, coulomb, define <i>pres indef</i> , determine <i>inf</i> , hydrogen, Lorentz, pion-nucleon, principle, quark, rays, strongly, takes, thickness, <i>p</i> -wave	42
623—636	calculate <i>inf</i> , conservation, crossing <i>pI</i> cutoff <i>n</i> , determination, equivalent, exists, intensity, means <i>n</i> , might <i>mod</i> , neutral, take <i>inf</i> , though, types <i>n</i>	41
637—651	according <i>a</i> , anomalous, close <i>a</i> , commutation, conclusion, correct <i>adj</i> , essentially, gas <i>n</i> , had <i>pres indef</i> , is (that), low-energy, pair <i>n</i> , ratios, rest <i>n</i> , show <i>inf</i>	40
652—659	e. g., explicit, integrals, last <i>adj</i> , negligible, parts <i>n</i> , shape <i>n</i> , trajectory	39
660—681	analytic, center-of-mass, certain, did <i>past ind</i> end <i>n</i> , expect <i>inf</i> , final-state, instead, interference, introduction, laboratory, Lagrangian, mechanism, notation, off <i>a</i> , pulse <i>n</i> , question <i>n</i> , rules <i>n</i> , sources, strength, unity, upon	38
682—692	allowed <i>pII</i> , bootstrap, easily, film <i>n</i> , involved <i>pII</i> , noted <i>pII</i> , relatively, resolution, sensitive, write <i>pres indef</i> , yields <i>pres indef</i>	37
693—707	differences, difficult, down, following <i>n</i> , interest <i>n</i> , left-hand, mentioned <i>pII</i> , neglected <i>pII</i> , positions, recoil <i>n</i> , relevant, similarly, suggested <i>pII</i> , theoretical, variable	36
708—718	basic, convenient, forms <i>n</i> , least, motion <i>n</i> , presence, replaced <i>pII</i> , runs <i>n</i> , spins, systems, too	35
719—735	choose <i>pres indef</i> , coincidence, correlation, correspond <i>pres indef</i> , definite, do <i>inf</i> , enough, gamma, generally, manner, models, normal, secondary, static, transitions, usually, virtual	34
736—750	antisymmetric, center <i>n</i> , define <i>inf</i> , gamma-ray, indicates, interesting <i>pI</i> , little, local, nearly, plotted <i>pII</i> , resulting <i>pI</i> , solve <i>inf</i> , useful, view, whereas	33
751—773	argument, arrangement, better, detectors, evidence, few, find <i>inf</i> , fits <i>n</i> , five, including <i>pI</i> , index, leptonic, nucleus, presented <i>pII</i> , previous, programme, projection, pure, scintillator, sense, typical, unit, upper	32
774—794	among, configuration, couplings detailed <i>pII</i> , earlier, estimate <i>n</i> , etc., generators, gets, lowest, quantity, represent <i>pres ind</i> , represents, simply, special, stable, starting <i>pI</i> , sufficient, uncertainty, unsubtracted <i>pII</i> , vectors	31
795—816	already, appear <i>pres indef</i> , assumption, denote <i>pres indef</i> , detection, discontinuity, done, during, entire, greater, increases <i>pres indef</i> , involving <i>pI</i> , normalized <i>pII</i> , obtains, parallel, sample, set <i>pII</i> , significant, size <i>n</i> , studied <i>pII</i> , valid, zeroes	29
817—837	almost, boundary, contamination, contribute <i>inf</i> , difficulty, dynamical, existence, features, introduced <i>pII</i> , isotopic, itself, massive, massless, once, outgoing <i>pI</i> , propagator, purpose, recently, requires, thick, triple	28
838—861	accuracy, approximations, become <i>pres indef</i> , best, denotes <i>pres indef</i> , detail, equivalent, expectation, immediately, kinematic, kinematics, meson-baryon, particularly, performed <i>pII</i> , recent, say <i>inf</i> , solenoid, spectrometer,	28

Таблица 2 (продолжение)

i	Словоформы	F
862—888	standard, taking <i>pI</i> , transform <i>pres indef</i> , treatment, variation, wish <i>pres indef</i> absolute, attenuation, broken, clearly, combination, complicated <i>pII</i> , computer, counting <i>pI</i> , depending <i>pI</i> , diagonal, eigenvalues, figures, found <i>past indef</i> , full, goes, indicated <i>pII</i> , loss, multiplets, nucleon-nucleon, obvious, occurs, requirement, scheme, separable, specific, suitable, thermal	27
889—920	able, analogous, application, coincidences, computed <i>pII</i> , consequently, consideration, employed <i>pII</i> , especially, estimated <i>pII</i> , every, idea, material, nonet, obviously, opposite, pairs, path, place <i>n</i> , placed <i>pII</i> , previously, problems, produce <i>inf</i> , proper, radius, represented <i>pII</i> , residue, rotation, seem <i>inf</i> , somewhat, subtraction, systematic	26
921—952	attention, bound-state, changes <i>n</i> , compare <i>inf</i> , considering <i>pI</i> , consists, correspondence, critical, deuterium, diffusion, dominance, flux, include <i>inf</i> , indeed, input <i>n</i> , instance, mean <i>adj</i> , need <i>pres indef</i> , orbital, outside, per, pions, reason <i>n</i> , singlets, solved <i>pII</i> , telescope, transforms <i>pres indef</i> , transmission, triplet, volume <i>n</i> , well-known, Yukawa	25
953—983	accurate, appear <i>inf</i> , coefficient, coherent, connected <i>pII</i> , connection, denominator, difficulties, diffraction, dominated <i>pII</i> , electronic, essential, fitted <i>pII</i> , hard, induced <i>pII</i> , interpretation, involves, lab., limits <i>n</i> , perhaps, probably, products <i>n</i> , radiative, six, straightforward, three-particle, throughout, transformations, unique, validity, via	24
984—1016	across, amount, arising <i>pI</i> , characteristic, coincident, common, commutator, construct <i>inf</i> , double <i>a</i> , eigenvalue, expect <i>pres indef</i> , feature, formulas, include <i>pres indef</i> , inside, irreducible, isovector, lead <i>pres indef</i> , lifetime, location, moment, multiple, origin, pulse-height, pulses, remaining, renormalization, separately, sets <i>n</i> , shifts <i>n</i> , slightly, <i>w</i> -spin, velocity	23
1017—1043	added <i>pII</i> , area, branch, carbon, containing <i>pI</i> , details, distance, evident, exist <i>inf</i> , forces, get <i>inf</i> , half <i>n</i> , importance, independently, left <i>pII</i> , methods, occur <i>inf</i> , particle-exchange, period, permutation, really, roughly, sides, suggests, suppose, variations, yet	22
1044—1090	absorptive, actually, additionally, advantage, allows, arises, calibration, carried <i>pII</i> , cascade, choose <i>inf</i> , consequence, considerably, creation, derivation, enhancement, equally, exchanged <i>pII</i> , explained <i>pII</i> , extension, fit <i>inf</i> , generator, height, his, homogenous, hyperon, interval, magnet, main, modified <i>pII</i> , muons, necessarily, ones, ordinary, original, overall, photoproduction, plus, predicts, proceeds, purely, rise <i>n</i> , separation, sufficiently, thin, units, vertices, vibrational	21
1091—1132	alone, annihilation, apart, arguments, assignment, authors, auxiliary, found <i>n</i> , branching <i>pI</i> , cancellation, corresponding <i>a</i> , dense, description, deviation, discuss <i>inf</i> ,	20

Таблица 2 (продолжение)

i	Словоформы	F
	easy, entirely, follow <i>pres indef</i> , gap, generalized <i>pII</i> , integrated <i>pII</i> , introduce <i>inf</i> , kind, let <i>inf</i> , minimal, phase-shift, pion-pion, polarized <i>pII</i> , prediction, provided <i>ej</i> , regions, signs <i>n</i> , single-particle, spark <i>n</i> , specified <i>pII</i> , techniques, treated <i>pII</i> , vanish <i>inf</i> , variational, vector-meson, wide, width	

Таблица 3

Распределение количества словоформ при $F < 20$

i	m	F	i	m	F
1132—1187	55	19	1882—2009	128	9
1188—1228	41	18	2012—2190	181	8
1229—1291	63	17	2191—2457	287	7
1292—1353	62	16	2458—2817	360	6
1354—1434	81	15	2818—3268	451	5
1435—1495	61	14	3269—3801	533	4
1496—1580	85	13	3802—4518	717	3
1581—1675	95	12	4519—6086	1568	2
1676—1782	107	11	6087—9051	2965	1
1783—1881	99	10			

Л. В. Малаховский

О ЧАСТОТНОМ РАСПРЕДЕЛЕНИИ АНГЛИЙСКИХ КОЛИЧЕСТВЕННЫХ ЧИСЛИТЕЛЬНЫХ ПЕРВОЙ СОТНИ

Интуитивно представляется, что частота употребления числительного как-то зависит от того места, которое занимает обозначаемое им число в натуральном ряду. В частности, числительные первого десятка употребляются в речи, вероятно, чаще, чем те, которые соответствуют большим числам, а числительные, обозначающие так называемые «круглые числа» (т. е. числа, оканчивающиеся на 0 или 5), — чаще, чем числительные «некруглые».¹

Специальных исследований в этой области, по-видимому, не было,² но имеется ряд наблюдений, косвенно подтверждающих это интуитивное представление. Так, например, материалы переписей населения свидетельствуют о склонности людей округлять свой возраст до ближайшего круглого числа, что проявляется на кривых возрастного состава населения в чрезвычайно сильных завышениях частот тех возрастов, которые соответствуют круглым числам (рис. 1).³ Замечено, что склонность эта тем сильнее, чем ниже образовательный уровень обследуемого контингента,⁴ что подтверждается сравнением кривых, полученных в переписях старых лет, с более поздними (например, 1897 г. в России и 1926 г. в СССР), а также данных обследования более грамотных и менее

грамотных слоев населения (например, мужчин и женщин). Подобные же завышения частот на «круглых» возрастах наблюдаются и при регистрации возраста умерших (табл. 1),⁵ причем и здесь уровень образования играет ту же роль (ср. графы «1935» и «1900», а также «Все население» и «Не-белое население»).

Можно предположить, что склонность к предпочтительному использованию в речи круглых числительных сказывается не только при ответе на вопрос о возрасте, но и в других сферах коммуникации, как устной, так и письменной. Гипотеза о большей частоте употребления круглых числительных по сравнению с некруглыми проверялась в настоящей работе на материале англий-

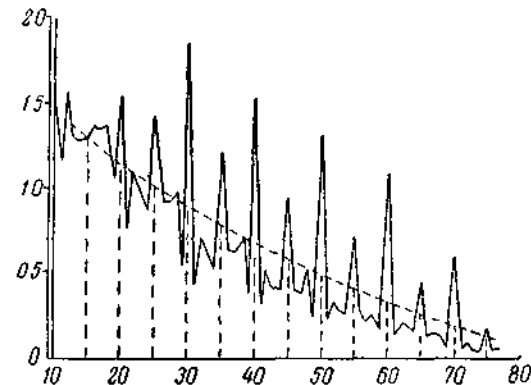


Рис. 1. Распределение женского населения России по возрастам (согласно данным переписи 1897 г.).

По оси абсцисс — возраст в годах, по оси ординат — число жителей в млн.

ского языка по данным двух частотных словарей — словаря Кучеры и Фрэнсиса (СК), построенного на американских текстах смешанного характера (публицистических, художественных, научно-технических и др.) объемом более 1 млн словоупотреблений [6], и словаря Ринсланда (СР), основанного на текстах письменных работ американских школьников общим объемом свыше 6 млн словоупотреблений [9]. Использование двух разных словарей имело целью определить роль фактора образованности, поскольку известно, что в выборку СК отбирались тексты, написанные образованными людьми — писателями, журналистами, учеными и т. п., тогда как в выборку СР вошли письменные работы школьников с 1-го по 8-й класс, не подвергавшиеся редакционной обработке.

⁵ Таблица заимствована из книги Линдера и Гроува [7, стр. 10] и дополнена расчетами, касающимися распределения числа умерших по «круглым» и «некруглым», четным и нечетным возрастам (нижняя часть таблицы).

¹ Ср., например, высказывание А. Е. Супруна (не основанное на статистических расчетах) о том, что круглые числительные «гораздо более вероятны, чем числительные типа пятьдесят девять или шестьдесят один» [4].

² Отдельные наблюдения, касающиеся частотного распределения числительных (на материале английского языка), содержатся в статье Р. Г. Пиотровского и Л. А. Турыгиной [2], однако характер зависимости частоты числительного от места соответствующего числа в натуральном ряду в ней не исследуется.

³ Рисунок воспроизводится по [5, стр. 139].

⁴ См. там же, стр. 134, а также [1].

Таблица 1

Частотное распределение числа умерших в США по последней цифре возраста (в процентах к общему числу умерших в каждой категории) *

Последняя цифра возраста	Количество умерших			
	в 1935 г.		в 1900 г.	
	все население	не-белое население	все население	не-белое население
0	11.10	15.03	11.81	15.62
1	8.84	7.65	8.58	7.46
2	9.97	9.82	10.04	9.58
3	9.65	8.52	9.54	9.13
4	10.03	9.15	9.71	8.81
5	11.14	13.54	10.77	13.10
6	9.53	8.64	9.48	8.48
7	9.75	8.68	9.90	8.87
8	10.16	9.91	10.52	9.80
9	9.83	9.06	9.65	9.15
Всего	100.00	100.00	100.00	100.00
В том числе:				
В круглых возрастах	22.24 (11.12)	28.57 (14.28)	22.58 (11.29)	28.72 (14.36)
В некруглых возрастах	77.76 (9.72)	71.43 (8.93)	77.42 (9.68)	71.28 (8.91)
Из них:				
В четных	39.69 (9.92)	37.52 (9.38)	39.75 (9.94)	36.67 (9.17)
В нечетных	38.07 (9.52)	33.91 (8.48)	37.67 (9.42)	34.61 (8.65)

Распределение английских количественных числительных первой сотни по частоте встречаемости в текстах СК и СР представлено в табл. 2. В отличие от обычных частотных списков числительные расположены здесь не в порядке убывания частоты, а в порядке натурального ряда обозначаемых ими чисел. Это делает возможным выявление тенденций к завышению (или занижению) частот на определенных участках натурального ряда.

Для удобства сравнения данных по двум разным словарям (СР и СК) в табл. 2 приводятся, помимо абсолютных (f_1 и f_2), также и относительные частоты (f'_1 и f'_2) тех же числительных и относительная разность этих частот $D_{f'}$, определяемая по формуле

$$D_{f'} = \frac{f'_1 - f'_2}{f'_2}$$

Из табл. 2 следует прежде всего, что общая частота встречаемости числительных в письменных работах школьников намного выше, чем в текстах СК (9987.7 против 6375.7 миллионов, т. е. на 57 % больше). Легко заметить, что эта повышенная встречаемость не распределена равномерно по всему рассматриваемому

* Таблица содержит сведения об умерших в возрасте от 7 до 96 лет. В нижней части таблицы для удобства сопоставления данных приведены суммарные (а в скобках — средние) частоты для «круглых» и «некруглых» возрастов, а также для некруглых четных и некруглых нечетных.

Таблица 2

Частота встречаемости английских количественных числительных в текстах СР и в текстах СК

n	i	f_1	f_2	f'_1	f'_2	$D_{f'}$
one	1	19214	2197	3195.8	2166.2	+0.5
two	2	11457	1412	1905.6	1392.2	+0.4
three	3	6235	610	1037.0	601.4	+0.7
four	4	3563	359	592.6	354.0	+0.7
five	5	3117	286	518.4	282.0	+0.8
six	6	2378	220	395.5	216.9	+0.8
seven	7	1414	113	235.2	111.4	+1.1
eight	8	1434	104	238.5	102.5	+1.3
nine	9	1132	81	188.3	79.9	+1.4
ten	10	1937	165	322.2	162.7	+1.0
eleven	11	508	40	84.5	39.4	+1.1
twelve	12	817	48	135.9	47.3	+1.9
thirteen	13	279	11	46.4	10.8	+3.3
fourteen	14	289	31	48.1	30.6	+0.6
fifteen	15	448	56	74.5	55.2	+0.3
sixteen	16	210	20	34.9	19.7	+0.8
seventeen	17	151	24	25.1	23.7	+0.1
eighteen	18	213	17	35.4	16.8	+1.1
nineteen	19	183	18	30.4	17.7	+0.7
twenty	20	757	80	125.9	78.9	+0.6
twenty-one	21	85	8	14.1	7.9	+0.8
twenty-two	22	47	8	7.8	7.9	0.0
twenty-three	23	23	7	3.8	6.9	-0.4
twenty-four	24	52	14	8.6	13.8	-0.4
twenty-five	25	160	25	26.6	24.6	+0.1
twenty-six	26	20	5	3.3	4.9	-0.3
twenty-seven	27	14	1	2.3	1.0	+1.3
twenty-eight	28	42	5	7.0	4.9	+0.4
twenty-nine	29	12	2	2.0	2.0	0.0
thirty	30	311	59	101.6	58.2	+0.7
thirty-one	31	12	3	2.0	3.0	-0.3
thirty-two	32	16	3	2.7	3.0	-0.1
thirty-three	33	1	2	0.2	2.0	-0.9
thirty-four	34	10	7	1.7	6.9	-0.8
thirty-five	35	30	14	5.0	13.8	-0.6
thirty-six	36	18	4	3.0	3.9	-0.2
thirty-seven	37	15	1	2.5	1.0	+1.5
thirty-eight	38	10	1	1.7	1.0	+0.7
thirty-nine	39	1	2	0.2	2.0	-0.9
forty	40	311	36	51.7	35.5	+0.5
forty-one	41	9	0	1.5	0.0	(+)
forty-two	42	6	2	1.0	2.0	-0.5

Таблица 2 (продолжение)

<i>n</i>	<i>i</i>	<i>f</i> ₁	<i>f</i> ₂	<i>f'</i> ₁	<i>f'</i> ₂	<i>D_f</i>
forty-three	43	8	2	1.3	2.0	-0.4
forty-four	44	7	6	1.2	5.9	-0.8
forty-five	45	43	7	7.2	6.9	+0.04
forty-six	46	21	2	3.5	2.0	+0.8
forty-seven	47	7	4	1.2	3.9	-0.7
forty-eight	48	26	1	4.3	1.0	+3.3
forty-nine	49	1	4	0.2	3.9	-0.9
fifty	50	503	68	83.7	67.0	+0.2
fifty-one	51	1	1	0.2	1.0	-0.8
fifty-two	52	11	2	1.8	2.0	-0.1
fifty-three	53	6	2	1.0	2.0	-0.5
fifty-four	54	7	1	1.2	1.0	+0.2
fifty-five	55	24	2	4.0	2.0	+1.0
fifty-six	56	49	0	8.2	0.0	(+)
fifty-seven	57	1	1	0.2	1.0	-0.8
fifty-eight	58	9	0	1.5	0.0	(+)
fifty-nine	59	7	1	1.2	1.0	+0.2
sixty	60	163	21	27.1	20.7	+0.3
sixty-one	61	1	8	0.2	7.9	-1.0
sixty-two	62	1	2	0.2	2.0	-0.9
sixty-three	63	1	0	0.2	0.0	(+)
sixty-four	64	10	0	1.7	0.0	(+)
sixty-five	65	18	9	3.0	8.9	-0.7
sixty-six	66	5	0	0.8	0.0	(+)
sixty-seven	67	1	1	0.2	1.0	-0.8
sixty-eight	68	1	1	0.2	1.0	-0.8
sixty-nine	69	1	1	0.2	1.0	-0.8
seventy	70	104	4	17.3	3.9	+3.4
seventy-one	71	1	0	0.2	0.0	(+)
seventy-two	72	5	1	0.8	1.0	-0.2
seventy-three	73	1	0	0.2	0.0	(+)
seventy-four	74	1	1	0.2	1.0	-0.8
seventy-five	75	48	1	6.0	1.0	+5.0
seventy-six	76	1	1	0.2	1.0	-0.8
seventy-seven	77	1	0	0.2	0.0	(+)
seventy-eight	78	1	1	0.2	1.0	-0.8
seventy-nine	79	1	0	0.2	0.0	(+)
eighty	80	188	11	31.3	10.8	+1.9
eighty-one	81	1	1	0.2	1.0	-0.8
eighty-two	82	1	0	0.2	0.0	(+)
eighty-three	83	5	3	0.8	3.0	-0.7
eighty-four	84	1	1	0.2	1.0	-0.8
eighty-five	85	8	2	1.3	2.0	-0.4
eighty-six	86	1	0	0.2	0.0	(+)
eighty-seven	87	1	0	0.2	0.0	(+)

Таблица 2 (продолжение)

<i>n</i>	<i>i</i>	<i>f</i> ₁	<i>f</i> ₂	<i>f'</i> ₁	<i>f'</i> ₂	<i>D_f</i>
eighty-eight	88	1	0	0.2	0.0	(+)
eighty-nine	89	1	1	0.2	1.0	-0.8
ninety	90	88	12	14.6	11.8	+0.3
ninety-one	91	1	0	0.2	0.0	(+)
ninety-two	92	1	0	0.2	0.0	(+)
ninety-three	93	4	0	0.7	0.0	(+)
ninety-four	94	1	0	0.2	0.0	(+)
ninety-five	95	1	1	0.2	1.0	-0.8
ninety-six	96	5	2	0.8	2.0	-0.6
ninety-seven	97	1	0	0.2	0.0	(+)
ninety-eight	98	10	1	1.7	1.0	+0.7
ninety-nine	99	1	3	0.2	3.0	-0.9
hundred	100	1396	171	232.2	168.6	+0.4
Суммарная частота числительных первой сотни		60055	6466	9937.7	6375.7	+0.57

Примечание. *n* — количественные числительные; *i* — соответствующие им числа натурального ряда; *f*₁, *f*₂ — абсолютные частоты этих числительных в СР и СК; *f'*₁ и *f'*₂ — относительные частоты (в миллионных). В связи с тем что в СР числительные с частотой меньше трех не приведены, всем таким числительным условно приписана частота 1 (среднеарифметическое от трех возможных частот — 2, 1 и 0). Для числительного *one*, частота которого в СР и СК приводится суммарно с частотой омофоничного ему местоимения *one*, дана приближительная оценка частоты на основе соотношения частот этих слов в частотно-семантическом словаре Лорджа и Торндайка [8]. В графе *D_f* положительные значения имеют те числительные, у которых относительная частота в школьных текстах выше, чем в текстах СК.

диапазону: числительные первых двух десятков употребляются в выборке СР значительно чаще, чем в выборке СК, у числительных третьего десятка превышение частот СР над частотами СК незначительно, а числительные последующих десятков встречаются чаще уже не в текстах СР, а в текстах СК (см. графу *D_f* в табл. 2).

Из табл. 2 следует далее, что частоты числительных быстро убывает по мере увеличения обозначаемых ими чисел. Это убывание частот особенно хорошо заметно при графическом представлении материала (рис. 2). Можно видеть, что кривая, соответствующая частотам числительных в СР (крестики), имеет в общем более сильный наклон к оси абсцисс, чем кривая, полученная по материалам СК (сплошная линия). Это говорит о том, что тенденция к убыванию частоты числительных по мере роста обозначаемых ими чисел выражена в текстах школьников сильнее, чем в текстах общего характера.

Наблюдающееся убывание частот отражает, по-видимому, тот факт, что человек в своей повседневной деятельности чаще сталкивается со сравнительно небольшими количествами предметов или явлений, вследствие чего в процессе коммуникации ему приходится обозначать словами преимущественно малые числа: чем больше число, тем реже возникает необходимость его обозначать. Для речевой практики ребенка это, по-видимому, особенно характерно.⁶

Легко заметить, что полученные нисходящие кривые не являются плавными: на них обнаруживаются всплески, соответствующие частотам тех числительных, которые обозначают круглые числа. В этом отношении данные кривые очень напоминают кривую возрастного состава населения, изображенную на рис. 1. Однако, если учесть, что на рис. 2 частоты даны в логарифмическом масштабе, то станет ясно, что всплески частот на круглых числах здесь во много раз больше: частота «круглого» числительного (например, fifty 'пятьдесят') может превосходить частоту соседнего «некруглого» (например, forty-nine 'сорок девять') или fifty-one 'пятьдесят один') в десятки и даже сотни раз.

Можно предложить, что в формировании таких сильных завышений частот участвуют как внутриязыковые, так и экстраязыковые факторы. С одной стороны, это отражение того реального факта, что многие предметы — товары, денежные суммы и т. п. — в силу существующей системы счисления объединяются в партии, содержащие именно «круглые» количества единиц. С другой стороны, играет роль определенная психологическая особенность человека, выражающаяся в стремлении обозначать количества, известные лишь приблизительно, ближайшими круглыми числительными, а также использовать их в тех ситуациях, когда количество известно точно, но в точном его выражении нет надобности. Эта особенность, по крайней мере в английском языке, связана в какой-то степени со стремлением к экономии речевых усилий: круглые числительные, соответствующие числам, оканчивающимся на 0 (т. е. являющиеся названиями десятков), короче и проще по структуре, чем некруглые, и при отсутствии необходимости в обозначении точного количества они предпочитают числительным некруглым. Однако полностью объ-

⁶ Высокая частотность числительных первого десятка объясняется отчасти еще и тем обстоятельством, что они используются также для обозначения всех чисел за пределами первой сотни, входя в состав сложных числительных типа two hundred 'двести', five hundred and sixteen 'пятьсот шестнадцать', seven thousand 'семь тысяч' и т. п. (по той же причине несколько завышается и частота числительного hundred, которое используется для обозначения не только 'ста', но и любого другого числа в пределах от 101 до 999). Однако частота числительных за пределами первой сотни (даже суммарная) настолько мала, что это не может существенно отразиться на форме рассматриваемой кривой.

яснить этим факт завышения частот у круглых числительных нельзя, поскольку круглые числительные, соответствующие числам, оканчивающимся на 5, тоже обладают высокой частотностью, хотя и не отличаются ни простотой структуры, ни краткостью. По-видимому, здесь играет роль и ряд других факторов, в частности широкое использование круглых числительных в «окачествленном» значении, например во фразеологических сочетаниях

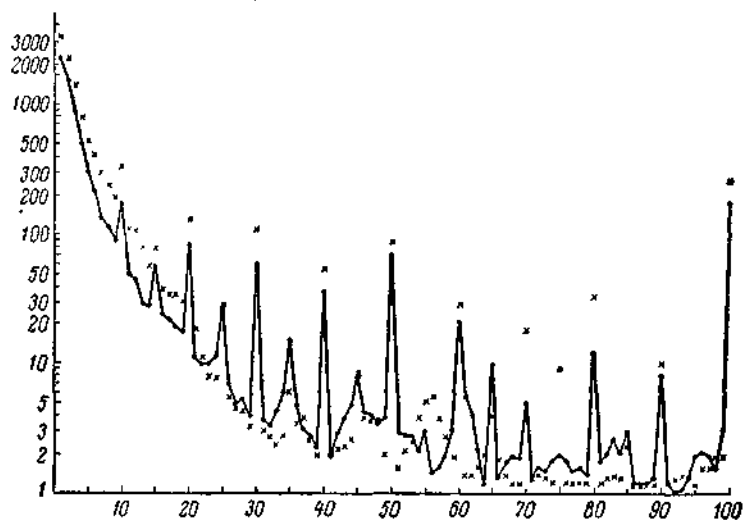


Рис. 2. Частотное распределение английских количественных числительных в текстах СК (сплошная линия) и в текстах СР (крестики).

По оси абсцисс — числительные, обозначенные соответствующими числами натурального ряда; по оси ординат — относительные частоты числительных (в миллионных) в логарифмическом масштабе. Точки для некруглых числительных получены путем сглаживания данных табл. 2 методом «скользящей средней».

типа ninety times I've told you 'сто раз я тебе говорила' (ср. [3, стр. 5]).

Из рис. 2 (как и из табл. 2) видно, что в текстах СР круглые числительные встречаются, как правило, чаще, чем в текстах СК. Однако это еще не позволяет делать какие-либо выводы о более сильном предпочтении к круглым числам в текстах школьников, поскольку и суммарная частота числительных в этих текстах выше. Необходимо ввести какую-то меру, которая позволила бы определять степень такого предпочтения независимо от общей частоты числительных в тексте. Такой мерой мог бы быть «коэффициент округления» (k), вычисляемый отдельно для каждого полудесятка числительных (круглого и четырех соседних с ним

некруглых — двух больших и двух меньших) по следующей формуле:

$$k_r = \frac{4f'_r - (f'_{r-2} + f'_{r-1} + f'_{r+1} + f'_{r+2})}{f'_{r-2} + f'_{r-1} + f'_{r+1} + f'_{r+2}}$$

где r — круглое числительное, а $r-2$, $r-1$, $r+1$ и $r+2$ — окружающие его числительные некруглые.

Вычисления, проведенные на материале СР и СК, показали, что во всех полудесятках, кроме первого, коэффициент округления выражается положительным числом, причем для круглых числительных типа r_0 (т. е. оканчивающихся на 0) величина его особенно высока (табл. 3). Оказалось также, что величина коэффициентов округления в текстах СР в большинстве случаев выше, чем в текстах СК. Это превышение особенно велико в случае числительных типа r_0 (кроме ten и twenty), а также у числительных twenty-five 'двадцать пять' и seventy-five 'семьдесят пять'. Таким образом, независимо от общей частоты употребления числительных, в текстах школьных письменных работ сильнее, чем в текстах общего характера, выражено предпочтение к тем круглым числительным, которые являются названиями десятков и четвертей сотни (включая и самое сотню). Все это позволяет сделать вывод о несомненной (и притом обратной) зависимости между стремлением к предпочтению круглых числительных и уровнем образования.

Внимательное изучение полученных кривых позволяет обнаружить еще одну тенденцию в распределении числительных — систематическое превышение частот четных числительных над частотами нечетных. Можно заметить, что частоты числительных, обозначающих числа, оканчивающихся на 2, 4, 6 и 8, в большинстве случаев выше частот соседних с ними числительных, обозначающих числа, оканчивающиеся на 1, 3, 7 и 9. Эта тенденция выражена значительно слабее, чем тенденция к предпочтению круглых числительных, но действие ее несомненно, особенно во второй и третьей четвертях исследуемого диапазона. Интересно, что и здесь уровень образования играет заметную роль: если в текстах СК тенденция к завышению четных числительных проявляется, главным образом, на втором—пятом десятках (рис. 3, нижняя часть), то в текстах СР она выступает заметно ярче и прослеживается на протяжении почти всей первой сотни (рис. 3, верхняя часть).⁷

Возможно, что значительно более высокая частотность круглых числительных типа r_0 (т. е. 10, 20, 30 и т. д.) по сравнению с круглыми же числительными типа r_s (15, 25 и т. д.) объясняется,

помимо уже названных факторов, также и тенденцией к предпочтительному употреблению четных чисел.

Для количественной характеристики этой тенденции может быть предложен «коэффициент предпочтения четности» (k_s), вычисляемый отдельно для каждого полудесятка числительных (группирующихся вокруг очередного круглого) и определяемый как отношение разности частот четных и нечетных числительных

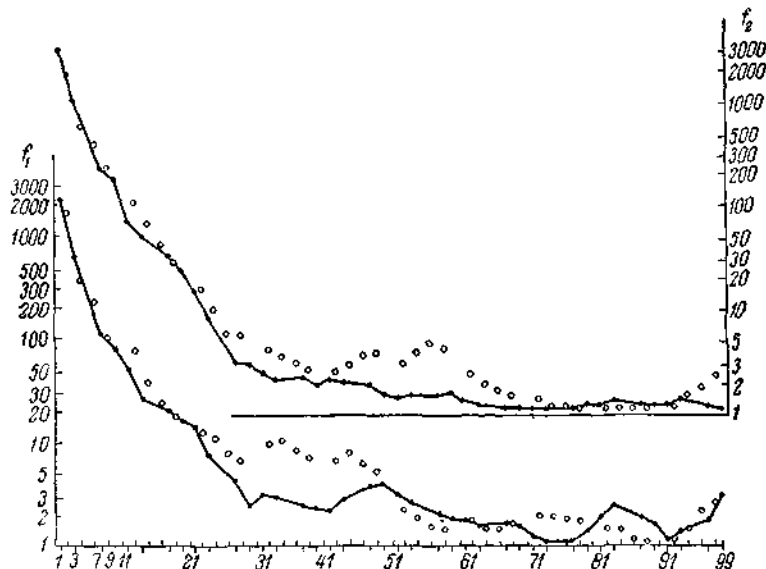


Рис. 3. Частотное распределение английских некруглых числительных в текстах СК (нижний график) и в текстах СР (верхний график).

Сплошная линия — нечетные числительные, кружочки — четные. На оси абсцисс, в целях экономии места, подписаны только отдельные нечетные числа.

к частоте этих же нечетных числительных, например (для первого полудесятка):

$$k_s = \frac{(f_4 + f_6) - (f_3 + f_7)}{f_3 + f_7} = \frac{f_4 + f_6}{f_3 + f_7} - 1.$$

Масштабы настоящей статьи не дают возможности привести здесь данные, характеризующие значения коэффициентов предпочтения четности в рассматриваемых словарях по всему изучаемому диапазону; ограничимся лишь указанием на то, что в большинстве случаев значения эти в текстах СР выше, чем в текстах СК (что и можно было ожидать по форме полученных кривых).

В отличие от тенденции к предпочтению круглых числительных, в формировании данной тенденции играют роль, вероятно, только экстралингвистические (в первую очередь психологиче-

⁷ В демографических материалах, приведенных выше (рис. 1, табл. 1), также можно обнаружить завышения частот на четных возрастах (что было в свое время отмечено Б. С. Ястремским [5]).

Таблица 3

Величина коэффициентов округления в текстах СР (k_1) и в текстах СК (k_2)

r_4	k_1	k_2	$k_1 - k_2$	r_0	k_1	k_2	$k_1 - k_2$
5	- 0.08	- 0.12	+ 0.04	10	+ 0.99	+ 1.42	- 0.43
15	+ 0.93	+ 1.60	- 0.67	20	+ 4.73	+ 5.27	- 0.54
25	+ 4.87	+ 2.70	+ 2.17	30	+ 28.80	+ 17.15	+ 11.65
35	+ 1.73	+ 3.00	- 1.27	40	+ 46.85	+ 13.40	+ 33.45
45	+ 3.00	+ 1.00	+ 2.00	50	+ 50.59	+ 33.00	+ 17.59
55	+ 0.52	+ 1.00	- 0.48	60	+ 35.22	+ 6.64	+ 28.58
65	+ 3.24	+ 71.00	- 67.76	70	+ 51.00	+ 4.33	+ 46.67
75	+ 47.00	+ 1.00	+ 46.00	80	+ 187.00	+ 21.00	+ 166.00
85	+ 3.00	+ 1.00	+ 2.00	90	+ 87.00	+ 47.00	+ 40.00
95	+ 1.75	+ 1.00	+ 0.75	100	+ 252.82	+ 84.50	+ 168.32

Примечание. r_5 — числа натурального ряда, оканчивающиеся на 5 и соответствующие круглым нечетным числительным; r_0 — числа натурального ряда, оканчивающиеся на 0 и соответствующие круглым четным числительным; k — коэффициент округления (см. объяснения в тексте). Положительное значение $k_1 - k_2$ соответствует случаям, когда предпочтение круглого числительного некруглым выражено в детских текстах сильнее, чем в текстах общего характера.

ские) факторы. Тенденция эта нуждается, несомненно, в более глубоком изучении в плане задач инженерной психологии, особенно в связи с отмечаемыми в научно-технической литературе фактами систематического превышения четных чисел по сравнению с нечетными при обработке результатов измерений. В плане же чисто лингвистических исследований заслуживает дальнейшего изучения первая из этих двух тенденций, ибо стремление к предпочтительному использованию круглых числительных самым тесным образом связано со структурой языка и может существенно различаться от одного языка к другому.

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

1. Боярский А. Я. и др. Курс демографии. М., 1967.
2. Пиотровский Р. Г. и Турыгина Л. А. Антиномия «язык — речь» и статистическая интерпретация нормы языка. СтРААТ, стр. 5—46.
3. Супрун А. Е. Обозначение неточных (приблизительных) количеств при помощи определенно-количественных числительных. Уч. зап. филол. факультета Киргизского гос. ун-ва, вып. 8, Фрунзе, 1962.
4. Супрун А. Е. Славянские числительные. (Становление числительных как части речи). Дисс. на соискание учен. степени доктора филол. наук. Фрунзе, 1965 (рукоп.).
5. Ястремский Б. С. Труды по статистике. М., 1937.
6. Kučera H. and Francis W. N. Computational analysis of present-day American English. Providence, Rhode Island, 1967.
7. Linder F. E., Grove R. D. Vital statistic rates in the United States, 1900—1940. Washington, 1943.
8. Lorge I. and Thorndike E. L. A semantic count of English words. New York, 1938.
9. Rinsland H. D. A basic vocabulary of elementary school children. New York, 1945.

Д. С. Демьянова

ЧАСТОТНЫЙ СЛОВАРЬ НЕМЕЦКОГО ПОДЪЯЗЫКА
АВТОМАТИЗАЦИИ

Словарь (табл. 2) составлен в результате сплошного расписывания 200 текстов по автоматизации на немецком языке по 1000 словоупотреблений каждый (всего 200 тыс. словоупотреблений). В выборке зарегистрировано 29 759 разных словоформ с учетом лексико-грамматической омографии, из которых в настоящий список включены 1005 словоформ¹ с частотами не менее 21. Тексты расписывались вручную. При их подборе (табл. 1) и организации выборки в качестве справочного материала использовались реферативные журналы, советские и зарубежные периодические издания, монографии и учебные пособия. Консультации были оказаны сотрудниками Киевского института автоматки, Берлинского института автоматки и Украинского научно-исследовательского института информации; всем этим учреждениям автор искренне признателен за полезные советы.

Таблица 1

Состав выборки			
№№	Раздел подъязыка	Количество словоупотреблений	% от общего объема выборки
1	Автоматизация в машиностроении	40 000	26%
2	Социальные проблемы и автоматизация управленческих работ	36 000	18%
3	Автоматизация в транспортном деле	30 000	15%
4	Автоматизация в металлургии	30 000	15%
5	Вычислительная и информационная техника	32 000	16%
6	Технические средства автоматизации и телемеханики	32 000	16%
Итого		200 000	100

¹ В приводимом словаре лексико-грамматические омографы отдельно не указываются.

Таблица 2

Частотный список словоформ

i	Словоформы	F
1	die	9050
2	der	8366
3	und	5937
4	in	3774
5	von	2669
6	den	2372
7	mit	2271
8	werden	2092
9	ist	2062
10	das	2060
11	zu	1965
12	für	1910
13	des	1863
14	eine	1800
15	auf	1648
16	wird	1466
17	sich	1365
18	bei	1333
19	ein	1261
20	durch	1214
21	dem	1184
22	auch	1139
23	sind	1083
24	im	1051
25	daß	1042
26	einer	1021
27	es	1016
28	an	1013
29	oder	892
30	als	856
31	nicht	813
32	kann	744
33	aus	717
34	so	709
35	zur	698
36	diese	692
37	man	684
38	einem	670
39	sie	661
40	nach	606
41	einen	589
42	können	556
43	über	553
44	wie	526
45	nur	522
46	um	506
47	dieser	462
48	wenn	425
49	bis	409
50	zum	405
51	noch	401

Таблица 2 (продолжение)

i	Словоформы	F
52	aber	357
53	eines	356
54	hat	350
55	vor	326
56—57	wurde, zwei	324
58	haben	316
59	muß	314
60	sehr	305
61	Maschine	286
62	sein	277
63	dann	270
64	zwischen	268
65	mehr	264
66—67	am, damit	261
68	Automatisierung	246
69	alle	240
70	unter	239
71	mm	234
72	er	230
73	ohne	226
74	Bild	214
75—76	immer, je	212
77—80	beim, hier, während, z. B.	208
81	vom	206
82	wir	203
83—84	möglich, müssen	197
85	sowie	196
86	lassen	195
87—88	also, ihre	191
89	jedoch	190
90	dabei	189
91—93	da, diesem, Entwicklung	187
94	wurden	183
95	besonders	177
96—97	einzelnen, heute	174
98	automatisch	173
99	Maschinen	171
100	diesen	169
101	denen	166
102—103	schon, Zeit	165
104	etwa	163
105	wobei	160
106	dieses	159
107	bereits	158
108	deren	157
109—110	neue, sondern	154
111	bzw.	151
112	Steuerung	150
113—114	Anlage, verwendet	149
115	soll	144
116	großen	142
117	gibt	141
118	Computer	137

Таблица 2 (продолжение)

i	Словоформы	F
119	war	136
120	keine	135
121	zeigt	134
122—124	dadurch, elektronischen, läßt	132
125—126	besteht, jeder	130
127—128	anderen, Art	129
129	beiden	128
130—132	Form, liegt, verschiedenen	127
133	Hilfe	126
134	automatische	125
135	System	123
136—138	große, seine, neuen	122
139	gleichzeitig	120
140	selbst	119
141—142	Betrieb, Werkstücke	118
143	Anlagen	115
144	Teil	114
145—147	allem, erfolgt, Rechner	113
148	Einsatz	112
149—151	Automation, erreicht, wieder	111
152	ihrer	109
153	usw.	108
154—155	elektronische, was	105
156—157	daher, nun	104
158	andere	101
169—161	dies, Produktion, Verfahren	100
162—163	Beispiel, ganz	99
164—165	Firma, Teile	98
166—168	Anwendung, Jahren, Verwendung	96
169—170	Bedeutung, technischen	95
171—172	ab, welche	94
173	automatischen	93
174—176	entsprechend, jede, viele	92
177—178	dazu, Weise	90
179	allen	89
180—182	arbeitet, drei, Leistung	88
183—189	Aufgaben, erst, Hand, innerhalb, wo, Zahl, zwar	87
190	elektrische	86
191—192	ihren, Werkstück	85
193—196	Anzahl, Lage, sollen, worden	84
197—200	entwickelt, Fall, neben, verschiedene	83
201—203	arbeiten, bisher, ermöglicht	82
204—206	beispielsweise, Daten, oft	81
207—208	eingesetzt, jeweils	80
209—214	außerdem, denn, d. h., ebenfalls, Fällen, waren	79
215	seiner	78
216—219	leicht, m, Modell, Technik	77
220—222	hohe, Probleme, uns	76
223—226	erforderlich, liegen, Menschen, Regelung	75
227—229	Aufgabe, deshalb, weniger	74
230—231	handelt, solche	73
232—237	aller, bestimmten, ersten, Möglichkeit, sowohl, Wert	72
238—242	dessen, dort, Größe, Messung, weil	71

Таблица 2 (продолжение)

i	Словоформы	F
243—246	gesamte, Lösung, vielen, zunächst	70
247—250	bestimmt, Programm, seit, vier	69
251—256	Bearbeitung, Geräte, jedem, kommt, Werkzeugmaschinen, wesentlich	68
257—258	Druck, Herstellung	67
259—264	Fertigung, gegeben, insbesondere, Jahre, machen, Temperatur	66
265—268	geht, Genauigkeit, Größen, Kontrolle	65
269—274	Bereich, durchgeführt, ergibt, genau, Grund, sei	64
275—286	ausgeführt, doch, gerade, jedes, jetzt, letzten, notwendig, praktisch, Regler, Reine, stehen, steht	63
287—295	allerdings, Gebiet, kleinen, kommen, Leistungen, mehrere, stellt, Verbindung, Verfügung	62
296—302	direkt, Gerät, größere, Datenverarbeitung, meist, natürlich, weitere	61
303—309	Arbeit, bestimmte, beträgt, darüber, Frage, Vorteil, würde	60
310—314	ausgerüstet, dient, IBM, Kosten, Vorteile	59
315—323	besondere, einmal, hohen, Industrie, Informationen, kein, kleine, somit, Stelle	58
324—330	elektrischen, entsprechende, entsprechenden, Möglichkeiten, ob, viel, weiter	57
331—334	gegenüber, großer, jeweiligen, solchen	56
335—342	dafür, einige, erste, ihr, möglichst, wäre, Werkzeuge, wirtschaftlichen	55
343—346	Ausführung, konnte, Länge, weitgehend	54
347—351	Eigenschaften, Impulse, relativ, Unternehmen, Werkstücke	53
352—364	Bauelemente, darauf, Einrichtungen, erforderlichen, gegen, Geschwindigkeit, hergestellt, ich, mittels, schließlich, stark, technische, wegen	52
365—370	fast, folgende, führt, gleichen, größeren, mechanische	51
371—373	Ende, entspricht, gemacht	50
374—380	beide, benötigt, dienen, Einrichtung, geeignet, gesamten, sollte	49
381—387	Elektronik, geben, Maße, meisten, Spannung, USA, Werkzeugmaschine	48
388—399	allein, ergeben, ermöglichen, führen, ganze, gut, konnten, Mensch, Prinzip, verwenden, vorgesehen, Wirtschaft	47
400—406	Änderung, Arbeiten, C, kleiner, Lochstreifen, Problem, seinen	46
407—415	allgemeinen, Anforderungen, Band, bleibt, erster, gilt, Jahr, Transport, weit	45
416—426	Abmessungen, bietet, ebenso, erzeugt, gemessen, gleiche, Konstruktion, pro, sogar, Transistoren, wesentlichen	44
427—463	Aufbau, Bedingungen, gebracht, gesteuert, lange, nämlich, Schaltungen, Typ, Unternehmens, Vorgänge	43
437—453	Angaben, Bandes, dar, Deutschland, Folge, groß, Gruppe, halten, häufig, mechanischen, neu, Praxis, selbsttätig, V., verbunden, versehen, Werte	42
454—467	bekannt, benutzt, besser, eigentlichen, einfach, eingestellt, Entscheidungen, ihnen, Luft, sofort, solcher, stellen, Weg, Zusammenhang	41
468—484	Antrieb, Aufwand, enthält, entweder, erhalten, erreichen.	40

Таблица 2 (продолжение)

i	Словоформы	F
485—498	finden, geführt, hierbei, höhere, ihm, ja, modernen, Rahmen, stets, Systems, untergebracht bestehen, darin, erfolgen, Ergebnisse, Forderungen, hierfür, Höhe, ihrem, recht, Schaltung, Stand, Stellung, steuern, zusammen	39
499—513	Ablauf, Ausbildung, bleiben, davon, Durchmesser, Elektronenrechner, erfordert, Förderer, Gruppen, insgesamt, Linie, Verhältnis, wenig, wichtig, wodurch	38
514—526	bewegt, dagegen, Einheiten, Fehler, gebaut, gezeigt, Messungen, Programmierer, Rationalisierung, Regelgröße, verschiedener, Wege, Zukunft	37
527—543	Abhängigkeit, automatisierten, befinden, besitzt, durchführen, einfachen, erhält, Falle, Firmen, Gehänge, hinaus, jeden, macht, notwendigen, Steuerungen, unmitelbar, wohl	36
544—562	aufgebaut, besonderen, darf, dargestellt, einfache, Einführung, eingebaut, erkennen, Erzeugnisse, Fabrbahn, Funktion, industriellen, Kreisförderer, mehreren, Programmierung, schnell, Schwierigkeiten, Überwachung, zweiten	35
563—573	deutlich, Drehzahl, Einstellung, folgenden, genaue, gute, laufend, Material, Seite, Sollwert, zugeführt	34
574—592	abhängig, Banken, bedeutet, ca., elektrisch, ferner, Forderung, früher, geringen, gleich, gleicher, hoch, infolge, Methode, Serien, ständig, Stellen, vorgenommen, Welt	33
593—609	Anfahren, betreffenden, einzelne, genommen, gewöhnliche, kaum, lediglich, nimmt, Rolle, schneller, Steigerung, Typen, unabhängig, vielfach, vorhanden, weiteren, zeigen	32
610—632	Abstand, arbeitende, bilden, eigenen, elektronischer, entstehen, erzielt, etwas, gewährleistet, Grundlage, hierzu, konstant, menschlichen, Qualität, sämtliche, Sicherheit, Werk, Widerstand, wiederum, will, wirkt, zusätzliche, Zweck	31
633—676	alles, Änderungen, Bänder, betrachtet, bringt, Erfahrungen, ermittelt, Fahrzeuge, fest, Fortschritte, ganzen, gehen, gewisse, geworden, GmbH, größten, hatte, her, kg, Kunden, Maß, Mechanisierung, Messen, Meßgeräte, nächsten, neuer, obwohl, Programme, Rechenanlage, Reglers, seinem, setzt, sobald, sogenannten, Strom, Teilen, Unternehmensführung, Untersuchungen, verschiedensten, voll, völlig, Voraussetzung, vorher, Zahlen	30
677—711	Analyse, angeschlossenen, Arbeitsgänge, automatisierte, Bedarf, befindet, berücksichtigt, bezeichnet, bringen, Dicke, Ebene, erheblich, Fabrik, festgelegt, festgestellt, gehört, gestellt, Information, könnte, Ländern, Markt, messen, miteinander, nehmen, nunmehr, Operationen, Regelabweichung, Richtung, Sechs, Temperaturen, Verwaltung, wirtschaftlich, wirtschaftliche, Zähler, zurück	29
712—742	ähnlich, allgemein, ändern, angeordnet, Arbeitsweise, außerordentlich, automatischer, beschrieben, besten, bewirkt, daraus, entsprechen, geringer, geschieht, Grenzen, größer, hängt, höheren, Impuls, Informationsverarbeitung, kurz, Maßnahmen, Planung, Rückführung	28

Таблица 2 (продолжение)

i	Словоформы	F
743—765	Sekunde, u. a., üblichen, unterscheiden, vielleicht, Werkzeug, Zwecke angebracht, Belastung, gering, Gewicht, gewünschten, grundsätzlich, Kette, laufen, liefert, Lochkarten, menschliche, Methoden, Signale, sogenannte, Spannungen, speziellen, Straße, Tätigkeit, Toleranzen, tritt, unsere, verhältnismäßig, Wirtschaftlichkeit	27
766—798	andererseits, Artikel, Automaten, Berechnung, Bewegung, bieten, Boden, Darstellung, derartige, durchaus, Durchführung, Eingang, elektrischer, elektronisch, entsteht, erforderliche, erfordern, erfüllen, findet, genannt, geringe, Gründen, höher, konstruiert, schaffen, selbstverständlich, spezielle, t, Tag, Umfang, umfaßt, vollautomatisch, zweite	26
799—846	angepaßt, Anordnung, Arbeiter, ausgestattet, Berücksichtigung, besonderer, Betriebs, Dr., Durchgang, Einfluß, eingeben, Erhöhung, erscheint, erwähnt, Fortschritt, gekennzeichnet, ins, Kern, Kombination, mathematische, Millionen, Mitarbeiter, New, Oberfläche, physikalischen, pneumatische, Produkt, Punkt, Rechenanlagen, Rechnung, rund, schnelle, sicher, speziell, Spiel, Stahl, Systeme, teilweise, Verbesserung, Vergleich, Verhältnisse, vermeiden, Vorgang, Wahl, Werkstücken, wichtigsten, wollen, Zustand	25
847—883	Abweichungen, außer, Behälter, beschränkt, besitzen, digitale, eingerichtet, einigen, erhebliche, erzielen, Faktoren, fünf, gar, geliefert, gelöst, gestattet, gewählt, größte, hinsichtlich, indem, Meßwerte, moderne, normalen, Schritt, Speicher, Stunden, Tatsache, trägt, vermieden, Verstärker, Vielzahl, vollautomatische, vorgegebenen, Vorschub, wenigen, zentralen, Ziel	24
884—919	anderer, Auswahl, Auswirkungen, bearbeitet, Beispiele, Berechnungen, bildet, digitalen, Dollar, dürfte, eingeführt, festen, Formen, gefunden, Gegensatz, geschaffen, gewissen, Hersteller, ihn, klar, klein, konventionellen, lösen, Mittel, Montage, oben, optimale, Streifen, Stückzahlen, übertragen, unserer, Untersuchung, Veränderung, weiteres, wesentliche, würden	23
920—961	Arbeitskräfte, Ausführungen, Bedienung, Behandlung, bessere, betrieblichen, bisherigen, Breite, Computers, Eingabe, enthalten, erlaubt, Fahrzeug, feststellen, frei, gefertigt, gehalten, gelegt, gesagt, Gewinn, guten, Interesse, Kapazität, läuft, lieferbar, Minuten, nachdem, numerisch, parallel, Produktivität, Regel, Rollenbahnen, stehenden, Tätigkeiten, treten, überhaupt, Unterlagen, Veränderungen, vielmehr, vorhandenen, weiterhin, wirklich	22
962—1005	AG, Arbeitsgang, aufgenommen, ausgebildet, Bau, bedient, beeinflusst, beginnt, bestimmter, betragen, Betriebe, Dauer, eignen, eignet, Erfolg, erhöht, folgt, Fördergut, führte, Gebieten, genauen, geschaltet, größerer, heißt, hinter, keinen, kleinere, Lager, Meßgerät, nahezu, neuerdings, Nutzen, Platz, rasch, Reihenfolge, rein schließt, sehen, Stücks, trotz, Verarbeitung, voneinander, voraus, weiterer	21

Э. Я. Рудник, А. Н. Галкин

ЧАСТОТНЫЙ СЛОВАРЬ НЕМЕЦКОГО ПОДЪЯЗЫКА ПОЧВОВЕДЕНИЯ

Словарь составлен на материале 100 текстов из немецких не-периодических изданий и журнала «Zeitschrift für Pflanzenernährung und Bodenkunde», изданных в ГДР в 1965—1970 гг. Каждый текст имеет длину в 1000 словоупотреблений; общий объем выборки равен 100 тыс. словоупотреблений.

В словаре (табл. 1) приводятся 547 словоформ с частотой не менее 20. Имена собственные, физические единицы измерения, химические символы, формулы и цифры исключались.

Список распределения количества словоформ с абсолютными частотами $F < 20$ приведен в табл. 2.

Частотный список словоформ

Таблица 1

i	Словоформы	F
1	der	4559
2	die	4431
3	und	2951
4	in	2140
5	von	1420
6	den	1401
7	auf	1389
8	ist	1164
9	des	1033
10	mit	1031
11	bei	1023
12	im	940
13	eine	902
14	zu	863
15	das	782
16	weiden	772
17	durch	767

i	Словоформы	F
18	als	717
19	auch	633
20	nicht	629
21	sich	627
22	daß	616
23	dem	605
24—25	für, aus	593
26	wird	562
27	an	557
28	Böden	504
29	es	486
30	ein	485
31	nach	480
32	oder	441
33	Boden	428
34	so	401
35	nur	391
36	einer	351
37	sie	326
38	aber	321
39	wie	305
40	diese	295
41	zur	294
42	man	293
43	kann	278
44	bis	273
45	einem	260
46	unter	255
47—48	dieser, um	227
49	Wasser	224
50	wir	214
51	wurde	210
52	einen	209
53—55	Bodens, noch, Pflanzen	207
56	vor	196
57—58	also, wurden	194
59	sehr	193
60	können	187
61	zwischen	179
62	da	172
63	Gehalt	169
64	hat	162
65	über	157
66	je	152
67	zum	147
68	haben	146
69	wenn	145
70	dann	138
71—72	beim, hier	127
73	am	122
74	Proben	120
75	sein	117
76	etwa	115

Таблица 1 (продолжение)

i	Словоформы	F
77—79	eines, er, muß	114
80—81	besonders, während	110
82	mehr	109
83—86	bzw., jedoch, Kalk, Teil	102
87	organischen	96
88—89	diesen, Düngung	94
90—91	Bedeutung, Werte	93
92—93	war, Wirkung	89
94—95	daher, Form	87
96—97	ab, infolge	84
98	ihrer	83
99	sondern	82
100—105	alle, Bildung, denen, liegt, vom, Zeit	81
106—108	deren, diesem, Humus	79
109—111	sowie, Ton, weniger	78
112	Einfluß	77
113—114	einzelnen, gut, schon	75
115—117	anderen, Klimastufe, Verwitterung	74
118—119	ohne, Untersuchungen	71
120—122	erfolgt, leicht, pH	69
123—127	dieses, hohen, kommt, Tabelle, weil	68
128—133	bereits, dagegen, dies, Pflanze, schon, Wert	66
134—136	Eigenschaften, Horizont, zwar	65
137	Horizonten	64
138	besteht	63
139—141	Ergebnisse, meist, wieder	62
142	läßt	61
143—146	allem, lassen, Poren, zeigt	60
147	Nährstoffe	59
148	wenig	58
149	deshalb	57
150	vorhanden	56
151—155	Abbildung, entsprechend, Gehalte, Mengen, selbst	55
156—160	Luft, möglich, oberen, Verhältnis, Zustand	54
161—165	bestimmt, Dünger, enthalten, geringe, keine	53
166—170	allen, Bestimmung, ihre, Sand, zwei	52
171—174	Bereich, großen, viel, wirkt	51
175—178	Anteil, dabei, damit, erst	50
179—182	beiden, relativ, Schichten, Tiefe	49
183—185	gebildet, große, liegen	48
186—192	Folge, immer, mineralischen, Oberfläche, Reaktion, Tab., z. B.	47
193—197	Anwendung, denn, meist, zeigen, zunächst	46
198—200	ganz, gegenüber, Zunahme	45
201—205	etwas, Menge, Phosphorsäure, Unterschiede, was	44
206—217	Ausmachung, beträgt, finden, gleichen, Jauche, Ionen, Klima, kommen, müssen, organischer, Probe, wegen	43
218—219	Beziehung, Material	42
220—226	bezeichnet, chemischen, Feuchtigkeit, Grund, höher, Lösung, Waldböden	41
227—236	dadurch, ebenfalls, erreicht, findet, Frage, organische, nimmt, seiner, Verbindungen, wohl	40
237—246	darauf, geht, Methode, seine, Vergleich, verschiedenen,	39

Таблица 1 (продолжение)

i	Словоформы	F
	verwendet, Verwendung, Wassergehalt, Zusammensetzung	
247—252	andere, erhalten, Fall, gibt, Löß, werden	38
253—261	doch, Durchschlammung, entsteht, größer, Lagerung, Messung, uns, Wasserleitfähigkeit, Wurzeln	37
262—268	allerdings, Bodenlösung, entstehen, sogar, schwach, Vegetation, wobei	36
269—279	Abb., Aggregate, bestimmten, chemische, Erde, Fraktionen, gleich, hohe, Kieselsäure, Mikroorganismen, Sandboden	35
280—292	Abhängigkeit, bilden, entnehmen, höheren, ihr, konnte, konnten, Konzentration, oft, sauren, Weise, wirken, wo	34
293—298	Absorption, allgemeinen, Bedingungen, gegeben, Kali, neben	33
299—310	Beziehungen, fast, nun, schnell, solche, Wald, waren	32
311—325	auftreten, bildet, Bodenbildung, d. h., größere, Grundwasser, hauptsächlich, hin, Humusgehalt, innerhalb, kaum, ob, oben, untersuchten, welche	31
326—336	Ammoniak, einige, handelt, hoch, Kalium, Regel, Schicht, sowohl, statt, Temperatur, weitere	30
337—347	erforderlich, gegen, Gegensatz, hierbei, Jahren, kein, Stroh, treten, Verfahren, wesentlich, würde	29
348—361	Arbeit, aufgenommen, Chloralhydrat, entstanden, ergibt, folgende, Gebieten, Hilfe, Horizonte, tritt, unten, untere, Untergrund, Zersetzung	28
362—374	abhängig, drei, Entnahme, Entwicklung, Ergebnis, festgestellt, höher, ihren, Methoden, Niederschläge, Verdunstung, Wassers, Zusammenhang	27
375—396	Art, Atmosphäre, Aufnahme, außerdem, bedingt, Fällern, Fixierung, führen, geringer, GHV, Höhe, ihrem, Kalkgehalt, Kulturpflanzen, möglichst, Tongehalt, unterscheiden, untersucht, Ursache, weit, zurückzuführen, zusammen	26
397—421	Ackerkrume, Aufbau, bezeichnen, Bodenreaktion, Braunerden, Entstehung, erhöht, ersten, Frühjahr, geben, gewisse, Größe, gute, häufig, heute, hinsichtlich, Hohlräume, physikalischen, Profit, pseudovergleiten, Rolle, sandigen, Verhältnissen, Verlauf, Versuchen	25
422—438	Abnahme, Aggregats, Anzahl, Auffassung, Austauschkapazität, besten, Bodenmaterial, Branntkalk, dazu, folgenden, führt, gemessen, geringen, kleiner, meisten, Unterschied, Veränderung	24
439—460	Anstieg, Berechnung, biologischen, dafür, darf, dort, ergeben, geeignet, gilt, ihnen, Leguminosen, leichten, Magnesium, Masse, niedrigen, offenbar, schwerer, Streu, Strontium, Veränderungen, ziemlich	23
461—492	bedeutet, bekannt, besitzen, besondere, bestimmte, bleibt, Calcium, demnach, direkt, Düngemittel, Eisen, Farbe, Flächen, gebunden, Gerät, geringeren, Gesteine, gleiche, Gleichung, groß, größeren, größten, insbesondere, Kationen, Lage, letzten, Mist, praktisch, setzt, teilweise, Versuche, Wasserbewegung, weiteren	22
493—525	ähnlich, berücksichtigt, beschriebenen, Bestandteile,	21

Таблица 1 (продолжение)

i	Словоформы	F
526—547	Bodenart, Bodenteilchen, Durchschlämmungen, Faktoren, feuchten, früher, gebraucht, gleichzeitig, grauen, Gründüngung, hatte, Jahre, kleinen, Kohlensäure, mineralisches, Möglichkeit, natürlich, somit, Stelle, stets, Stickstoffdüngung, tief, Tonminerale, trockenen, ungesättigten, unterschiedlich, unterschiedlichen, verschieden Bindung, biologisch, bisher, Bodenproben, darin, dessen, Düngern, durchgeführt, Durchmesser, dürfen, ermittelt, fünf, Gefahr, gleicher, langsam, Maße, notwendig, Phosphor, Schluff, Wassermenge, Zusatz	20

Таблица 2

Распределение количества словоформ при $F < 20$

i	m	F	i	m	F
548—589	42	10	1210—1360	151	9
590—622	33	18	1361—1556	196	8
623—664	42	17	1557—1765	209	7
665—720	56	16	1766—2118	353	6
721—768	48	15	2119—2564	446	5
769—825	57	14	2565—3254	690	4
826—902	77	13	3255—4461	1207	3
903—986	84	12	4462—6767	2306	2
987—1080	94	11	6768—14434	7667	1
1081—1209	129	10			

С. Е. Мазарская, Л. И. Просветова

ЧАСТОТНЫЙ СЛОВАРЬ СУЩЕСТВИТЕЛЬНЫХ НЕМЕЦКОГО ПОДЪЯЗЫКА НЕОРГАНИЧЕСКОЙ ХИМИИ

Словарь базируется на выборке объемом 100 тыс. случаев употребления существительных в 100 текстах из журнала «Zeitschrift für anorganische und allgemeine Chemie» (ГДР) за 1969—1970 гг. Длина каждого текста равна 1000 употреблений существительных или 3600—3700 всех словоупотреблений. Общий объем выборки, таким образом, составляет около 360—370 тыс. всех словоупотреблений, относящихся ко всем частям речи.

Тексты распределяются по разделам подъязыка неорганической химии следующим образом:

- | | |
|---|---------|
| 1) строение вещества и химическая связь | — 15%; |
| 2) закономерности протекания химических реакций | — 15%; |
| 3) жидкие и твердые растворы | — 10%; |
| 4) химия металлов | — 15%; |
| 5) химия неметаллов | — 15%; |
| 6) комплексные соединения элементов | — 30%; |
| Итого | — 100%. |

Ниже (табл. 1) приводится список 779 существительных в их исходной форме (им. пад. ед. ч.) с частотами не менее 20. В число учитываемых существительных входят субстантивированные глаголы и прилагательные, а также имена собственные. На весь объем выборки приходится 1179 имен собственных.

Таблица 1

Частотный список словоформ

i	Словоформы	F
1	Verbindung	1964
2	Lösung	1514
3	Abbildung	1384
4	Tabelle	1318
5	Spektrum	1012
6	Reaktion	972
7	Wert	945
8	Mol	817
9	Untersuchung	610
10	Substanz	601
11	Temperatur	588
12	Umsetzung	569
13	Komplex	556
14	Stunde	554
15	Bildung	535
16	Bande	524
17	Phase	518
18	Gruppe	499
19	System	491
20	Ion	476
21	Wasser	466
22	Struktur	458
23	Atom	421
24	Ergebnis	413
25	Fall	410
26	Menge	383
27	Produkt	379
28—29	Bindung, Zusammensetzung	363
30	Bereich	354
31	Versuch	348
32	Gleichung	344
33	Messung	336
34	Lösungsmittel	331
35	Ausbeute	330
36	Säure	328
37	Grund	326
38	Darstellung	316
39	Probe	305
40	Kristall	303
41—43	Abstand, Vergleich, Verhältnis	300
44	Benzol	296
45	Analyse	289
46	Teil	281
47	Form	280
48	Äther	278
49	Reaktionsprodukt	272
50	Zersetzung	270
51	Reflex	267
52	Abhängigkeit	261
53	Gemisch	257

Таблица 1 (продолжение)

i	Словоформы	F
54	Intensität	255
55	Gitterkonstante	250
56	Bestimmung	245
57	Aufnahme	240
58	Salz	238
59—60	Konzentration, Vakuum	237
61	Typ	236
62—63	Präparat, Verschiebung	227
64—66	Anion, Signal, Verhalten	226
67—68	Phosphor, Rückstand	222
69	Stickstoff	216
70	Valenzschwingung	214
71	Arbeit	207
72	Methode	201
73	Eigenschaft	199
74	Schwingung	198
75	Kurve	196
76	Bedingung	195
77	Schmelze	193
78	Hydrolyse	191
79	Sauerstoff	189
80—81	Beispiel, Übergang	186
82	Oxid	185
83	Weise	181
84—85	Erhitzen, Molverhältnis	178
86	Kation	174
87	Zeit	172
88	Linie	170
89	Zuordnung	164
90—91	Formel, Reihe	161
92	Luft	157
93	Ligand	153
94	Schicht	152
95	Reduktion	151
96	Gitter	150
97—100	Lage, Molekel, Raumgruppe, Zimmertemperatur	149
101	Proton	148
102	Oxydation	146
103	Absorption	145
104	Chlorid	144
105—107	Abspaltung, Raumtemperatur, Rühren	143
108—109	Minute, Synthese	142
110	Anteil	141
111—112	Verlauf, Übereinstimmung	140
113—114	Parameter, Symmetrie	139
115	Effekt	137
116	Zusammenhang	136
117—119	Gehalt, Strahlung, Verfahren	132
120	Maximum	131
121	Angabe	130
122—123	Annahme, Torr	129
124	Schwefel	127
125	Leitfähigkeit	125

Таблица 1 (продолжение)

i	Словоформы	F
126	Datum	122
127—128	Gegensatz, Stufe	121
129	Modell	120
130	Einfluß	119
131—136	Abweichung, Diskussion, Richtung, Farbe Niederschlag, Pulver	118
137	Wasserstoff	117
138—140	Filtrat, Frequenz, Oberfläche	114
141—144	Absorptionsbanbe, Auftreten, Methanol, Zugabe	113
145	Gleichgewicht	111
146—147	Metall, Schmelzpunkt	109
148—150	Aktivität, Hilfe, Überschuß	107
151—152	Aufspaltung, Mischung	106
153—156	Diagramm, Dichte, Strom, Tetrahydrofuran	105
157—158	Autor, Möglichkeit	104
159—161	Äthanol, Dichloräthan, Verwendung	103
162—164	Abbau, Löslichkeit, Hexan	102
165	Anordnung	101
166—168	Berechnung, Größe, Zustand	100
169	Zahl	98
170—171	Stelle, Tag	99
172	Druck	97
173—177	Aceton, Beziehung, Komponent, Stabilität, Wechselwirkung	96
178—179	Elementarzelle, Jod	94
180	Spaltung	93
181—183	Fluorid, Frage, Koordinationszahl	92
184—185	Auswertung, Chrom	91
186	Zerfall	89
187	Infrarotspektrum	88
188—190	Ammoniak, Beobachtung, Teilchen	87
191	Literatur	86
192	Achse	85
193—197	Apparatur, Funktion, Gegenwart, Ring, Weg	84
198—199	Oxydationszahl, Unterschied	83
200—201	Hochvakuum, Katalysator	82
202—203	Phosphorsäure, Substituent	81
204—205	Kopplung, Modifikation	80
206—207	Änderung, Erwärmen	79
208—209	Essigsäure, Wellenzahl	78
210—211	Gasphase, Reaktionsgemisch	77
212—215	Aussage, Existenz, Umwandlung, Überschuß	76
216—218	Addukt, Einkristall, Phosphorsäure	75
219	Acetonitril	74
220—221	Debeyogramm, Oxydationszone	73
222	Reaktionszeit	72
223—230	Ammonolyse, Analogie, Bodenkörper, Brom, Carbonat, Element, Phosphin, Rest	71
231—237	Abkühlen, Einwirkung, Elektron, Flüssigkeit, Kette, Veränderung, Zusatz	70
238—241	Gas, Kupfer, Röntgendiagramm, Stehen	69
242—245	Befund, Kristallstruktur, Nickelhydroxid Rückfluß	68
246—247	Hinweis, Oxydationsstufe	65
248	Austausch	64

Таблица 1 (продолжение)

i	Словоформы	F
249—255	Chlorwasserstoff, Entwicklung, Molekulargewicht, Pyridin, Rechnung, Tatsache	63
256—257	Maß, Trocken	62
258—259	Eisen, Pentan	61
260—261	Seite, Titration	60
262—267	Fällung, Gerät, Konstante, Länge, Uran, Verunreinigung	59
268—276	Aufbau, Diacetylmonoxin, Charakter, Einklang, Molybdän, Oktaeder, Phosphoratom, Schwefelsäure, Vorliegen	58
277—280	Differenz, Disproportionierung, Gruppierung, Temperaturbereich	57
281—285	Deformationsschwingung, Destillation, Folge, Pulverdiagramm, Rohr	56
286—289	Abschnitt, Elektronenspektrum, Nitrat, Trifluorophosphin	55
290—295	Art, Atmosphäre, Derivat, Formeleinheit, Nickel, Reaktionstemperatur	54
296—299	Herstellung, Inhaltsübersicht, Koordination, Wirkung	53
300—307	Berücksichtigung, Betrachtung, Diäthyläther, Fraktion, Lithium, Punktlage, Sauerstoffatom, Tetra	52
308—315	Amin, Anzahl, Argon, Erhöhung, Kopplungskonstante, Schwierigkeit, Umkristallisieren, Wolfram	51
316—319	Kraftkonstante, Molgewicht, Schulter, Zwischenprodukt	50
320—326	Addition, Ausgangsprodukt, Ladung, Mitteilung, Mittelwert, Position, Temporn	49
327—332	Abscheidung, Chlor, Dublett, Interferenz, Natrium, Verfügung	48
333—342	Alkohol, Anwesenheit, Blei, Chelat, Ebene, Ende, Energie, Pulveraufnahme, Stellung, Stoff	47
343—346	Dissoziation, Gewicht, Reaktionsbedingung, Salzsäure	46
347—354	Abnahme, Entfernung, Erfahrung, Festkörper, Feuchtigkeit, Geschwindigkeit, Halogenid, Winkel	45
355—366	Absorptionsspektrum, Calcit, Einwaage, Faktor, Gebiet, Kalium, Kobalt, Kolben, Rückflußkühler, Triplett, Zelle, Zustandsdiagramm	44
367—373	Aktivierungsenergie, Antimon, Gemenge, Masse, Molekül, Substitution, Triphenylphosphinoxid	43
374—383	Anwendung, Ausgangsverwendung, Einbau, Gerade, Grenze, Mechanismus, Ofen, Reaktionslösung, Suspension, Verteilung	42
384—393	Ausbildung, Deutung, Mischkristall, Natronlauge, Orbital, Potenzial, Projektion, Punkt, Schluß, Sinn	41
394—403	Fehlordnung, Fritte, Hand, Isolierung, Isotherme, Molvolumen, Nachweis, Resonanz, Rolle, Summe	40
404—415	Anstieg, Auflösung, Brücke, Doppelbindung, Einleitung, Feld, Kontakt, Kühlung, Ursache, Verknüpfung, Vorgang, Zwischenstufe	39
416—428	Ansatz, Ausgangssubstanz, Bedeutung, Einheit, Endprodukt, Fehler, Interesse, Legierung, Phasengrenze, Reinigung, Sicherheit, Sulfat, Valenzschwingsbande	38
429—438	Atomabstand, Ausnahme, Beschreibung, Cyclohexan, Hälfte, Kohlenstoff, Ordnung, Petroläther, Sulfid, Transport	37
439—444	Alkalimetallion, Indizierung, Kernresonanzspektrum, Konstitution, Strukturbestimmung, Thallium	36

Таблица 1 (продолжение)

i	Словоформы	F
445—457	Ähnlichkeit, Atomverhältnis, Citronensäure, Einzelheit, Jodzahl, Medium, Nacht, Ölpumpenvakuum, Patrialdruck, Pyrrol, Reaktionsgeschwindigkeit, Reduktionsstufe, Salpetersäure	35
458—468	Amid, Ausgangsmaterial, Auslöschung, Ausschuß, Dissoziationskonstante, Komplexverbindung, Metallatom, Moment, Umsatz, Widerstand, Zinkboranat	34
469—483	Absorption, Aktivierung, Base, Betracht, Chloroform, Fluor, Halbstenpotential, Phosphid, Reflexionsspektrum, Röhrchen, Schwingungsspektrum, Sieden, Spinell, Überlegung, Voraussetzung	33
484—497	Abgabe, Alkalimetall, Ammoniakat, Beginn, Chromacetat, Hochtemperaturphase, Metallion, Mutterlauge, Nebenreaktion, Schritt, Silylphosphin, Spalte, Temperaturfaktor, Tetrachlorkohlenstoff, Verfeinerung	32
498—517	Abtrennung, Anlagerung, Bariumsulfat, Behandlung, Chemie, Extinktion, Gefäß Genauigkeit, Kapillare, Keal, Kristallgitter, Magnesium, Mittel, Öl, Siedepunkt, Stickstoffatom, Wellenlänge, Zentrum, Zone	31
518—532	Donator, Feuchtigkeitsschub, Glas, Höhe, Komplexbildung, Kondensation, Reinheit, Spur, Strukturvorschlag, Tetraederlücke, Trennung, Tropftrichter, Vorkühlen, Wassergehalt, Zunahme	30
533—550	Alkaliion, Besetzung, Bombenrohr, Dioxan, Dipolmoment, Durchführung, Ermittlung, Ersatz, Größenordnung, Isomer, Meßpunkt, Schema, Silizium, Solvat, Tetraeder, Triazin, Volumen, Wasserstrahlvakuum	29
551—569	Barium, Beendigung, Bindungsabstand, Chloridion, Durchmesser, Erklärung, Konfiguration, Kristallwasser, Küvette, Literaturangabe, Palladium, Perchlorat, Phenylphosphor, Resultat, Schutzgas, Strukturfaktor, Strukturtyp, Wärme	27
570—595	Argonit, Betrag, Charakterisierung, Enthalpie, Erscheinung, Fläche, Formulierung, Hydrid, Hydrolysat, Konzentrationsbereich, Kristallisation, Mangan, Multiplert, Oktaederlücke, Prisma, Rahmen, Reaktionsgleichung, Schiffchen, Summenformel, Threonin, Tieftemperaturphase, Überführungszahl, Umkristallisieren, Verzerrung, Zahl, Ziel	26
596—616	Analysenergebnis, Beweis, Bor, Hydrolyseprodukt, Intensitätsverhältnis, Leitfähigkeitsmessung, Molekulargewichtsbestimmung, Nadel, Oktaederplatz, Platin, Prinzip, Quecksilber, Radius, Reflexintensität, Schmelzwärme, Standard, Überleiten, Ultramarin, Vorschrift, Zentralatom, Zink	25
617—649	Analysenwert, Aufarbeitung, Beitrag, Benzolsäure, Bindungsverhältnis, Carbonation, Elektrolyse, Elektronendichte, Fehlergrenze, Filtration, Gitterschwingung, Hydrat, Hydrazin, Hydrobromid, Ionenradius, Lauf, Material, Mitarbeiter, Nujoll, Preßling, Regel, Schmelze, Störung, Tetrachloräthan, Tiegel, Toluol, Triäthylamin, Trockung, Umgebung, Versuchsergebnis, Vorstellung, Wahrscheinlichkeit, Zersetzungspunkt	25

Таблица 1 (продолжение)

i	Словоформы	F
650—672	Ablauf, Bromid, Dampf, Drehhalskolben, Drehkristallaufnahme, Elektrofluorierung, Entstehung, Gasentwicklung, Gewinnung, Gleichgewichtsdruck, Kennnis, Linienbreite, Näherung, Problem, Reaktionsgefäß, Reaktionspartner, Reihenfolge, Rührer, Selen, Solvolyse, Stabilisierung, Stadium, Suszeptibilität	24
673—690	Beweglichkeit, Bindungsenergie, Chlorwasserstoffkonzentration, Eiswasser, Elektronenpaar, Fluorierung, Hexacycloheptadistann, Identifizierung, Kugelpackung, Lanthanidchlorid, Meßergebnis, Meßwert, Methylgruppe, Phenol, Phosphat, Spektrophotometer, Verdünnung, Versuchsbedingung	23
691—720	Ammoniumsalz, Angriff, Arsin, Auffassung, Aufstellung, Autoklav, Bindungsgrad, Carbosilan, Diphosphatriazin, Entführung, Hinblick, Lösen, Mischkristallphase, Nachbar, Nähe, Phenylgruppe, Punktgruppe, Reaktionsablauf, Röntgenaufnahme, Schmelztemperaturphase, Schwerpunkt, Singulett, Standardabweichung, Strukturformel, Sublimation, Sulfurychlorid, Variation, Verlust, Vermutung, Wolframtransport	22
721—749	Additionsverbindung, Absorptionsmaximas, Ausdruck, Bariumsulfat, Einschlußrohr, Empfindlichkeit, Erniedrigung, Existenzbereich, Germanium, Hahn, Halogen, Jodid, Kälte, Klammer, Kochen, Kombination, Länge, Luftsauerstoff, Nebengruppe, Nebenprodukt, Phasenbreite, Resonanzsignal, Schwefeldioxid, Stärke, Strukturmodell, Tropfen, Urannitrid, Wasserstoffbrücke, Zwischenschicht	21
750—779	Alkali, Alkoholat, Bromid, Diffusionsgrenzstrom, Disilacyclopentan, Elektronegativität, Entwässerung, Erfolg, Harnstoff, Hauptgruppe, Koordinate, Korrektur, Massenspektrum, Minimum, Phosphortrichlorid, Präparation, Quarzrohr, Ringsystem, Schlußfolgerung, Schwefeltrioxid, Spektrometer, Stabilitätskonstante, Temperaturabhängigkeit, Temperaturgefälle, Trockenrohr, Übergangsmetall, Umlagerung, Variante, Zinn, Zwischengitterplatz	20

Таблица 2

Распределение количества словоформ при $F \leq 20$

i	m	F	i	m	F
780—812	33	19	1429—1556	128	9
813—854	42	18	1557—1798	152	8
855—907	53	17	1709—1905	197	7
908—961	54	16	1906—2175	270	6
962—1008	47	15	2176—2526	351	5
1009—1072	64	14	2527—3083	557	4
1073—1150	78	13	3084—3959	876	3
1151—1234	84	12	3960—5800	1841	2
1235—1323	89	11	5801—10303	4503	1
1324—1428	105	10			

С. А. Шубик

О РАЗМЕРАХ ПРЕДЛОЖЕНИЯ В СОВРЕМЕННОМ НЕМЕЦКОМ ЯЗЫКЕ

Распределение размеров предложения обычно не совпадает в различных функциональных стилях; в пределах одного функционального стиля оно может не совпадать в различные исторические эпохи.

В немецком языке XX в. объем предложения стал значительно меньше по сравнению с XIX, XVIII и особенно XVII ст. Так, согласно выборочным обследованиям, средний размер предложения в художественной прозе XVII в.¹ составляет 36,3 слова,² в XVIII в. — 26,2,³ в первой половине XIX в. — 30,3,⁴ во второй половине XIX в. — 23,4,⁵ в XX в. — 19,3.⁶

¹ Термином «художественная проза» здесь и далее обозначается только сплошное (авторское) повествование.

² При исчислении этой величины учтены выборочные средние у следующих авторов: И. Шупп — 24,7; Х. Рейтер — 31,1; Х. Вейзе — 34,6; Г. Циглер — 37,5; Д. Лознштейн — 40,1; Г. Гриммельсхаузен — 49,7. Данные показатели взяты из кн.: В. Г. Адмони. Развитие структуры предложения в период формирования немецкого национального языка. Л., 1966, стр. 192, 193.

³ Средние по авторам: Х. Геллерт — 15,2, И. Юнг-Штиллинг — 19,5, И. Хермес — 27,0, И. В. Гете («Театральное призвание Вильгельма Мейстера») — 28,0, И. Вецель — 41,2; см.: В. Г. Адмони, ук. соч., стр. 198, 199, 205; Н. Агенса. *Verborgene Ordnung. Die Beziehungen zwischen Satzlänge und Wortlänge in deutscher Erzählprosa vom Barock bis heute.* Düsseldorf, 1965, S. 54, 55.

⁴ Средние по авторам: О. Людвиг — 24,0, К. Гуцков — 25,7, Л. Тик — 27,0, И. Эйхендорф, А. Дросте-Гюльсгоф — 27,3, Э. Т. А. Гофман — 27,4, И. В. Гете («Избирательное сродство») — 30,8, Г. Гейне — 36,3, Г. Клейст — 47,0; см.: В. Г. Адмони, ук. соч., стр. 205; Н. Агенса, ук. соч., стр. 54, 55; С. А. Шубик. Размер предложения в немецкой художественной прозе. Сб. статей по методике преподавания иностранных языков и филологии, вып. 4, Л., 1969, стр. 77.

⁵ Средние по авторам: Д. Лилленкрон — 11,4, В. Поленц — 13,5, М. Крөцнер — 17,7, Г. Зудерман — 19,9, Ф. Шпильгаген — 29,2; Т. Фонтане — 30,2 («Штехлин» — 25,1, «Эффи Брист» — 29,6, «Поппеншудль» — 35,7), В. Раабе — 31,7; Г. Келлер — 33,3; см.: Н. Агенса, ук. соч., стр. 54; С. А. Шубик, ук. соч., стр. 77.

⁶ См.: Н. Меиер. *Deutsche Sprachstatistik.* Hildesheim, 1964, S. 191.

Аналогичным образом происходило развитие размеров предложения и в научной литературе. По имеющимся данным, средний размер предложения в XIX в. равнялся в общественнонаучной прозе 34,3 слова, в естественнонаучной — 28,5;⁷ между тем в XX в. он опустился в общественнонаучной прозе, представленной преимущественно философскими текстами, до 27,8,⁸ а в естественнонаучной — до 20,0.⁹ Сокращение объема предложения наблюдается и в научно-популярной и в научно-художественной речи.¹⁰

Ниже излагаются полученные нами результаты изучения размеров предложения в немецкой художественной и научно-популярной прозе 50—60-х гг. нашего столетия. Художественный и научно-популярный стили широко распространены в общественно-речевой практике и во многом определяют специфику стилистической системы языка в целом.

Наш материал по художественному стилю охватывает 10 тыс. элементарных предложений (ЭП),¹¹ отобранных из произведений двадцати пяти писателей; из одного или двух произведений каждого из них взято 400 ЭП.¹² Материал по научно-популярному стилю состоит из 5 тыс. ЭП, отобранных из сочинений двадцати пяти ученых; здесь на долю каждого автора приходится 200 ЭП.¹³ В выборке по художественной литературе насчитывается 5212 предложений, по научно-популярной — 2673. Из научно-популярных текстов мы взяли меньше материала для обследования, поскольку эти тексты менее разнообразны в языковом отношении.

В художественном стиле современного немецкого языка предложение в общем короче, чем в научно-популярном. Средний размер предложения у писателей — 17,7 слова,¹⁴ у популяризаторов науки — 21,5.¹⁵ Резкие различия наблюдаются в употреблении

⁷ См.: E. Beneš. Syntaktische Besonderheiten der deutschen wissenschaftlichen Fachsprache. In: *Deutsch als Fremdsprache*, 1966, Hft. 3, S. 27—28.

⁸ См.: Н. Меиер, ук. соч., стр. 186, 191.

⁹ См.: E. Beneš, ук. соч., стр. 27—28.

¹⁰ См.: Н. Eggers. Sprache und Gesellschaft im zwanzigsten Jahrhundert. In: *Deutsche Akademie für Sprache und Dichtung, Darmstadt, Jahrbuch 1969.* Heidelberg—Darmstadt, 1970, S. 30.

¹¹ Элементарным считается простое предложение как в самостоятельном употреблении, так и в составе сложного предложения.

¹² Перечень обследованных произведений см. в конце статьи.

¹³ Перечень обследованных сочинений см. в конце статьи.

¹⁴ Средние по авторам: Э. Штригматтер — 7,9, Э. М. Ремарк — 9,9, Б. Апиц, В. Бредель — 11,6, Д. Нолль — 11,8, М. Грюн — 12,6; А. Зегерс — 12,9, Ф. Фюман — 13,0, Г. Нахбар — 13,2, Э. Клаудиус — 14,6, К.-Х. Якобс — 15,0, С. Андрес — 16,7, Г. Гайзер — 17,3, В. Шнурре — 17,4, Л. Фейхтвангер — 17,7, И. Бобровски — 19,3, П. Вейс — 19,5, Л. Франк — 20,3, М. Вальзер — 20,4, Г. Кант — 20,5, А. Андерш — 21,7, А. Цвейг — 25,0, Г. Бёльль — 27,1, Г. Грасс — 31,2, Т. Манн — 34,6.

¹⁵ Средние по авторам: Х.-Г. Фогт — 14,6, О. Рюле — 15,6, Х. Ливинер — 16,3, В. Будденброк — 16,8, П. Куххенн — 17,6, Ф. Альтгейм —

коротких предложений (1—9 слов), с одной стороны, и длинных (16 слов и более) — с другой. Доля коротких предложений в художественной прозе (34.6%) в три раза больше, чем в научно-популярной (11.4%). Чем короче предложение, тем отчетливее стиливая дифференциация: предложения от 1 до 3 слов составляют в художественных текстах 4.6%, в научно-популярных — 0.3,¹⁶ от 4 до 6 слов — 15.1 и 3.6% соответственно, от 7 до 9 слов — 14.9 и 7.5%.

Что касается отдельных авторов, то у писателей доля коротких предложений обычно не опускается ниже 18%, причем у многих из них она превышает 36%, между тем у популяризаторов процент таких предложений, как правило, не достигает 18, опускаясь нередко ниже 9 (см. табл. 1).

Таблица 1

Доля коротких предложений (1—9 слов) у писателей и у популяризаторов науки

Доля в %	Число (абсолютное и в %)			
	писатели		популяризаторы	
До 8.9	—	—	8	32%
9—17.9	3	12%	16	64%
18—26.9	4	16%	—	—
27—35.9	7	28%	1	4%
36—44.9	5	20%	—	—
45 и более	6	24%	—	—
Всего	25	100%	25	100%

Данные по авторам в %.

Писатели: Т. Мани — 9.0, Г. Грасс — 13.4, А. Цвейг — 14.0, П. Вейс — 19.8, Г. Каят — 21.3, Г. Бёльль — 21.8, Л. Фейхтгаангер — 24.9, С. Андрес — 28.5, А. Андерш — 28.8, Л. Франк — 29.0, Г. Гайзер — 30.3, К.-Х. Якобс — 31.1, В. Шпурре — 33.2, А. Зегерс — 35.8, М. Вальзер — 36.0, Э. Клаудиус — 37.7, Г. Нахбар — 38.8, И. Бобровски — 43.8, Ф. Фюман — 44.2, В. Аппц — 45.7, В. Бредель — 46.2, М. Грюн — 47.5, Д. Нолль — 52.4, Э. М. Ремарк — 64.0, Э. Штриматтер — 68.0.

Популяризаторы науки: П. Хофштеттер — 2.6, Д. Ваттенберг, Г. Штейгер — 4.3, Г. Рёмп — 4.9, Х. Рунге — 5.5, Э. Хильшер — 6.8, К. Бюлов — 7.7, В. Вестфаль — 8.7, Г. Хартман — 9.0, Р. Борман — 10.2, В. Кайзер — 10.8, Ф. Маттес — 11.2, Э. Круг — 11.6, Ф. Альтгейм, Х. Швенкель — 11.7, Х. Линднер — 12.4, Э. Грасси — 12.8, П. Кунхенн — 13.3, К. Берингер — 14.5, В. Равальд — 15.1, Е. Липс — 15.7, В. Будденброк — 15.9, О. Рюле — 16.5, В. Грейлинг — 17.8, Х.-Г. Фогт — 30.4.

17.7, Э. Круг — 18.9, Р. Борман, В. Равальд — 19.8, К. Берингер — 20.0, В. Грейлинг — 20.6, Э. Грасси — 21.0, Ф. Маттес — 21.4, Х. Рунге — 22.2, Г. Рёмп — 22.8, Э. Хильшер — 23.3, В. Вестфаль — 23.4, Г. Хартман — 23.7, Х. Швенкель — 23.9, Д. Ваттенберг — 24.8, В. Кайзер — 24.9, Е. Липс — 25.6, К. Бюлов — 26.0, П. Хофштеттер — 26.5, Г. Штейгер — 29.6.

¹⁶ В двадцати выборках популяризаторов такие предложения не встретились ни разу, в остальных пяти выборках их доля равняется 0.9 (К. Берингер — 1.0 (В. Равальд), 1.1 (Э. «Грасси, Ф. Маттес) и 3.5 (Х.-Г. Фогт). Среди выборки у писателей указанные предложения не зафиксированы только в одной (Г. Грасс), у десяти авторов их доля колеблется от 5.6 до 9.9% (Г. Бёльль — 5.6; М. Грюн — 6.3; М. Вальзер — 7.2; Э. М. Ремарк — 7.7; Д. Нолль — 8.0; В. Аппц — 8.7; Ф. Фюман — 8.8; И. Бобровски — 9.0; Э. Штриматтер — 9.5; Э. Клаудиус — 9.9).

Длинные предложения (16—50 слов) значительно преобладают в научно-популярном стиле: их доля здесь 62.3%, в художественном стиле — 37.7. Предложения от 16 до 24 слов составляют у популяризаторов 33.5%, у писателей 20.9, от 25 до 50 слов — 28.8 и 16.8% соответственно.¹⁷

Различия в удельном весе коротких и длинных предложений в художественном и научно-популярном стиле являются несомненно дифференциальным признаком этих стилей.

Предложения среднего размера (10—15 слов) распространены в исследуемых стилях примерно одинаково: художественный стиль — 23.2%,¹⁸ научно-популярный — 23.9. Удельный вес данных предложений представляется возможным рассматривать в качестве нейтрального признака художественного и научно-популярного стилей. Правда, предложения от 10 до 12 слов употребляются писателями несколько чаще, а предложения от 13 до 15 слов несколько реже, чем популяризаторами, но расхождения здесь невелики (12.9 против 10.5% в первом случае и 10.3 против 13.4% — во втором); невелика к тому же и разница между размерами предложений. Существенно, с другой стороны, то обстоятельство, что распределения авторов по количеству используемых предложений среднего размера весьма близки (см. табл. 2).

Объем предложения непосредственно определяется двумя величинами: комплексностью, т. е. числом составляющих его элементарных предложений, и размерами этих последних.

Наблюдающиеся расхождения в размерах предложения между художественным и научно-популярным стилями обусловлены в основном различиями в размерах ЭП. Средний размер ЭП в художественной прозе — 8.6 слова,¹⁹ в научно-популярной;

гер), 1.0 (В. Равальд), 1.1 (Э. «Грасси, Ф. Маттес) и 3.5 (Х.-Г. Фогт). Среди выборки у писателей указанные предложения не зафиксированы только в одной (Г. Грасс), у десяти авторов их доля колеблется от 5.6 до 9.9% (Г. Бёльль — 5.6; М. Грюн — 6.3; М. Вальзер — 7.2; Э. М. Ремарк — 7.7; Д. Нолль — 8.0; В. Аппц — 8.7; Ф. Фюман — 8.8; И. Бобровски — 9.0; Э. Штриматтер — 9.5; Э. Клаудиус — 9.9).

¹⁷ Сверхдлинные предложения (более 50 слов) в целом несвойственны как научно-популярному, так и художественному стилю. В выборках восемнадцати популяризаторов и пятнадцати писателей такие предложения либо вовсе не встретились, либо их встречаемость оказалась менее 3%. Заметную роль сверхдлинные предложения играют лишь у отдельных авторов, в частности среди писателей — у Т. Мани (21.4%), Г. Бёльля (16.2%), Г. Грасса (15.3%), И. Бобровски (9.6%), М. Вальзера (8.3%), А. Андерша (7.5%); среди популяризаторов — у Е. Липс (11.2%), Г. Штейгера (9.4%), В. Кайзера — (7.1%). Три самых длинных предложения насчитывают в художественных текстах 208, 174 и 166 слов, в научно-популярных — 125, 86 и 81.

¹⁸ Почти такова же доля предложений среднего размера и в материале по художественной прозе у Х. Мейера — 22.6% (см.: Н. Мейер, ук. соч., стр. 186).

¹⁹ Средние по авторам: Э. Штриматтер — 6.9; М. Грюн, Э. М. Ремарк, Ф. Фюман — 7.0; К.-Х. Якобс — 7.3; Г. Гайзер — 7.6; Д. Нолль — 7.7;

Таблица 2

Доля предложений среднего размера (10—15 слов)
у писателей и у популяризаторов науки

Доля в %	Число (абсолютное и в %)			
	писатели		популяризаторы	
До 13.9	2	8%	4	16%
14—17.9	2	8%	—	—
18—21.9	6	24%	6	24%
22—25.9	8	32%	8	32%
26—29.9	3	12%	2	8%
30 и более	4	16%	5	20%
Всего	25	100%	25	100%

Данные по авторам в %.

Писатели: Т. Манн — 8.2; Г. Грасс — 13.3; Г. Кант — 14.4; И. Бобровски — 16.8; Л. Франк — 19.3; А. Цвейг — 19.5; Э. Клаудиус — 20.2; Д. Нолль — 21.5; Г. Гайзер — 21.7; Г. Белль — 21.8; А. Андерш, Э. Штриматтер — 22.5; М. Грюн — 22.9; М. Вальзер, Л. Фейхтвангер — 23.9; Э. М. Ремарк — 24.0; П. Вейс — 25.2; В. Шнурре — 25.7; Ф. Фюман — 27.0; Б. Алиц — 27.3; С. Андрес — 28.0; В. Бредель, Г. Нахбар — 30.6; К.-Х. Якобс — 33.7; А. Зегерс — 35.8.

Популяризаторы науки: П. Хофштеттер — 10.4; Г. Штейгер — 11.1; К. Бюлов — 11.5; Е. Липс — 12.4; В. Равальд — 18.2; Х. Рунге — 18.3; Х. Швенкель — 18.5; В. Кайзер — 19.1; Г. Хартман — 19.6; Д. Ваттенберг — 21.1; В. Грейлинг — 22.1; Ф. Маттес — 22.5; В. Вестфаль — 24.0; Р. Борман — 25.0; Э. Хильшер — 25.2; Х.-Г. Фогт — 25.3; Э. Грасси — 25.5; Э. Круг — 25.9; К. Берингер — 26.3; Г. Рёмп — 29.4; В. Будденброк — 33.6; Ф. Альтгейм — 34.3; П. Кунхени — 36.0; О. Рюле — 39.3; Х. Линднер — 40.5.

11.4^ю Доля коротких ЭП (1—6 слов) у писателей почти вдвое больше, чем у популяризаторов (42.7 и 21.6% соответственно). Особенно резко исследуемые стили противостоят друг другу по числу ЭП от 1 до 3 слов — их доля в художественных текстах (9.2%) приблизительно в три раза выше по сравнению с научно-популярными (3.3%).

Писатели употребляют обычно от 30 до 60% коротких ЭП, популяризаторы — менее 30 (см. табл. 3).

В. Бредель, Г. Нахбар — 7.8; А. Зегерс — 8.0; И. Бобровски — 8.1; Г. Кант, Э. Клаудиус — 8.2; С. Андрес, Б. Алиц — 8.4; Г. Белль — 8.6; А. Андерш, В. Шнурре — 8.7; М. Вальзер — 9.2; П. Вейс, Л. Фейхтвангер — 9.3; Л. Франк — 10.3; А. Цвейг — 11.6; Г. Грасс — 11.9; Т. Манн — 12.6.

^ю Средние по авторам: Х.-Г. Фогт — 8.5; Ф. Альтгейм — 9.1; В. Будденброк, Ф. Маттес — 9.5; Э. Грасси, Х. Линднер, В. Равальд — 9.9; К. Бюлов — 10.2; П. Хофштеттер — 10.3; В. Кайзер — 10.5; Э. Круг — 10.7; К. Берингер — 11.0; П. Кунхени, О. Рюле — 11.3; Е. Липс — 11.4; Г. Рёмп — 11.6; Д. Ваттенберг — 11.8; В. Грейлинг — 12.0; Х. Рунге — 12.2; В. Вестфаль — 12.3; Х. Швенкель — 12.5; Р. Борман — 13.0; Э. Хильшер — 13.9; Г. Хартман — 14.5; Г. Штейгер — 17.4.

Таблица 3

Доля коротких элементарных предложений (1—6 слов)
у писателей и у популяризаторов науки

Доля в %	Число (абсолютное и в %)			
	писатели		популяризаторы	
До 9.9	—	—	1	4%
10—19.9	—	—	10	40%
20—29.9	2	8%	11	44%
30—39.9	7	28%	3	12%
40—49.9	10	40%	—	—
50 и более	6	24%	—	—
Всего	25	100%	25	100%

Данные по авторам в %.

Писатели: А. Цвейг — 25.8; Т. Манн — 29.5; П. Вейс, Г. Грасс — 30.8; Л. Фейхтвангер — 31.5; М. Вальзер, Л. Франк — 36.0; С. Андрес — 39.0; Б. Алиц — 39.5; А. Андерш — 40.8; Г. Белль — 41.5; А. Зегерс — 43.5; В. Шнурре — 44.0; В. Бредель — 44.3; Г. Нахбар — 44.8; Э. Клаудиус — 45.0; Г. Кант — 45.5; Г. Гайзер — 48.3; Д. Нолль — 48.5; И. Бобровски — 50.8; К.-Х. Якобс — 51.5; М. Грюн — 51.8; Э. Штриматтер — 54.8; Э. М. Ремарк — 56.8; Ф. Фюман — 57.3.

Популяризаторы науки: Г. Штейгер — 9.5; Э. Хильшер — 10.0; Г. Хартман — 12.5; Р. Борман — 14.5; Х. Рунге — 15.0; Г. Рёмп, О. Рюле — 16.0; Д. Ваттенберг — 16.5; В. Вестфаль, Х. Швенкель — 17.5; К. Берингер — 19.5; П. Кунхени — 20.0; В. Грейлинг, Э. Круг — 20.5; Е. Липс — 21.5; П. Хофштеттер — 24.0; Х. Линднер — 25.0; В. Кайзер — 27.0; К. Бюлов — 27.5; В. Будденброк, Ф. Маттес — 28.0; В. Равальд — 28.5; Э. Грасси — 30.0; Ф. Альтгейм — 36.0; Х.-Г. Фогт — 38.5.

Совершенно иначе обстоит дело с длинными ЭП (11—30 слов): они составляют в художественных текстах 24.0%, в научно-популярных — 44.4 (ЭП от 11 до 20 слов — 21.5 и 37.8% соответственно, от 21 до 30 слов — 2.5 и 6.6%).²¹

По количеству ЭП среднего размера (7—10 слов) исследуемые стили расходятся незначительно. На их долю в художественной прозе приходится 32.2%, в научно-популярной — 32.4 (семи- и восьмисловные ЭП составляют 19.4 и 16.9% соответственно, девяти- и десятисловные — 12.8 и 15.5%). У подавляющего большинства как писателей, так и популяризаторов процент данных ЭП заключен в узком интервале от 30 до 38 (см. табл. 4).

Итак, удельный вес коротких и длинных элементарных предложений в художественном и научно-популярном стилях можно

²¹ Сверхдлинные ЭП (более 30 слов) употребляются крайне редко. У двадцати двух писателей и у двадцати одного популяризатора они либо вовсе не зафиксированы, либо представлены единичными случаями (менее 3%). Более 5% сверхдлинных ЭП мы насчитали у двух писателей (Т. Манн — 6.3, Г. Грасс — 5.3) и одного популяризатора (Г. Штейгер — 11.5). Три самых длинных ЭП содержат в художественных текстах 65, 64 и 63 слова, в научно-популярных — 74 и дважды по 63 слова.

Таблица 4

Доля элементарных предложений среднего размера (7—10 слов)
у писателей и у популяризаторов науки

Доля в %	Число (абсолютное и в %)			
	писатели		популяризаторы	
До 21.9	—	—	1	4%
22—25.9	1	4%	2	8%
26—29.9	6	24%	2	8%
30—33.9	9	36%	9	36%
34—37.9	8	32%	10	40%
38 и более	1	4%	1	4%
Всего	25	100%	25	100%

Данные по авторам в %.

Писатели: Т. Манн — 25.3; И. Бобровский — 27.8; Г. Грасс — 28.3; Ф. Фюман — 28.5; Э. Клаудиус — 29.3; Э. М. Ремарк — 29.5; А. Цвейг — 29.8; Л. Франк — 30.0; А. Андерш — 30.5; Э. Штригматтер — 30.8; Г. Бёль — 31.0; Д. Нолль — 31.5; В. Шнурре — 32.0; Г. Кант — 32.5; К.-Х. Якобс — 32.8; Б. Апиц — 33.5; М. Грюн — 34.0; М. Вальзер, А. Зегерс — 34.5; Г. Гайзер — 35.3; Г. Нахбар — 35.8; В. Бредель — 36.0; Л. Фейхтвангер — 36.5; П. Вейс — 37.5; С. Андрес — 38.0.

Популяризаторы науки: Г. Штейгер — 18.0; Э. Хильшер — 23.5; Г. Хартман — 25.0; Р. Борман — 26.0; В. Вестфаль — 26.5; Д. Ваттенберг — 31.0; Э. Грасси — 31.5; Э. Круг, Х. Швенкель — 32.0; В. Грейлинг — 32.5; В. Кайзер, П. Кунхенн, Г. Рёмп, П. Хофштеттер — 33.5; Е. Липс, Х. Рунге, О. Рюле — 34.0; Ф. Альтгейм, Х.-Г. Фогт — 34.5; К. Бюлов — 35.0; В. Равальд — 36.0; К. Берингер — 36.5; Х. Линднер, Ф. Маттес — 37.0; В. Будденброк — 40.0.

отнести к числу дифференциальных признаков этих стилей, а удельный вес элементарных предложений среднего размера — к числу их нейтральных признаков.

Комплексность предложения в отличие от размеров ЭП обнаруживает в художественном и научно-популярном стилях большое сходство. Средняя комплексность у писателей — 2.0 ЭП,²² у популяризаторов — 1.9.²³

²² Данные по авторам: Э. Штригматтер — 1.2; Б. Апиц, Э. М. Ремарк — 1.4; В. Бредель, Д. Нолль — 1.5; А. Зегерс — 1.6; Г. Нахбар — 1.7; М. Грюн, Э. Клаудиус — 1.8; Л. Фейхтвангер, Ф. Фюман — 1.9; С. Андрес, Л. Франк, В. Шнурре, К.-Х. Якобс — 2.0; П. Вейс — 2.1; М. Вальзер, А. Цвейг — 2.2; Г. Гайзер — 2.3; И. Бобровский — 2.4; А. Андерш, Г. Кант — 2.5; Г. Грасс — 2.6; Т. Манн — 2.8; Г. Бёль — 3.1.

²³ Данные по авторам: О. Рюле — 1.4; Р. Борман — 1.5; П. Кунхенн, Г. Хартман — 1.6; В. Грейлинг, Х. Линднер, Х.-Г. Фогт, Э. Хильшер, Г. Штейгер — 1.7; К. Берингер, В. Будденброк, Э. Круг, Х. Рунге — 1.8; В. Вестфаль, Х. Швенкель — 1.9; Ф. Альтгейм, В. Равальд, Г. Рёмп — 2.0; Д. Ваттенберг, Э. Грасси — 2.1; Е. Липс, Ф. Маттес — 2.2; В. Кайзер — 2.4; К. Бюлов, П. Хофштеттер — 2.5.

В обоих стилях ведущее положение занимают простые предложения, т. е. структуры, состоящие из одного ЭП. Их численность в художественных текстах — 48.5%, в научно-популярных — 42.8.²⁴ Писатели в подавляющем большинстве используют от 30 до 60% простых предложений, популяризаторы — от 30 до 50 (см. табл. 5).

Таблица 5

Доля простых предложений у писателей и у популяризаторов науки

Доля в %	Число (абсолютное и в %)			
	писатели		популяризаторы	
До 29.9	1	4%	3	12%
30—39.9	5	20%	9	36%
40—49.9	10	40%	8	32%
50—59.9	4	16%	2	8%
60—69.9	2	8%	3	12%
70 и более	3	12%	—	—
Всего	25	100%	25	100%

Данные по авторам в %.

Писатели: Т. Манн — 21.4; Г. Кант — 31.3; Г. Бёль — 31.5; А. Цвейг — 33.5; Г. Гайзер, Г. Грасс — 36.0; А. Андерш — 40.0; П. Вейс — 41.4; С. Андрес — 43.5; В. Шнурре — 44.7; Л. Фейхтвангер — 45.0; К.-Х. Якобс — 45.4; И. Бобровский — 46.1; Л. Франк — 46.3; М. Грюн — 48.9; Ф. Фюман — 49.3; М. Вальзер — 51.1; Г. Нахбар — 54.0; Э. Клаудиус — 54.3; А. Зегерс — 56.1; Д. Нолль — 64.8; В. Бредель — 66.0; Б. Апиц — 70.2; Э. М. Ремарк — 74.2; Э. Штригматтер — 79.7.

Популяризаторы науки: П. Хофштеттер — 18.2; К. Бюлов — 25.6; В. Кайзер — 29.8; Д. Ваттенберг — 31.6; Е. Липс — 32.6; Ф. Альтгейм — 34.3; В. Равальд — 36.4; Ф. Маттес — 37.1; Э. Грасси — 37.2; Г. Рёмп — 38.2; Э. Круг — 39.3; Х. Швенкель — 39.8; В. Вестфаль — 41.3; К. Берингер — 45.5; Х. Рунге — 45.9; В. Будденброк — 46.0; Х. Линднер — 47.1; Х.-Г. Фогт, Г. Штейгер — 48.7; Э. Хильшер — 49.6; Г. Хартман — 50.8; Р. Борман — 55.5; В. Грейлинг — 60.2; П. Кунхенн — 61.7; О. Рюле — 69.0.

На втором месте по частоте употребления стоят двухкомпонентные сложные предложения, т. е. структуры, состоящие из двух ЭП. Но если простые предложения несколько более характерны для художественных текстов, то двухкомпонентные — для научно-популярных. Их употребительность у писателей колеблется, как правило, от 18 до 36%, у популяризаторов — от 30 до 42 (см. табл. 6). Процент двухкомпонентных предложений по художественным текстам в целом равен 27.0, по научно-популярным — 36.1.

²⁴ Наш показатель по научно-популярным текстам хорошо согласуется с соответствующим показателем у Г. Эггерса — 43.5% (употребляемый нами термин «простое предложение» включает две структуры Г. Эггерса — *Einfachsätze* и *Setzungen*). См.: H. Eggert. Zur Syntax der deutschen Sprache der Gegenwart. In: Studium generale, 1962, Hft. 1, S. 53.

Таблица 6

Доля двухкомпонентных сложных предложений у писателей и у популяризаторов науки

Доля в %	Число (абсолютное и в %)			
	писатели		популяризаторы	
До 17.9	1	4%	—	—
18—23.9	5	20%	1	4%
24—29.9	12	48%	2	8%
30—35.9	7	28%	9	36%
36—41.9	—	—	10	40%
42 и более	—	—	3	12%
Всего	25	100%	25	100%

Данные по авторам в %.

Писатели: Э. Штригматтер — 17.5; Э. М. Ремарк — 18.0; И. Бобровски — 19.8; В. Бредель — 21.6; М. Вальзер — 22.2; Б. Апиц — 23.9; Д. Ноэль — 24.9; Т. Манн — 25.5; Э. Клаудиус — 25.6; Г. Грасс — 26.0; Г. Белль — 26.6; В. Шнурре — 27.1; А. Андерш — 27.5; Г. Кант, Л. Франк — 28.1; Г. Нахбар — 28.9; П. Вейс, Ф. Фюман — 29.8; Л. Фейхтвангер — 30.1; А. Зегерс — 30.9; С. Андрес — 31.0; К.-Х. Якобс — 31.1; М. Грюн — 31.8; Г. Гайзер — 33.1; А. Цвейг — 35.1.

Популяризаторы науки: В. Грейлинг — 22.9; О. Рюле — 25.5; П. Кунхенн — 25.8; Ф. Маттес — 31.5; Э. Грасси, Х.-Г. Фогт — 33.0; В. Кайзер — 33.3; Г. Штейгер — 34.2; К. Бюлов — 34.6; Е. Липс — 34.8; В. Будденброк — 35.4; К. Берингер — 35.5; Г. Хартман — 36.9; Э. Хильшер — 37.0; Х. Рунге — 37.6; Х. Швенкель — 38.8; Р. Борман — 39.1; Г. Ремп — 39.2; В. Вестфаль, В. Равальд — 40.4; Ф. Альтхейм — 41.2; Х. Линднер — 41.3; Д. Ваттенберг — 42.1; П. Хофштеттер — 42.9; Э. Круг — 46.4.

Общая совокупность простых и двухкомпонентных предложений в художественном стиле составляет 75.4%, в научно-популярном — 78.9.

Трех- и четырехкомпонентные сложные предложения встречаются в обоих стилях почти с одинаковой частотой: художественный стиль — 18.7%, научно-популярный — 18.4 (доля трехкомпонентных структур — 13.1 и 13.5% соответственно, четырехкомпонентных — 5.6 и 4.9%). Ряды распределения данных предложений по авторам в сущности параллельны (см. табл. 7).

Сложные предложения, состоящие из 5 ЭП и более, употребляются очень редко, особенно в научно-популярном стиле. Здесь на их долю приходится всего лишь 2.7%, в художественном стиле 5.9. У пятнадцати популяризаторов и десяти писателей такие структуры либо отсутствуют, либо их число не достигает 2%. Круг авторов, сравнительно широко использующих данные предложения, невелик: один популяризатор (К. Бюлов — 12.8) и семь писателей (Г. Белль — 21.8, Г. Грасс — 14.7, И. Бобровски — 13.2, Т. Манн — 12.4, А. Андерш — 11.9, М. Вальзер, Г. Кант — 10.6).

Таблица 7

Доля трех- и четырехкомпонентных сложных предложений у писателей и у популяризаторов науки

Доля в %	Число (абсолютное и в %)			
	писатели		популяризаторы	
До 11.9	4	16%	4	16%
12—16.9	4	16%	6	24%
17—21.9	9	36%	7	28%
22—26.9	6	24%	6	24%
27 и более	2	8%	2	8%
Всего	25	100%	25	100%

Данные по авторам в %.

Писатели: Э. Штригматтер — 2.8; Б. Апиц — 4.8; Э. М. Ремарк — 6.0; Д. Ноэль — 8.4; А. Зегерс — 12.2; В. Бредель — 12.3; Г. Нахбар — 15.7; М. Вальзер — 16.1; М. Грюн — 17.5; Ф. Фюман — 17.7; Э. Клаудиус, К.-Х. Якобс — 18.4; Г. Белль — 20.1; С. Андрес — 20.5; А. Андерш — 20.7; И. Бобровски — 21.0; Л. Франк — 21.7; Г. Гайзер — 22.3; П. Вейс, Л. Фейхтвангер — 23.0; Г. Грасс — 23.3; В. Шнурре — 23.6; А. Цвейг — 26.5; Г. Кант — 30.1; Т. Манн — 40.6.

Популяризаторы науки: Р. Борман — 4.7; О. Рюле — 5.5; Х. Линднер — 11.5; П. Кунхенн — 11.7; Г. Хартман — 12.3; Э. Хильшер — 13.5; Э. Круг — 14.3; В. Вестфаль — 14.4; Х. Рунге — 15.6; В. Грейлинг — 16.1; Г. Штейгер — 17.1; К. Берингер — 17.2; В. Будденброк — 17.7; Х.-Г. Фогт — 18.3; Х. Швенкель — 19.4; В. Равальд — 20.2; Г. Ремп — 20.6; Д. Ваттенберг — 23.2; Ф. Альтхейм, Э. Грасси — 24.5; Ф. Маттес — 24.7; Е. Липс — 25.9; К. Бюлов — 26.9; В. Кайзер — 28.6; П. Хофштеттер — 31.2.

В общем, комплексность предложения вносят в дифференциацию художественного и научно-популярного стилей гораздо менее существенный вклад, чем объем элементарного предложения. В основном это объясняется, очевидно, следующими причинами.

В выборе той или иной комплексности как художнику слова, так и популяризатору науки предоставлена большая свобода, поскольку язык богат параллельными способами выражения смысловых отношений, с одной стороны — между ЭП в составе сложного предложения, а с другой — между отдельными предложениями, в том числе и простыми. В связи с этим характерная для художественного и научно-популярного стилей тенденция к сокращению объема предложения легко приводит в области комплексности этих стилей к весьма близким результатам, тем более что, во-первых, указанная тенденция складывается отчасти под действием одних и тех же факторов (экономия языковых средств, проникновение элементов обиходно-разговорной речи) и, во-вторых, синтаксический строй художественной прозы оказывает известное влияние на структуру предложения в научно-популярном стиле.²⁵

²⁵ Из других факторов развития предложения в сторону уменьшения его размеров важнейшим в научно-популярных текстах является установка

Иначе обстоит дело с размерами ЭП. По сравнению с писателями возможности использования небольших ЭП у популяризаторов сильно ограничены. Объективное, точное и непротиворечивое раскрытие научных понятий очень часто требует большой затраты слов, особенно в присубстантивном употреблении. Поэтому сокращение объема предложения в научно-популярном стиле не может происходить в значительных масштабах за счет размеров ЭП, как это происходит в художественной прозе.

ОБСЛЕДОВАННАЯ ЛИТЕРАТУРА

1. Художественные произведения

A. Andersch. Die Rote. Olten-Freiburg/B., 1960; S. Andres. Der Knabe im Brunnen. München, 1966; B. Apitz. Nackt unter Wölfen. Halle/S., 1959; J. Bobrowski. Levins Mühle. Berlin, 1964; H. Böll. Billard um halbzehn. Köln—Berlin, 1960; W. Bredel. Ein neues Kapitel. Berlin, 1960; E. Claudius. Menschen an unserer Seite. Moskau, 1953; L. Feuchtwanger. Jefta und seine Tochter. Berlin, 1957; L. Frank. Links wo das Herz ist. Berlin, 1955; F. Fühmann. Kameraden. In: König Ödipus. Gesammelte Erzählungen. Berlin—Weimar, 1966; G. Gaiser. 1) Gianna aus dem Schatten. In: Einmal und oft. Erzählungen. München, 1956; 2) Gazelle, grün. In: Gazelle, grün. Erzählungen und Aufzeichnungen. München, 1965; G. Grass. Die Blechtrommel. Frankfurt/M.—Hamburg, 1965. M. Grün. Irrlicht und Feuer. Berlin—Weimar, 1967; K.-H. Jakobs. Beschreibung eines Sommers. Berlin, 1967; H. Kant. 1) Das Kennwort; 2) Die Trompete. In: Ein bißchen Südsee. Berlin, 1967; Th. Mann. Die Betrogene. In: Gesammelte Werke, Bd. 9. Berlin—Weimar, 1965; H. Nachbar. Die Hochzeit von Länneken. Berlin—Weimar, 1967; D. Noll. Die Abenteuer des Werner Holt. Roman einer Jugend. Berlin—Weimar, 1965; E. M. Remarque. Zeit zu leben und Zeit zu sterben. Berlin, 1957; A. Seghers. Die Entscheidung. Berlin, 1960; W. Schürre. 1) Ein Fall für Herrn Schmidt, 2) Steppenköpp. In: Ein Fall für Herrn Schmidt. Erzählungen. Stuttgart, 1962; E. Strittmatter. Ole Bienköpp. Berlin—Weimar, 1964; M. Walser. Ehen in Philippsburg. Frankfurt/M., 1958; P. Weiss. Fluchtpunkt. Frankfurt/M., 1965; A. Zweig. Die Zeit ist reif. Berlin—Weimar, 1967.

2. Научно-популярные сочинения

F. Althelm. Der unbesiegte Gott. Heidentum und Christentum. Hamburg, 1957; C. Ch. Beringer. Vulkanismus und andere Tiefenkräfte der Erde. Stuttgart, 1953; R. Bormann. Jugend und Liebe. Leipzig—Jena—Berlin, 1966; W. Buddenbrock. Wie orientieren sich die Tiere?

на доступность изложения для постоянно растущего круга читателей, интересующихся научной картиной мира. В художественных произведениях короткое предложение часто используется для показа крупным планом тех неярких, будничных сторон действительности, которые, как это выясняется при ближайшем рассмотрении, играют большую роль в сцеплении событий. Другой существенный источник их — небывалое стремление к художественному лаконизму: вместо развернутого описания предмета дается характерный штрих, броская деталь, красноречивая подробность. Огромное значение приобрел также подтекст, выразительность которого обычно тем сильнее, чем проще структура текста. К короткому предложению ведут и такие противоположные, но одинаково присущие современной художественной прозе явления, как эмоциональная насыщенность, с одной стороны, и деловая сухость — с другой.

Stuttgart, 1956; K. Bülow. Die Entstehung der Kontinente und Meere. Stuttgart, 1963; E. Grassi. Kunst und Mythos. Hamburg, 1957; W. Greiling. Mehr Brot für mehr Menschen. Stuttgart, 1963; G. M. Hartmann. Der Materialismus in der Philosophie der griechisch-römischen Antike. Berlin, 1959; E. Hilscher. Thomas Mann. Berlin, 1966; P. R. Hofstätter. Gruppendynamik. Die Kritik der Massenpsychologie. Reinbek bei Hamburg, 1962; W. Kayser. Die Wahrheit der Dichter. Reinbek bei Hamburg, 1961; E. Krug. Radioastronomie. Stuttgart, 1962; P. Kunhenn. Pygmäen und andere Primitivvölker. Stuttgart, 1952; H. Lindner. Ströme. Felder. Elektronen. Leipzig—Jena—Berlin, 1967; E. Lips. Das Indianerbuch. Leipzig, 1967; F. Matthes. Elemente, Atome, Isotope. Leipzig—Jena, 1961; W. Rawald. Pflanze, Umwelt und Natur. Leipzig—Jena, 1959; H. Römp. Isotope. Stuttgart, 1963; O. Rühle. Brot für sechs Milliarden. Leipzig—Jena—Berlin, 1963; H. Runge. Das Wetter und wir. Gotha—Leipzig, 1967; H. Schwenkel. Die Landschaft als Natur und Menschenwerk. Stuttgart, 1957; G. Steiger. Aufbruch. Urburschenschaft und Wartburgfest. Leipzig—Jena—Berlin, 1967; H.-H. Vogt. Der Nürnberger Trichter. Lernmaschinen für Ihr Kind? Stuttgart, 1966; D. Wattenberg. Mars, der rote Planet. Leipzig—Jena, 1956; W. H. Westphal. Die Relativitätstheorie. Stuttgart, 1955.

А. В. Грошева

ЧАСТОТНОСТЬ КЛАССОВ ЛАТИНСКИХ ФОНЕМ

Исследование частоты встречаемости звуков или букв проводилось со времени начала применения статистических методов в лингвистике. Для латинского языка первые результаты были получены Э. Фёрстеманном в исследовании, посвященном количественным соотношениям звуков в греческом, латинском и немецком языках.¹ Работа Фёрстеманна страдала многими недостатками: не назван исследуемый латинский автор и объем используемого материала; не проведено различие между звуком и буквой; не различались долгие и краткие гласные и т. п. Г. К. Ципф, который много лет спустя также занимался подсчетами латинских согласных фонем на материале писем Цицерона (объем его выборки — 10 тыс. фонем),² использовал данные Э. Фёрстеманна для частичного контроля своих результатов и отмечал большое сходство в процентном выражении частот, полученных им самим и предыдущим исследователем.

Однако все осуществленные до сих пор подсчеты латинских звуков, или фонем, в связном тексте проводились без учета занимаемого этими единицами положения в слове. Такое сплошное исчисление, не лишённое само по себе смысла, тем не менее препятствует выяснению роли статистического фактора в звуковых изменениях, которые, как известно, протекают различно в зависимости от структуры слова.

Наше исследование, задачей которого является изучение частотности различных классов фонем в латинском языке,³ мы предпочли ограничить двумя основными позициями — начала и ис-

¹ E. Försteman n. Numerische Lautverhältnisse im Griechischen, Lateinischen und Deutschen. Zeitschrift für vergleichende Sprachforschung, Bd. I, 1852, S. 163—179.

² G. K. Z i p f. Relative Frequency as a Determinant of Phonetic Change. Harvard Studies in Classical Philology, v. 40, 1929, p. 60.

³ Фонемы объединены в классы в соответствии с наличием у них определенного дифференциального признака (ДП). Матрицу фонем и перечень ДП см. в статье: А. В. Грошева. Система фонем латинского языка классического периода. УЗК, т. 54, 1969, стр. 152.

хода слова. Материалом для подсчетов послужили речи Цицерона, из которых мы взяли 10 выборок размером 1 авторский лист каждая. Таким образом было обследовано всего около 58 тыс. начальных и столько же конечных фонем. Однако в статье мы будем оперировать не цифрами абсолютной частотности и не их процентным выражением а, главным образом, полученными на их основании качественно новыми характеристиками — коэффициентами частотности (*КЧ*) и избирательности (*КИ*).⁴ Поясним эти термины.

КЧ по существу является мерилем функциональной загруженности фонем, т. е. степени их использования. Если бы нагрузка распределялась совершенно равномерно между всеми фонемами латинской фонологической системы, средняя частотность, найденная из условия равной вероятности фонем, была бы равна 100% : 29—3.45%. Соотнося фактическую частотность фонемы в исследуемой позиции со средней частотностью, получаем *КЧ*, который показывает, во сколько раз функциональная нагрузка (*ФН*) данной фонемы больше или меньше средней, равной 1.

При вычислении *КИ* частотность фонемы в определенной позиции сравнивается с частотностью этой фонемы в речи в целом. *КИ* показывает, какую позицию в слове «предпочитает» данная фонема, а какую «отвергает». Если вдуматься в суть этого показателя, то для такого языка, как латинский, с его богато развитой системой склонения и спряжения, *КИ* может сигнализировать о том, какой роли — лексической (начало слова) или грамматической (исход слова) — отдает предпочтение данная фонема.

Самым всеобъемлющим ДП, разделяющим фонемы на два больших класса, является «вокальность — консонантность». В начале латинского слова встречаются 11 (i, ī, ē, ē, ā, ā, ò, ò, u, ū, æ) из 14 гласных, входящих в фонологическую систему (отсутствуют ŷ, ŷ, œ) и 14 (k, s, p, q, n, m, t, d, w, f, g, l, g, h) из 15 согласных (отсутствует g^h). По нашим подсчетам, в этой позиции гласные фонемы составляют 33.6%, согласные — 66.4%, т. е. на 1 слово, начинающееся с гласной, приходится 2 слова, начинающиеся с согласной.

Если при анализе частотности отдельных фонем *КЧ* прямо пропорционален частотности, выраженной в процентах, то при анализе частотности различных классов фонем *КЧ* дают качественно иную характеристику, чем частотность в процентах.

Учитывая разницу в количестве начальных гласных и согласных фонем, определяем *КЧ* этих двух классов:

$$KЧ_{гл} = \frac{33.6\%}{3.45\% \cdot 11} = 0.9; \quad KЧ_{сог} = \frac{66.4\%}{3.45\% \cdot 14} = 1.4.$$

⁴ Подробно об этих коэффициентах см.: А. М. К а щ е е в а. К вопросу о методике количественного фонологического анализа. Вестник ЛГУ, № 20, вып. 4, 1968, стр. 138—146.

Коэффициенты частотности показывают, что *ФН* согласных фонем выше *ФН* гласных в начале латинского слова не в 2 раза, как об этом можно заключить, исходя из абсолютной частотности, а лишь в 1.5 раза.

В конце латинского слова практически могут выступать те же гласные фонемы, что и в начале, а количество согласных сокращается до 9 (помимо *g^u*, отсутствуют *p, q^u, w, f, g*). Однако это изменение почти не отразилось на частотности конечных фонем: согласные по-прежнему количественно преобладают над гласными (61.8% и 38.2% соответственно). Но *ФН* согласных стала вдвое выше средней (*КЧ*=2.0). Нагруженность гласных фонем в обоих исследуемых положениях средняя (*КЧ*=0.9, 1.0). Следовательно, роль согласных фонем в оформлении крайних позиций латинского слова значительно выше, чем роль гласных. Этот вывод подтверждается и коэффициентами избирательности, которые свидетельствуют, что согласные фонемы охотнее занимают позиции начала и исхода слова (*КИ*=1.2 и 1.1), чем гласные (*КИ*=0.7 и 0.8), которые, оказывается, предпочитают середину слова (*КИ*=1.1). Согласные, напротив, появляются в этой позиции реже, чем в крайних (*КИ*=0.9).

Другой ДП, обладающий значительным объемом в латинской фонологической системе, — «низкая тональность — высокая тональность» (охватывает 25 фонем из 29). В крайних позициях встречаются 13 низких по тону фонем: *ā, ū, ū, ō, ō, (g)^u, (q^u) (p), b, (f), (w), m* и 8 высоких: *ae, ī, ī, ē, ē, t, d, n, s*. Таким образом, число высоких фонем в крайних позициях не меняется, а набор низких в исходе сильно ограничен (8 из 13). Собственно говоря, все согласные фонемы, невозможные в конечной позиции латинского слова, относятся к числу низких, которые, следовательно, практически исключены из участия в грамматическом формообразовании.

В начале слова относительная частота низких и высоких фонем примерно одинакова, но низкие имеют среднюю *ФН*, высокие — в 1.5 раза выше. В исходе слова, где относительная частота высоких возрастает настолько, что становится в 2 раза больше частоты низких, *ФН* высоких фонем также увеличивается вдвое (*КЧ*=2.0), в то время как *ФН* низких остается по-прежнему средней (*КЧ*=1.0). Таким образом, в обоих крайних позициях латинских слов низкие имеют нормальную загруженность, а высокие перегружены, особенно в исходе слова. Что касается выбора позиций, то низкие фонемы охотно появляются в начале (*КИ*=1.3), но гораздо менее охотно — в исходе латинского слова (*КИ*=0.8). Высокие, напротив, не слишком охотно занимают начальную позицию в слове (*КИ*=0.8), предпочитая исход (*КИ*=1.1). Следовательно, фонемы низкой тональности в оформлении лексического облика латинского

слова играют более значительную роль, чем в обозначении его морфологических характеристик: для фонем высокой тональности (особенно для конечных *s, t*) в некоторой степени более характерна морфологическая роль.

Из 20 фонем,⁶ обладающих в латинской фонологической системе ДП «компактность — диффузность», в крайних позициях представлены 17 — 6 компактных: *ā, ā, ae, k, (g), (q^u)* и 11 диффузных: *i, ī, ū, ū, t, d, (p), b (f), (w), s*. Хотя относительная частота говорит о значительном преобладании диффузных в начале слова (на 1 компактную фонему приходится примерно 2 диффузных), однако оба класса фонем имеют примерно равную *ФН* в этой позиции, близкую к средней (*КЧ* компактных — 0.9, *КЧ* диффузных — 1.1). В исходе слов разрыв частот двух данных классов фонем становится еще более ощутимым (на 1 компактную фонему приходится 4 диффузных). *ФН* компактных снижается до 0.6, а *ФН* диффузных возрастает до 1.7.

Диффузные более или менее равномерно встречаются в обеих крайних позициях (*КИ*=0.9 и 1.1); компактные гораздо охотнее занимают начальную позицию в слове (*КИ*=1.4), чем конечную (*КИ*=0.8), что свидетельствует об их предпочтительном участии в оформлении лексического облика латинского слова.

Всеобъемлющим ДП для гласных фонем является «долгота — краткость». Из 6 кратких гласных латинской фонологической системы в крайних позициях представлены *ā, ē, ī, ō, ū*, из 8 долгих — *ā, ā, ī, ō, ū, ae*. Подсчеты показали необыкновенно высокую частотность кратких гласных в сравнении с долгими в начальной позиции: в среднем на 10 латинских слов, начинающихся с краткой гласной, приходится 1 слово, начинающееся с долгой. *КЧ* показывает, что *ФН* долгих гласных фонем в начале слова совершенно незначительна (*КЧ*=0.2), в то время как нагрузка кратких выше средней в 2.5 раза.

В конечной позиции дело обстоит иначе: здесь долгие гласные количественно преобладают над краткими в отношении 1.7 : 1.0. Естественно, что *ФН* долгих заметно увеличилась (*КЧ*=1.5), а у кратких она стала средней (*КЧ*=1.0).

КИ показывают, что долгие гласные, играющие в латинском языке большую роль в оформлении различных грамматических категорий, в исходе слова встречаются в 2 раза чаще, чем в речи

⁶ В области вокализма происходит расщепление ДП «компактность — диффузность» на «компактность — некомпактность» и «диффузность — недиффузность» из-за фонем *e* и *o*, которые являются некомпактными и недиффузными. Гласные *ā, ā* и *ae* по существу являются компактными, *ī* и *ū* — диффузными. Для удобства подсчетов проводя и в области гласных четкую линию разделения фонем на компактные и диффузные, мы исключили из рассмотрения частотность начальных *e* и *o*. Можно было бы решить этот вопрос иначе: включить частотность *e* и *o* при вычислении частотности как компактных фонем, так и диффузных, но это только усложнило бы расчеты, не повлияв на существо дела (процентное соотношение от этого не изменится).

⁵ В круглые скобки заключены фонемы, не встречающиеся в исходе латинского слова.

вообще, но избегают начала слова ($KИ=0.3$); краткие, напротив, явно предпочитают начальную позицию ($KИ=1.3$) конечной ($KИ=0.5$).

Прочие ДП характеризуют только согласные латинские фонемы.

ДП «звонкость — глухость» в латинской фонологической системе обладают 4 пары фонем: $t - d$, $p - b$, $k - g$, $q^h - g^h$. Последняя пара не представлена в начале латинского слова, а в исходе противопоставление по данному признаку сохраняет только пара $t - d$.⁷ В начальной позиции звонкие имеют довольно низкую $ФН$ ($KЧ=0.6$), зато глухие несут двойную нагрузку ($KЧ=2.0$). В исходе слова, где одна глухая фонема t как бы кумулирует в себе весь объем «глухости», $ФН$ глухих достигает необычайной высоты ($KЧ=6.2$). Фонема d соответственно кумулирует в себе всю «звонкость», также принимая большую нагрузку ($KЧ=1.8$).

Что касается позиций в слове, здесь поведение звонких и глухих фонем совершенно своеобразно по сравнению с другими классами фонем: у звонких и глухих нет излюбленных позиций; и те, и другие равномерно распределяются между началом, исходом, а также центром слова. Соотношение частот звонких и глухих, равное 1 : 3, остается постоянным во всех позициях.

ДП «прерывность — непрерывность», как и предыдущий, охватывает только согласные фонемы: прерывные — t , (p) , b , g , непрерывные — (t) , (w) , l , s . Незначительное преобладание частот непрерывных фонем над прерывными в начале слова не отражается на $ФН$, которая оказывается одинаково средней у обоих классов и в силу равного по количеству состава фонем. В исходе слова отношение непрерывных к прерывным равно 4 : 3. $ФН$ прерывных в этой позиции возрастает незначительно ($KЧ=1.2$), а у непрерывных, несмотря на отсутствие f и w , увеличивается в 2.2 раза по сравнению со средней, что вызвано необычайно высокой частотностью конечного s .

Прерывные встречаются в начале и исходе слова менее часто (в обеих позициях $KИ=0.8$), чем в речи вообще, предпочитая, до некоторой степени, середину слова ($KИ=1.1$). Это объясняется тем, что из фонем данного класса только p выполняет более или менее значительную роль в оформлении начала латинского слова, а роль t и g очень невелика, не говоря уже о начальном b . В исходе

слова значительный удельный вес имеет лишь конечное t , являясь показателем 3-го лица ед. и мн. ч. глаголов всех времен и наклонений действительного залога. Непрерывные также равномерно встречаются в крайних позициях ($KИ=1.2$), но они, напротив, предпочитают эти позиции середине слова ($KИ=0.9$), причем s преобладает в исходе слова, f и w — в начале.

ДП «назальный — неназальный» в латинской фонологической системе обладают две пары фонем: $m - n$ и $b - d$. Если в начале слова отношение назальных к неназальным по частоте равно 2 : 1, то в исходе слова разрыв между двумя группами фонем становится еще более заметным, сменяясь отношением 6 : 1. Это изменение самым непосредственным образом отражается на загруженности данных классов фонем. Если в начальной позиции $ФН$ назальных выше средней ($KЧ=1.3$), то в конечной позиции, в основном за счет m , она возрастает до 1.7. Соответственным образом, невысокая загруженность неназальных в начале слова ($KЧ=0.7$) в исходе становится почти ничтожной ($KЧ=0.3$). Таким образом, носовые в крайних позициях явно перегружены, носовые — очевидно недогружены. Интересно, что подобная картина наблюдается и в середине латинского слова ($KЧ$ назальных = 1.4, $KЧ$ неназальных = 0.6), что дает возможность сделать вывод о большой неравномерности в нагрузке указанных классов латинских фонем в речи вообще: высокая употребительность назальных сочетается с весьма низкой употребительностью неназальных.

Назальные отдают заметное предпочтение исходу слова ($KИ=1.2$) по сравнению с началом ($KИ=0.9$), что связано с необычайно высокой частотностью конечного m , входящего в состав многих именных и глагольных флексий; напротив, неназальные, отвергая конечную позицию ($KИ=0.5$), встречаются одинаково часто в начале и в середине слова ($KИ=1.2$), играя, таким образом, роль лексических оформителей.

ДП «плавность — неплавность» делит все латинские согласные на два неравных класса: плавные — g и l , неплавные — все остальные. Употребительность g и l в крайних позициях очень невелика: в тексте из 100 начальных согласных только 8 являются плавными, из 100 конечных — только 6. Естественно, что $KЧ$ плавных в начале слова равен всего 0.6, в исходе — 0.4. Плавные явно недоиспользуют свои возможности в этих позициях, зато в середине слова они берут реванш ($KЧ=1.8$), ибо обе плавные фонемы являются компонентами многих именных суффиксов, а одна g — также и глагольных.

Величина $ФН$ неплавных мало отличается от соответствующих характеристик согласных вообще (см. «вокальность — ковсонантность»).

$KИ$ подтверждают показания $KЧ$: латинские плавные предпочитают середину слова ($KИ=1.4$) в ущерб крайним позициям ($KИ=0.5$ и 0.4).

⁷ По всей вероятности латинский язык классической эпохи еще сохранял значимость оппозиции $t - d$ в конце слова, хотя уже с середины I в. до н. э. надписи отражают начавшееся смешение и взаимозамену конечных t и d . Правильное употребление t и d в таких словах, как at и ad , sed , $quit$ и $quid$, $quot$ и $quod$ и т. д. неоднократно было предметом обсуждения у грамматиков. Утверждение Э. Бенвениста (см. его статью: Répartition des consonnes et phonologie du mot. Travaux de Cycle Linguistique de Prague, 1939, № 8, p. 28), что фонологическая нагрузка оппозиций по звонкости в исходе латинского слова почти равна нулю, по-видимому, верно лишь для фонологической системы латинского языка; в системе речи $ФН$ оппозиции «звонкий — глухой» в указанном положении больше в силу высокой частоты отдельных компонентов минимальных пар.

Итак, в фонологической системе латинского языка классического периода имеющими высокую функциональную загруженность в обеих крайних позициях в слове оказались прежде всего согласные фонемы в противоположность гласным, имеющим нормальную среднюю нагрузку. Из числа согласных фонем особенно интенсивно используются в речи глухие и назальные согласные. Фонемы высокой тональности, как гласные, так и согласные, также относятся к высоконагруженным в обеих крайних позициях.

Слабо нагруженными, т. е. употребляющимися реже, чем допускает теоретически рассчитанная норма, в исследуемых позициях латинского слова оказались плавные фонемы.

Только в начальной позиции высокой функциональной нагрузкой обладают краткие гласные, низкой нагрузкой — долгие гласные.

Только в конечной позиции высокой функциональной нагрузкой обладают долгие гласные, непрерывные согласные и диффузные гласные и согласные.

Сравнение классов фонем с точки зрения их излюбленных позиций в латинском слове показало, что если фонемы низкой тональности отдают предпочтение началу слова, то их коррелят — фонемы высокой тональности — исходу слова. Подобным же образом распределяются группы компактных и диффузных, неназальных и назальных согласных, кратких и долгих гласных. Но у некоторых классов фонем такого «размежевания» по излюбленным позициям не наблюдается. Так, согласные господствуют как в начале, так и в исходе латинского слова, а гласные не оказывают предпочтения ни одной из крайних позиций. Оказалось, что излюбленным местом гласных является середина слова ($KИ=1.1$ по сравнению с 0.7 в начале слова и 0.8 в исходе); согласные в некоторой степени реже встречаются в середине слова по сравнению с краями ($KИ=0.9$ в середине, 1.2 — в начале, 1.1 — в исходе). Особым образом ведут себя также прерывные и непрерывные согласные: первые сконцентрированы в середине слова, последние — по краям. Плавные, избегая крайних позиций, отдают предпочтение центральной части слова.

Необходимо подчеркнуть, что все указанные различия в поведении классов латинских фонем обусловлены различной ролью составляющих их фонем в оформлении лексического и морфологического облика латинских слов. Отсюда вытекает, например, что краткие гласные выполняют по преимуществу роль лексических оформителей начала слова, а долгие — морфологических, участвуя в многообразных и высокочастотных флексиях. Точно так же на долю низких, компактных и неназальных фонем выпало главным образом оформление лексического облика латинских слов, на долю высоких, диффузных и назальных — выражение различных морфологических категорий.

Полученные результаты надлежало проверить на другом материале, поскольку нет твердой уверенности в том, зависит или не зависит частотность отдельных фонем, а также классов фонем от стилистических особенностей текста. Для сравнения было взято произведение совершенно иного жанра и другого автора, а именно «Записки о Галльской войне» Цезаря, относящиеся к роду комментариев. Из книг V, VI и VII были сделаны три выборки того же объема, что из речей Цицерона, — по одному авторск. листу каждая.

Хотя вопрос о частотности отдельных фонем не является предметом изложения в данной статье, надо отметить, что полного совпадения частот как начальных, так и конечных фонем в текстах двух авторов-современников не оказалось. И это не удивительно, ибо фонемная частотность является итогом действия ряда самых разнообразных факторов.

Напротив, частотность классов фонем в текстах Цицерона и Цезаря как в начальной, так и в конечной позициях оказалась весьма сходной. Так, при сравнении частотности классов фонем, выделяющихся по самому всеобъемлющему ДП «вокальность — консонантность», оказалось, что количество гласных в начале слова у Цезаря больше, а согласных меньше, чем у Цицерона, однако КЧ показали, что разница в функциональной загруженности этих классов у обоих авторов совершенно ничтожна:

Классы	КЧ	
	Цицерон	Цезарь
Гласные	0.9	1.0
Согласные	1.4	1.3

Различие во встречаемости начальных долгих и кратких гласных фонем состоит в том, что у Цицерона на 1 долгую фонему приходится 10 кратких, у Цезаря — 9. $ФН$ этих классов фонем тем не менее у обоих авторов одинакова: долгие имеют крайне низкую загруженность ($КЧ=0.2$), краткие — в 2.5 раза выше средней.

Если у Цицерона на 1 компактную фонему в начале слова приходилось примерно 2 диффузных, то у Цезаря отношение стало равно 1 : 1.5. Но это изменение отразилось на $ФН$ компактных и диффузных у Цезаря самым незначительным образом:

Классы	КЧ	
	Цицерон	Цезарь
Компактные	0.9	1.0
Диффузные	1.1	1.0

В классах начальных низких и высоких фонем, звонких и глухих, прерывных и непрерывных у обоих авторов нет существенных различий в частотности.

Количество назальных в начале слова у Цезаря снизилось по сравнению с текстом Цицерона. Одна из причин этого изменения — почти полное отсутствие местоименных форм 1-го л. ед. ч. (*mē, mīhi, meus*) в «Записках», поскольку Цезарь пишет о себе как военачальнике в 3-м л. Как и в предыдущих случаях, снижение количества назальных мало отразилось на *ФН* этого класса фонем: *КЧ* с 1.3 в речах Цицерона снизился в комментариях Цезаря до 1.2.

Плавные в начале слова у Цезаря встречаются несколько чаще, чем у Цицерона. Если в ораторской прозе на 100 начальных согласных фонем приходится 8 плавных, то в комментариях — 10, что дало увеличение *КЧ* всего на 0.1.

Таким образом, путем сравнения мы пришли к выводу, что *ФН* классов начальных латинских фонем, рассчитанная для текста Цезаря, мало отличается от *ФН* соответствующих классов фонем в тексте Цицерона, а иногда совершенно одинакова у обоих авторов.

Если частотность классов начальных фонем оказалась в значительной мере сходной в речах Цицерона и комментариях Цезаря, то с тем большим основанием можно было ожидать сходства частот классов конечных фонем в рассматриваемых текстах в соответствии с высказыванием Хердана о том, что «грамматические характеристики. . . склонны быть в большей степени не зависящими от стиля (по сравнению с лексическим составом и длиной предложения, — *A. G.*) и таким образом обусловленными языковой системой».⁸

И действительно, *КЧ* однородных классов конечных фонем у обоих авторов колеблется в пределах ± 0.1 .

На основании проведенного сравнения можно, следовательно, сделать вывод о том, что различие в жанре двух текстов в основном мало отражается на различии в частотности классов фонем, а тем более на функциональной нагрузке этих классов. Частотные колебания отдельных фонем при объединении последних в классы в соответствии с их противопоставленностью по тому или другому ДП как бы нивелируются.

Часть II. ЛЕКСИКО-СЕМАНТИЧЕСКИЙ МАШИННЫЙ ПЕРЕВОД

*Е. Б. Бектася, Л. П. Белоцерковская,
А. Д. Борисевич, [В. А. Букович], Н. С. Булашева,
В. В. Гончаренко, М. В. Данейко, Т. Р. Зуева,
Н. И. Исабекова, О. А. Нехай,
Р. Г. Пиотровский, В. А. Соркина*

СТАТИСТИЧЕСКОЕ ВЫДЕЛЕНИЕ РУССКИХ МАШИННЫХ ОБОРОТОВ И ПОСТРОЕНИЕ ДВУЯЗЫЧНЫХ АВТОМАТИЧЕСКИХ СЛОВАРЕЙ

Каждый текст состоит из ключевых единиц (словоупотреблений и их комбинаций), несущих основную смысловую нагрузку, и единиц заполнения, как правило, не связанных непосредственно с его содержанием, а выполняющих синтаксико-стилистическую и композиционную функцию организации текста и его ключевых единиц.

Разные виды переработки текста предусматривают различное обращение с единицами заполнения. При автоматическом аннотировании и реферировании текста единицы заполнения обычно не подвергаются машинной переработке; все внимание сосредотачивается здесь на выявлении ключевых единиц и их смысловых и логических связей. Напротив, при машинном переводе единицы заполнения должны перерабатываться наряду с ключевыми элементами.

Одной из эффективных форм переработки единиц заполнения при МП является перевод их с помощью автоматического словаря оборотов (АСО) без обращения к грамматическому анализу. Такой подход с позиции «грубой силы», называемый иногда икоическим подходом, является эффективным приемом устранения многозначности отдельных словоформ и словосочетаний, а также средством грамматической организации текста. В итоге использования АСО мы получаем вполне нормативный и стилистически удовлетворительный перевод некоторых входных сегментов заполнения [4, стр. 26—33].

⁸ G. Herdan, *The Calculus of Linguistic Observations*, s'-Gravenhage, 1962, p. 23.

К сожалению, малая память современных ЭВМ заметно лимитирует объем АСО. Поэтому в АСО должны отбираться лишь наиболее типичные для данного подъязыка отрывки-штампы. Отбор этих типичных сегментов заполнения осуществляется путем статистического обследования как входных текстов, так и текстов выходного языка. Статистический отбор выходных сегментов важен при построении таких алгоритмов МП, которые должны выдавать потребителю не полные переводы, но расширенные рефераты-переводы, организованные по принципу информационного бланка. В этом бланке заранее задаются стандартные формы элементов заполнения и отводятся места для перевода ключевых единиц.

Прототипом такого перевода-реферата является бланк индексации и аннотации-перевода, который выдает система лингвистического обслуживания, созданная в группе «Статистика речи» [10, стр. 66—71].

Наиболее эффективным приемом массового выбора оборотов является машинное или ручное выделение из текста наиболее частых сегментов фиксированной длины (триады, тетрады), опирающихся на заранее заданные ядерные (опорные) словоформы [26, стр. 200—210], с последующим преобразованием их в машинные обороты.

Весь процесс выделения формальных сегментов, преобразования их в обороты, а затем использование этих последних для МП опирается на следующие соображения. Каждый текст (предложение, словосочетание) можно рассматривать как результат некоторого развертывания, стоящего в начале текста словоупотребления или словосочетания путем так называемых $(k+1)$ — вероятностных приближений, описанных Н. Хомским и Дж. Миллером [12, стр. 426—430; ср. также: 9, стр. 56; 14, стр. 176—185]. Так, например, текст: *Сила тока термоэлектронной эмиссии тем больше, чем выше температура катода* (Н. И. Кацов. *Электроника*. М., 1956, стр. 19) может быть представлен как развертывание начальной триады *Сила тока* путем ее дополнения триадами *электронной эмиссии* | *тем больше* | *чем выше* | *температура катода*.

Этот же текст можно представить как теоретико-множественную сумму триад, образуемых путем сдвига вправо на одну словоформу:

Δ сила тока
 сила тока термоэлектронной
 тока термоэлектронно эмиссии
 термоэлектронной эмиссии тем
 эмиссии тем больше Δ
 тем больше Δ
 тем больше чем
 чем выше температура
 выше температура катода

сила тока термоэлектронно эмиссии тем больше, чем выше температура катода

Возникает вопрос: нельзя ли использовать методику развертывания для вероятной «сборки» машинных оборотов из ранее выделенных триад и тетрад? Необходимость в такой «сборке» возникает в связи с тем, что из-за большого объема работы не удается получить достоверную статистику словосочетаний, состоящих более чем из трех-четырех словоупотреблений (ср. [1, 2, 3, 5, 7, 8, 11]). Между тем достаточно большое количество устойчивых и употребительных словосочетаний, в том числе сложных терминов, состоит из пяти и более словоупотреблений.

Как показали Н. Хомский и Дж. Миллер [12, стр. 428 и сл.], образование текстов путем вероятностного развертывания начального словосочетания с помощью k -ограниченного марковского источника не может быть использовано для выделения машинных оборотов. Увеличение k не ведет к выделению множества грамматически правильных сегментов. Хотя число высоковероятных грамматически правильных сегментов растет, число маловероятных грамматически правильных последовательностей растет еще быстрее. Кроме того, при любом конечном k возникнут такие неправильные, с точки зрения нормы языка, сегменты, состоящие более чем из k символов, которые ЭВМ не сможет опознать как неправильные.

Чтобы получить достаточно большое число типичных для данного подъязыка оборотов длиной в пять и более словоупотреблений, целесообразно собирать их путем напластования достаточно частых «заготовок» — триад и тетрад. Нижний порог употребительности этих заготовок принимается нами равным 10^{-5} (такой порог соответствует нижнему порогу частотности слов и словоформ, включаемых обычно в автоматические словари).

Если сборка оборота осуществлена при условии, что каждая предыдущая триада (тетрада) имеет по крайней мере две общие словоформы с последующей триадой (тетрадой), а получаемый в результате теоретико-множественного сложения оборот является вполне осмысленным, можно предполагать, что такой оборот реально встречается в текстах и его можно включить условно в АСО.

Например, имеются четыре триады:

- 1) в зависимости от $(3,3 \cdot 10^{-4})$; 1
- 2) зависимости от напряжения $(1,6 \cdot 10^{-4})$;
- 3) от напряжения на $(2,7 \cdot 10^{-5})$;
- 4) напряжения на аноде $(3,1 \cdot 10^{-5})$,

теоретико-множественное сложение которых дает вполне осмысленный, с точки зрения радиоэлектроники, сегмент: *в зависимости от напряжения на аноде*. Верхний порог потенциальной частоты этого сегмента, определяемый всегда по относительной частоте наиболее редкой триады, составляет $1,6 \cdot 10^{-5}$.

¹ В скобках указывается относительная частота триады или тетрады.

Путем использования указанного приема напластования из частотных списков наиболее частых триад,² предложных триад и тетрад,³ именных⁴ и глагольных⁵ триад, было выделено 160 оборотов, объединенных в три списка машинных оборотов: а) именных (табл. 1); б) глагольных (табл. 2); в) наречных, предложных и союзных (табл. 3).

Путем напластования триад мы можем построить целое предложение, которое по своему синтаксису не отличается от реальных предложений, взятых из соответствующих текстов. Например, теоретико-множественное сложение трех триад:

- 1) анодный ток возрастает ($1.1 \cdot 10^{-5}$);
- 2) возрастает с увеличением ($1.4 \cdot 10^{-6}$);
- 3) с увеличением температуры (10^{-5})

дает осмысленное, нормативно построенное квазипредложение: анодный ток возрастает с увеличением температуры. Это квазипредложение обладает достаточно высокой вероятностью — 10^{-5} . Учитывая сказанное, мы можем включить его в выходной список АСО.

Прием напластования триад может быть использован и при построении программ предсказуемого синтеза [13, стр. 1—29; 15, стр. 47—48].

Например, имеем триады:

- 1) Δ в зависимости ($1.3 \cdot 10^{-4}$);
- 2) в зависимости от ($3.3 \cdot 10^{-4}$);
- 3) зависимости от величины ($3 \cdot 10^{-5}$);
- 4) от величины поля ($7 \cdot 10^{-6}$);
- 5) от величины сопротивления ($7 \cdot 10^{-6}$);
- 6) от величины тока ($9 \cdot 10^{-6}$).

Напластование триад 1—3 дает достаточно частый ($3 \cdot 10^{-5}$) для включения в АСО сегмент в зависимости от величины. Однако дальнейшее его развертывание на лексическом уровне нецелесообразно, поскольку находящиеся справа от него словоупотребления имеют слишком малую вероятность. Поэтому целесообразно объединить их в одну группу, указывая в АСО грамматический инвариант этой группы — существительное ед. ч. род. пад. В итоге получаем следующий оборот с управлением:

в зависимости от величины + сущ. ед. ч. род. пад.

Используя уже упомянутые списки триад и тетрад (см. ниже, стр. 111—114), мы выделили 101 машинный оборот с левыми и правыми грамматическими валентностями (табл. 4—7).

² Составитель Н. С. Булашева. Объем выборки 700 тыс. словоупотреблений; ср. [3].

³ Составитель Т. Р. Зуева. Объем выборки 2 млн. словоупотреблений; ср. [7].

⁴ Составитель Н. И. Исабекова. Объем выборки 2.5 млн. словоупотреблений; ср. [8].

⁵ Составитель Л. И. Белоцерковская; ср. [2].

Список именных машинных оборотов

№№	Обороты	f·10 ⁻⁶
1	анодный ток в лампе	1.1
2	анодный ток в цепи	2.7
3	анодный ток в цепи анода	1.0
4	анодный ток в цепи коллектора	3.0
5	анодный ток в цепи эмиттера	2.7
6	анодного тока в цепи анода	1.0
7	анодного тока в цепи коллектора	1.9
8	анодного тока в цепи сетки	1.0
9	анодного тока в цепи эмиттера	1.0
10	анодного тока лампы	1.0
11	в зависимости от величины	1.6
12	в зависимости от напряжения на аноде	1.6
13	в зависимости от способа	1.0
14	в зависимости от температуры	1.6
15	в соответствии с выражением	1.0
16	в соответствии с формулой	1.0
17	в соответствии с этим	2.0
18	ветвь вольтамперной характеристики диода	1.9
19	входного и выходного напряжения	1.0
20	движение электронов в зоне проводимости	5.0
21	для этого случая	2.0
22	как и в схеме	1.0
23	коэффициент усиления имеет вид	3.5
24	коэффициент усиления по току	1.8
25	напряжение на выходе схемы	1.6
26	напряжение на выходе усилителя	2.8
27	напряжения на выходе схемы	1.6
28	напряжения на выходе усилителя	2.8
29	напряжения на газоразрядном диоде	1.0
30	одно и то же время	2.0
31	одного и того же порядка	1.0
32	одной и той же	4.0
33	одном и том же рисунке	1.0
34	одну и ту же величину	1.0
35	от катода к аноду	1.9
36	от напряжения на коллекторе	2.0
37	от напряжения на сетке	2.0
38	от напряжения на электродах	2.0
39	от температуры катода	1.0
40	по отношению к катоду	1.6
41	под действием внешнего электрического поля	1.8
42	под действием электрического поля	5.6
43	при более высоких температурах	1.1
44	раз больше, чем с	1.0
45	раз больше, чем это	1.0
46	распределение электронов по энергиям	2.3
47	так и в схеме	1.0
48	ток в цепи анода	2.5
49	ток в цепи базы	1.0
50	ток в цепи коллектора	2.4
51	ток в цепи сетки	1.6

Таблица 1 (продолжение)

№№	Обороты	f 10'
52	ток во внешней цепи	1.5
53	тока в цепи анода	2.0
54	тока в цепи базы	1.0
55	тока в цепи коллектора	2.4
56	тока в цепи сетки	1.6
57	токи и напряжения на аноде	1.0
58	токи и напряжения на сетке	1.0
59	фотоэлементов с внешним эффектом	1.2
60	что и в случае	1.0

Таблица 2

Список глагольных машинных оборотов

№№	Обороты	f 10
1	будем считать, что	2.3
2	будет иметь вид	4.3
3	будет равно	1.0
4	будет установлено, что	4.5
5	в качестве примера рассмотрим	1.7
6	говорит о том, что	1.1
7	дело в том, что если	1.0
8	для этого нужно	1.2
9	должен быть равен	1.0
10	должна быть равна	1.4
11	должно быть равно	1.0
12	должны быть равны	1.0
13	достаточно для того, чтобы	1.0
14	зависит от амплитуды	1.0
15	зависит от давления	2.0
16	зависит от напряжения на аноде	1.4
17	зависит от напряжения на переходе	1.6
18	зависит от напряжения на сетке	1.0
19	зависит от температуры	4.0
20	зависит от тока эмиттера	1.0
21	зависит от частоты сигнала	1.0
22	зависят от значения	1.0
23	заключается в том, что	5.9
24	из формулы следует, что	2.2
25	изменяется в пределах от С до С	1.4
26	как говорилось выше	1.0
27	как отмечалось выше	1.0
28	как показано на рис.	1.0
29	как это видно	1.0
30	как это имеет место	1.0
31	как это сделано	1.0
32	колеблется в пределах от С до С	1.1
33	лежат в пределах от С до С	1.0
34	лежит в пределах от С до С	2.3

Таблица 2 (продолжение)

№№	Обороты	f 10 ²
35	могут быть использованы	2.1
36	могут быть найдены	1.0
37	могут быть определены	1.0
38	могут быть получены	1.3
39	могут быть построены	1.1
40	могут быть представлены	1.6
41	может быть выражена	1.9
42	может быть записано	1.7
43	может быть значительно больше, чем	1.0
44	может быть найдено	1.0
45	может быть получено	2.0
46	может быть представлено	1.0
47	можно видеть, что	2.0
48	можно думать, что	1.0
49	можно заключить, что	1.0
50	можно заметить, что	1.1
51	можно записать в виде	1.2
52	можно записать в следующем виде	1.7
53	можно использовать в качестве	1.1
54	можно найти, что	2.0
55	можно написать в виде	1.4
56	можно написать, что	1.4
57	можно ожидать, что	1.9
58	можно определить, что	2.0
59	можно показать, что	6.0
60	можно полагать, что	1.2
61	можно предположить, что	1.1
62	можно представить в виде	1.5
63	можно сказать, что	2.0
64	можно считать, что	7.0
65	не может быть	5.5
66	не может быть больше	1.0
67	представлены на рис.	2.8
68	приведены на рис.	4.5
69	принимая во внимание, что во многих случаях	1.0
70	свидетельствует о том, что	1.2
71	связаны друг с другом	1.0
72	соединены друг с другом	1.0
73	состоит в том, что	4.5
74	состоит в том, чтобы	4.0
75	увеличивается примерно в С раз	1.0
76	уменьшается примерно в С раз	1.0
77	чтобы получить выражение для	1.0

Итак, получено 263 русских машинных оборота (включая одно квазипредложение), которые в среднем покрывают около 3% наугад взятого текста (на страницу текста ими покрываются цепочки общей длиной в 9—10 словоупотреблений). Разумеется, покрываемость текста нашим выловным словником оборотов пока невелика, однако она может быть увеличена путем расширения самого словника. Но не это главное.

Таблица 3

Список наречных, предложных и союзных машинных оборотов

№№	Обороты	$f \cdot 10^5$
1	<i>а вместе с тем</i>	1.6
2	<i>в зависимости от того</i>	2.9
3	<i>в зависимости от того, где</i>	2.7
4	<i>в зависимости от того, как</i>	2.6
5	<i>в несколько раз больше, чем в</i>	1.9
6	<i>в результате чего</i>	5.5
7	<i>в результате этого</i>	3.0
8	<i>в соответствии с этим</i>	2.0
9	<i>в то время как</i>	7.0
10	<i>в то же время</i>	7.5
11	<i>в это же время</i>	1.0
12	<i>ввиду того, что</i>	1.0
13	<i>все более и более</i>	1.9
14	<i>вследствие того, что</i>	1.4
15	<i>дело в том, что</i>	2.2
16	<i>до тех пор, пока</i>	6.0
17	<i>как в этом случае</i>	1.9
18	<i>лишь в том случае</i>	1.0
19	<i>несмотря на то, что</i>	2.4
20	<i>тем больше, чем больше</i>	2.7
21	<i>тем меньше, чем больше</i>	2.4
22	<i>только в этом случае</i>	2.0
23	<i>что в этом случае</i>	1.6

Как уже говорилось, наш словарь может служить источником отбора типичных штампов для составления перевода-аннотации или перевода-реферата, построенных по принципу информационного бланка. Так, например, двум входным английским оборотам целесообразно соотнести только один русский эквивалент из нашего словаря — можно записать в виде ($7.5 \cdot 10^{-5}$). Одновременно выходной машинный оборот *как показано на рис.* (6.5×10^{-5}) может быть использован в качестве эквивалента таких английских сегментов, как

it is shown in Fig. C. ($1.2 \cdot 10^{-5}$),
this is shown in Fig. C. ($1.2 \cdot 10^{-5}$);
as shown in Fig. C. ($22 \cdot 10^{-5}$).

обладающих суммарной частотностью $f=0.00024$ (ср. 1, 5, 11). В переводе, выдаваемом машиной, рассматриваемый русский сегмент будет давать значительно большую покрываемость текста, чем это можно было ожидать, опираясь на статистику русских текстов по радиоэлектронике.

Вместе с тем, сопоставляя частые машинные обороты (штампы) русского текста с отобранными статистическим путем машинными

Таблица 4

Список именных машинных оборотов с правой грамматической валентностью

№№	Обороты	$f \cdot 10^5$
1	<i>анодный ток лампы e-сущ. предл. пад.</i>	2.7
2	<i>анодный ток лампы будет+инф.</i>	1.0
3	<i>анодный ток при+сущ. ед. ч. предл. пад.</i>	1.0
4	<i>в зависимости от величины+сущ. ед. ч. род. пад.</i>	4.0
5	<i>в зависимости от напряжения на+сущ. ед. ч. род. пад.</i>	2.5
6	<i>в качестве примера+спр. ф.</i>	1.2
7	<i>в результате этого+сущ. ед. ч. род. пад.</i>	3.0
8	<i>величина которого зависит от+им. ф. род. пад.</i>	1.0
9	<i>во время работы+сущ. ед. ч. род. пад.</i>	2.0
10	<i>как и в области+сущ. ед. ч. род. пад.</i>	1.5
11	<i>как и в случае+сущ. ед. ч. род. пад.</i>	3.0
12	<i>напряжение на выводе+сущ. ед. ч. род. пад.</i>	1.0
13	<i>напряжения на выводе+сущ. ед. ч. род. пад.</i>	1.0
14	<i>но и от величины+сущ. ед. ч. род. пад.</i>	1.0
15	<i>но и от температуры+сущ. ед. ч. род. пад.</i>	1.0
16	<i>от напряжения на входе в+сущ. ед. ч. вин. пад.</i>	2.0
17	<i>под действием которой+спр. ф.</i>	1.0
18	<i>при подаче на+сущ. ед. ч. вин. пад.</i>	1.8
19	<i>распределение потенциала в+сущ. ед. ч. предл. пад.</i>	1.2
20	<i>распределение потенциала в+сущ. ед. ч. предл. пад.</i>	1.3
21	<i>так и в области+сущ. ед. ч. род. пад.</i>	1.5
22	<i>так и в случае+сущ. ед. ч. род. пад.</i>	3.0
23	<i>ток в цепи+сущ. ед. ч. род. пад.</i>	1.5
24	<i>тока в цепи+сущ. ед. ч. род. пад.</i>	1.5
25	<i>токи и напряжение на+им. ф. предл. пад.</i>	1.0
26	<i>характеристика для схемы с+им. ф. твор. пад.</i>	1.0

Таблица 5

№№	Обороты	$f \cdot 10^5$
1	<i>будет зависеть от+им. ф. род. пад.</i>	1.0
2	<i>будут зависеть от+им. ф. род. пад.</i>	1.0
3	<i>выполняется в виде+сущ. ед. ч. род. пад.</i>	1.0
4	<i>для этого необходимо+инф.</i>	1.0
5	<i>зависит от величины+сущ. ед. ч. род. пад.</i>	1.0
6	<i>зависит от концентрации+сущ. ед. ч. род. пад.</i>	1.0
7	<i>зависит от напряжения на+сущ. ед. ч. предл. пад.</i>	1.0
8	<i>зависит от напряженности+им. ф. род. пад.</i>	1.0
9	<i>зависит от температуры+сущ. ед. ч. род. пад.</i>	1.0
10	<i>зависит от частоты+сущ. ед. ч. род. пад.</i>	1.0
11	<i>зависят от величины+сущ. ед. ч. род. пад.</i>	1.0
12	<i>зависят от температуры+сущ. ед. ч. род. пад.</i>	1.0
13	<i>записано в виде+сущ. ед. ч. род. пад.</i>	1.6
14	<i>записывается в виде+сущ. ед. ч. род. пад.</i>	1.0
15	<i>как было показано в+сущ. ед. ч. предл. пад.</i>	5.0
16	<i>как видно из+сущ. ед. ч. род. пад.</i>	7.0
17	<i>как следует из+сущ. ед. ч. род. пад.</i>	3.0
18	<i>как это показано в+сущ. ед. ч. предл. пад.</i>	5.0

Таблица 5 (продолжение)

№№	Обороты	f 10 ⁵
19	можно было бы+инф.	2.0
20	можно записать в виде+сущ. ед. ч. род. пад.	4.0
21	можно использовать для+сущ. ед. ч. род. пад.	1.0
22	можно найти из+сущ. ед. ч. род. пад.	1.0
23	можно определить из+сущ. ед. ч. род. пад.	1.0
24	можно определить по+сущ. ед. ч. дат. пад.	1.0
25	можно рассматривать как+сущ. ед. ч. вин. пад.	1.0
26	не зависят от+сущ. ед. ч. род. пад.	1.0
27	переписать в виде+сущ. ед. ч. род. пад.	1.6
28	представить в виде+сущ. ед. ч. род. пад.	5.4
29	при отсутствии внешнего+сущ. ед. ч. род. пад.	1.1
30	приводит к изменению+сущ. ед. ч. род. пад.	1.7
31	приводит к увеличению+сущ. ед. ч. род. пад.	2.2
32	принимая во внимание, что во многих+им. ф. предл. пад.	3.0
33	сильно зависит от +им. ф. род. пад.	2.0
34	слабо зависит от+им. ф. род. пад.	3.0
35	уменьшается с ростом+сущ. ед. ч. род. пад.	1.6
36	чтобы получить+сущ. ед. ч. вин. пад.	2.6

Таблица 6

Список наречных, предложных и союзных машинных оборотов с правой грамматической валентностью

№№	Обороты	f 10 ⁵
1	даже в случае+сущ. ед. ч. род. пад.	1.4
2	до того, как+сп. ф. глагола	1.1
3	как в этом+сущ. ед. ч. предл. пад.	1.0
4	по отношению к+сущ. ед. ч. дат. пад.	1.5
5	по сравнению с+им. ф. твор. пад.	1.0
6	точно так же, как и в+им. ф. предл. пад.	2.0
7	точно так же, как и для+им. ф. род. пад.	1.0
8	точно так же, как и при+им. ф. предл. пад.	1.0
9	точно так же, как и у+им. ф. род. пад.	1.0
10	это следует из+им. ф. род. пад.	1.0

оборотами входного текста, принадлежащего к тому же подязыку, можно получить достаточно надежный с переводческой и статистической точки зрения словарь оборотов для нужд МП.

Используя английские сегменты, полученные В. А. Буковичем, В. В. Гончаренко, М. В. Данейко [5], О. А. Нехай [11], В. А. Соркиной [11], а также описанный выше словарь вместе с сегментами частотных списков, упомянутых на стр. 110, мы построили пробный англо-русский словарь машинных оборотов подязыка электроники, который приводится ниже (нижний вероятностный порог цепочек составляет $f=10^{-5}$). Покрываемость этого словаря достигает 5%.

Таблица 7

Список машинных оборотов с левой и правой грамматической валентностью

№№	Обороты	f 10 ⁵
1	глагол+в виде+сущ. род. пад.	1.0
2	глагол+в зависимости от+сущ. ед. ч. род. пад.	2.6
3	глагол+в зависимости от типа	1.8
4	глагол+в качестве+им. ф. род. пад.	5.0
5	глагол+в результате диффузии	1.2
6	глагол+в результате ионизации	1.0
7	глагол+в результате+им. ф. ед. ч. род. пад.	8.0
8	глагол+в течение+сущ. ед. ч. род. пад.	2.0
9	глагол+в течение некоторого времени	1.4
10	глагол+в течение нескольких+сущ. мн. ч. род. пад.	4.5
11	глагол+во много раз+наречие сравнительной степени	1.5
12	глагол+до высокой температуры	1.0
13	глагол+до нескольких десятков	1.5
14	глагол+за счет+сущ. ед. ч. род. пад.	1.0
15	глагол+за счет диффузии	1.4
16	глагол+за счет изменения	1.7
17	глагол+за счет увеличения	1.5
18	глагол+за счет уменьшения	1.0
19	глагол+за счет энергии	1.5
20	глагол+от нуля до	1.7
21	глагол+по экспоненциальному закону	2.2
22	глагол+через переход	2.8
23	глагол+через электронно-дырочный переход	1.0
24	глагол+электрона из валентной зоны	1.5
25	глагол+электрона из зоны проводимости	1.2
26	прич.+от анода к катоду	1.0
27	прич.+от катода к аноду	2.5
28	сущ.+вольтамперной характеристики диода	1.4

АНГЛО-РУССКИЙ СЛОВАРЬ МАШИНЫХ ОБОРОТОВ (ПОДЪЯЗЫК ЭЛЕКТРОНИКИ)

1 access circuits	схемы выборки
2 access time	время выборки
3 according to the equation	в соответствии с формулой
4 according to the expression	в соответствии с выражением
5 according to the formula	в соответствии с формулой
6 amplification factor	коэффициент усиления
7 amplitude changes	изменения амплитуды
8 as a function of the cathode temperature	в зависимости от температуры катода
9 at the same time it is assumed that	при этом предполагается, что
10 at the same time it is necessary	при этом необходимо
11 at this point	в этом месте
12 and in a more general case	а в более общем случае
13 and in this equation	и в этом уравнении
14 and in this layer	и в этом слое
15 and in this place	и в этом месте

16 and in this relation	и в этом отношении	61 as in this ratio	как в этом отношении
17 and in this respect	и в этом отношении	62 as in this relation	как в этом отношении
18 and the electrode voltage	и напряжение на электродах	63 as indicated above	как отмечалось выше
19 and the grid voltage	и напряжение на сетке	64 as is done	как это сделано
20 and the voltage across the diode	и напряжение на диоде	65 as is seen	как видно
21 and the voltage in the tube	и напряжение на лампе	66 as is shown in	как это показано в
22 anode current in the circuit	анодный ток в цепи	67 as is the case	как это имеет место
23 anode current in the circuit of the emitter	анодный ток в цепи эмиттера	68 as is well known	как известно
24 anode current in the collector circuit	анодный ток в цепи коллектора	69 as may be seen from Fig.	как видно из рис.
25 anode current in the tube	анодный ток в лампе	70 as mentioned above	как отмечалось выше
26 anode currents and voltages	токи и напряжения на аноде	71 as seen from Fig.	как это видно из рис.
27 are connected	соединены друг с другом;	72 as shown in Fig.	как показано на рис.
28 are given in Fig.	связаны друг с другом;	73 as the magnitude of the alternating current in the circuit of the grid	так как величина переменного тока в цепи сетки
29 are given in Table	представлены на рис.;	74 as the resistance of the current	так как сопротивление тока
30 are shown in Fig.	представлены на фиг.	75 as the voltage on the electrodes	так как напряжение на электродах
	представлены на табл.	76 as was pointed out	как указывалось
	приведены на рис.;	77 as was shown	как было показано
	приведены на фиг.;	78 as well as in germanium	как и в германии
	показаны на рис.;	79 as well as in the base region	как и в области базы
	показаны на фиг.	80 as well as in the temperature range	как и в области температур
31 are shown in the Table	приведены в таблице	81 as will be seen	как будет видно
32 are used as sources	используются в качестве источников	82 as will be shown in Fig.	как будет показано на рис.
33 as	в качестве	83 as will be shown in the table	как будет показано в таблице
34 as a function of	в зависимости от	84 at higher temperatures	при более высоких температурах
35 as a function of temperature	в зависимости от температуры	85 at least	при крайней мере
36 as a function of the anode voltage	в зависимости от напряжения на аноде	86 at room temperature	при комнатной температуре
37 as a function of the input resistance	в зависимости от величины сопротивления на входе	87 at the expense of	за счет
38 as a function of the input voltage of the amplifier	в зависимости от напряжения на входе усилителя	88 at the output of the amplifier	на выходе усилителя
39 as a function of the output voltage of the amplifier	в зависимости от напряжения на выходе усилителя	89 at the output of the circuit	на выходе схемы
40 as a function of the output voltage of the circuit	в зависимости от напряжения на выходе схемы	90 at the same time	в то же время
41 as a function of voltage	в зависимости от напряжения	91 at these frequencies	на этих частотах
42 as a result	в результате этого	92 before	до того, как
43 as a result of diffusion	в результате диффузии	93 block diagram	блок-схема
44 as a result of ionization	в результате ионизации	94 branch of the current voltage characteristic of the diode	ветвь вольтамперной характеристики диода
45 as a rule	как правило	95 but at the same time	а вместе с тем
46 as an example	в качестве примера	96 but in this case	но в этом случае
47 as an example, consider	в качестве примера рассмотрим	97 by the action of the electric field	действием электрического поля
48 as an example, let us consider	в качестве примера рассмотрим	98 by the exponential law	по экспоненциальному закону
49 as at this point	как в этом месте	99 can be built	могут быть построены
50 as can be seen from Fig.	как видно из рис.	100 can be calculated by X	может быть вычислена по ϕ .
51 as compared to	по сравнению с	101 can be calculated from X	может быть вычислена по ϕ .
52 as compared with	по сравнению с	102 can be considerably greater than	может быть значительно больше, чем
53 as described above	как описано выше	103 can be considered as	можно рассматривать как
54 as follows	следующим образом; как	104 can be constructed	могут быть построены
55 as in the case of the application	как и в случае применения	105 can be defined	могут быть определены
56 as in the circuit in Fig.	как и в схеме на рис.	106 can be determined	могут быть определены
57 as in the region	как и в области	107 can be determined by	можно определить по
58 as in the space charge region	как и в области пространственного заряда	108 can be determined from	можно определять из
	так как при этом	109 can be determined from the equation	можно определить из уравнения
	как в этом слое	110 can be expressed	может быть выражена
59 as in this case		111 can be found	могут быть найдены;
60 as in this layer			может быть найдено

112 can be found by	можно найти по	159 depends on the frequency of the signal	зависит от частоты сигнала
113 can be found from	можно найти из	160 depends on the input voltage of the amplifier	зависит от напряжения на входе усилителя
114 can be greater than	может быть больше, чем	161 depends on the magnitude	зависит от величины
115 can be neglected	можно пренебречь	162 depends on the magnitude of the current in the anode circuit	зависит от величины тока в цепи анода
116 can be obtained	могут быть получены; может быть получена; может быть получено; можно получить	163 depends on the magnitude of the current in the circuit of the emitter	зависит от величины тока в цепи эмиттера
117 can be obtained from equation,	можно определить по формуле	164 depends on the magnitude of the current in the grid circuit	зависит от величины тока в цепи сетки
118 can be quite	может быть весьма	165 depends on the voltage at the input of the amplifier	зависит от напряжения на входе усилителя
119 can be regarded as	можно рассматривать как	166 depends strongly on	сильно зависит от
120 can be represented	могут быть представлены; может быть представлено	167 differ from each other	отличаются друг от друга
121 can be represented as a sum	может быть представлено в виде суммы	168 digital computer	цифровая вычислительная машина
122 can be represented by a sum	может быть представлено в виде суммы	169 discrimination circuit	схема различения
123 can be rewritten	можно переписать	170 do not depend on	не зависят от
124 can be shown	может быть показано	171 does not depend on	не зависит от
125 can be used	могут быть использованы	172 drops to a point	падает до точки
126 can be used as	можно использовать в качестве	173 during	в течение
127 can be used for	можно использовать для	174 during the transit time of the electron in the atom	за время пробега электрона в атоме
128 can be written	может быть записано; можно записать	175 during the transit time of the electron in the electric field	за время пробега электрона в электрическом поле
129 can be written as follows	можно записать в следующем виде	176 during the transit time of the electron in the magnetic field	за время пробега электрона в магнитном поле
130 can be written in the form	можно написать в виде	177 electron bombardment	электронная бомбардировка
131 can occur	может иметь место	178 even in the case	даже в случае
132 can take place	может иметь место	179 external circuit current	ток во внешней цепи
133 cannot be	не могут быть; не может быть	180 for all gases	для всех газов
134 cannot be greater than	не может быть больше, чем	181 for all values	для всех значений
135 cannot be represented	не могут быть представлены	182 for some	в течение нескольких
136 cathode-ray tube	электроннолучевая трубка	183 for some time	в течение некоторого времени
137 clock system	система синхронизации	184 for the characteristic of the cathode temperature	для характеристики температуры катода
138 collector current	ток коллектора	185 for this case	для этого случая
139 communication changes	каналы связи	186 for this purpose it is necessary	для этого необходимо; для этого нужно
140 communications field	область связи	187 for this purpose it is necessary to keep in mind that his paper	для этого необходимо иметь в виду, что его статья
141 compared with	по сравнению с	188 for this purpose one has to bear in mind that to obtain all gases	для этого необходимо иметь в виду, что для получения всех газов
142 could be expected	можно было бы ожидать	189 from the cathode to the anode	от катода к аноду
143 could be obtained	можно было бы получить	190 from zero to	от нуля до
144 could be obtained from the expression	можно было бы получить из выражения	191 gives X	дает X
145 could be obtained from the equation	можно было бы получить из формулы	192 goes to zero	стремится к нулю
146 current gain	коэффициент усиления по току	193 greatly depends on	сильно зависит
147 current in the external circuit	ток во внешней цепи	194 grid currents and voltages	токи и напряжения на сетке
148 currents and voltages	токи и напряжения на сетке	195 has a maximum value	имеет максимальное значение
149 cycle time	время цикла	196 has the bulk value	имеет большую величину
150 decay time	время затухания	197 has the form	имеет вид
151 decreases with increasing	уменьшается с увеличением	198 has the form of the curve	имеет вид кривой
152 depending on the magnitude	в зависимости от величины	199 have a higher	обладают более высокой
153 depending on the output voltage of the system	в зависимости от напряжения на выходе системы	200 however in the case	однако в случае
154 depending on the resistance	в зависимости от величины сопротивления	201 if at the same time for increasing the voltage	если при этом для увеличения напряжения
155 depending on the type	в зависимости от типа		
156 depends greatly on	сильно зависит от		
157 depends on the anode voltage	зависит от напряжения на аноде		
158 depends on the distance between	зависит от расстояния между		

202 if at the same time the voltage is not greater than at the anode
203 if the output voltage
204 in all directions
205 in case
206 in operation
207 in order to obtain an expression for
208 in relation to the cathode
209 in spite of the fact that
210 in the anode circuit
211 in the base
212 in the base region
213 in the case of
214 in the circuit in Fig.
215 in the circuit with a common base
216 in the circuit with a common emitter
217 in the collector circuit
218 in the course of
219 in the form of ions
220 in the high voltage region
221 in the last case
222 in the space charge region
223 in the temperature range
224 in these devices
225 increases at the expense of
226 increases from zero to a maximum value
227 increases, which
228 increases, which makes it possible to consider that
229 increases with increasing temperature
230 insertion-loss characteristic
231 is a function of current in the circuit of the emitter
232 is a function of distance between
233 is a function of energy
234 is a function of frequency
235 is a function of frequency of the signal
236 is a function of temperature
237 is a function of voltage
238 is defined by the expression
239 is delayed in relation to the input signal
240 is determined by the condition
241 is determined by the equation
242 is determined by the formula
243 is effected by electrons
244 is fed to the grid of the tube
245 is formed as a result of bombardment
246 is formed as a result of collision ionization
247 is formed as a result of diffusion

если при этом напряжение не больше, чем на аноде
если напряжение на выходе во всех направлениях в случае
во время работы
чтобы получить выражение для
по отношению к катоду
несмотря на то, что
в цепи анода
в базе
в области базы
в случае
в схеме на рис.
в схеме с общей базой
в схеме с общим эмиттером
в цепи коллектора
в течение
в виде ионов
в области высоких напряжений
в последнем случае
в области пространственного заряда
в области температур
в этих приборах
увеличивается за счет
возрастает от нуля до максимального значения
возрастает, что
возрастает, что позволяет считать, что
возрастает с увеличенным температурой
характеристика потерь на вводе
зависит от величины тока в цепи эмиттера
зависит от расстояния между
зависит от энергии
зависит от частоты
зависит от частоты сигнала
зависит от температуры
зависит от напряжения
определяется выражением
запаздывает по отношению к входному сигналу
определяется из условия
определяется по уравнению
определяется по формуле
производится при помощи электронов
подается на сетку лампы
образуется в результате бомбардировки
образуется в результате ударной ионизации
образуется в результате диффузии

248 is heated to a high temperature
249 is large compared to
250 is large compared with
251 is observed for several hours
252 is obtained as a result of collision
253 is important
254 is not important
255 is of great importance
256 is of importance
257 is of no importance
258 is of particular importance
259 is possible only in the region
260 is produced as a result of bombardment
261 is seen from
262 is seen from Fig.
263 is seen from the table
264 is sufficient to
265 is that
266 is that in order to
267 is to
268 is used as
269 is X
270 is written as
271 it appears that
272 it can be assumed that
273 it can be calculated that
274 it can be concluded that
275 it can be considered
276 it can be considered that
277 it can be determined that
278 it can be expected that
279 it can be observed that
280 it can be said that
281 it can be seen that
282 it can be shown that
283 it follows from the equation that
284 it follows that
285 it is known that
286 it is noted that
287 it is observed that
288 it is seen that
289 it is shown that
290 it is possible that
291 it may be anticipated that
292 it may be assumed that
293 it may be concluded that
294 it may be determined that

нагревается до высокой температуры
велико по сравнению с
велико по сравнению с
наблюдается в течение нескольких часов
получается в результате соударения
имеет большое значение;
играет важную роль
не играет никакой роли
имеет большое значение;
играет большую роль
имеет важное значение;
играет важную роль
не играет никакой роли
играет существенную роль
возможно только в области
образуется в результате бомбардировки
видно из
видно из фиг.
видно из таблицы
достаточно для того, чтобы
заключается в том, что;
состоит в том, что
заключается в том, чтобы
состоит в том, чтобы
используется в качестве
имеет вид ϕ .;
равно ϕ .
записано в виде
записывается в виде
оказывается, что
можно предположить, что
можно найти, что
можно заключить, что
его можно рассматривать
можно считать, что
можно определить, что
можно ожидать, что
можно заметить, что
можно сказать, что
можно видеть, что
можно показать, что
из формулы следует, что
из этого следует, что
известно, что
отмечается, что
наблюдается, что
видно, что
показано, что
возможно, что
можно ожидать, что
можно допустить, что;
можно предполагать, что;
можно предположить, что
можно заключить, что
можно определить, что

295 it may be expected that
 296 it may be mentioned that
 297 it may be noted that
 298 it may be remarked that
 299 it may be said that
 300 it may be seen that
 301 it may be shown that
 302 it may be supposed that
 303 it might be thought that
 304 it should be noted that
 305 it turns out that
 306 it was assumed that
 307 it was established that
 308 it was found that

309 it was known that
 310 it was noted that
 311 it was noticed that
 312 it was observed that
 313 it was pointed out that
 314 it was proved that
 315 it was realized that
 316 it was seen that
 317 it was shown that
 318 it was stated that
 319 it will be found that
 320 junction capacitance
 321 just as for
 322 just as in
 323 leads to
 324 makes it possible to determine
 325 makes it possible to find
 326 makes it possible to obtain
 327 many times
 328 may be calculated from X
 329 may be considered as
 330 may be constructed
 331 may be expressed
 332 may be found by
 333 may be greater than in the case
 334 may be neglected
 335 may be obtained
 336 may be obtained from equation
 337 may be quite
 338 may be regarded as
 339 may be represented by a sum

340 may be rewritten
 341 may be used as
 342 may be used for
 343 may be written as follows
 344 may be written in the form
 345 may occur
 346 memory access

347 might be expected
 348 must be equal to

можно ожидать, что
 можно отметить, что
 можно отметить, что
 можно отметить, что
 можно сказать, что
 можно видеть, что
 можно показать, что
 можно полагать, что
 можно думать, что
 необходимо отметить, что
 оказывается, что
 было принято, что
 было установлено, что
 было выяснено, что;
 было найдено, что;
 было обнаружено, что
 было известно, что
 было отмечено, что
 было замечено, что
 было замечено, что
 было указано, что
 было доказано, что
 было понятно, что
 было видно, что
 было показано, что
 было сказано, что
 будет установлено, что
 емкость коллекторного перехода
 точно так же, как и для
 точно так же, как и в
 приводит к
 дает возможность определять
 дает возможность найти
 дает возможность получить
 во много раз
 может быть вычислена по ϕ .
 можно рассматривать как
 могут быть построены
 может быть выражена
 можно найти по
 может быть больше, чем в случае
 можно пренебречь
 можно получить
 могут быть получены по формуле
 может быть весьма
 можно рассматривать как
 может быть представлено в виде
 суммы
 можно переписать
 можно использовать в качестве
 можно использовать для
 можно записать в следующем виде
 можно записать в виде
 может иметь место
 выборка из запоминающего устрой-
 ства
 можно было бы ожидать
 должен быть равен;
 должна быть равна;

546

349 must be greater than X

350 must be large enough
 351 must be X

352 must be zero
 353 must have the form
 354 occurs as a result
 355 of charge carriers in the form
 356 of collector current at constant
 voltage
 357 of electrons in a unit
 358 of electrons in it
 359 of electrons in the atom
 360 of electrons in the band
 361 of electrons in the crystal
 362 of electrons in the metal
 363 of electrons in the semiconductor
 364 of holes and electrons in the band
 365 of holes and electrons in the mag-
 netic field

366 of holes and electrons in the se-
 miconductor
 367 of holes in the base
 368 of holes in the region
 369 of the anode current and voltage
 370 of the anode voltage
 371 of the concentration of current
 carriers
 372 of the current in the anode circuit
 373 of the current in the collector
 circuit
 374 of the current source
 375 of the electric field in the base
 376 of the emitter to the collector
 377 of the external electric field
 378 of the gas discharge voltage
 379 of the input and output voltage

380 of the input voltage of the ampli-
 fier
 381 of the negative value
 382 of the output voltage of the ampli-
 fier
 383 of the resistance in the anode
 circuit
 384 of the reverse current
 385 of the same order
 386 of the signal at the amplifier
 output
 387 of the signal at the output
 388 of the voltage across the gas-
 discharge diode

должно быть равно;
 должны быть равны
 должен быть больше ϕ .;
 должна быть больше ϕ .;
 должно быть больше ϕ .
 должен быть достаточно большим
 должен быть равен ϕ .;
 должна быть равна ϕ .;
 должно быть равно ϕ .;
 должна иметь вид ϕ .
 должна быть равна нулю
 должна иметь вид
 происходит в результате
 носителей заряда в форме
 коллекторного тока при постоян-
 ном напряжении
 электронов в единице
 электронов в нем
 электронов в атоме
 электронов в зоне
 электронов в кристалле
 электронов в металле
 электронов в полупроводнике
 дырок и электронов в зоне
 дырок и электронов в магнитном
 поле

дырок и электронов в полупро-
 воднике
 дырок в базе
 дырок в области
 анодного тока и напряжения
 напряжения на аноде
 концентрации носителей тока
 тока в цепи анода
 тока в цепи коллектора

источника тока
 электрического поля в базе
 эмиттера к коллектору
 внешнего электрического поля
 напряжения газового разряда
 входного и выходного напряже-
 ния
 напряжения на входе усилителя
 отрицательного значения
 напряжения на выходе усилителя
 сопротивления в цепи анода
 обратного тока
 одного и того же порядка
 сигнала на выходе усилителя
 сигнала на выходе
 напряжения на газоразрядном
 диоде

389 on the cathode temperature	от температуры катода	431 storage system	система памяти
390 on the collector voltage	от напряжения на коллекторе	432 store output	мощность запоминающего устройства
391 on the surface of the cathode	на поверхности катода	433 takes place	имеет место
392 on the surface of the metal	на поверхности металла	434 takes place as a result of	происходит в результате
393 on the voltage at the electrodes	от напряжения на электродах	435 taking into account that in many cases	принимая во внимание, что во многих случаях
394 one can show that	можно показать, что	436 that as a result of ionization	что в результате ионизации
395 one may conclude that	можно заключить, что	437 that at the same time	что при этом
396 one may note that	можно отметить, что	438 that can be	которая может быть
397 one may say that	можно сказать, что	439 that if at the same time the current increases	что если при этом ток увеличивается
398 only in that case	только в том случае	440 that if at the same time the input voltage increases	что если при этом напряжение на входе возрастает
399 only in this case	только в этом случае	441 that in a number of	что в ряде
400 operating time	время работы	442 that in germanium	что в германии
401 oscillator frequency	частота генератора	443 that in that	что в том
402 otherwise	в противном случае	444 that in the base region	что в области базы
403 owing to	за счет	445 that in the case of	что в случае
404 owing to the decrease	за счет уменьшения	446 that in the circuit	что в цепи
405 owing to the increase	за счет увеличения	447 that in the circuit of Fig.	что в схеме рис.
406 owing to the change	за счет изменения	448 that in the first	что в первом
407 potential distribution in	распределение потенциала в	449 that in the following	что в дальнейшем
408 ranges from X to X	лежит в пределах от ϕ . до ϕ .	450 that in the high temperature range	что в области высоких температур
409 ranges from X to X	изменяется в пределах	451 that in the present case	что в данном случае
410 saturation current	ток насыщения	452 that in the space charge region	что в области пространственного заряда
411 search time	время поиска	453 that in this	что в этом
412 several times greater than expected	в несколько раз больше, чем это предполагалось	454 that in this case the electrode voltage	что при этом напряжение на электродах
413 several times greater than in the case of	в несколько раз больше, чем в случае	455 that in this case the grid voltage	что при этом напряжение на сетке
414 shift time	время сдвига	456 that in this case the voltage at the collector	что при этом напряжение на коллекторе
415 should be such that	должна быть такой, что	457 that in this case the voltage at the input	что при этом напряжение на входе
416 shows that	показывает, что;	458 that the relation	что зависимость
417 since all valent electrons can pass to higher	говорит о том, что;	459 that the value is given here	что значение дается здесь
418 since between the electrons	свидетельствует о том, что	460 that to obtain all the values	что для получения всех значений
419 since in the case of	так как все валентные электроны могут переходить на более высокие	461 the amplification factor has the form	коэффициент усиления имеет вид
420 since the anode voltage	так как между электронами	462 the anode voltage	напряжение на аноде
421 since the capacitor voltage	так как в случае	463 the average current value in the collector circuit	среднее значение тока в цепи коллектора
422 since the electrode voltage	так как напряжение на аноде	464 the case is that	дело в том, что
423 since the electrons	так как напряжение на конденсаторе	465 the circuit of the grid	цепь сетки
424 since the grid voltage of the tube	так как напряжение на электродах	466 the current in the anode circuit	ток в цепи анода
425 since the input voltage	так как электроны	467 the current in the base circuit	ток в цепи базы
426 since the magnitude of the alternating current in the anode circuit	так как напряжение на сетке лампы	468 the current in the collector circuit	ток в цепи коллектора
427 since the magnitude of the alternating current in the circuit of the emitter	так как величина переменного тока в анодной цепи	469 the current in the grid circuit	ток в цепи сетки
428 since the magnitude of the alternating current in the grid circuit	так как величина переменного тока в цепи эмиттера	470 the electrode voltage	напряжение на электродах
429 since the voltage at the collector	так как величина переменного тока в цепи сетки	471 the electron energy distribution	распределение электронов по энергии
430 since with increasing the voltage	так как напряжение на коллекторе	472 the electrons will move	электроны будут двигаться
	так как с увеличением напряжения	473 the equation shows that	из формулы следует, что
		474 the grid voltage of the tube	напряжение на сетке лампы
		475 the higher the concentration of holes in the base	чем выше концентрация дырок в базе

476 the lifetime of nonequilibrium holes	время жизни неравновесных дырок	524 was shown in	было показано в
477 the magnitude of the current can be	величина тока может быть	525 we can assume that	можно предположить, что
478 the magnitude of the current can be greater than	величина тока может быть больше, чем	526 we can conclude that	можно заключить, что
479 the output voltage	напряжение на выходе	527 we can expect that	можно ожидать, что
480 the output voltage of the amplifier	напряжение на выходе усилителя	528 we can see that	можно видеть, что
481 the output voltage of the circuit	напряжение на выходе схемы	529 we have that	имеем, что
482 the point is that	дело в том, что	530 we find that	находим, что
483 the rise time	время нарастания	531 we obtain the differential equation	получаем дифференциальное уравнение
484 the temperature of the diode	температура диода	532 we obtain the equation	получаем уравнение
485 the transit time of electrons in the atom	время пролета электронов в атоме	533 we obtain the following equation	получаем следующее уравнение
486 the transition of electrons from region to region	переход электронов из области в область	534 we obtain the following expression	получаем следующее выражение
487 the transition of the electron from the valence band	переход электрона из валентной зоны	535 we obtain X	получаем ϕ .
488 the tunnel diode capacitance	емкость туннельного диода	536 we see that	видно, что
489 the tunnel diode characteristic	характеристика туннельного диода	537 we shall consider that	будем считать, что
490 the tunnel diode circuit	схема на туннельных диодах	538 we shall obtain the equation for	получим выражение для
491 the voltage at the input and output of the amplifier	напряжение на входе и выходе усилителя	539 when fed to	при подаче на
492 the voltage at the output	напряжение на выходе	540 whereas	в то время как
493 the voltage of the collector	напряжение коллектора	541 whereas in the circuit	в то время как в схеме
494 then for determining	то для определения	542 whereas in the circuit	в то время как в цепи
495 this follows from	это следует из	543 whereas in this case	в то время как при этом
496 this follows from Fig.	это следует из рис.	544 which can be	которое может быть
497 this takes place	это имеет место	545 while	в то время как
498 through the electron-hole junction	через электронно-дырочный переход	546 will be a function of	будет зависеть от
499 through the emitter junction	через эмиттерный переход	547 will be determined by the equation	будет определяться формулой
500 through the junction	через переход	548 will be determined by the value	будет определяться величиной
501 times greater than for the tube	раз больше, чем для лампы	549 will be equal to zero	будет равен нулю
502 times greater than from the cathode surface	раз больше, чем с поверхности катода	550 will be greater than	будет больше, чем
503 times greater than this	раз больше, чем это	551 will be in the range	будет находиться в пределах
504 times greater than with	раз больше, чем с	552 will be seen from Table	будет видно из табл.
505 times larger than on the electrodes	раз больше, чем на электродах	553 will be seen in Fig.	будет видно из фиг.
506 to a high temperature	до высокой температуры	554 will be shown below	будет показано ниже
507 to give rise to	вызвать	555 will be shown in	будет показано в
508 to obtain an expression for	чтобы получить выражение для	556 will be shown later	будет показано далее
509 under these conditions	в этом режиме	557 will be X	будет равно ϕ .;
510 until	до тех пор, пока	558 will be zero	будет равно ϕ .;
511 voltage drop	падение напряжения	559 will decrease	будет равен ϕ .;
512 voltage gain	коэффициент усиления напряжения	560 will depend on	будет иметь вид ϕ .
513 voltage-current characteristic	вольт-амперная характеристика	561 will increase	будет равен нулю
514 volumetric charge regions	области объемного заряда	562 will lie in the range	будет уменьшаться
515 was carried out	было проведено	563 will occur	будет зависеть от
516 was detected in	было обнаружено в	564 will take place	будет увеличиваться
517 was described in	было описано в	565 will vary from	будет находиться в пределах
518 was done in	было сделано в		будет иметь место
519 was found in	было выяснено в		будет иметь место
520 was given	было дано		будет изменяться от
521 was investigated	было исследовано		
522 was observed in	было отмечено в		
523 was obtained	было получено		

Статистическое построение словарей машинных оборотов имеет не только прагматический смысл, но может рассматриваться и в русле экспликации лингвистической интуиции человека-переводчика. Действительно, наблюдение над поведением человека, переводящего иностранный текст, особенно в условиях синхронного перевода, показывает, что переводчик не так уж часто обращается к схемам порождения предложения, заложенным в системе языка. Чаще всего он использует готовые, освященные нормой

штампы, не всегда точно соответствующие правилам системы. Каков объем этого словаря штампов? Насколько гибко и эффективно он может использоваться в переводе? На все эти вопросы помогает ответить статистическое исследование и машинное моделирование единиц наполнения текста.

ЛИТЕРАТУРА

1. Алексеев П. М., Нехаёв О. А. Статистические характеристики единиц английского языка в текстах по радиоэлектронике. ЭЯСтР.
2. Белоцерковская Л. И. Статистика трехсловных сочетаний с опорной глагольной словоформой или словоформой, принадлежащей к категории состояния (на материале русских текстов по радиоэлектронике). СТ I.
3. Булашева Н. С. Статистика трехсловных сочетаний с опорным частотным словом из текстов по русской радиоэлектронике. СТ I.
4. Вертель В. А., Вертель Е. В., Кричевич В. С., Пиотровский Р. Г., Трибис Л. И. Автоматические словари в системе бинарного вероятностного МП. ИЛ.
5. Дәнейко М. В. Статистика глагольных словосочетаний в английских текстах по радиоэлектронике. СТ I.
6. Дәнейко М. Н., Машкина Л. Е., Нехаёв О. А., Соркина В. А., Шарайда А. Н. Статистическое исследование дистрибуции словоформы. СтР.
7. Зуева Т. Р. Статистическая характеристика четырехсловных сочетаний русского текста по электронике. СТ I.
8. Псабекова Н. П. Именные словосочетания в русских текстах по радиоэлектронике. СТ I.
9. Моль А. Теория информации и эстетическое восприятие. М., 1966.
10. Рахубо Н. П., Тарасова Е. С., Попеску А. Н. К вопросу об автоматическом индексировании и аннотировании научно-технических текстов. УЗК.
11. Соркина В. А. Статистика именных словосочетаний в английских текстах по радиоэлектронике. СТ I.
12. Chomsky N., Miller C. A. Finitary Models of Language Users. In: Handbook of Mathematical Psychology, vol. 2, New York, 1963 (русск. пер. см.: Кибернетический сб., новая серия, вып. 4, М., 1967).
13. Kuno S. and Oettinger A. G. Multipath Syntactic Analyser. In: Mathematical Linguistics and Automatic Translation, Computation Laboratory of the Harvard University, Report № NSF-8, Cambridge Mass., 1963 (русск. пер.: АП).
14. Miller G. A., Selfridge I. A. Verbal Context and the Recall of Meaningful Material. American Journal of Psychology, vol. 63, 2, 1950.
15. Rhodes I., Alt F. L. Hindsight Technique in Machine Translation of Natural Languages. Journal of Research of National Bureau of Standards, B. Mathematics and Mathematical Physics, 66 B, 1962, № 2 (русск. пер.: АП).

Л. И. Трибис

ОБ ОДНОЙ МОДЕЛИ РАСПОЗНАВАНИЯ ЛЕКСИЧЕСКИХ ЗНАЧЕНИЙ НЕОДНОЗНАЧНЫХ СЛОВ

Передача информации в коммуникативной системе включает порождение сообщения отправителем, передачу его от отправителя к адресату, восприятие адресатом смысла сообщения. Неотъемлемым условием последнего этапа является распознавание лексических значений единиц сообщения. Если языковой нормой за некоторой единицей закреплено несколько лексических значений, задача адресата заключается в устранении лексической неоднозначности.

Распознавание лексических значений неоднозначных единиц производится человеком автоматически, на основе некоторых имплицитных правил. Поскольку отправитель и адресат сообщения в одинаковой степени владеют этими правилами, лексическая неоднозначность не препятствует их общению.

Иначе обстоит дело в системах автоматического перевода, аннотирования, индексирования, реферирования, представляющих коммуникативный комплекс «человек — машина — человек». В указанных системах участником процесса коммуникации выступает ЭВМ, которая не обладает знанием правил декодирования значений. Эти правила необходимо машине задать. Следовательно, устранение лексической неоднозначности в системе «человек — машина — человек» предполагает моделирование соответствующего «человеческого» процесса.

В современном языкознании разграничиваются внутриязыковой и межъязыковой планы исследования. Соответственно этому будем различать два типа лексической неоднозначности: одноязычную и двуязычную. Одноязычной лексической неоднозначностью назовем одно-многозначное соотношение между означающим и означаемым языкового знака, рассматриваемого как единица лексического уровня языка. Двуязычной лексической неоднозначностью назовем одно-многозначное соотношение между входным

языковым знаком и выходными знаками, соотносимыми с ним на лексическом уровне в двуязычной ситуации.¹

Определим своей целью построение лингвистической модели устранения двуязычной лексической неоднозначности и исследование путей ее реализации в системе «человек — машина — человек».

Часть I. ЛИНГВИСТИЧЕСКИЙ АНАЛИЗ ЕДИНИЦ, ОБЛАДАЮЩИХ ДВУЯЗЫЧНОЙ ЛЕКСИЧЕСКОЙ НЕОДНОЗНАЧНОСТЬЮ

§ 1. Критерии и методика отбора материала исследования

При МП выдвигается требование устранения лексической неоднозначности не отдельных слов, а целых классов единиц. Ставится задача разработки наиболее общих методов, которые позволяют оперировать группами диагностирующих признаков (ДП).

В соответствии с этим материалом исследования выбрана группа английских общетехнических словоформ, выделенная на основе формально-семантических критериев. Исходили из предположения, что сходство семантических характеристик исследуемых единиц проявляется в некоторых сходных контекстных условиях их реализации. Предполагалось, что это обстоятельство позволит оперировать классами ДП. Таким образом, ставилась задача разработки типового подхода к снятию неоднозначности.

Общетехническими называем термины и единицы полутерминологического характера,² употребительные в технических текстах и имеющие семантические характеристики, не меняющиеся от подязыка к подязыку.³

Учитывая изложенное, можно свести нашу задачу к выделению группы лексических единиц, обладающих следующими свойствами: 1) употребительность в технических текстах, достаточная частотность;⁴ 2) неспециализированный, общетехнический ха-

рактер значений; 3) независимость семантических характеристик от тематики текстов.

Обследование всех текстов по технике невыполнимо. Ограничиваемся рассмотрением английских подязыков автомобилестроения, вычислительной техники, строительных материалов и электронных приборов. Выделение общетехнических единиц проведено по методике, разработанной в группе «Статистика речи». Сравнивались словники частотных словарей по указанным подязыкам. Употребительными мы считали словоформы, зафиксированные в каждом из частотных словарей. Из числа слов, обладающих этим качеством, были исключены единицы, являющиеся общеязыковыми элементами или терминами специальных отраслей знания.⁵ Этот этап работы проводился по данным словарей и специалистов-переводчиков.

Из полученного списка отбирались слова, обладающие идентичностью общетехнических значений в исследуемых подязыках. Идентичность значений устанавливалась сначала по данным словарей, а затем проверялась анализом текста. Для исследования было отобрано 137 словоформ, представляющих 50 английских слов:⁶ adaptation, adjust, adjustment, allowance, arrange, arrangement, aspect, attachment, behaviour (behavior), capability, capacity, classify, compensate, compensation, conduction, conductivity, construction, definition, density, derive, detect, detection, determination, difference, efficiency, evaluate, evaluation, excitation, excite, failure, indicate, indication, induce, install, installation, integrate, leakage, maintain, maintenance, multiply, multiplication, operate, operation, penetrate, performance, production, ratio, resolve, saturate, saturation.

§ 2. Анализ исследуемых слов с точки зрения одноязычной ситуации

Для исследования отобранных слов был проведен лингвистический эксперимент. Из четырех массивов текста по 1,5 млн словоупотреблений в каждом из английских подязыков автомобилестроения, вычислительной техники, строительных материалов и электронных приборов произведена сплошная выборка 30 тыс. предложений, реализующих значения анализируемых слов. Собранный материал был переведен на русский язык специалистами-переводчиками по соответствующим отраслям техники. Анализ проводился в два этапа. На первом этапе рассматривались харак-

⁵ Термином считали лексическую единицу, которая выражает и классифицирует определенное научное понятие.

⁶ При исследовании письменной речи различаем словоупотребление (букваряд, заключенный между двумя пробелами в тексте), словоформу (полностью совпадающее словоупотребление) и слово (морфологически производные от одной лексической основы словоформы).

¹ Определение двуязычной ситуации приводится в работе: Р. Г. П и о т р о в с к и и В. А. Ч и ж а к о в с к и й. О двуязычной ситуации. СтРААТ, стр. 444.

² А. В. Ф е д о р о в. Основы общей теории перевода. М., 1968, стр. 279.

³ Из рассмотрения исключаются случаи, когда однозначному термину входного языка соответствует несколько переводных эквивалентов, являющихся терминами-дублетами (например, английскому ботаническому термину *Alerro grass* соответствуют русские термины-дублиеты: 1) гумай; 2) Джонсова трава, 3) сорго алепское.

⁴ Стремление рассматривать частотные единицы закономерно, так как между частотой употребления слова и его неоднозначностью существует прямая зависимость.

теристики слов с точки зрения одноязычной ситуации. На втором этапе учитывалась двуязычная ситуация. Результат первой стадии анализа отражен в таблице (стр. 134, фрагмент), которая указывает для каждого слова:

1) значения, реализуемые в текстах данных тематик (для наблюдения условий одноязычной ситуации значения описываются по английскому толковому словарю);

2) употребительность значений по подъязыкам (величина, характеризующаяся отношением числа употреблений слова в данном значении к общему числу употреблений этого слова во всех значениях).

Семантическая структура исследуемых слов

№ №	Английские слова	Значения	Употребительность значения в подъязыках (в %)			
			автомобильные средства	строительные материалы	вычислительная техника	электронные приборы
1	adjust	put in order	0.1	5.5	0	0
		set right	93.6	89	87.7	91.1
		make a thing fit	6.3	5.5	12.3	8.9
2	arrange	put in order	64	49.3	61.5	48.9
		settle, organize	16	46.2	34.9	40.0
		make a thing fit	18.7	4.5	1.2	11.1
		make plans	1.3	0	2.4	0
3	classify	arrange in classes	100	100	100	100

Исследуемые существительные обладают одноязычной лексической неоднозначностью. Среди глаголов однозначны только два.

Данные таблицы показывают, что рассматриваемые лексические единицы не являются терминами специальных отраслей знания. При этом некоторые единицы передают, наряду с нетерминологическими, также терминологические общетехнические значения (detect, integrate, multiply, saturate).

Рассмотрим характер лексических значений глаголов. Они обозначают действия в самом широком, обобщенном смысле: действия с целью упорядочения, систематизации, улучшения (adjust, arrange, classify, compensate); воздействия с целью изменения количества предметов или соотношения их частей (integrate, multiply); воздействия с целью изменения качества (excite, saturate, resolve); действие поддержания свойства (maintain); перемещение предметов (install, penetrate); получение, обнаружение (derive, induce, detect); функционирование (operate); абстрактные действия (evaluate, indicate). По признаку переходности глаголы делятся на две группы: а) глаголы переходного значения (classify, detect, excite, induce, install, integrate); б) глаголы с переходными и непереходными значениями (adjust, arrange, compensate, derive, evaluate, indicate, maintain, multiply, operate, penetrate, resolve, saturate).

Исследуемые существительные принадлежат к абстрактной лексике. Они представлены двумя основными группами.

Первую группу составляют имена действия со словообразовательными суффиксами -ion, -ment, -ance (-ence), -age, -ure: adaptation, adjustment, allowance, arrangement, attachment, compensation, conduction, construction, definition, detection, determination, evaluation, excitation, failure, indication, installation, leakage, maintenance, multiplication, operation, performance, production, saturation).

Эти отглагольные существительные обладают значениями процесса (действия), состояния и результата действия.

Вторая группа представлена абстрактными существительными, обозначающими свойства, качества, состояния, отвлеченные от предмета: aspect, behaviour (behavior), capability, capacity, conductivity, density, difference, efficiency, ratio. Широта семантики этих существительных проявляется в чистом виде в условиях изоляции. В речи она получает известное сужение и конкретизацию. В сочетании с разными классами слов реализуются разные значения неоднозначных единиц. Лексическая неоднозначность и широкая понятийная отнесенность — взаимообусловленные характеристики.

Специфической чертой исследуемых слов является общетехнический характер их лексических значений. Семантические характеристики каждого слова остаются постоянными в текстах разной тематической направленности. Это позволяет считать, что исследуемые единицы представляют лексическую группу, обладающую общностью внутриязыковых семантических характеристик.

§ 3. Анализ исследуемых слов с точки зрения двуязычной ситуации

Анализ показал, что исследуемые единицы обладают одноязычной лексической неоднозначностью. При переводе, как известно, проблему представляет снятие двуязычной лексической неоднозначности. Поскольку между одноязычной и двуязычной неоднозначностью может не быть прямой зависимости, необходимо рассмотреть семантические характеристики интересующих нас единиц в двуязычной (англо-русской) ситуации.

Тридцать тысяч английских предложений с исследуемыми словами были сопоставлены с их переводами на русский язык. Карточка была систематизирована. Карточки, представляющие употребление одного слова, классифицировались по подъязыкам, внутри каждого подъязыка — по значениям, внутри каждого значения — по переводам. Были получены данные об употребительности каждого перевода относительно значения, которое он представляет, и относительно других подъязиков. Для каждого

английского анализируемого слова составлен конкорданс, т. е. списки контекстов и их переводов на русский язык (с указанием частот для каждого подязыка), ориентированные по значениям английского слова, а внутри значения — по переводам. Данный конкорданс может быть назван двуязычным, так как представляет двуязычную (англо-русскую) ситуацию. На основе информации конкордансов составлены таблицы и вычерчены графы, представляющие семантическую структуру единиц в одноязычной и двуязычной ситуациях. Таблицы включают частоты переводов в каждом из подязыков. Англо-русские соответствия в таблицах свидетельствуют о многовариантности русских переводов, соответствующих одному и тому же значению английского слова. Многовариантность переводов сама по себе не служит критерием двуязычной неоднозначности, так как переводы могут быть синонимичными и представлять двуязычную квази-неоднозначность. Переводы проверялись на синонимичность. В русский аналог английского предложения вместо эквивалента неоднозначного слова подставлялись все русские переводы, указанные для него в таблице. Взаимозаменяемость слов в контексте оценивалась специалистами-переводчиками. Переводы признавались синонимичными, если могли заменять друг друга хотя бы в одном контексте, не нарушая смысла высказывания. Синонимичные переводы сводились в ряды и отделялись от несинонимичных. Каждый синонимический ряд заменялся одним эквивалентом, доминантой ряда. Количество несинонимичных переводов служило критерием двуязычной характеристики входного слова. Анализ показал, что все исследуемые английские слова обладают двуязычной лексической неоднозначностью.

§ 4. Диагностирующие признаки

Анализ выявил специфику одноязычной и двуязычной лексических характеристик исследуемых входных единиц. Эта специфика состоит в их известной самостоятельности и независимости. Действительно, если обратиться к таблицам, представляющим семантическую структуру анализируемых единиц, то можно сделать некоторые выводы. Внутриязыковые значения каждой единицы неадекватно представлены переводными эквивалентами.⁷ Эта неадекватность также проявляется в различии способов разграничения как внутриязыковых значений (одноязычная неодно-

⁷ Адекватность понимается как одно-однозначное соответствие между значением и переводом: одному значению соответствует один и только один переводной эквивалент. Исследуемые слова хорошо подтверждают мысль Л. В. Щербы о том, что слова разных языков находятся в весьма сложных и многообразных отношениях; см.: Л. В. Щ е р б а. Русско-французский словарь. Изд. 4-е, М., 1955, стр. 4.

значность), так и переводов (двуязычная неоднозначность). У существительного *adjustment* три значения: *an action of adjusting*, *a condition of being adjusted*, *a means of adjusting*. Разграничение значений может быть проведено с помощью дистрибутивного метода. Так, временные предлоги, предшествующие существительному *adjustment*, диагностируют первое значение. Неопределенный артикль диагностирует третье значение. Однако эти формальные средства ничего не дают для разграничения переводов, так как один и тот же перевод ('регулировка', 'юстировка' и т. п.) может представлять все три внутриязыковые значения.⁸ Признаками, диагностирующими переводные эквиваленты, являются семантические типы слов, сочетающихся с исследуемыми неоднозначными единицами. Так, всем трем значениям существительного *adjustment* соответствует русский эквивалент *юстировка*, если оно сочетается со словами, обозначающими названия приборов. Изложенное приводит к мысли о том, что в качестве метода снятия двуязычной лексической неоднозначности анализируемых единиц должен быть принят анализ лексической сочетаемости.⁹

Известно, что в границах сочетания слов формализация контекста возможна путем составления списка всех сочетающихся слов. Ставим своей целью выявление (по данным наших конкордансов) списков признаков, диагностирующих каждый из русских переводов английских слов.

Общность семантических характеристик входных единиц проявляется в сходстве контекстных условий их реализации. Значительная часть общетехнических словоформ реализует свои русские переводы в сочетании со словами некоторых семантических групп. Каждая такая группа диагностирует переводы ряда английских слов. Число неоднозначных словоформ, русские переводы которых диагностируются одной и той же семантической группой слов, достигает 38. Эмпирически выделено 47 семантических групп лексических диагностирующих признаков: «устройство», «деталь», «механизм», «элемент», «средство передвижения», «подъемный механизм», «вход—выход», «численные характеристики», «данные», «арифметические действия», «формула», «планирование», «предписание», «умственное восприятие», «способ управления», «физические характеристики», «тип конструкции», «напряжение», «мощность», «сигнал», «ток», «ограниченное пространство», «группировка», «форма», «частицы», «материал», «жидкость—газ», «тепло-

⁸ Среди технических терминов, обладающих неоднозначностью типа «предмет — процесс», есть и такие, значения которых могут быть идентифицированы русскими переводами; ср. *opening* — 'отверстие', 'открывание'.

⁹ Дистрибутивный метод автоматического снятия двуязычной лексической неоднозначности был реализован для английских глаголов более конкретного характера. При этом он применялся в сочетании с анализом лексической сочетаемости. См.: Ю. Н. М а р ч у к. Об алгоритмическом разрешении лексической многозначности. АКД, М., 1968.

вой», «степень влажности», «испытание», «определение», «меры длины», «меры площади», «меры объема», «память», «резервуар», «здание», «то, что имеет пропускную способность», «период времени», «графическое представление», «знак», «книга», «собрание», «агент действия».

Двуязычная лексическая неоднозначность исследуемых единиц устраняется также учетом сочетаемости с отдельными словоформами и списками слов, не образующих тематических групп. Используются несколько списков морфологических ДП. Морфологическая информация выступает для нас как семантическая. Например, суффиксы -ег, -ог в некоторых технических текстах указывают на название приборов.

При описании русских переводов в терминах ДП учитывались следующие обстоятельства.

В лексике любого языка взаимодействуют правила, порождающие сочетания слов, и правила, ограничивающие сочетаемость. Ограничение сочетаемости лексических единиц обеспечивается законами сложения смыслов, благодаря чему в речи функционируют сочетания, обладающие правильной семантической структурой. Вследствие открытого характера языковой системы, развивающейся во времени, меняется понятие правильности семантической структуры. Метафорическое употребление соединяет единицы, ранее семантически несовместимые: «космическая целина», «звездные братья» и т. п. Таким образом, список сочетающихся слов языка является открытым. Это свойство находится в противоречии с закрытым характером описания языка, вводимого в ЭВМ. Учитывая это обстоятельство, мы делаем списки ДП, предназначенные для ввода в ЭВМ, конечными — отбираем самые частые единицы. Для лексических ДП входного слова, не зафиксированных в наших списках, предусмотрен некоторый частый «общий» перевод.

При МП важное значение имеет указание возможной зоны поиска ДП в предложении. Определим своей задачей выяснение такой зоны.

Семантически реализуемое слово и ДП связаны сложными и многообразными отношениями. Они могут находиться как в контактной, так и дистантной дистрибуции.¹⁰ Между семантически сочетающимися словами может быть синтаксическая связь, но ее может и не быть. При определении минимального отрезка текста для поиска ДП приходится учитывать эти обстоятельства, а также особенности восприятия речи человеком.

¹⁰ Связию двуязычной неоднозначности в условиях контактной дистрибуции посвящены исследования: А. В. Зубов. Переработка текста естественного языка в системе «человек — машина» — человек. АҚД. Л., 1969; О. А. Нехай. Статистика и автоматический анализ текста (на материале английских текстов по электронике). АҚД. Минск, 1968; А. Н. Шранда. Выделение типовых контекстов и автоматический перевод. СТ II, стр. 33—153.

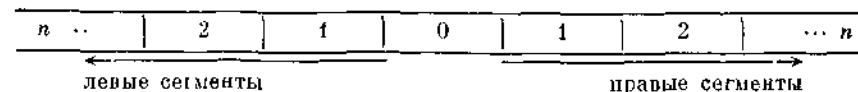
Предполагается, что при восприятии речи человек осмысливает поступающую информацию некоторыми «квантами», синтагмами. Процесс перевода может быть представлен как «селективное» понимание. Поэтому естественно считать синтагму также единицей перевода.

Автоматическая переработка речевой информации вносит своеобразие в трактовку единицы перевода. За эту единицу принимают сегмент — отрезок текста, вычленимый ЭВМ на основе учета формальных языковых средств. Сигнализаторами границ сегментов служат делимитаторы — формальные показатели синтаксических отношений. Так, артикли и предлоги являются левой границей субстантивных групп. Одновременно они служат определителями отношений с предшествующими или последующими глагольными группами. Формальным сигналом появления глагольной группы часто является вспомогательный или модальный глагол. С точки зрения человека (а не ЭВМ) сегменты являются семантически и логически оформленными синтаксическими конструкциями, имеющими формальные признаки подлежащего, сказуемого, беспредложного и предложного дополнений, атрибутивных, комплетивных и адвербиальных синтагм.¹¹

Позиционно различаем следующие типы сегментов, вычленимые относительно нулевого (сегмента с исследуемой неоднозначной словоформой):

- 1) нулевой сегмент;
- 2) первый левый сегмент, примыкающий к нулевому слева;
- 3) второй левый сегмент, примыкающий к первому левому сегменту слева; третий левый и т. д.;
- 4) первый правый сегмент, примыкающий к нулевому справа;
- 5) второй правый сегмент, примыкающий к первому правому справа; третий правый и т. д.

Размещение сегментов в предложении относительно нулевого можно представить схемой:



По предварительным данным, в английском научно-техническом тексте наиболее частыми являются сегменты длиной в две-три словоформы. По количеству сегментов самыми частыми являются предложения, состоящие из пяти-семи сегментов.

Был проведен эксперимент. По программе автоматической сегментации М. В. Данейко и В. М. Петровской обработан массив

¹¹ См.: М. В. Данейко, В. М. Петровская. Алгоритм автоматической сегментации английского предложения. СТ II, стр. 216—219.

технического английского текста в 16 тыс. словоупотреблений. Он включал 1000 предложений с исследуемыми неоднозначными единицами. Приводим фрагмент сегментации.

1. (A contribution) (to studies) (for obtainig) (a good realization) (would be) (the definition) (of synthesis procedures).

2. (The current calculation) (in itself) (is not a very accurate indicator) (of power carability).

Подсчеты, произведенные с рассегментированной выборкой, позволяют сделать некоторые выводы. Для неоднозначных глаголов и существительных наибольшей диагностирующей силой обладает зона, включающая нулевой сегмент и по два сегмента слева и справа от него. Число случаев, когда ДП обнаружены и за пределами этой зоны, незначительно (6.1%). Внутри зоны ДП располагаются следующим образом: в 50.4% обследованных предложений — в нулевом сегменте; в 37.7% предложений — в первом левом и первом правом; в 11.9% — во втором левом и втором правом. Иными словами, наибольшей диагностирующей силой обладает зона, находящаяся в непосредственной близости от неоднозначной словоформы. По мере удаления от нее вправо и влево диагностическая сила окружения ослабевает. Этот вывод согласуется с утверждением о том, что степень близости расположения слов в предложении отражает степень их смысловой связанности.¹²

§ 5. Некоторые итоги лингвистического анализа

Лингвистический анализ 137 английских словоформ проведен на материале 30 тыс. предложений и их русских переводов. Одноязычный анализ показал, что исследуемые единицы имеют общетехнический характер, так как отвечают сформулированному нами определению: употребительны в технических текстах, обладают общетехническими значениями. Их одноязычные семантические характеристики не меняются от подъязыка к подъязыку. В двуязычном плане они являются полутерминологическими элементами, так как среди русских переводов есть синонимичные единицы.

Характеризуясь широкой понятийной основой и абстрактностью лексических значений, исследуемые единицы обладают двуязычной лексической неоднозначностью. Устранение ее при переводе происходит благодаря наличию во входном тексте ДП. Исследованы ДП, определена примерная область их поиска во входных предложениях. Сходство семантических характеристик неоднозначных единиц проявляется в сходных контекстуальных условиях их реализации. Это создает возможность учета лексических ДП на уровне семантических классов.

¹² А. Я. Шайкевич. Распределение слов в тексте и выделение семантических полей. Иностранные языки в высшей школе, 1963, № 2, стр. 58.

Часть II. МОДЕЛИРОВАНИЕ УСТРАНЕНИЯ ДВУЯЗЫЧНОЙ ЛЕКСИЧЕСКОЙ НЕОДНОЗНАЧНОСТИ В СИСТЕМЕ «ЧЕЛОВЕК — МАШИНА — ЧЕЛОВЕК»

§ 1. Выбор системы моделей

Необходимым условием осуществления двуязычной автоматической переработки текста является снятие машинного типа двуязычной лексической неоднозначности. Выполнение этой задачи связывается с исследованием устранения двуязычной лексической неоднозначности человеком. Механизм информационных процессов, протекающих в мозгу при переводе неоднозначной единицы, не дан в непосредственном наблюдении. Исследователь сталкивается с ситуацией «черного ящика», на входе которого текст с неоднозначной единицей, на выходе — ее перевод. В этих условиях единственным способом исследования является метод опосредствованного познания, моделирование.

Определим своей задачей построение модели,¹³ которая получая входное предложение, находит единицы, обладающие двуязычной лексической неоднозначностью, определяет их адекватный перевод на выходной язык.

Согласно современной теории мышления в процессе решения проблемы человек обнаруживает существенные, но первоначально скрытые признаки элементов, из которых складывается проблемная ситуация. Это происходит путем повторных обращений к элементам проблемы.¹⁴

Можно предположить, что процесс перевода входной единицы, обладающей двуязычной лексической неоднозначностью, состоит из нескольких обращений к контексту. Вначале происходит распознавание внутриязыкового значения на основе ДП контекста. Определив его, человек обращается к памяти (или переводному словарю), где хранится система соответствий входного и выходного языков, которая, вероятно, имеет сложное строение. Она «записывается» как комплекс внутриязыковых значений, которым поставлены в соответствие выходные единицы (схема 1).

Определив внутриязыковое значение, человек сокращает поиск среди множества переводов ($\Phi_1^i, \Phi_2^i, \dots, \Phi_n^i$). Он переходит к тому подмножеству эквивалентов, которое соотносится только с данным внутриязыковым значением. Дальнейшее сокращение поиска

¹³ Модель понимаем как систему некоторых объектов, поведение которой воспроизводит функцию оригинала; см.: Р. Г. Пиотровский. Моделирование фонологических систем и методы их сравнения. Л., 1966, стр. 16.

¹⁴ В. Н. Пушкин, Д. А. Поспелов, Е. И. Ефимов. Психологическая теория мышления и некоторые пути развития кибернетики. ВП, 1971, № 2, стр. 66—77.

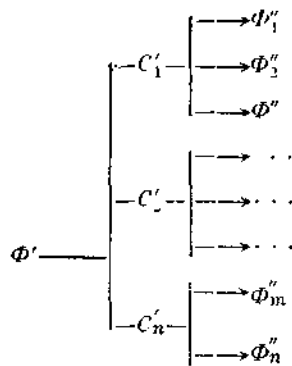


Схема 1.

Обозначения на схеме: Φ' — означающее входной единицы; C'_1, C'_2, \dots, C'_n — варианты означаемого входной единицы; $\Phi''_1, \Phi''_2, \dots, \Phi''_n$ — означающие выходных единиц.

среди переводов этого подмножества происходит при новом обращении к контексту. При выборе оптимального перевода учитываются контекстные условия не только входного, но и выходного текста. Так, определенные ограничения продиктованы правилами сочетания смыслов и идиоматикой выходного языка.

Следовательно, условием процесса устранения двуязычной неоднозначности является наличие: 1) двуязычной суперструктуры, т. е. системы соответствий, позволяющей переходить от означающего входной неоднозначной единицы ко множеству означающих ее выходных переводов; 2) описания каждого перевода в терминах ДП контекста.

Перевод неоднозначной единицы есть переход от гнезда переводов к одному эквиваленту. Он происходит при наличии в контексте признаков, воспринимаемых как диагностирующие. Схематически это можно представить следующим образом (рис.1).

На вход модели поступает контекст, содержащий входную неоднозначную единицу Φ' и некоторые ДП. Система, устраняющая

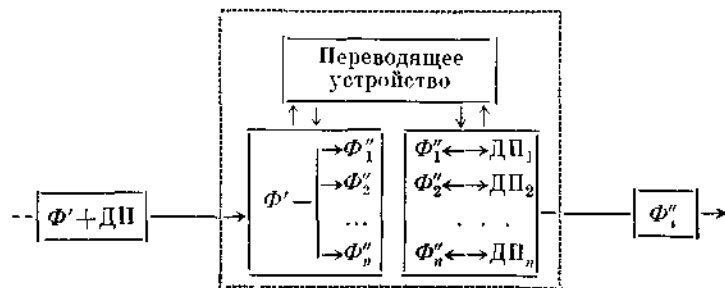


Рис. 1.

неоднозначность, представлена переводящим устройством, словарем соответствий между входной и выходными единицами, а также описанием выходных единиц в терминах ДП. На выходе системы — перевод Φ''_i , определенный на основе нахождения в контексте соответствующих ДП.

Модель предназначена для МП, реализуемого в системе «человек — машина — человек». Необходимо рассмотреть вопрос об участии каждого из элементов системы в реализации модели. Создание модели включает три основных этапа. На первом, доалгоритмическом этапе отбираются входные единицы, двуязычная лексическая неоднозначность которых должна быть устранена моделью. Создается АС, в котором каждая входная единица соотносится с гнездом переводов. Выходные эквиваленты описываются в терминах ДП. Второй этап — создание алгоритма, предусматривающего операции анализа текста, выполнение которых приводит к определению перевода неоднозначной единицы. Третий этап представляет реализацию построенной модели в системе «человек — машина — человек» и оценку ее эффективности.

Первые два этапа выполняются человеком. На третьем этапе принимает участие ЭВМ. В соответствии с этим модель устранения двуязычной неоднозначности строится как система двух моделей — доалгоритмической и алгоритмической.

§ 2. Модель доалгоритмической реализации

Доалгоритмическая модель включает формирование АС, т. е. системы соответствий между входными и выходными единицами. Существенной частью доалгоритмической модели является «фильтр» компрессии синонимии. Его применение обусловлено существованием в языке синонимичных средств, приводящих в двуязычном плане к синонимии переводов. Несовместимость памяти человека и ЭВМ обязывает экономно использовать языковые средства. Это требование не соблюдается, если АС включает синонимичные переводы.

«Фильтр» работает на этапе отбора в АС выходных единиц. Работа «фильтра» организована в два этапа. На первом этапе русские эквиваленты, зафиксированные конкордансами, проверяются на синонимичность. Синонимичность определяется по их способности заменять друг друга в ряде контекстов без нарушения смысла высказывания. Несинонимичные эквиваленты отделяются от синонимичных. Последние организуются в ряды. На втором этапе обрабатываются синонимические ряды. Каждый ряд заменяется доминантой. Доминантой считаем единицу ряда, покрывающую своим семантическим полем значения всех слов ряда и обладающую высокой частотой употребления. Формируется окончательный список несинонимичных переводов. По числу несинонимичных

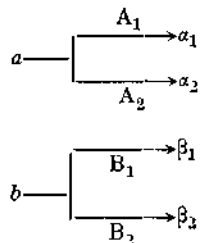
переводов входная единица идентифицируется как однозначная или неоднозначная (в двуязычном плане). Входные неоднозначные словоформы отбираются в АС.

Модель доалгоритмической реализации включает также описание выходных эквивалентов в терминах ДП. ДП организованы в 90 списков, представленных следующими группами:

- 1) список формально-грамматических средств (формы глагола *be* и детерминативы имени существительного, т. е. два списка);
- 2) списки морфологических ДП (5 списков).
- 3) списки лексических ДП (83 списка), включающие: а) семантические группы слов (например, список № 9 — семантическая группа «деталь»); б) слова, которые не образуют одной тематической группы (например, список № 61 — *field, concrete, road, aircraft*).

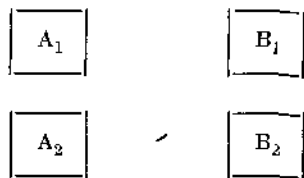
Для рационального размещения списков в памяти произведено их объединение.

Объединение производилось при соблюдении некоторых условий. Пусть *a* и *b* — входные неоднозначные единицы; α_1 , α_2 и β_1 , β_2 — их выходные эквиваленты, детерминированные списками A_1 , A_2 и B_1 , B_2 :



Если представить списки в виде множеств, то отношения между ними могут быть описаны следующими ситуациями.

Ситуация 1. Ни одно слово списка A_1 не входит в список B_1 , и ни одно слово списка A_2 не принадлежит списку B_2 :



Ситуация 2. Несколько слов списка A_1 входят в список B_1 , несколько слов списка A_2 принадлежат списку B_2 :



Ситуация 3. Все слова списка A_1 являются словами списка B_1 , все единицы списка A_2 принадлежат и списку B_2 . Причем $A_1=B_1$, $A_2=B_2$:



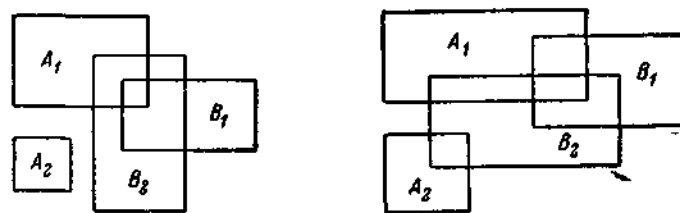
Ситуация 4. Все слова списка A_1 являются словами списка B_1 , все единицы списка A_2 принадлежат и списку B_2 . Причем $B_1 > A_1$, $B_2 > A_2$:



Ситуация 5. Все слова списка B_1 являются словами списка A_1 , все слова списка B_2 являются словами списка A_2 . Причем $A_1 > B_1$, $A_2 > B_2$.



Ситуация 6. Сложные пересечения множеств A_1 , A_2 , B_1 , B_2 :



и т. д.

Укрупнение списков производили, когда отношения между списками описывались ситуациями 1, 2, 4, 5. Действительно, все эти ситуации могут обслуживаться укрупненными списками:

$$A_1 \cup B_1 = A_1 B_1$$

$$A_2 \cup B_2 = A_2 B_2$$

После формального укрупнения списки проверялись на соблюдение дополнительного условия: объединение списков A_1 и B_1 , A_2 и B_2 не изменяет переводов α_1 , α_2 , β_1 , β_2 .

В результате укрупнения получены списки двух типов: 1) семантическая группа + семантическая группа (список № 21 — «резервуар» + «жидкость»); 2) семантическая группа + отдельные словоформы (список № 37 — «книга» + *flaw, gask*).

§ 3. Алгоритмическая модель

1. Принципиальная блок-схема работы алгоритмической модели

Модель алгоритмической реализации (рис. 2) обеспечивает ввод в ЭВМ английского текста, поиск в нем неоднозначных словоформ и устранение их двуязычной лексической неоднозначности. Анализ контекста сводится к различного рода проверкам на наличие ДП в сегментах текста.



Рис. 2. Принципиальная блок-схема работы алгоритмической модели.

Рассмотрим, каким образом работает алгоритмическая модель. В ЭВМ предложениями вводится незнакомый английский текст. С помощью программы сегментации предложение автоматически делится на сегменты. Каждое словоупотребление предложения проверяется на вхождение в список неоднозначных входных словоформ. Найдя такие словоупотребления, ЭВМ анализирует контекст по подпрограмме анализа для найденного словоупотребления, определяет перевод. Выпечатывается предложение, неоднозначная словоформа, ДП и русский перевод. Предложение прове-

ряется до конца на наличие неоднозначных словоформ. Если таковые обнаружены, цикл повторяется.

Из блок-схемы видно, что алгоритмическая модель¹⁵ состоит из «внешнего» и «внутреннего» компонентов. «Внешний» компонент обеспечивает ввод текста, поиск неоднозначных словоформ в предложении, печать предложения, анализируемой единицы, ДП и перевода. «Внутренний» компонент модели представлен алгоритмами анализа текста, которые обеспечивают перевод 137



Рис. 2 (продолжение).

неоднозначных словоформ. Словоформы, представляющие одно слово, анализируются по одному и тому же алгоритму (подпрограмме). Так, словоформы operate, operated, operates, operating анализируются по подпрограмме для глагола operate. Словоформы colour, colours, color, colors — по подпрограмме для существительного colour. Программа, реализующая алгоритмическую модель, строится как комплекс 50 подпрограмм, анализирующих соответственно 50 английских слов. Отсылка к «внутреннему» компоненту модели — в блоке 7. Все остальные блоки представляют «внешний» компонент алгоритмической модели.

¹⁵ Алгоритмическая модель реализована в программе для ЭВМ «Минск-22», составленной старшим инженером-программистом Л. А. Славиной, которой автор выражает искреннюю признательность.

2. Поиск ДП в предложении

Работа «внутреннего» компонента модели состоит в обнаружении признаков, диагностирующих адекватный русский перевод английской словоформы. ДП расположены на МЛ в списках в порядке убывания частотности. Весь массив списков организован в три зоны. После вызова нужной зоны в МОЗУ поиск списка внутри зоны производится перебором.

При МП существенное значение имеет указание зоны поиска ДП. Известны следующие подходы: 1) поиск ДП среди непосредственно контактирующих слов в так называемом линейном контексте; 2) поиск по всему предложению с пропуском слов, обладающих некоторыми синтаксическими функциями; 3) поиск ДП после выявления синтаксической структуры предложения.¹⁶ В этом случае снятие неоднозначности ставится в зависимость от построения дерева предложения.

Наша модель осуществляет поиск ДП в сегментах. Сегменты выступают не как синтаксические единицы, а как речевые отрезки, позволяющие экономно организовать поиск ДП. Это связано с тем, что наша модель оперирует в основном лексическими ДП. Лексическое детерминирование осуществляется как единицами, синтаксически связанными с неоднозначным словом, так и единицами, синтаксически с ним не связанными.

В подавляющем большинстве случаев (93,9%) ДП расположены в зоне, включающей нулевой сегмент и по два сегмента слева и справа от него. Алгоритм предусматривает их поиск в несколько этапов: от наиболее вероятных сегментов (нулевого, первых правого и левого) к менее вероятным (вторым правому и левому). В редких случаях обращаются к сегментам, лежащим вне этой зоны.

Выбор указанной области поиска не случаен. Если рассмотреть лексическую сочетаемость английского существительного, то можно отметить следующее.

Английское существительное способно последовательно «наращивать» определители в препозиции, расширять именную группу, т. е. нулевой сегмент, влево.

Ср (the capacity) — ‘мощность’; (the load capacity) — ‘грузоподъемность’; (the load carrying capacity) — ‘грузоподъемность’.

Увеличение именного сегмента влево до каких угодно размеров не меняет того факта, что определители существительного обнаруживаются внутри все того же нулевого сегмента, слева от анализируемой неоднозначной именной словоформы.

¹⁶ См. указанные работы А. В. Зубова, О. А. Нехай, А. Н. Шаранды, Ю. Н. Марчука, а также: Automatische Sprachubersetzung Russisch-Deutsch, Berlin, 1969.

Для случаев, когда диагностирующим лексическим признаком является глагол, управляющий неоднозначной именной словоформой, необходим просмотр сегментов, примыкающих слева к нулевому.

Английское существительное способно посредством предлогов присоединять определители справа (предложная именная группа):

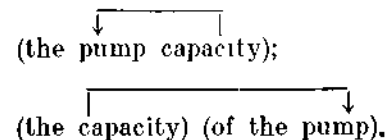
(the pump capacity), (the capacity) (of the pump) ‘мощность насоса’.

Поскольку предлог служит делимитатором, вводящим именной сегмент, определители находятся в первом правом сегменте, а не в нулевом.

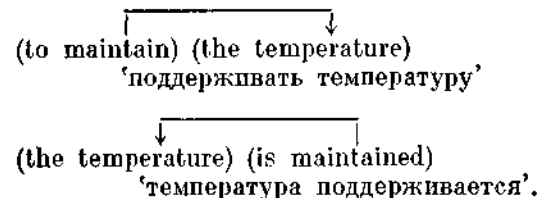
В ряде случаев поиск диагностирующего признака должен быть продолжен вправо за пределы первого правого сегмента, как, например, в случае:

(the capacity) (of the pump) (is. . .) ‘мощность’;
(the capacity) (of the pump) (to. . .) ‘способность’.

Все сказанное убеждает нас в том, что поиск диагностирующего признака для неоднозначной именной словоформы должен включать проверки как влево, так и вправо от исследуемого слова:



Мысль о двух направлениях поиска в равной мере относится и к анализу неоднозначных глагольных словоформ. В этом случае направление поиска диагностирующих признаков (вправо, влево) определяется активностью (пассивностью) глагольной структуры, ср.



Указанные свойства существительных и глаголов наводят на мысль о создании системы стандартных операторов поиска ДП в предложении. Все операции анализа текста при определении перевода неоднозначного слова могут быть сведены к следующему списку стандартных операторов, работающих с автоматически извлекаемыми сегментами.

№№	Код	Оператор
1	ПП	проверка анализируемой неоднозначной словоформы на признак
2	ПС1 (ПС2)	проверка соседней правой (левой) словоформы на входжение в список диагностирующих словоформ или на отдельную словоформу
3	ПП1 (ПП2)	проверка соседней правой (левой) словоформы на признак
4	ППС	поиск диагностирующей словоформы (списка) в сегменте, где обнаружен искомый делimitатор
5	ППП	поиск признака в сегменте, где обнаружен искомый делimitатор
6	ПС01 (ПС02)	поиск диагностирующей словоформы (списка) направо (налево) до границы нулевого сегмента
7	ПП01 (ПП02)	поиск признака направо (налево) до границы нулевого сегмента
8	ПС 11 (ПС 12)	поиск диагностирующей словоформы (списка) в первом правом (левом) сегменте
9	ПП11 (ПП12)	поиск признака в первом правом (левом) сегменте
10	ПС21 (ПС22)	поиск диагностирующей словоформы (списка) во втором правом (левом) сегменте
11	ПП 21(ПП22)	поиск признака во втором правом (левом) сегменте
12	НПС	проверка начальной словоформы нулевого сегмента на входжение в список диагностирующих словоформ или на отдельную диагностирующую словоформу
13	НПП	проверка начальной словоформы нулевого сегмента на признак
14	НПС11 (НПС12)	проверка начальной словоформы первого правого (левого) сегмента на входжение в список диагностирующих словоформ или на отдельную диагностирующую словоформу
15	НПП11 (НПП12)	проверка начальной словоформы первого правого (левого) сегмента на признак
16	НПС21 (НПС22)	проверка начальной словоформы второго правого (левого) сегмента на входжение в список
17	НПП21 (НПП22)	проверка начальной словоформы второго правого (левого) сегмента на признак
18	П5, П20, П... П70	перевод определен: выбрать в гнезде переводов тот перевод, в информационную ячейку которого входит данный код (5, 20, ... 70)

Примечание.

1) признаком называем диагностирующий морфологический признак;
2) цифры в кодах оператора имеют следующие значения: а) первая цифра указывает номер сегмента, в котором ведется поиск (0 — нулевой, 1 — первый, 2 — второй);

3) вторая цифра обозначает расположение сегмента относительно нулевого (1 — справа, 2 — слева); так, код 21 обозначает второй правый сегмент, а 12 — первый левый.

Описываемая система не воспроизводит, а моделирует поведение человека при переводе входных словоформ, обладающих двуязычной лексической неоднозначностью. Особенностью моделирования как метода научного познания является изучение объекта через посредство его аналога. Сущность любой аналогии — неполнота. Поэтому моделирование неизбежно связано с упрощением, абстрагированием от ряда свойств изучаемого объекта.

Эта черта позволила при создании нашей модели абстрагироваться от грамматических явлений и исследовать только лексические англо-русские соответствия. В связи с этим наш АС отличается от АС обычного типа тем, что содержит только лексико-семантическую информацию. Он включает: 1) словарь входных (английских) словоформ, обладающих двуязычной лексической неоднозначностью; 2) словарь выходных (русских) эквивалентов; 3) таблицы, по которым после анализа контекста входным единицам отыскиваются соответствия в выходном языке.

Входные словоформы записаны на МЛ четырьмя порциями. Каждая входная словоформа соотносится с набором кодов.

1) Входной код — номер слова, которое эта словоформа представляет. Словоформы одного слова имеют один и тот же код. Например:

```
operate ( 01
operated 01
operates 01
operating 01
```

Число входных кодов в АС соответствует числу неоднозначных исследуемых слов, т. е. 50.

2) Выходные коды — адреса ячеек, в которых записаны русские эквиваленты с данным входным кодом.

Переход от входного кода к гнезду переводов, соотносимому с этим кодом, происходит по таблице.

Словник выходных лексических единиц представлен словоформами, причем одна словоформа представляет одно слово. Словоформы являются «нулевыми» формами соответствующих слов. Существительные приводятся в форме им. падежа ед. числа, глаголы — в инфинитиве. Словник выходных словоформ расположен в МОЗУ и упорядочен по частоте.

Каждая выходная словоформа имеет семантический код, т. е. машинную идентификацию соответствующего семантического признака.

Машинная лексическая неоднозначность входной словоформы представлена в АС одно-многозначным соотношением входного и выходных кодов. Ее устранение достигается, когда в результате анализа контекста реализуется одна из логических ветвей алгоритма, завершающаяся кодом семантического признака. Иными

словами, нахождение определенных эксплицитных (лексических и морфологических) признаков в окружении неоднозначной словоформы переносит анализ на уровень соотношения имплицитных признаков входного слова и его переводов. Выработав код семантического признака, машина выбирает среди переводов гнезда тот, которому приписан этот код. Так, в гнезде переводов для входного слова performance

'характеристика'	70
'выполнение'	8
'осуществление'	7
'функция'	5
'действие'	11
'применение'	55
'показатель'	39
'качества'	65
'приемистость'	63
'производительность'	38
'эффективность'	66
'эксплуатация'	6

машина выберет перевод 'применение', если она «выработает» выход на код 35, 'эффективность' — на 66, 'действие' — на 11 и т. д.

Для нашего материала эмпирически было отобрано 66 семантических признаков. Эти признаки являются типовыми значениями соответствующих групп русских эквивалентов. Семантическое кодирование используется для идентификации разных переводов одного и того же входного слова. Классификация русских переводов по семантическим группам производилась на основе общности типовых значений. При этом соблюдалось условие: одним и тем же семантическим кодом не должны описываться переводы одного и того же входного слова. Некоторые группы включают слова, принадлежащие к различным частям речи.

4. Алгоритмы устранения двузвучной лексической неоднозначности английских слов (в полуоператорной записи)

Содержание столбцов полуоператорной записи: первый столбец — номер строки (оператора) по порядку в данном алгоритме, второй — код оператора (содержание кодов операторов см. на стр. 150), третий — переход по «да», четвертый — переход по «нет». Коды типа П5 (П20, П15) означают выбор из гнезда переводов (ГП), соотносимого с анализируемым неоднозначным словом, такого русского эквивалента, в информационной ячейке которого есть семантический код 5 (или 20, 15). Числа в третьем и четвертом столбцах (вне сочетания с буквой П) обозначают номер строки (в данном алгоритме) перехода по «да» и «нет».

ADAPTATION

1) НПС 11 сп. 81	П15 2	7) НПС 11 of	9	8
2) НПС 21 сп. 81	П15 7	8) НПС 21 of	9	3
3) ПСО 2 сп. 9	П12 4	9) ППС сп. 9	П12	10
4) ПСО 2 сп. 18	П12 5	10) ППС сп. 18	П12	11
5) ПСО 2 сп. 28	П12 6	11) ППС сп. 28	П12	12
6) ПСО 2 сп. 29	П12 П13	12) ППС сп. 29	П12	П13

ГП

приспособление	— 15
усовершенствование	— 12
адаптация	— 13

ADJUST

1) НПС 11 to	П15 2	10) ПП 12 сп. 83	П14	11
2) НПС 21 to	П15 3	11) ПСО 2 account	П10	12
3) НПС 11 for	5 4	12) ПС 12 account	П10	13
4) НПС 21 for	5 6	13) ПСО 1 сп. 4	П23	14
5) ППП сп. 82	П15 П12 14	14) ПС 11 сп. 4	П23	15
6) ПП-ed	7 13	15) ППО 1 сп. 83	П14	16
7) ПСО 2 сп. 4	П23 8	16) ПП 11 сп. 83	П14	17
8) ПС 12 сп. 4	П23 9	17) ПСО 1 account	П10	18
9) ППО 2 сп. 83	П14 10	18) ПС 11 account	П10	П13

ГП

регулировать	— 13
приспосабливать	— 15
сверять	— 10
юстировать	— 14
настраивать	— 12
устанавливать	— 23

ADJUSTMENT

1) ПП -s	П15 8	7) НПС 21 сп. 81	П15	П13
2) ПСО 2 сп. 3	П15 3	8) НПС 11 of	10	9
3) ПСО 2 сп. 4	П23 4	9) НПС 21 of	10	2
4) ПСО 2 сп. 5	П12 5	10) ППС сп. 4	П23	11
5) ППО 2 сп. 83	П14 6	11) ППС сп. 5	П12	12
6) НПС 11 сп. 81	П15 7	12) ППП сп. 83	П14	П13

ГП

регулировка	— 13
юстировка	— 14
поправка	— 15
установка	— 23
настройка	— 12

ALLOWANCE

1) ПСО 2 depreciation	П69	2
2) ПСО 2 сп. 20	П54	3
3) ПСО 2 full	П18	4
4) ПСО 2 standard	П15	5
5) НПС 11 for	7	П37

6) НПС 21 for	7	П37
7) ППП сн. 82	П15	П37

ГП

учет	— 37
скидка	— 69
возможности	— 54
поправка	— 15
допущение	— 18

ARRANGE

1) ПС 11 that	П59	2	30) НПС 21 for	П53	П33
2) ПС 11 to	П59	3	31) НПС 11 for	П59	32
3) ПС 21 to	П59	4	32) НПС 21 for	П59	33
4) ПП -ed	5	31	33) ПСО 1 reinforcement	П21	34
5) ПСО 2 reinforcement	П21	6	34) ПСО 1 hanger	П22	35
6) ПСО 2 hanger	П22	7	35) ПСО 1 matters	П15	36
7) ПСО 2 matters	П15	8	36) ПСО 1 circuit	П24	37
8) ПСО 2 circuit	П24	9	37) ПСО 1 сн. 17	П10	38
9) ПСО 2 сн. 17	П10	10	38) ПСО 1, situation	П51	39
10) ПСО 2 situation	П51	11	39) ПСО 1 сн. 4	П23	40
11) ПСО 2 сн. 4	П23	12	40) ПСО 1 сн. 6	П42	41
12) ПСО 2 сн. 6	П42	13	41) ПСО 1 сн. 29	П56	42
13) ПСО 2 сн. 29	П56	14	42) ПСО 1 сн. 9	П25	43
14) ПСО 2 сн. 9	П25	15	43) ПСО 1 сн. 10	П55	44
15) ПСО 2 сн. 10	П55	16	44) ПСО 1 сн. 8	57	45
16) ПСО 2 сн. 8	29	17	45) ПС 11 reinforcement	П21	46
17) ПС 12 reinforcement	П21	18	46) ПС 11 hanger	П22	47
18) ПС 12 hanger	П22	19	47) ПС 11 matters	П15	48
19) ПС 12 matters	П22	20	48) ПС 11 circuit	П24	49
20) ПС 12 circuit	П24	21	49) ПС 11 сн. 17	П10	50
21) ПС 12 сн. 17	П10	22	50) ПС 11 situation	П51	51
22) ПС 12 situation	П51	23	51) ПС 11 сн. 4	П23	52
23) ПС 12 сн. 4	П23	24	52) ПС 11 сн. 6	П42	53
24) ПС 12 сн. 6	П42	25	53) ПС 11 сн. 29	П56	54
25) ПС 12 сн. 29	П56	26	54) ПС 11 сн. 9	П25	55
26) ПС 12 сн. 9	П25	27	55) ПС 11 сн. 10	П55	56
27) ПС 12 сн. 10	П55	28	56) ПС 11 сн. 8	58	57
28) ПС 12 сн. 8	30	31	57) НПС 11 for	П53	П33
29) НПС 11 for	П53	П33	58) НПС 21 for	П53	П33

ГП

расположить	— 29	уладить	— 15
уложить	— 21	создать	— 51
закрепить	— 22	рассчитать	— 42
подобрать	— 10	организовать	— 33
установить	— 23	монтажировать	— 25
составить	— 56	классифицировать	— 55
назначить	— 53	компоновать	— 24
предусмотреть	— 59		

ARRANGEMENT

1) ПП -s	2	17	4) НПС 11 for	П48	5
2) НПС 11 with	П15	3	5) НПС 21 for	П48	17
3) НПС 21 with	П15	4	6) ПСО 2 сн. 7	П15	7

7) ПСО 2 сн. 11	П33	8	17) НПС 11 of	19	18
8) ПСО 2 сн. 15	П50	9	18) НПС 21 of	19	6
9) ПСО 2 сн. 13	П25	10	19) ППС сн. 29	П10	20
10) ПСО 2 сн. 14	П28	11	20) ППС сн. 13	П25	21
11) ПСО 2 сн. 10	П55	12	21) НПС сн. 12	П26	22
12) ПСО 2 circuit	П24	13	22) ППС circuit	П24	23
13) ПСО 2 сн. 12	П26	14	23) ППП сн. 82	П26	24
14) ППО 2 сн. 83	П26	15	24) НПС 11 for	26	25
15) ППО 2 сн. 82	П26	16	25) НПС 21 for	26	П29
16) ППО 2 сн. 86	П26	24	26) ППП сн. 82	П26	П29

ГП

устройство	— 26	распределение	— 25
соглашение	— 15	структура	— 28
мероприятие	— 33	классификация	— 55
приготовление	— 48	компоновка	— 24
упорядочение	— 10	схема	— 50
расположение	— 29		

АСПЕКТ

1) ПС1 ratio	П67	2	9) НПС 11 of	11	10
2) НПС from	П58	8	10) НПС 21 of	11	3
3) НПСО 2 сн. 14	П55	4	11) ППС сн. 14	П55	12
4) НПСО 2 сн. 16	П65	5	12) ППС сн. 16	П65	13
5) ННПО 2 сн. 83	П65	6	13) ППС сн. 17	П27	14
6) ННПО 2 сн. 86	П57	7	14) ППП сн. 82	П53	15
7) ННПО 2 сн. 82	П53	8	15) ППП -ics	П53	16
8) ННПО 2 -ics	П53	П39	16) ППП сн. 83	П65	П39

ГП

аспект	— 39	особенность	— 65
формат	— 67	проблема	— 53
вид	— 55	точка зрения	— 58
сторона	— 57	положение	— 27

ATTACHMENT

1) ПП -s	П15	2
2) ПС 1 сн. 9	П24	3
3) ПСО 2 сн. 2	П15	П22

ГП

прикрепление	— 22
приспособление	— 15
соединительный	— 24

BEHAVIOUR

1) НПС 11 of	3	2	8) ПСО 2 сн. 40	П7	9
2) НПС 21 of	3	7	9) ПСО 2 сн. 73	П7	10
3) ППС сн. 45	П5	4	10) ПСО 2 сн. 45	П5	11
4) ППС сн. 18	П6	5	11) ПСО 2 сн. 18	П6	12
5) ППС сн. 9	П6	7	12) ПСО 2 сн. 9	П6	13
6) ППП сн. 83	П6	П70	13) ППО 2 сн. 82	П7	14
7) ПСО 2 сн. 39	П7	8	14) ППО 2 сн. 83	П6	П70

ГП	
режим	— 7
характеристика	— 70
работа	— 6
поведение	— 5

СARABILITY

1) ПП -s	П54	2	8) ППП сн. 83	П64	9
2) НПС 11 сн. 81	П65	3	9) ППП сн. 82	П65	П70
3) НПС 21 сн. 81	П65	4	10) ПСО 2 сн. 20	П19	11
4) НПС 11 of	6	5	11) ПСО 2 сн. 18	П64	12
5) НПС 21 of	6	10	12) ППО 2 сн. 82	П65	13
6) ППС сн. 20	П19	7	13) ППО 2 сн. 83	П64	П70
7) ППС сн. 18	П64	8			

ГП

способность	— 65
возможности	— 54
емкость	— 19
мощность	— 64
характеристика	— 70

СARACITY

1) ПП-ies	П54	2	23) НПС in	П53	24
2) НПС 11 сн. 81	П65	3	24) ПСО 2 per	П38	25
3) НПС 21 сн. 81	П65	4	25) ПС 12 per	П38	26
4) ПП 1 сн. 82	П63	29	26) ППО 2 сн. 83	П64	27
5) ПСО 2 junior	П27	6	27) ПС 1 сн. 20	П19	28
6) ПСО 2 flow	П69	6	28) ПС 1 сн. 22	П19	П65
7) ПСО 2 electrostatic	П19	8	29) НПС 11 of	31	30
8) ПСО 2 seating	П20	9	30) НПС 21 of	31	5
9) ПСО 2 strength	П22	10	31) ППС сн. 21	П62	32
10) ПСО 2 сн. 3	П5	11	32) ППС сн. 26	П66	33
11) ПСО 2 сн. 11	П53	12	33) ППС сн. 18	П64	34
12) ПСО 2 сн. 13	П66	13	34) ППС сн. 27	П18	35
13) ПСО 2 сн. 21	П62	14	35) ППС сн. 20	П19	36
14) ПСО 2 сн. 26	П66	15	36) ППС сн. 22	П19	37
15) ПСО 2 сн. 18	П64	16	37) ППС сн. 23	П20	38
16) ПСО 2 сн. 27	П18	17	38) ППС сн. 25	П61	39
17) ПСО 2 сн. 74	П60	18	39) ППС сн. 24	П57	40
18) ПСО 2 сн. 20	П19	19	40) ППС heat	П60	41
19) ПСО 2 сн. 22	П19	20	41) ППП сн. 83	П64	42
20) ПСО 2 сн. 25	П61	21	42) ПС 11 per	П38	43
21) ПСО 2 сн. 24	П57	22	43) ПС 21 per	П38	П65
22) ПСО 2 сн. 87	П64	23			

ГП

способность	— 65	функция	— 5
положение	— 27	должность	— 53
оптимальный	— 63	объем	— 62
возможности	— 54	грузоподъемность	— 66
мощность	— 64	пропускная способность	— 18
расход	— 69	теплоемкость	— 60
емкость	— 19	емкостительность	— 20
площадь	— 61	расстояние	— 57
производительность	— 38	прочность	— 22

CLASSIFY

1) ПП -ed	2	3	6) ПП 12 -s	П55	7
2) ПСО 2 сн. 1	3	7	7) ППО 1 -s	9	8
3) ПС 11 as	П42	4	8) ПП 11 -s	10	П42
4) ПС 21 as	П42	5	9) ПСО 1 as	П42	П55
5) ППО 2 -s	П55	6	10) ПС 11 as	П42	П55

ГП

определять	— 42
классифицировать	— 55

COMPENSATE

1) ПП -ed	2	6	6) ПСО 1 сн. 23	П13	7
2) ПСО 2 сн. 23	П13	3	7) ПС 11 сн. 23	П13	8
3) ПС 12 сн. 23	П13	4	8) ПСО 1 сн. 10	П12	9
4) ПСО 2 сн. 10	П12	5	9) ПС 11 сн. 10	П12	П14
5) ПС 12 сн. 10	П12	6			

ГП

компенсировать	— 14
выравнивать	— 13
корректировать	— 12

COMPENSATION

1) НПС 11 of	3	2	4) ППС сн. 23	П13	П14
2) НПС 21 of	3	5	5) ПСО 2 сн. 10	П12	6
3) ППС сн. 10	П12	4	6) ПСО 2 сн. 23	П13	П14

ГП

компенсация	— 14
корректировка	— 12
выравнивание	— 13

CONDUCTION

1) ПС 2 electrical	П59	2	5) НПС 21 of	6	7
2) ПС 2 magnetic	П19	3	6) НПС heat	П60	7
3) ПС 2 сн. 74	П60	4	7) НПС during	П7	П18
4) НПС 11 of	7	5			

ГП

теплопроводность	— 60	проводимость	— 18
электропроводность	— 59	прохождение	— 7
проницаемость	— 19		

CONDUCTIVITY

1) ПС 2 electrical	9	2	7) НПС 21 of	8	П60
2) ПС 2 сн. 74	6	3	8) ППС Z	П61	П60
3) НПС 11 of	5	4	9) НПС 11 of	11	10
4) НПС 21 of	5	П18	10) НПС 21 of	11	П59
5) ППС heat	П60	П18	11) ППС Z	П40	П59
6) НПС 11 of	8	7			

ГП			
проводимость	— 18	электропроводность	— 59
теплопроводность	— 60	удельная электропроводность	— 40
удельная теплопроводность	— 61		

CONSTRUCTION

1) ПП -s	П26	2	14) ППО 2 сп. 88	П26	15
2) ПСО 2 сп. 2	П26	3	15) ППО 2 сп. 86	П26	16
3) НПС of	4	7	16) ППО 2 -ed	П26	17
4) ППС сп. 17	П52	5	17) ПСО 1 сп. 17	П52	18
5) ППС сп. 29	П52	6	18) ПСО 1 сп. 29	П52	19
6) ППС сп. 39	П52	7	19) ПСО 1 сп. 32	П52	20
7) НПС in	8	9	20) ПСО 1 сп. 39	П52	21
8) ПС 12 сп. 33	П26	П52	21) НПС 11 of	23	22
9) ПСО 2 сп. 28	П50	10	22) НПС 21 of	23	П51
10) ПСО 2 сп. 33	П26	11	23) ППС сп. 28	П50	24
11) ПСО 2 сп. 29	П25	12	24) ППС сп. 29	П25	25
12) ПСО 2 сп. 30	П52	13	25) ППС сп. 31	П52	П51
13) ПСО 2 сп. 31	П52	14			

ГП

построение	— 50	составление	— 25
создание	— 51	строительство	— 52
конструкция	— 26		

DEFINITION

1) НПС 11 of	3	2	5) ПСО 2 сп. 5	П46	6
2) НПС 21 of	3	4	6) НПС 11 for	П46	7
3) ППС сп. 5	П46	П42	7) НПС 21 for	П46	П42
4) ПСО 2 сп. 34	П64	5			

ГП

определение	— 42
объяснение	— 46
четкость	— 64

DENSITY

1) НПС 11 of	3	2	8) ППС сп. 36	П19	П66
2) НПС 21 of	3	9	9) ПСО 2 сп. 27	П18	10
3) ППС сп. 27	П18	4	10) ПСО 2 сп. 35	П59	11
4) ППС сп. 35	П59	5	11) ПСО 2 сп. 4	П20	12
5) ППС сп. 4	П20	6	12) ПСО 2 Z	П40	13
6) ППС Z	П40	7	13) ПСО 2 сп. 15	П41	14
7) ППС сп. 15	П41	8	14) ПСО 2 сп. 36	П19	П66

ГП

плотность	— 66	густота	— 20
индукция	— 59	объемный вес	— 41
удельный вес	— 40	концентрация	— 19
пропускная способность	— 18		

DERIVE

1) ПП -ed	2	12	3) ПСО 2 сп. 38	П40	4
2) ПСО 2 сп. 28	П50	3	4) ПСО 2 сп. 17	П51	5

5) ПСО 2 сп. 29	П56	6	17) ПС 12 сп. 23	П54	П7
6) ПСО 2 tube	П57	7	18) ПСО 2 сп. 37	П52	19
7) ПС 12 сп. 28	П50	8	19) ПС 12 сп. 37	П52	20
8) ПС 12 сп. 38	П40	9	20) ПС 11 сп. 2	22	21
9) ПС 12 сп. 17	П51	10	21) ПП 11 -s	П32	22
10) ПС 12 сп. 29	П56	11	22) ПС 11 tube	П57	23
11) ПС 12 tube	П57	12	23) ПС 11 сп. 28	П50	24
12) НПС 11 from	45	13	24) ПС 11 сп. 38	П40	25
13) НПС 21 from	15	14	25) ПС 11 сп. 17	П51	26
14) ПСО 2 сп. 1	П32	18	26) ПС 11 сп. 29	П56	27
15) ПСО 2 сп. 1	П32	16	27) ПС 11 сп. 37	П52	П32
16) ПСО 2 сп. 23	П54	17			

ГП

присходить	— 7	отвести	— 57
вывести	— 40	составить	— 56
разработать	— 51	привести	— 52
получить	— 32	основываться	— 54
построить	— 50		

DETECT

1) ПП -ed	2	8	8) ПСО 1 сп. 39	П47	9
2) ПСО 2 сп. 39	П47	3	9) ПС 11 сп. 39	П47	10
3) ПС 12 сп. 39	П47	4	10) ПСО 1 сп. 41	П14	11
4) ПСО 2 сп. 40	П10	5	11) ПС 11 сп. 41	П14	12
5) ПС 12 сп. 40	П10	6	12) ПСО 1 сп. 40	П10	13
6) ПСО 2 сп. 41	П14	7	13) ПС 11 сп. 40	П10	П46
7) ПС 12 сп. 41	П14	8			

ГП

обнаруживать	— 46
выпрямлять	— 14
измерять	— 47
детектировать	— 10

DETECTION

1) НПС 11 of	3	2	6) ПСО 2 сп. 37	П14	7
2) НПС 21 of	3	6	7) ПСО 2 сп. 15	П32	8
3) ППС сп. 37	П14	4	8) ПСО 2 сп. 40	П10	9
4) ППС сп. 15	П32	5	9) ПС 1 coefficient	П10	П46
5) ППС сп. 40	П10	П46			

ГП

обнаружение	— 46
дефектоскопия	— 14
прием	— 32
детектирование	— 10

DETERMINATION

1) НПС 11 to	П58	2	5) ППС сп. 6	П41	6
2) НПС 21 to	П58	3	6) ППС сп. 14	П46	П42
3) НПС 11 of	5	4	7) ПСО 2 сп. 6	П41	8
4) НПС 21 of	5	7	8) ПСО 2 сп. 14	П46	П42

ГП	
определение	— 42
вычисление	— 41
установление	— 46
решимость	— 58

DIFFERENCE

1) НПС of	2	4	16) ППО 1 сп. 83	П69	17
2) ПС 12 сп. 6	П69	3	17) ПСО 1 сп. 79	П69	18
3) ПС 12 сп. 14	П69	4	18) ПС 11 сп. 79	П69	19
4) ПСО 1 Z	П69	5	19) НПС 11 with	П30	20
5) ПС 11 Z	П69	6	20) НПС 21 with	П30	21
6) ПСО 2 сп. 43	7	8	21) НПС 11 сп. 89	23	22
7) ПС 12 сп. 44	П58	8	22) НПС 21 сп. 89	23	29
8) ПС 12 сп. 79	П69	9	23) НПС сп. 6	П69	24
9) ПСО 2 сп. 42	П69	10	24) НПС сп. 38	П69	25
10) ПСО 2 brightness	П64	11	25) ППС сп. 40	П69	26
11) ПСО 2 сп. 24	П30	12	26) ППС сп. 41	П69	27
12) ПСО 1 сп. 17	П69	13	27) ППС Z	П69	28
13) ПСО 1 сп. 38	П69	14	28) ППС сп. 24	П30	П70
14) ПСО 1 сп. 40	П69	15	29) ПСО 1 Z	П69	30
15) ПСО 1 сп. 41	П69	16	30) ПС 11 Z	П69	П70

ГП

разность	— 69	расхождение	— 30
различие	— 70	значения	— 58
контраст	— 64		

EFFICIENCY

1) ПС 2 full	П64	2	15) НПС 21 to	П65	16
2) ПС 2 structural	П22	14	16) НПС 11 of	18	17
3) ПСО 2 сп. 27	П18	4	17) НПС 21 of	18	3
4) ПСО 2 сп. 45	П63	5	18) ППС сп. 27	П18	19
5) ПСО 2 сп. 9	П38	6	19) ППС сп. 45	П63	20
6) ПСО 2 сп. 18	П38	7	20) ППС сп. 9	П38	21
7) ПСО 2 сп. 75	П38	8	21) ППС сп. 18	П38	22
8) ПСО 2 сп. 39	П31	9	22) ППС сп. 39	П31	23
9) ПСО 2 сп. 41	П31	10	23) ППС сп. 41	П31	24
10) ПСО 2 сп. 44	П40	11	24) ППС сп. 44	П40	25
11) ППО 2 сп. 83	П38	12	25) ППП сп. 83	П38	26
12) ППО 2 сп. 86	П38	13	26) ППП -ion	П40	27
13) ППО 2 -ion	П40	П66	27) ППП -ing	П65	П66
14) НПС 11 to	П65	15			

ГП

эффективность	— 66	пропускная способность	— 18
работоспособность	— 63	отдача	— 31
способность	— 65	коэффициент	— 40
коэффициент полезного действия	— 38	мощность	— 64
прочность	— 22		

EVALUATE

1) ПСО 1 сп. 76	П38	2	3) НПС 11 from	П55	4
2) ПС 11 сп. 76	П38	3	4) НПС 21 from	П55	5

5) ПП -ed	6	14	14) ПСО 1 сп. 6	П42	15
6) ПСО 2 сп. 6	П42	7	15) ПС 11 сп. 6	П42	16
7) ПС 12 сп. 6	П42	8	16) ПСО 1 сп. 9	П13	17
8) ПСО 2 сп. 9	П13	9	17) ПС 11 сп. 9	П13	18
9) ПС 12 сп. 9	П13	10	18) ПСО 1 сп. 18	П13	19
10) ПСО 2 сп. 18	П13	11	19) ПС 11 сп. 18	П13	20
11) ПС 12 сп. 18	П13	12	20) ПСО 1 сп. 38	П58	21
12) ПСО 2 сп. 38	П58	13	21) ПС 11 сп. 38	П58	П37
13) ПС 12 сп. 38	П58	14			

ГП

определить	— 37	проверить	— 13
выразить	— 38	отнести	— 55
решить	— 58	вычислить	— 42

EVALUATION

1) ПП -ed	3	2	7) ППО 2 сп. 83	П13	П9
2) ПСО 2 сп. 2	3	8	8) НПС 11 of	10	9
3) ПСО 2 empirical	П67	4	9) НПС 21 of	10	12
4) ПСО 2 сп. 46	П37	П37	10) ППС сп. 6	П41	11
5) ПСО 2 сп. 19	П42	6	11) ППП сп. 83	П13	П9
6) ПСО 2 сп. 6	П41	7	12) ПС 1 сп. 65	П13	5

ГП

анализ	— 9	вычисление	— 41
величина	— 67	проверка	— 13
показатель	— 39	оценка	— 37
расчет	— 42		

EXCITATION

1) НПС 11 of	3	2	4) ПСО 2 core	П11	5
2) НПС 21 of	3	4	5) ПСО 2 enertia	П59	П14
3) ППС core	П11	П14			

ГП

возбуждение	— 14
возмущение	— 59
намагничивание	— 11

FAILURE

1) НПС 11 to	П64	2	15) ПСО 2 сп. 48	П11	16
2) НПС 21 to	П64	3	16) ПСО 2 сп. 41	П7	17
3) ПС 2 to	4	6	17) ПСО 2 сп. 29	П6	18
4) ПСО 2 сп. 13	П66	5	18) ППО 2 сп. 83	П11	19
5) ПС 12 сп. 13	П66	6	19) ПСО 2 open-circuit	П68	20
6) ПП -s	7	26	20) ПСО 2 short-circuit	П22	21
7) ПСО 2 successes	П30	8	21) ПСО 1 сп. 13	П28	22
8) ПС 12 successes	П30	26	22) НПС 11 in	24	23
9) ПСО 2 signal	П69	10	23) НПС 21 in	24	28
10) ПСО 2 сп. 17	П65	11	24) ППС сп. 16	П30	25
11) ПСО 2 сп. 46	П5	12	25) ППС сп. 29	П6	П5
12) ПСО 2 сп. 47	П26	13	26) НПС 11 of	28	27
13) ПСО 2 сп. 9	П11	14	27) НПС 21 of	29	9
14) ПСО 2 сп. 18	П11	15	28) НПС 21 to	П64	30

29) НПС 31 to	П64	30	34) ППС сп. 9	П11	35
30) ППС сп. 17	П65	31	35) ППС сп. 18	П11	36
31) ППС сп. 47	П26	32	36) ППС сп. 48	П11	37
32) ППС signal	П69	33	37) ППС сп. 29	П6	38
33) ППС сп. 41	П7	34	38) ППП сп. 83	П11	П5

ГП

неудача	— 30	несостоятельность	— 65
разрушение	— 26	неисправность	— 5
затухание	— 69	отсутствие	— 7
отказ	— 11	замыкание	— 22
размыкание	— 68	предел	— 66
сбой	— 6	предельный	— 28
неспособность	— 64		

INDICATE

1) НПС 11 by	3	2	17) ПС 11 сп. 39	П70	18
2) НПС 21 by	3	6	18) ПСО 2 pattern	П37	19
3) ППС сп. 18	П37	4	19) ПС 12 pattern	П37	20
4) ППС сп. 49	П47	5	20) ПСО 2 error	П46	21
5) ППП сп. 83	П37	П49	21) ПС 12 error	П46	П49
6) НПС 11 with	8	7	22) ПСО 1 error	П46	23
7) НПС 21 with	8	9	23) ПС 11 error	П46	24
8) ППС список 49	П47	П49	24) ПСО 1 pattern	П37	25
9) ПП -ed	13	10	25) ПС 11 pattern	П37	26
10) ПП -ing	11	22	26) ПСО 1 revolutions	П38	27
11) ПП1 сп. 83	П70	12	27) ПС 11 revolutions	П38	28
12) ПС 1 сп. 18	П70	П49	28) ПСО 2 список 10	П48	29
13) ПСО 2 сп. 1	18	14	29) ПС 12 список 10	П48	30
14) ПСО 1 сп. 48	П70	15	30) ПСО 2 список 49	П47	31
15) ПС 11 сп. 48	П70	16	31) ПС 12 список 49	П47	П49
16) ПСО 1 сп. 39	П70	17			

ГП

указать	— 49	распознать	— 37
обнаружить	— 46	отсчитать	— 38
обозначить	— 47	измерить	— 37
индикаторный	— 70	показать	— 48

INDICATION

1) НПС 11 that	П39	2	11) ППП сп. 82	П39	12
2) НПС 21 that	П39	3	12) ППП сп. 84	П39	П48
3) НПС 11 of	5	4	13) ПСО 2 « »	П46	14
4) НПС 21 of	5	13	14) ПСО 2 failure	П46	15
5) ППП « » (кавычки)	П46	6	15) ПСО 2 strong	П54	16
6) ППС сп. 18	П49	7	16) ПСО 2 сп. 18	П49	17
7) ППС сп. 39	П49	8	17) ПСО 2 сп. 39	П49	18
8) ППС сп. 49	П47	9	18) ПСО 2 сп. 50	П49	19
9) ППС сп. 16	П39	10	19) ПСО 2 сп. 49	П47	20
10) ППП сп. 83	П49	11	20) ППО 2 сп. 83	П49	П48

ГП

указание	— 48	основание	— 54
сигнал	— 46	показатель	— 39
индикация	— 49	обозначение	— 47

INDUCE

1) НПС 11 through	П20	2	11) ППО 2 сп. 82	П32	12
2) НПС 21 through	П20	3	12) ПП 12 сп. 82	П32	13
3) НПС 11 into	П20	4	13) ППО 2 сп. 84	П32	14
4) НПС 21 into	П20	5	14) ПП 12 сп. 84	П32	15
5) НПС 11 to	7	6	15) ПСО 1 сп. 41	П8	16
6) НПС 21 to	7	8	16) ПС 11 сп. 41	П8	17
7) ППС список 2	П20	8	17) ППО 1 сп. 82	П32	18
8) ПП -ed	9	15	18) ПП 11 сп. 82	П32	19
9) ПСО 2 сп. 41	П8	10	19) ППО 1 сп. 84	П32	20
10) ПС 12 сп. 41	П8	11	20) ПП 11 сп. 84	П32	П51

ГП

вызывать	— 32	вводить	— 20
заставлять	— 11	создавать	— 51
индуцировать	— 8		

INSTALL

1) НПС 11 as	П20	2	10) ПСО 2 сп. 17	П35	11
2) НПС 21 as	П20	3	11) ПС 12 сп. 17	П35	12
3) НПС 11 with	П25	4	12) ПС 02 сп. 48	П29	13
4) НПС 21 with	П25	5	13) ПС 12 сп. 48	П29	14
5) НПС 11 into	П21	6	14) ПСО 1 сп. 17	П35	15
6) НПС 21 into	П21	7	15) ПС 11 сп. 17	П35	16
7) ПП -ed	8	14	16) ПСО 1 сп. 48	П29	17
8) ПСО 2 newly	П20	9	17) ПС 11 сп. 48	П29	П23
9) ПС 12 newly	П20	10			

ГП

установить	— 23	вмонтировать	— 21
ввести	— 20	расположить	— 29
смонтировать	— 25	применить	— 35

INSTALLATION

1) ПП -s	П23	2	9) ПСО 2 сп. 2	П23	10
2) НПС 11 of	4	3	10) ПСО 2 computer	П36	11
3) НПС 21 of	4	9	11) ПСО 2 full	П20	12
4) ППС сп. 9	П25	5	12) ПСО 2 сп. 9	П25	13
5) ППС сп. 18	П25	6	13) ПСО 2 сп. 18	П25	14
6) ППС сп. 31	П20	7	14) ПСО 2 сп. 31	П20	15
7) ППС сп. 45	П19	8	15) ПСО 2 сп. 45	П19	16
8) ППС сп. 17	П35	П23	16) ПСО 2 сп. 17	П35	П23

ГП

установка	— 23	ввод	— 20
внедрение	— 35	монтаж	— 25
введение	— 19	ЭВМ	— 36

INTEGRATE

1) ПП -ed	2	13	5) ПП 1 сп. 83	П24	6
2) ПС 1 сп. 51	П24	3	6) ПСО 2 сп. 1	7	13
3) ПС 1 сп. 41	П24	4	7) ПСО 2 сп. 6	П45	8
4) ПС 1 сп. 9	П25	5	8) ПС 12 сп. 6	П45	9

- 9) ПСО 2 сп. 39
10) ПС 12 сп. 39
11) ПСО 2 сп. 38
12) ПС 12 сп. 38
13) ПСО 1 сп. 6

П45 10
П45 11
П45 12
П45 13
П45 14

- 14) ПС 11 сп. 6
15) ПСО 1 сп. 39
16) ПС 11 сп. 39
17) ПСО 1 сп. 38
18) ПС 11 сп. 38

П45 15
П45 16
П45 17
П45 18
П45 П56

- 11) ПС 1 сп. 54
12) ПС 1 сп. 77
13) ПС 1 сп. 29
14) НПС 11 of
15) НПС 21 at

П6 12
П34 13
П34 14
15 П34
П17 16

- 16) НПС 31 at
17) НПС curriculum
18) НПС сп. 53
19) ППП сп. 82
20) ППП сп. 84

П17 17
П8 18
П16 19
П17 20
П17 П34

ГП

объединять — 56
интегрировать — 45
интегральный — 24
встроенный — 25

LEAKAGE

- 1) НПС 11 сп. 52
2) НПС 21 сп. 52
3) НПС 11 to

П30 2
П30 3
П62 4

- 4) НПС 21 to
5) НПС 11 between
6) НПС 21 between

П62 5
П68 6
П68 П69

ГП

утечка — 69
протекание — 30
натекание — 62
неплотность — 68

MAINTAIN

- 1) ПП -ed
2) НПС 11 at
3) НПС 21 at
4) НПС 11 in
5) НПС 21 in
6) ПСО 2 size
7) ПС 12 size
8) ПСО 2 сп. 41
9) ПС 12 сп. 41
10) ПСО 2 сп. 18
11) ПС 12 сп. 18
12) ПСО 2 сп. 29
13) ПС 12 сп. 29
14) ПСО 2 сп. 53
15) ПС 12 сп. 53
16) НПС 11 that

2 16
П62 3
П62 4
П62 5
П62 6
П67 7
П67 8
П18 9
П18 10
П34 11
П34 12
П8 13
П8 14
П16 15
П16 16
П58 17

- 17) НПС 21 that
18) НПС 11 at
19) НПС 21 at
20) НПС 11 in
21) НПС 21 in
22) ПСО 1 size
23) ПС 11 size
24) ПСО 1 сп. 41
25) ПС 11 сп. 41
26) ПСО 1 сп. 18
27) ПС 11 сп. 18
28) ПСО 1 сп. 53
29) ПС 11 сп. 53
30) ПСО 1 сп. 29
31) ПС 11 сп. 29

П58 18
П62 19
П62 20
П62 21
П62 22
П67 23
П67 24
П18 25
П18 26
П34 27
П34 28
П16 29
П16 30
П8 31
П8 П17

ГП

поддерживать — 17
выдерживать — 67
эксплуатировать — 34
утверждать — 58
содержать — 63
проводить — 18
вести — 27
выполнять — 8

MAINTENANCE

- 1) ПСО 2 curriculum
2) ПСО 2 сп. 53
3) ПСО 2 сп. 34
4) ПСО 2 сп. 29
5) ПСО 2 сп. 77

П8 2
П16 3
П69 4
П8 5
П34 6

- 6) ПСО 2 сп. 85
7) ППО 2 сп. 82
8) ППО 2 сп. 84
9) ПС 1 сп. 34
10) ПС 1 сп. 85

П12 7
П17 8
П17 9
П32 10
П12 11

обслуживание — 34
поддержание — 17
эксплуатация — 6
дотация — 32

ГП

выполнение — 8
ведение — 16
ремонт — 12
эксплуатационные расходы — 69

MULTIPLICATION

- 1) НПС 11 of
2) НПС 21 of
3) НПС сп. 40
4) НПС сп. 41

3 2
3 5
П43 4
П43 П45

- 5) ПСО 2 сп. 52
6) ПСО 2 сп. 40
7) ПСО 2 сп. 41

П44 6
П43 7
П43 П45

ГП

умножение — 45
увеличение — 44
усиление — 43

MULTIPLY

- 1) ПП -ed
2) ПСО 2 сп. 41
3) ПС 12 сп. 11
4) ПСО 2 сп. 40
5) ПС 12 сп. 40
6) ПСО 2 сп. 41
7) ПС 12 сп. 41
8) ПСО 1 сп. 11
9) ПС 11 сп. 11
10) ПСО 1 сп. 40

2 8
П44 3
П44 4
П43 5
П43 6
П43 7
П43 8
П44 9
П44 10
П43 11

- 11) ПС 11 сп. 40
12) ПСО 1 сп. 41
13) ПС 11 сп. 41
14) НПС 11 by
15) НПС 21 by
16) НПС 11 out
17) НПС 21 out
18) НПС 11 together
19) НПС 21 together

П43 12
П43 13
П43 14
П45 15
П45 16
П37 17
П37 18
П37 19
П37 П45

ГП

усилить — 43
умножить — 45
перемножить — 37
увеличить — 44

OPERATE

- 1) ПП -ed
2) ПП -ing
3) ПСО 1 сп. 10
4) ПС 11 сп. 10
5) ПСО 1 сп. 13
6) ПС 11 сп. 13
7) ПСО 1 сп. 39
8) ПС 11 сп. 39
9) ПСО 1 сп. 41
10) ПС 11 сп. 41
11) ПСО 1 сп. 57
12) ПС 11 сп. 57

15 2
3 18
П5 4
П5 5
П5 6
П5 7
П5 8
П5 9
П5 10
П5 11
П36 12
П36 13

- 13) ПСО 1 сп. 45
14) ПС 11 сп. 45
15) ПС 2 сп. 78
16) ПС 2 сп. 55
17) ПСО 2 сп. 1
18) НПС 11 by
19) НПС 21 by
20) НПС 11 with
21) НПС 21 with
22) НПС 11 on
23) НПС 21 on
24) НПС 11 upon

П33 14
П33 П6
П34 16
П34 17
44 18
П54 19
П54 20
П8 21
П8 22
26 23
26 24
26 25

25) НПС 21 урон	26	28	41) ПС 12 сп. 17	П35	42
26) ППС сп. 56	П8	27	42) ПСО 2 сп. 29	П35	43
27) ППС сп. 58	П6	П11	43) ПС 12 сп. 29	П35	П6
28) ПСО 1 сп. 17	П51	29	44) НПС 11 by	П34	45
29) ПС 11 сп. 17	П51	30	45) НПС 21 by	П34	46
30) ПСО 1 сп. 29	П51	31	46) ПСО 2 сп. 55	П34	47
31) ПС 11 сп. 29	П51	32	47) ПС 12 сп. 55	П34	48
32) ПСО 1 school	П49	33	48) ПС 2 сп. 78	П34	49
33) ПС 11 school	П49	34	49) ПС 11 сп. 78	П34	50
34) ПСО 1 permutations	П52	35	50) НПС 11 on	54	51
35) ПС 11 permutations	П52	36	51) НПС 21 on	54	52
36) НПС 11 список 2	П34	37	52) НПС 11 урон	54	53
37) НПП 11 -s	П34	38	53) НПС 21 урон	54	П6
38) ПСО 2 process	П7	39	54) ППС сп. 58	П6	55
39) ПС 12 process	П7	40	55) ПС 12 сп. 56	П8	П6
40) ПСО 2 сп. 17	П35	41			

ГП

работать	— 6	управляться	— 54
применяться	— 35	руководить	— 49
влять	— 11	управлять	— 34
применять	— 51	производить	— 52
оперировать	— 8	эксплуатационный	— 36
рабочий	— 5	обслуживающий	— 33
протекать	— 7		

OPERATION

1) ПП -s	25	2	14) НПС 11 урон	П11	15
2) ПСО 2 сп. 2	21	3	15) НПС 21 урон	П11	16
3) ПС 2 full	П64	4	16) НПС 11 of	18	17
4) ПСО 2 сп. 59	П11	5	17) НПС 21 of	18	П6
5) ПСО 2 сп. 60	П8	6	18) ППС сп. 59	П11	19
6) ПСО 2 сп. 62	П34	7	19) ППС сп. 60	П8	20
7) ПСО 2 сп. 17	П35	8	20) ППС сп. 17	П35	П6
8) ПСО 2 colour	П32	9	21) ПСО 2 сп. 59	П11	22
9) НПС сп. 90	П11	10	22) НПС 11 of	24	23
10) НПС 11 by	П34	11	23) ППС 21 of	24	П7
11) НПС 21 by	П34	12	24) ППС сп. 59	П11	П7
12) НПС 11 on	П11	13	25) ПС 2 сп. 61	П6	26
13) НПС 21 on	П11	14	26) ПС 2 scientific	П9	24

ГП

выполнение	— 8	применение	— 35
работа	— 6	мощность	— 64
действие	— 11	управление	— 34
исследование	— 9	изображение	— 32
операция	— 7		

PENETRATE

1) НПС 11 with	П49	2
2) НПС 21 with	П49	3
3) ПСО 1 сп. 63	П58	4
4) ПС 11 сп. 63	П58	П21

ГП

проникать	— 21
понимать	— 58
пронитывать	— 19

PERFORMANCE

1) НПС 11 on	П11	2	13) ПСО 2 сп. 64	П5	14
2) НПС 21 on	П11	3	14) ПСО 2 сп. 66	П65	15
3) НПС 11 of	5	4	15) ПСО 2 сп. 67	П63	16
4) НПС 21 of	5	11	16) ПСО 2 сп. 45	П38	17
5) ППС сп. 60	П8	6	17) ПСО 2 сп. 68	П38	18
6) ППС сп. 80	П7	7	18) ПСО 2 сп. 17	П35	19
7) ППС сп. 17	П35	8	19) ППО 2 сп. 82	П66	20
8) ППС сп. 66	П65	9	20) ППО 2 сп. 84	П39	21
9) ППС сп. 67	П63	10	21) ПСО 1 сп. 10	П38	22
10) ППС сп. 45	П38	П70	22) ПСО 1 сп. 65	П6	23
11) ПСО 2 сп. 60	П8	12	23) ПСО 2 сп. 17	П66	П70
12) ПСО 2 сп. 80	П7	13			

ГП

характеристика	— 70	выполнение	— 8
осуществление	— 7	функция	— 5
действие	— 11	применение	— 35
показатель	— 39	качество	— 65
приемистость	— 63	производительность	— 38
эффективность	— 66	эксплуатация	— 6

PRODUCTION

1) ПСО 2 full	П64	2	15) ППС сп. 29	П25	16
2) ПСО 2 сп. 34	П38	3	16) ППС сп. 39	П32	17
3) НПС of	4	5	17) ППС сп. 41	П32	18
4) ПС 12 precision	П11	5	18) ППС сп. 68	П31	П52
5) ПСО 1 сп. 9	П56	6	19) ПСО 2 defence	П36	20
6) ПСО 1 сп. 18	П56	7	20) ПСО 2 torque	П39	21
7) НПС 11 of	9	8	21) ПСО 2 сп. 36	П51	22
8) НПС 21 of	9	19	22) ПСО 2 сп. 40	П51	23
9) ППС drawing	П8	10	23) ПСО 2 сп. 29	П25	24
10) ППС art	П37	11	24) ПСО 2 сп. 39	П32	25
11) ППС torque	П39	12	25) ПСО 2 сп. 41	П32	26
12) ППС сп. 35	П51	13	26) ПСО 2 сп. 35	П51	27
13) ППС сп. 36	П51	14	27) ПСО 2 сп. 68	П31	28
14) НПС сп. 40	П51	15	28) ПС 12 сп. 68	П31	П52

ГП

производство	— 52	производительность	— 38
мощность	— 64	промышленность	— 36
обработка	— 11	выполнение	— 8
образование	— 51	получение	— 32
составление	— 25	произведение	— 37
выработка	— 39	серийный	— 56
продукция	— 31		

RATIO

1) НПС 11 of	3	2	3) ППС mixture	П26	4
2) НПС 21 of	3	10	4) ППС reinforcement	П37	5

словоформ. Назовем результатом работы переводной эквивалент, определенный нашей моделью на основе анализа контекста. Положительным будем считать результат, совпадающий с «человеческим» переводом. Отрицательным назовем результат, не совпадающий с «человеческим».

Коэффициентом соответствия модели оригиналу считаем отношение числа положительных результатов к общему числу результатов.

Для получения предварительной оценки эффективности модели проведен эксперимент. Из текстов по автомобилестроению, вычислительной технике, строительным материалам и электронным приборам выбраны предложения, представляющие 1000 словоупотреблений исследуемых неоднозначных словоформ. Массив обработан по нашему алгоритму (рис. 3). Этот же массив предложений был предъявлен четырем специалистам-переводчикам. Информанты перевели неоднозначные английские словоформы на русский язык. Их переводы сравнивались с переводами, полученными в результате проверки алгоритма.

Совпадение переводов фиксировалось как положительный результат работы модели. Несовпадение считалось отрицательным результатом. Отношение количества положительных результатов к общему числу результатов составило 0.933. Следовательно, эффективность модели равна 93.3%.

Экспериментальная проверка модели выявила некоторые неточности, связанные с вероятностным подходом к отбору материала. В соответствии с этим подходом учитывались наиболее частые ДП. Для ДП, не учтенных доалгоритмической моделью, предусмотрен «общий выход». Им является частый эквивалент с наиболее общим значением. В ряде случаев «общий выход» является неадекватным переводом.

Некоторые неточности перевода наблюдаются, когда ДП удалены от принятой зоны поиска.

Зафиксировано несколько ошибок, возникших вследствие несоблюдения критериев объединения списков ДП.

Неточности, выявленные в ходе описанного эксперимента, не имеют принципиального характера. Они будут устраняться по мере проверки работы программы на больших массивах текста.

Предварительная проверка модели позволяет считать, что модель имеет практическую ценность и, следовательно, может быть использована в системе МП.

ПРИЛОЖЕНИЕ

Списки диагностирующих признаков

Сп. 1		Сп. 2
1) was	3) are	1) a
2) is	4) be	2) an
	5) were	
	6) been	

3) the	10) averages	Сп. 8
4) this	11) mean	1) congress
5) these	12) number	2) congresses
6) those	13) numbers	3) conference
7) each	14) fraction	4) conferences
8) single	15) fractions	5) session
9) every	16) square	6) sessions

Сп. 3

1) minor	17) squares	7) meeting
2) permissible	18) root	8) meetings
3) executive	19) roots	9) symposium
4) non-executive	20) sum	10) symposia
5) professional	21) sums	11) trip
	22) product	12) trips
	23) products	13) gathering
	24) series	14) gatherings

Сп. 4

1) size	25) vector	15) visit
2) sizes	26) vectors	16) visits
3) level	27) gain	17) courses
4) levels	28) gains	18) school
5) angle	29) quadratics	19) schools
6) angles	30) column	20) calling
7) clearance	31) columns	21) callings
8) clearances	32) statistics	22) interview
9) balance	33) constant	23) interviews
10) balances	34) constants	
11) lever	35) variable	
12) levers	36) variables	
13) smoke	37) component	
14) spot	38) components	
15) spots	39) element	
16) feedback	40) elements	
	41) polynomial	
	42) polynomials	
	43) derivative	
	44) derivatives	
	45) integrand	
	46) integrands	
	47) integral	
	48) integrals	
	49) increment	
	50) increments	
	51) proportion	
	52) proportions	
	53) implicant	
	54) implicants	
	55) quantity	
	56) quantities	
	57) geometry	

Сп. 5

1) zero	7) flange	1) joint
2) null	8) flanges	2) joints
3) infinity	9) bracket	3) unit
4) camera	10) brackets	4) units
5) cameras	11) screw	5) bolt
6) aerial	12) screws	6) bolts
7) aerials	13) clip	7) flange
8) abbreviation	14) clips	8) flanges
9) abbreviations	15) cam	9) bracket
10) words	16) cams	10) brackets
11) word	17) lip	11) screw
12) phrase	18) lips	12) screws
13) phrases	19) belt	13) clip
	20) belts	14) clips
	21) wheel	15) cam
	22) wheels	16) cams
	23) top	17) lip
	24) tops	18) lips
	25) plant	19) belt
	26) plants	20) belts
	27) crancase	21) wheel
	28) crancases	22) wheels
	29) arm	23) top
	30) arms	24) tops
	31) boom	25) plant
	32) booms	26) plants

Сп. 6

1) value	1) trust
2) values	2) trusts
3) parameter	3) marketing
4) parameters	4) credit
5) coefficient	5) credits
6) coefficients	6) private
7) ratio	7) lease
8) ratios	8) loan
9) average	9) loans

Сп. 7

1) congress
2) congresses
3) conference
4) conferences
5) session
6) sessions
7) meeting
8) meetings
9) symposium
10) symposia
11) trip
12) trips
13) gathering
14) gatherings
15) visit
16) visits
17) courses
18) school
19) schools
20) calling
21) callings
22) interview
23) interviews

Сп. 9

1) joint
2) joints
3) unit
4) units
5) bolt
6) bolts
7) flange
8) flanges
9) bracket
10) brackets
11) screw
12) screws
13) clip
14) clips
15) cam
16) cams
17) lip
18) lips
19) belt
20) belts
21) wheel
22) wheels
23) top
24) tops
25) plant
26) plants
27) crancase
28) crancases
29) arm
30) arms
31) boom
32) booms

33) conductor
34) conductors
35) chart
36) charts
37) bumper
38) bumpers
39) jack
40) jacks
41) throat
42) throats
43) pindle
44) pindles
45) shaft
46) shafts
47) member
48) members
49) taper
50) tapers
51) cushion
52) cushions
53) duct
54) ducts
55) link
56) links
57) tube
58) tubes
59) gate
60) gates
61) bulb
62) bulbs
63) cell
64) cells
65) bearing
66) bearings
67) anchorage
68) form
69) forms
70) shell
71) shells
72) module
73) modules
74) piston
75) pistons
76) ring
77) rings
78) stamping
79) stampings
80) electrode
81) electrodes
82) bank
83) banks

Cn. 10

1) data
2) error
3) errors
4) result
5) results
6) characteristic

7) characteristics
8) defect
9) defects
10) condition
11) conditions
12) fact
13) facts
14) response
15) responses
16) test
17) tests
18) example
19) examples
20) experiment
21) experiments
22) study
23) studies
24) experience
25) figure
26) figures
27) measure
28) measures
29) measurement
30) measurements
31) standard
32) standards
33) criterium
34) criteria
35) requirement
36) requirements
37) index
38) indices

Cn. 11

1) organizational
2) administrative
3) engineering
4) advisory
5) consultative
6) magnitude
7) magnitudes
8) capabilities
9) rate
10) rates

Cn. 12

1) input
2) inputs
3) output
4) outputs
5) switch
6) switches
7) drive
8) memory
9) memories
10) multiaccess
11) lag
12) lags

13) block
14) blocks
15) shade
16) shades
17) photocell
18) photocells
19) turbine
20) turbines
21) brake
22) brakes
23) trive
24) trives
25) intake
26) intakes
27) test
28) tests

Cn. 13

1) load
2) loaded
3) loading
4) fee
5) ultimate
6) stress

Cn. 14

1) matrix
2) display
3) man
4) response
5) presence
6) cause
7) causes
8) inverse
9) inverses
10) vector
11) vectors
12) direct

Cn. 15

1) logical
2) functional
3) schematic
4) self-switching
5) wet
6) loose
7) bulk
8) dry
9) heterodyne
10) superheterodyne

Cn. 16

1) efficiency
2) existence
3) operation
4) operations
5) application

6) applications
7) applicability
8) use
9) work
10) performance
11) performances
12) life
13) behavior
14) behaviour

Cn. 17

1) method
2) methods
3) problem
4) problems
5) subject
6) subjects
7) theory
8) theories
9) procedure
10) procedures
11) technology
12) technique
13) treatment
14) principle
15) techniques
16) treatments
17) principles
18) idea
19) ideas
20) concept
21) concepts
22) approach
23) approaches
24) definition
25) definitions

Cn. 18

1) device
2) devices
3) equipment
4) machine
5) machines
6) engine
7) engines
8) vehicle
9) vehicles
10) model
11) models
12) car
13) cars
14) truck
15) trucks
16) lorry
17) lorries
18) motor
19) motors
20) system

21) systems
22) lamp
23) lamps
24) gauge
25) gauges
26) gage
27) gages
28) board
29) boards
30) battery
31) batteries
32) tool
33) tools
34) instrument
35) instruments
36) line
37) lines
38) tire
39) tires
40) mechanism
41) mechanisms
42) hardware
43) software
44) press
45) antenna
46) specific
47) rated
48) idle
49) reserve
50) productive
51) available
52) diode
53) diodes
54) aggregate
55) aggregates
56) kilowatt
57) watt
58) mw
59) kw
60) w

Cn. 19

1) statistical
2) mathematical
3) economic
4) numerical
5) accurate

Cn. 20

1) storage
2) store
3) memory
4) memories
5) cable
6) cables
7) bit
8) bits
9) digit
10) digits

11) information
12) tape
13) tapes
14) drum
15) drums
16) disc
17) discs
18) element
19) elements

Cn. 21

1) brain
2) paper
3) papers
4) magazine
5) magazines
6) void
7) voids
8) gear
9) gears
10) production
11) cylinder
12) cylinders
13) air
14) fuel
15) oil
16) petrol
17) liquid

Cn. 22

1) receiver
2) receivers
3) canister
4) canisters
5) reservoir
6) reservoirs
7) carrier
8) carriers
9) boot
10) boots
11) system
12) systems
13) sump
14) container
15) sumps
16) containers
17) silo
18) silos
19) tank
20) tanks
21) mixer
22) mixers
23) net
24) ton
25) tons
26) gallon
27) gallons
28) gal

29) pint	Cn. 25	Cn. 28			
30) pints		1) scheme	30) statement	30) stadia	15) supervisor
31) pt	1) in ²	2) schemes	31) classification	31) ship	16) regulation
32) litre	2) ft ²	3) lattice	32) classifications	32) ships	17) regulations
33) litres	3) yd ²	4) lattices	33) dictionary	33) pile	18) activities
34) lt	4) m ²	5) tree	34) dictionaries	34) piles	19) joint
35) yd ²	5) cm ²	6) trees		35) viaduct	20) joints
36) quart	6) sq.	7) graph	Cn. 30	36) viaducts	21) tour
37) quarts	7) square	8) graphs	1) hydrotechnical	37) deck	22) tours
38) qt	8) Z in×Z in	9) chain	2) domestic	38) decks	23) site
39) ft ²	9) Z yd×Z yd	10) chains	3) masonry	39) structure	24) sites
40) in ²	10) Z cm×Z cm	11) chart	4) engineering	40) structures	25) technical
41) cu	11) Z ft×Z ft	12) charts	5) wood	41) cathedral	26) lift
42) cm ³	12) Z m×Zm	13) diagram	6) timber	42) cathedrals	27) lifts
43) m ³		14) diagrams	7) concrete	43) shell	28) depth
	Cn. 23	15) nomogram	8) multistorey	44) shells	29) depths
1) room		16) nomograms	9) apartment	45) dock	
2) rooms	Cn. 26	17) meridian	10) underground	46) docks	Cn. 33
3) hall	1) ton	18) meridians	11) water	47) arch	1) sturdy
4) halls	2) tons	19) array	12) winter	48) arches	2) aluminum
5) hand	3) kip	20) arrays	13) weather	49) block	3) monocoque
6) hands	4) kips	21) vertex	14) marine	50) blocks	4) sandwich
7) bend	5) lb	22) curve	15) building	51) subway	5) robust
8) bends	6) bolt	23) curves	16) city	52) subways	6) rigid
9) wave	7) bolts	24) line	17) composite	53) tunnel	7) interior
10) waves	8) tire	25) lines	18) assembly-line	54) tunnels	8) handy
11) connector	9) tires	26) graphical	19) industrialized	55) tank	9) solid-state
12) connectors	10) body		20) resistant	56) tanks	10) angle-bracket
13) pulley	11) bodies	Cn. 29		57) scaffold	11) sliding
14) pulleys	12) crane	1) program	Cn. 31	58) scaffolds	12) waffle
15) deflection	13) cranes	2) programs	1) project	59) hopper	13) unitary
16) evaluation	14) spring	3) routine	2) projects	60) hoppers	14) honey-comb
17) evaluations	15) springs	4) routines	3) centre	61) runway	15) radial
18) contribution	16) bearing	5) algorithm	4) centres	62) culvert	16) bias
19) emission	17) bearings	6) algorithms	5) foundation	63) station	
	18) lifting	7) schedule	6) foundations	64) stations	Cn. 34
	19) cargo	8) schedules	7) basement	65) pavilion	1) high
		9) scheduling	8) basements	66) pavilions	2) low
		10) timetable	9) chimney	67) silo	3) grant
		11) timetables	10) chimneys	68) silos	
		12) table	11) wall	69) slab	Cn. 35
		13) tables	12) walls	70) slabs	1) field
		14) rule	13) dwelling	71) span	2) fields
		15) rules	14) dwellings	72) spans	3) flux
		16) planning	15) mill		4) magnetic
		17) list	16) mills	Cn. 32	5) range
		18) address	17) factory	1) industry	6) ranges
		19) instruction	18) factories	2) engineer	7) region
		20) manual	19) plant	3) engineers	8) regions
		21) lists	20) plants	4) management	9) limit
		22) addresses	21) works	5) machinery	10) limits
		23) instructions	22) office	6) procedure	
		24) manuals	23) offices	7) procedures	
		25) specification	24) metro	8) information	
		26) specifications	25) dam	9) firm	Cn. 36
		27) agreement	26) dams	10) firms	1) electron
		28) agreements	27) airport	11) worker	2) electrons
		29) statements	28) airports	12) workers	3) hole
			29) stadium	13) conference	4) holes
				14) conferences	

- 5) carrier
- 6) carriers
- 7) point
- 8) points
- 9) ion
- 10) ions
- 11) photon
- 12) photons
- 13) donor
- 14) donors
- 15) trap
- 16) traps
- 17) net
- 18) nets
- 19) doping
- 20) acceptor
- 21) acceptors
- 22) access
- 23) defect
- 24) defects
- 25) impurity

Cr. 37

- 1) book
- 2) chapter
- 3) books
- 4) chapters
- 5) appendix
- 6) flaw
- 7) flaws
- 8) crack
- 9) cracks

Cr. 38

- 1) expression
- 2) expressions
- 3) formula
- 4) formulas
- 5) equation
- 6) equations
- 7) theorem
- 8) theorems
- 9) proof
- 10) relationship
- 11) relationships
- 12) Z
- 13) task
- 14) approximation
- 15) tasks
- 16) correlation

Cr. 39

- 1) frequency
- 2) frequencies
- 3) quality
- 4) speed
- 5) speeds

- 6) mode
- 7) modes
- 8) temperature
- 9) temperatures
- 10) stresses
- 11) stress
- 12) pressure
- 13) pressures
- 14) time
- 15) force
- 16) forces
- 17) velocity
- 18) velocities
- 19) power
- 20) powers
- 21) weight
- 22) weights
- 23) density
- 24) densities
- 25) distance
- 26) distances
- 27) length
- 28) lengths
- 29) height
- 30) heights
- 31) period
- 32) periods
- 33) phase
- 34) phases
- 35) resistance
- 36) cycle
- 37) cycles
- 38) condition
- 39) conditions
- 40) rate
- 41) rates
- 42) sequence
- 43) degree
- 44) regime
- 45) process
- 46) processes
- 47) systems
- 48) systems
- 49) energy
- 50) energies
- 51) sequences

Cr. 40

- 1) signal
- 2) pulse
- 3) signals
- 4) pulses
- 5) wave
- 6) waves
- 7) diode
- 8) diodes
- 9) linear
- 10) multigrid

- 11) vapour
- 12) plate
- 13) plates
- 14) peak
- 15) peaks
- 16) output
- 17) modulation
- 18) modulations
- 19) input
- 20) component
- 21) components
- 22) coherent
- 23) slope
- 24) slopes
- 25) envelope
- 26) envelopes
- 27) parabolic
- 28) radiation

Cr. 41

- 1) current
- 2) currents
- 3) ac
- 4) dc
- 5) a. c.
- 6) d. c.
- 7) voltage
- 8) voltages
- 9) circuit
- 10) circuits
- 11) network
- 12) networks
- 13) image
- 14) images

Cr. 42

- 1) order
- 2) orders
- 3) finite
- 4) forward
- 5) central
- 6) normal
- 7) Boolean
- 8) extreme
- 9) numerical
- 10) proportional
- 11) partial
- 12) reciprocal
- 13) divided
- 14) squared
- 15) inverted
- 16) nominal
- 17) small
- 18) smaller
- 19) scale
- 20) scales
- 21) mean
- 22) means

- 23) sum
- 24) sums
- 25) percentage

Cr. 43

- 1) some
- 2) any
- 3) no
- 4) much
- 5) little
- 6) slight
- 7) great
- 8) a

Cr. 44

- 1) make
- 2) makes
- 3) made
- 4) transfer
- 5) transport
- 6) supply
- 7) aperture
- 8) beam
- 9) beams
- 10) light

Cr. 45

- 1) engineer
- 2) engineers
- 3) operator
- 4) operators
- 5) worker
- 6) workers
- 7) driver
- 8) drivers
- 9) officer
- 10) officers
- 11) interpreter
- 12) interpreters
- 13) personnel
- 14) pilot
- 15) pilots
- 16) astronaut
- 17) astronauts

Cr. 46

- 1) quantitative
- 2) qualitative
- 3) electrical
- 4) environmental
- 5) mechanical

Cr. 47

- 1) stress
- 2) stresses
- 3) strength
- 4) strengths

- 5) strain
- 6) strains
- 7) load
- 8) loads
- 9) compression
- 10) instability
- 11) shear
- 12) shears
- 13) connection
- 14) brittle
- 15) diagonal
- 16) flexural
- 17) plastic
- 18) bending
- 19) tensile
- 20) static
- 21) compressive
- 22) bearing
- 23) splitting
- 24) buckling
- 25) ultimate
- 26) tension
- 27) starting
- 28) connector
- 29) connectors
- 30) anchorage
- 31) frame
- 32) frames
- 33) formwork
- 34) bond
- 35) bonds
- 36) web
- 37) webs
- 38) form
- 39) forms
- 40) concrete
- 41) concretes

Cr. 48

- 1) supply
- 2) drive
- 3) spark
- 4) sparks
- 5) ignition
- 6) electronic
- 7) efficiency
- 8) work
- 9) output
- 10) horsepower
- 11) service
- 12) centre
- 13) centres
- 14) surface
- 15) surfaces

Cr. 49

- 1) bracket
- 2) brackets

- 3) letter
- 4) letters
- 5) comma
- 6) commas
- 7) circle
- 8) circles
- 9) sign
- 10) signs
- 11) subscript
- 12) subscripts
- 13) numeral
- 14) numerals
- 15) index
- 16) indices
- 17) line
- 18) lines
- 19) graph
- 20) graphs
- 21) curve
- 22) curves
- 23) relation
- 24) relations
- 25) Z
- 26) ellipse
- 27) scale
- 28) scales
- 29) ellipses

Cr. 50

- 1) visible
- 2) null
- 3) sensitive
- 4) immediate
- 5) direct
- 6) rev/min
- 7) remote
- 8) accurate
- 9) back
- 10) left-right
- 11) nominal
- 12) output
- 13) continuous
- 14) electronic
- 15) visual
- 16) rough
- 17) signal
- 18) signals
- 19) ampere
- 20) amperes
- 21) volt
- 22) volts
- 23) state
- 24) states

Cr. 51

- 1) array
- 2) arrays

- 3) supply
- 4) element
- 5) elements
- 6) package
- 7) packages
- 8) flux
- 9) mode
- 10) unit
- 11) units
- 12) gate
- 13) matrix
- 14) structure
- 15) structures
- 16) component
- 17) components
- 18) microsystem
- 19) microsystems
- 20) trajectory
- 21) trajectories

Cn. 52

- 1) along
- 2) through
- 3) at
- 4) dominant
- 5) strength
- 6) torque

Cn. 53

- 1) records
- 2) record
- 3) check
- 4) checks
- 5) file
- 6) files

Cn. 54

- 1) cost
- 2) costs
- 3) charge
- 4) charges
- 5) engineer
- 6) engineers
- 7) repair
- 8) inspection

Cn. 55

- 1) cam
- 2) cams
- 3) solenoid
- 4) solenoids
- 5) pilot
- 6) pilots
- 7) valve
- 8) valves
- 9) voice
- 10) cell

- 11) cells
- 12) battery
- 13) batteries
- 14) chopper
- 15) choppers
- 16) mains
- 17) vacuum
- 18) pedal
- 19) pushrod
- 20) pedals
- 21) pushrods
- 22) DRO

Cn. 56

- 1) column
- 2) columns
- 3) data
- 4) image
- 5) images
- 6) list
- 7) lists
- 8) array
- 9) arrays
- 10) address
- 11) addresses
- 12) curve
- 13) curves
- 14) function
- 15) functions
- 16) row
- 17) rows
- 18) axis
- 19) feature
- 20) features
- 21) forms
- 22) streams
- 23) values
- 24) stream
- 25) value
- 26) sequence
- 27) sequences
- 28) signal
- 29) signals
- 30) number
- 31) numbers
- 32) information
- 33) results
- 34) result
- 35) item
- 36) items
- 37) word
- 38) words
- 39) element
- 40) elements
- 41) command
- 42) commands
- 43) code
- 44) codes

- 45) record
- 46) records
- 47) file
- 48) files
- 49) portion
- 50) portions
- 51) stack
- 52) stacks
- 53) block
- 54) blocks
- 55) segment
- 56) segments
- 57) expression
- 58) expressions
- 59) tuple

Cn. 57

- 1) manual
- 2) manuals
- 3) instruction
- 4) instructions
- 5) life
- 6) cost
- 7) costs
- 8) status
- 9) tolerance
- 10) advantage
- 11) advantages
- 12) limitations
- 13) capability
- 14) capabilities
- 15) requirement
- 16) requirements
- 17) procedure
- 18) procedures
- 19) performance
- 20) performances
- 21) efficiency
- 22) efficiencies
- 23) data
- 24) economy
- 25) principle
- 26) principles
- 27) regulation
- 28) regulations
- 29) feature
- 30) features

Cn. 58

- 1) fuel
- 2) oil
- 3) principle
- 4) street
- 5) road
- 6) highway
- 7) principles
- 8) streets
- 9) highways
- 10) roads

Cn. 59

- 1) arithmetic
- 2) arithmetics
- 3) arithmetical
- 4) addition
- 5) additions
- 6) additive
- 7) subtractive
- 8) subtraction
- 9) subtractions
- 10) divisions
- 11) division
- 12) multiplication
- 13) multiplications
- 14) multiplicative
- 15) feedback
- 16) false
- 17) difference
- 18) differences
- 19) logic
- 20) logical
- 21) point
- 22) points
- 23) matrix
- 24) numerical
- 25) column
- 26) columns
- 27) elementary
- 28) row
- 29) rows

Cn. 60

- 1) tasks
- 2) task
- 3) step
- 4) steps
- 5) function
- 6) functions
- 7) budget
- 8) work
- 9) program
- 10) programs
- 11) plan
- 12) plans
- 13) process
- 14) processes

Cn. 61

- 1) field
- 2) concrete
- 3) road
- 4) aircraft

Cn. 62

- 1) hand
- 2) manual
- 3) automatic
- 4) direct

- 5) external
- 6) remote
- 7) electric

Cn. 63

- 1) logic
- 2) plan
- 3) meaning
- 4) idea
- 5) thought
- 6) theory
- 7) plans
- 8) meanings
- 9) ideas
- 10) thoughts
- 11) theories

Cn. 64

- 1) cognitive
- 2) psychomotor

Cn. 65

- 1) test
- 2) tests
- 3) testing
- 4) trial
- 5) trials
- 5) reliability

Cn. 66

- 1) construction
- 2) material
- 3) concrete
- 4) specimen
- 5) component
- 6) sealant
- 7) flight
- 8) aircraft
- 9) materials
- 10) coneretes
- 11) components
- 12) sealants
- 13) flights
- 14) aircrafts

Cn. 67

- 1) car
- 2) cars
- 3) vehicle
- 4) vehicles
- 5) truck
- 6) trucks
- 7) tire
- 8) tires
- 9) lorry
- 10) lorries

Cn. 68

- 1) human
- 2) labour
- 3) daily
- 4) day
- 5) annual
- 6) total
- 7) week
- 8) month
- 9) year
- 10) minute
- 11) delivery
- 12) volume
- 13) range
- 14) unit

Cn. 69

- 1) fuel
- 2) cover
- 3) transfer
- 4) wave
- 5) waves
- 6) Poisson's
- 7) steel

Cn. 70

- 1) problem
- 2) problems
- 3) detail
- 4) details
- 5) race
- 6) races
- 7) assignment
- 8) contradiction
- 9) contradicting
- 10) time
- 11) logic
- 12) area
- 13) areas

Cn. 71

- 1) patterns
- 2) pattern
- 3) shade
- 4) shades

Cn. 72

- 1) cluster
- 2) clusters
- 3) set
- 4) sets
- 5) group
- 6) groups

Cn. 73

- 1) stable
- 2) unstable

- 3) transient
- 4) gain
- 5) drift
- 6) oscillatory
- 7) transitional
- 8) switch
- 9) store
- 10) storage

Сп. 74

- 1) thermal
- 2) heat

Сп. 75

- 1) net
- 2) overall
- 3) absolute

Сп. 76

- 1) numerically
- 2) terms

Сп. 77

- 1) engineering
- 2) little
- 3) minimum
- 4) free
- 5) area
- 6) areas
- 7) shop
- 8) shops
- 9) personnel
- 10) scheduled
- 11) preventive
- 12) routine
- 13) field
- 14) fields
- 15) queue
- 16) queues

Сп. 78

- 1) electrically
- 2) mechanically

- 3) hydrolically
- 4) pneumatically
- 5) manually
- 6) remotely
- 7) independently
- 8) efficiently

Сп. 79

- 1) computer
- 2) computers
- 3) evaluate
- 4) evaluates
- 5) evaluated
- 6) evaluating
- 7) calculate
- 8) calculates
- 9) calculated
- 10) calculating
- 11) determine
- 12) determines
- 13) determined
- 14) determining

Сп. 80

- 1) control
- 2) process

Сп. 81

- 1) for
- 2) to

Сп. 82

- 1) -ion
- 2) -ing
- 3) -ment

Сп. 83

- 1) -er
- 2) -or
- 3) -ers
- 4) -ors

Сп. 84

- 1) -th
- 2) -ance

- 3) -ence
- 4) -cy
- 5) -ty
- 6) -ness
- 7) -age

Сп. 85

- 1) overall
- 2) progressive
- 3) periodic
- 4) period
- 5) periods
- 6) interval
- 7) intervals
- 8) party
- 9) parties
- 10) work
- 11) tool
- 12) van
- 13) vans
- 14) tools

Сп. 86

- 1) -al
- 2) -ic

Сп. 87

- 1) full
- 2) beyond

Сп. 88

- 1) -ar
- 2) -less
- 3) -ous

Сп. 89

- 1) in
- 2) of
- 3) between

Сп. 90

- 1) to
- 2) into

Ю. Н. Марчук

ОПЫТ МАШИННОЙ РЕАЛИЗАЦИИ ДИСТРИБУТИВНОЙ МЕТОДИКИ ОПРЕДЕЛЕНИЯ ЛЕКСИЧЕСКИХ ЗНАЧЕНИЙ

§ 1. ЦЕЛЬ И ПОСТРОЕНИЕ РАБОТЫ

Теория основных направлений прикладного языкознания так или иначе включает в себя проблему лексической многозначности. В автоматизации перевода текста это перевод многозначных слов, в информационном поиске, автоматическом реферировании и индексировании — разрешение многозначности ключевых слов, в обучении иностранным языкам — нахождение для многозначных слов таких признаков в предложении, которые детерминируют определенные значения этих слов.

Поэтому исследование многозначности, предпринимаемое в той или другой работе, должно вестись не стихийно, а с позиций некоторой теории, пригодной для решения всей проблемы в целом. Таким образом, возникает вопрос о построении единой теории снятия лексической омонимии. Решение этого вопроса, по-видимому, целесообразно начать с уточнения некоторых понятий, связанных с постановкой проблемы.

Допустим, перед нами стоит задача снимать омонимию слов при машинном переводе с английского языка на русский. Посмотрим, что предполагает постановка этой задачи.

Например, у нас есть билингвы, составленные человеком, владеющим английским и русским языками. Приняв английские тексты за исходные, а русские — за их перевод, мы можем этот перевод рассматривать в качестве некоторой системы отсчета и в ее терминах записать значения слов в исходных текстах. Указанный принцип эталонирования даст нам возможность подойти к проблеме многозначности следующим образом. Однозначными мы будем называть такие английские слова, которым при переводе сопоставляется одна (и только одна) лексическая единица, зафиксированная в словаре эталонного языка, а многозначными — те, которым сопоставляется несколько лексических

единиц. Например, лексическим эквивалентом английского слова *consider* явится, для определенных контекстов, русское словосочетание *принять во внимание*, если последнее имеется в русском словаре как одна лексическая единица, т. е. всё словосочетание записано под одним номером. В других контекстах лексическими эквивалентами слова *consider* могут выступать слова *учесть, рассмотреть* и т. п.

Естественно предположить, что выбор эквивалента из списка возможных переводов осуществляется не случайно, а детерминруется употреблением переводимого слова в той или иной конкретной ситуации, т. е. особенностями того отрезка текста, в котором оно встречается. Некоторые особенности могут быть связаны с самим контрольным словом (например, если на перевод данного слова влияет то, в какой грамматической форме оно стоит), другие особенности являются принадлежностью контекста (например, если перевод данного слова обуславливается тем, какого соседа оно имеет в тексте).

Другими словами, каждому из значений, фиксируемых посредством отдельного слова эталонного языка, в английском языке должна соответствовать некоторая форма, диагностирующая его: опираясь на форму, диагностирующую в каждом конкретном случае то или иное значение многозначного слова в английском языке, переводчик и выбирает каждый раз строго определенное значение в русском языке. Таким образом, русские переводы помогают нам эксплицировать то различие в значимостях, которое фиксирует английский язык, пользуясь в каждой конкретной ситуации строго определенным набором языковых единиц.

Очевидно, что если бы нам удалось построить словарь, в котором для каждого многозначного слова фиксировался набор диагностических форм, разводящих его значения, и набор переводов, однозначно соотнесенных с этими формами, задачу снятия омонимии можно было бы поручить ЭВМ. Для этого надо было бы еще построить алгоритмы опознавания диагностических форм в тексте.

Главная трудность на этом пути — то, что а priori нельзя задать те исходные данные, на которых базируется словарь. Понять его общую структуру нетрудно, для этого специальное исследование проводить не приходится, но словарные статьи таким путем не построить — их можно составить только а posteriori.

При этом можно использовать разные методы. Одному из таких методов посвящена данная работа. Его суть состоит в следующем.

За исходное примем словарь, составленный на основании билингвы таким образом, что в нем представлена дистрибуция слов, частота их употребления и все имеющиеся в данном объеме текстов значения каждого из слов, образующие словарные статьи. Назовем такой словарь конкордансом. Построим алгоритмы разведения омонимов на основании конкорданса. Реализуем машин-

ную проверку этих алгоритмов на текстах, не вошедших в исходные билингвы. Исследуя полученные результаты, будем совершенствовать конкорданс в двух отношениях: 1) будем увеличивать количество переводов многозначных слов (если полисемия данного слова в конкордансе отражена недостаточно) и при этом будем фиксировать соответствующие им диагностические формы; 2) будем дополнять уже имеющиеся в словаре диагностические формы, если их недостаточно для опознавания тех значений, которые зафиксированы в словаре.

Очевидно, что в результате этой процедуры мы получим новый конкорданс. Этот конкорданс будет лучше, чем предыдущий. — с его помощью, несомненно, будет вскрываться больше таких значений полисемантических слов, которые релевантны для переводчика. Однако ряд значений, релевантных для информанта, останется «за бортом». Некоторые из них всплывут в ходе исследования результатов новой машинной реализации, и мы сможем включить их в следующий конкорданс. Но некоторые снова попадут в неформализованный остаток и т. д., пока не будет составлен идеальный конкорданс, дающий возможность решать задачу о распознавании всех омонимов, которые может распознать человек с помощью ЭВМ.

Конечно, путь аппроксимаций, о котором идет речь, — долгий и трудоемкий. Но более легкого пути, по-видимому, нет. Останемся на этом вопросе подробнее.

На первый взгляд кажется, что в качестве исходного конкорданса можно взять обычный переводной двуязычный словарь. Однако такой словарь не может служить основой для построения специального машинного словаря по следующим причинам:

а) переводной словарь дает не переводы, а ряды синонимов, поясняя значения; лишь в частных случаях значение может совпасть с переводом;

б) отдельные значения слов в переводном словаре даются вне сочетаемости; указания на дистрибуцию случайны, и по ним не могут быть построены дистрибутивные модели; восстановление дистрибуции каждого лексико-семантического варианта представляется невыполнимой задачей;

в) переводной словарь объединяет употребление слов в самых различных стилях и жанрах; между тем в современных прикладных задачах, как правило, ограничиваются одной областью языка (подъязыком);

г) в переводных словарях нет статистических данных, которые позволили бы выбрать главное значение, отделить его от второстепенного, определить наиболее частое, и пр.¹

¹ В построении формальных правил различения значений слов по контексту переводной словарь обычного типа выступает как ценное вспомогательное средство, своего рода информант. Значения слова в данном наборе контекстов можно и весьма целесообразно сравнивать со словарными значениями,

В качестве исходного материала для конкорданса нельзя принять и различные дистрибутивные формулы, зафиксированные теми или иными представителями структурной семантики.

Под дистрибутивной формулой лексико-семантического варианта слова понимается структурная модель словосочетания, в котором оно встречается, и обобщенная формула сочетаемости (см., например, [4]). Структурная модель считается заданной списком по завершении синтаксического анализа предложения. Тем самым разрешение лексической многозначности производится после и в зависимости от предварительного полного синтаксического анализа предложения. Известно, что при современном состоянии науки автоматический синтаксический анализ дает для естественных языков лишь приближенные решения (см., например, [16]). Кроме того, независимо от эффективности процедуры анализа в собственном смысле существует синтаксическая омонимия, когда одному предложению можно поставить в соответствие несколько правильных синтаксических структур. Синтаксическая омонимия разрешается лишь последующим учетом семантических факторов; так, разные синтаксические деревья одного предложения, подающиеся на вход семантической модели Катца—Фодора [14], подвергаются воздействию компонентов этой модели — словаря и «правил проекции», после чего, как правило, остается одно дерево. Поскольку в речи имеется множество эксплицитных признаков, характеризующих семантику (лексические единицы сами по себе и их сочетаемость), и лексический контекст играет гораздо большую роль в разрешении лексической многозначности, чем синтаксический [3], нет причин полагать, что разрыв замкнутого круга «синтаксис»—«семантика» должен обязательно происходить по с л е звена «синтаксис».

Таким образом, ни один из рассмотренных выше «более легких» путей не ведет к результатам, которые необходимы, чтобы построить автоматический словарь, нужный для решения задачи об автоматическом снятии полисемии многозначных слов. В перспективе остаются только конкордансы, составляемые на основании билингв,² уточняемые посредством обычных двуязычных словарей и информантов и проверяемые с помощью ЭВМ.

что позволяет сделать вывод о полноте представления слова в данном конкордансе. Нередко оказывается, что контекстные значения слова более разнообразны и многочисленны, чем значения, зафиксированные в основных словарях, причем это относится и не к узко специальным текстам с преобладанием терминов.

² Исследование контекста как метод разрешения лексической многозначности было предложено в ранних работах по машинному переводу А. Каплана [13], А. Кутсудаса [15], А. Лукьяновой [17], И. К. Бельской [5] и других. А. Каплан получил экспериментальные результаты, показывающие, какой тип (объем) контекста наиболее эффективен. Однако для ранних экспериментальных исследований в области лексической многозначности характерно отсутствие проверки правил разрешения многозначности на сколько-нибудь широком материале.

Соответственно возрастает значение методики составления автоматических контекстологических словарей и, в частности, аппроксимационной методики, предложенной выше. Остановимся на этом вопросе — опишем полностью один из витков спирали, по которой идет аппроксимация. Исследование этой проблемы составляет задачу данной статьи. Ее план можно охарактеризовать нижеследующим образом.

Сначала решаются вопросы, связанные с построением исходного конкорданса и его структурой. Затем следует описание работы со статьями этого словаря. Выделяются составляющие диагностических форм (в работе они названы детерминантами) и описываются правила опознавания детерминант в тексте. Затем показывается, как строятся алгоритмы распознавания значений многозначных слов с помощью метода детерминант, выясняется структура этих алгоритмов, возможности их типизации, систематизации, а также приводятся примеры распознавания значений многозначных слов. После этого обсуждаются результаты машинных экспериментов и исследования алгоритмов с целью усовершенствования исходного конкорданса. Материалом для типизации алгоритмов является выборка, обусловленная нижеследующими соображениями.

Алгоритмы перевода составлялись на материале газетных информационных и публицистических текстов на английском языке общим объемом около одного миллиона словоупотреблений.³ Многозначные слова распределены по лексико-грамматическим классам. Этих слов в исходном словаре несколько тысяч: если принять общее количество многозначных слов за 100%, то по отдельным лексико-грамматическим классам многозначные слова распределяются так, как показано в табл. I (приложение). Больше всего в словаре многозначных существительных, затем глаголов и прилагательных. Многозначные слова составляют около 30% всех слов входного словаря. Внутри отдельных лексико-грамматических классов число многозначных слов различно. Коэффициент многозначности — отношение числа многозначных слов данного лексико-грамматического класса к общему числу слов этого класса — имеет следующие значения:

существительные	— 0.21;
прилагательные	— 0.45;
глаголы	— 0.51;
наречия	— 0.21;
предлоги и союзы	— 1.0

Полная многозначность предлогов и союзов объясняется высокой степенью их грамматикализации. Существительных больше,

³ Первоначальный конкорданс был составлен по выборке в 630 833 словоупотребления, затем был использован дополнительный конкорданс, составленный по выборке объемом примерно 200 тыс. словоупотреблений.

чем прилагательных, однако многозначных среди них меньше, чем среди прилагательных. Наиболее многозначны глаголы; это обстоятельство, а также важная роль глагола в синтаксическом строении предложения определяют выбор этой части речи в качестве предмета нашего рассмотрения.

В частотном словаре исходных текстов глаголы занимают следующее место. Встретившихся один раз глагольных словоформ — 411. Среди первых 1500 наиболее частых словоформ имеется 455 глагольных. Получается, что:

20%	всех глаголов имеют абсолютную частоту выше 57
60%	» » » » » в интервале 57÷1
20%	» » » » » 1.

Отсюда видно, что доля глаголов, имеющих абсолютную частоту 1, невелика по сравнению с числом глаголов, обладающих большей частотой. Уже одно это обстоятельство дает основание сделать вывод о постоянстве глагольного состава в исходном словаре,⁴ что представляется необходимым условием для значимости качественных выводов. Сравнение глагольного состава исходного словаря с глаголами других известных частотных словарей английского языка при помощи специальных статистических методов представляется затруднительным из-за разной величины исходной выборки (по сравнению с существующими словарями), различных исходных материалов и разных критериев организации выборки. Однако ориентировочное сопоставление частотного словаря исходных текстов с глаголами частотного словаря электроники [1], а также количественное и качественное сравнение всего исходного глагольного списка с английскими глаголами, приведенными в англо-русском словаре В. К. Мюллера, позволяют сделать общий и достаточный для поставленной задачи вывод о том, что глаголы в исходном списке достаточно представительны.

Именно этим и объясняется ориентация выборки на глагольные алгоритмы. Дальнейшее сужение поля исследования алгоритмов было вызвано нижеследующими соображениями.

В частотном словаре всех слов исходных текстов верхняя часть — ранги от 13 до 100 — принадлежит модальным и вспомогательным глаголам. Модальные глаголы в конкордансе однозначны. Особое место среди вспомогательных глаголов занимают *to be* и *to have*. Они могут считаться многозначными как в тех случаях, когда они выступают как знаменательные, (*to be* 'быть', 'являться', 'находиться'), так и тогда, когда они играют роль связки в составном сказуемом. Если во втором случае многозначность приписывать глаголу и разрешать ее в алгоритме перевода этого глагола, то алгоритм получается весьма громоздким.

⁴ Данные об устойчивости различных лексико-грамматических классов в отношении изменения словарного состава при составлении частотного словаря см. в работе П. М. Алексеева [2].

Альтернативным решением является перевод сочетаний этих глаголов со словами других классов в алгоритмах перевода других слов (прилагательных, существительных и т. п.). В нынешнем варианте алгоритма реализовано первое решение. Вследствие громоздкости и сложности алгоритмов перевода эти глаголы в выборку включены не были. Построение оптимальных алгоритмов перевода этих глаголов является самостоятельной задачей. Далее в частотном словаре следуют наиболее многозначные знаменательные глаголы, занимающие ранги от 100 до 200 и имеющие абсолютную частоту в пределах 1000—500. Это глаголы *to make*, *to do*, *to take* и др. Они немногочисленны, высоко многозначны и обладают большой индивидуальностью в отношении лексической сочетаемости и синтаксических характеристик. Как и глаголы первой группы, каждый из этих глаголов заслуживает самостоятельного изучения в структурном и алгоритмическом планах, поэтому они также не были включены в выборку, целью которой является репрезентировать основную массу глаголов. Последние ранги в частотном словаре — от 1500 до 1200 для глаголов — принадлежат глаголам с абсолютной частотой 1. Все они, за редкими исключениями, были признаны однозначными и в выборку не вошли. Основная масса многозначных глаголов занимает ранги от 200 до 1500 и имеет абсолютную частоту в пределах 500—2. Число этих глаголов — примерно 1200. Из них было отобрано 194, алгоритмы перевода которых и образовали исходную выборку. Отбор производился с соблюдением по возможности следующих требований: а) отбираемые глаголы должны более или менее равномерно распределяться в списке всех глаголов; б) объем алгоритма перевода должен приближаться к среднестатистическому объему алгоритма перевода многозначного слова — 10 стандартных операторов; в) должны быть представлены по возможности все разнообразные структуры глагольных алгоритмов. Не всегда удавалось удовлетворить в равной мере эти достаточно противоречивые требования; поэтому выборка, возможно, не лишена недостатков. Тем не менее, можно считать, что она представляет собой основной массив многозначных глаголов. В количественном отношении она составляет примерно одну шестую часть всех многозначных глаголов.

§ 2. ИСХОДНЫЙ КОНКОРДАНС

Основным исходным материалом при составлении автоматического словаря, как отмечалось выше, является конкорданс, т. е. алфавитно упорядоченный список контекстов фиксированного объема, ориентированный по основному слову и содержащий указания на текст или часть текста, откуда взят данный контекст. В таком конкордансе представлена дистрибуция слова, частота употребления и все имеющиеся в данном объеме текстов значе-

ния этого слова. Естественно, что полнота представления значений зависит от объема текстов, по которым составлен конкорданс. В настоящей работе для составления первоначальных вариантов алгоритмов перевода многозначных слов использовались конкордансы, сделанные по выборке объемом 630 833 словоупотребления английских газетных текстов публицистического характера (сообщения и комментарии на политические, экономические и военные темы).

Перевод с исследуемого языка на эталонный в данной работе производит переводчик, знающий английский и русский языки (последний как родной) и предмет, о котором идет речь в текстах. Переводчик получает конкорданс, составленный из разных употреблений исследуемого слова в исходном объеме текстов. Пользуясь своим знанием языка и предмета и в случае необходимости консультируясь с другими переводчиками, а также обращаясь к толковым и переводным словарям и к справочным пособиям (географическому атласу, словарям сокращений и пр.), переводчик дает перевод слова в каждом его употреблении, зафиксированном в конкордансе, и, если необходимо, указывает на значения слова, почему-либо не отраженные в конкордансе.

После того как произведен перевод, составитель алгоритма выполняет следующие операции:

а) отделяет однозначные в указанном выше смысле слова от многозначных и в дальнейшем рассматривает только последние;

б) перечисляет в выходном языке все слова, которые в заданном конкордансе могут служить при переводе лексическими эквивалентами исследуемого;⁵

в) устанавливает, какие факторы влияют на выбор лексического эквивалента в каждом конкретном случае;

г) определяет, как эти факторы взаимосвязаны между собой;

д) указывает типы дистрибуции, по которым устанавливается однозначное соответствие между исследуемым и переводящим словом;

е) представляет процедуру идентификации типа дистрибуции и выбора лексического эквивалента для перевода в виде алгоритма перевода многозначного слова.

Разрешение лексической многозначности происходит в дальнейшем как работа на тексте составленного таким образом алгоритма. Множество отдельных алгоритмов перевода многозначных слов подвергается затем специальному изучению с целью выявления закономерностей, общих для всех алгоритмов или некоторого их подмножества.

⁵ Разрешается в порядке исключения добавлять только весьма распространенные и очевидные переводы, отсутствие которых в конкордансе можно рассматривать как реализацию крайне редкого события. При этом требуется, чтобы отсутствующий перевод мог быть или легко детерминирован, или годился бы для «общего выхода» алгоритма.

Алгоритмическим путем, т. е. с помощью специально для этой цели построенных алгоритмов, существо которых в настоящей работе не рассматривается, из текста до разрешения лексической многозначности получается нижеследующая информация.

1) Морфологические признаки слова. Сами эти признаки (аффиксы и окончания) и информация о них являются результатом полного морфологического анализа. Вследствие того, что алгоритм перевода бинарный, т. е. рассчитан на перевод только с английского языка на русский и не предусматривает особых межъязыковых преобразований или языка-посредника, словоизменяемые категории определяются в грамматических категориях выходного языка.

2) Информация о принадлежности слова к лексико-грамматическому классу. Однозначность этой информации обеспечивается решением проблем омографии, т. е. принадлежности одного слова к разным лексико-грамматическим классам.

3) Сведения о синтаксических функциях слов в предложении. Не устанавливая полной картины синтаксических зависимостей, можно для каждого слова найти его собственную синтаксическую функцию (СФ). Роль СФ в последующем разрешении лексической многозначности сводится к тому, что СФ: а) помогает определить значение слова в ситуации, когда определяющим является так называемый синтаксический контекст; б) позволяет организовать поиск детерминант и установить границы определяющего контекста.

В описываемом алгоритме приняты следующие лексико-грамматические классы для слов входного языка: существительное, прилагательное, глагол, предлог, наречие, союз, знак препинания. Местоимения и числительные разделены между существительными и прилагательными. Знаки препинания, как это принято в автоматической обработке текстов, считаются самостоятельными словами. Внутри класса существительных по семантическому и функциональному признакам выделены следующие подклассы: местоимения, числительные, одушевленные, географические названия, имена собственные, названия документов и т. п. Внутри класса прилагательных выделены подклассы: местоименные прилагательные, числительные, качественные, географические названия и пр. Принадлежность одного слова к разным лексико-грамматическим классам называется омографией. Слова-омографы фиксируются в словаре как таковые, разбиваются на классы омографии и, по мере появления их в тексте, проходят обработку по специальной программе для каждого класса омографии. В результате работы этой программы каждый омограф в тексте получает признак того лексико-грамматического класса, к которому он принадлежит. Каждая программа разбора омографии представляет собой реализацию определенного лингвистического алгоритма, исследующего морфологический состав слова-омографа и синтаксическое строе-

ние контекста, но не в терминах СФ, которые еще не определены, а в терминах классов окружающих омограф слов.

Синтаксические функции слов устанавливаются по английскому тексту и для английских слов. Каждому слову в предложении, включая и знаки препинания, ставится в соответствие одна СФ из следующего списка: подлежащее, сказуемое, определение, дополнение, обстоятельство, предложная, разделительная. Принципы определения СФ и распределения СФ между лексико-грамматическими классами описаны подробно в работах [8] и [9]. Отметим, что СФ еще не позволяют создать полную картину синтаксических связей в предложении, в частности потому, что они определяются без привлечения семантических критериев. Кроме того, семь синтаксических функций — это, видимо, слишком простая система для такой сложной задачи. Описание синтаксиса предложения в терминах синтаксических функций слов можно считать лишь первым, приближенным описанием, пригодность которого определяется следующей задачей общего алгоритма — разрешением лексической многозначности.

Неразложимые фразеологические сочетания выделяются специальным словарем оборотов.

§ 3. ХАРАКТЕРИСТИКА ДЕТЕРМИНАНТ, ИХ ПОИСК В ТЕКСТЕ И СИСТЕМАТИЗАЦИЯ

Детерминирование значений слов контекстными признаками является сложным лингвистическим фактом и допускает разные степени детализованности в рассмотрении даже в пределах одной задачи.

Лексические значения многозначных слов детерминируются информацией о следующих явлениях: а) сочетаемости многозначного слова в исследуемом языке со словами определенных лексико-грамматических классов (подклассов); б) сочетаемости слова со списками некоторых слов или отдельными словами; в) участии слова в синтаксических конструкциях или СФ самого слова; г) наличии у слова морфологических признаков или морфем, или комбинации из перечисленных явлений, каждая из которых однозначно определяет один перевод.

Назовем *детерминант* — лексико-грамматический класс или подкласс, список слов или отдельное слово, синтаксический, морфологический признак или морфему, которые по отдельности или в некоторой совокупности однозначно определяют одно значение многозначного слова исследуемого языка как требующее отдельного лексического эквивалента в эталонном языке. *Определяющим контекстом* будем называть контекст, в котором отмечена хотя бы одна детерминанта.

Выделим следующие виды детерминант.

Лексико-грамматическая детерминанта (ЛГ-детерминанта) представляет собой лексико-грамматический класс или

подкласс слов, характеризующихся общим семантическим содержанием и общими особенностями дистрибуции. Кроме традиционных классов (части речи) и подклассов слов, выделяются специальные: например, в классе существительных, включающем также местоимения-существительные, имеются подклассы: «одушевленные существительные», «географические названия», «деньги», «документы», «имена собственные» и т. п. Слово, включенное в один из подклассов, уже не входит в другой, поэтому при необходимости применяются объединения подклассов. Представляя непересекающиеся классы слов (лексико-грамматическая ономия разрешена), ЛГ-детерминанта является лингвистически наиболее интересным видом детерминирования.

Лексическая детерминанта (Л-детерминанта) — отдельное слово или список слов. Л-детерминанта характеризует индивидуальную сочетаемость слова. Поскольку лексический контекст играет весьма важную роль в определении лексических значений слов, Л-детерминанта чрезвычайно распространена.

Для ЛГ-детерминанты и Л-детерминанты определен признак «направление». Детерминанты, находящиеся справа от исследуемого слова, называются правыми, слева — левыми.

Синтаксическая детерминанта (С-детерминанта) представляет собой синтаксическую функцию слова в предложении. В качестве С-детерминанты может выступать синтаксическая функция либо самого исследуемого слова, либо ЛГ- или Л-детерминанты, поэтому для С-детерминанты, как и для двух последующих, определен признак «принадлежность».

Морфемная детерминанта (М-детерминанта) — морфема, помогающая различить лексическое значение. Чаще всего в виде М-детерминанты выступают суффиксы и окончания.

Морфологическая детерминанта (МЛ-детерминанта) представляет собой грамматическую категорию в эталонном языке, релевантную выбору лексического эквивалента в этом языке. Например, если глаголы, образующие видовую пару в русском языке, взятом как эталонный, записаны в выходном словаре этого языка отдельно, то грамматическая категория «вид», определенная до перевода, позволяет сделать правильный выбор лексического эквивалента.

Пять перечисленных выше детерминант называются *детерминантами первого порядка*, или простыми детерминантами. Приведем примеры детерминант первого порядка.

ЛГ-детерминанта. Английский глагол *to forestall* при наличии справа от него одушевленного существительного переводится 'помешать'.⁶ В других контекстах переводы другие.

⁶ Именно такие переводы, а не другие и т. п., могут подвергаться сомнению. В связи со сказанным выше здесь исследуются переводы, сделанные переводчиками для словоупотреблений заданного конкорданса.

Одушевленное существительное является правой ЛГ-детерминантой глагола to forestall.

Л-детерминанта. Глагол to inflate в сочетании с последующим существительным price переводится 'вздувать', во всех других контекстах — 'произвести инфляцию'. Существительное price является правой лексической детерминантой глагола to inflate.

С-детерминанта. Прилагательное ill в синтаксической функции «сказуемое» переводится 'больной', в других контекстах переводы другие. Синтаксическая функция «сказуемое» служит С-детерминантой прилагательного ill.

М-детерминанта. Существительное consequence с окончанием -s переводится 'последствия', в остальных случаях переводы другие.

Окончание мн. ч. -s является М-детерминантой для существительного consequence в этом переводе.

МЛ-детерминанта. Глагол to guide с признаком «пассив» переводится 'руководствоваться', без такого признака — 'руководить'. Признак «пассив» является МЛ-детерминантой глагола to guide в указанном переводе.⁷

ЛГ- и Л-детерминанты будем называть «лексически-выраженными», а три остальные — «грамматически-выраженными». Свойство эквивалентности заключается в том, что две разные детерминанты дают один и тот же результат. Например, М-детерминанта s для существительного consequence эквивалентна МЛ-детерминанте «число» для этого же перевода. Свойство перемещения характерно для лексически-выраженных детерминант. По этому свойству, например, правая лексическая детерминанта активных конструкций становится для пассивных конструкций левой.

Обнаружение детерминанты первого порядка соответствует определенной стандартной лингвистической операции, которая в алгоритме выполняется стандартным оператором, т. е. набором команд, являющимся стандартной подпрограммой.

Детерминанты второго порядка представляют конъюнкции детерминант первого порядка. Установлено, что число детерминант в таких конъюнкциях редко превышает четыре. Максимальное число детерминант требуется для нахождения всех компонентов идиоматических словосочетаний, не вошедших в словарь оборотов. Наиболее часты детерминанты второго порядка, составленные из двух простых детерминант, реже встречаются детерминанты из трех. Среди детерминант второго порядка будем различать составные детерминанты, являющиеся объ-

единениями только лексически-выраженных детерминант, и сложные, компонентами которых могут быть любые простые детерминанты. Чаще всего сложные детерминанты состоят из двух компонентов. У сложных детерминант первого типа грамматически-выраженные детерминанты, входящие в объединение, приписаны исследуемому слову. У сложных детерминант второго типа эти детерминанты относятся к лексически-выраженным детерминантам. Приведем примеры указанных детерминант второго порядка.

Составная детерминанта: глагол to lose в значении 'упустить из виду' имеет составную детерминанту sight + of (Л-детерминанта + Л-детерминанта).

Сложная детерминанта первого типа: нулевой перевод глагола to interpret в сочетании interpreting personnel (слово personnel получает перевод 'переводчики', детерминируется С-детерминантой «определение» к слову personnel и Л-детерминантой personnel).

Сложная детерминанта второго типа: прилагательное above, в переводе 'вышеупомянутый', детерминируется последующим глаголом to mention и синтаксической функцией «определение» этого глагола (Л-детерминанта + С-детерминанта).

Существуют и другие виды сложного детерминирования. Многие из них допускают замены. Первыми во избежание ложного детерминирования в алгоритме должны следовать детерминанты второго порядка (принцип longest match) Это единственное ограничение, налагаемое на порядок расположения детерминант в алгоритме.

Правила поиска детерминант указывают, какие слова нужно пропустить, когда отыскиваются лексически-выраженные детерминанты. Правила поиска по существу определяют границы контекста, в котором производится поиск детерминант. Поскольку разрешающая сила лексического контекста зависит от его объема, правила поиска имеют существенное значение для эффективности алгоритма. Необходимость в правилах поиска возникает ввиду отсутствия структурных моделей и формул дистрибуции в классическом виде. Исходная информация позволяет сформулировать следующие основные правила поиска.

1) Не пропускать никаких слов. Лексически выраженные детерминанты образуют непосредственное окружение исследуемого слова.

2) Пропускать слова, принадлежащие к определенным лексикограмматическим классам. Пример: при поиске вправо ЛГ-детерминанты «существительное» пропускать слова классов «прилагательное» и «наречие».

3) Пропускать слова, имеющие определенную синтаксическую функцию. Пример: при поиске вправо ЛГ-детерминанты «существительное» пропускать слова, имеющие СФ «определение».

⁷ Алгоритмы анализа, работающие до перевода (см. стр. 189) по английскому тексту, дают информацию, релевантную для выбора переводящего эквивалента. Например, «пассив» может определяться по наличию перед глаголом to be плюс окончание -ed и т. г., «вид» — по формам Continuous и Perfect.

4) Пропускать слова в заданных границах. Пример: при поиске правой лексической детерминанты пропускать все слова в предложении до точки. Определяющим контекстом в таком поиске обычно являются:

- а) простое предложение;
- б) фраза (сложное или составное предложение);
- в) весь текст в оперативной памяти машины.

В каждой из основных перечисленных групп, кроме группы 1, назначаются отдельные частные правила пропуска (поиска). Частные правила поиска внутри разных основных групп могут быть эквивалентны. Так, правило «пропускать при поиске ЛГ-детерминанты прилагательные и наречия» эквивалентно правилу «пропускать слова, имеющие СФ „определение“ и „обстоятельство“».

Введем понятие «элемент контекстуальной модели перевода». Обозначим так некоторый типовой набор информации о слове-детерминанте и связи ее с переводимым основным словом. В этот набор входят следующие данные:

- а) принадлежность детерминанты к лексико-грамматическому классу;
- б) правила поиска детерминанты;
- в) участие в трансформациях;
- г) состав лексической единицы-эквивалента в эталонном языке;
- д) указание на то, переводится ли детерминанта вместе с исследуемым словом.

Пункты «а» и «б» позволяют выбрать правило поиска. Пункт «в» определяет направление лексически-выраженной детерминанты (например, в пассивных конструкциях правая лексическая детерминанта становится левой). Пункт «г» нужен для синтеза переводящей словоформы, пункт «д» для правильного выбора оператора перевода. Грамматически-выраженные детерминанты в элемент контекстуальной модели не включаются. Состав информации, образующей элемент контекстуальной модели перевода (КМП), различен для слов разных лексико-грамматических классов, по

Таблица 1*

Элемент контекстуальной модели перевода

Лексико-грамматический класс детерминанты. Обозначение элемента КМП	Промежуточные слова	Участие в трансформациях	Чем переводится исследуемое слово	Переводится ли детерминанта	Пример
Сущ. ЭС 1	прил. или прич.	пассив	одним глаголом	нет	heighten tension 'усиливать напряженность'

* Английские глаголы приводятся в табл. без to.

этому для каждого из них имеются собственные элементы КМП. В алгоритме перевода многозначного слова каждый элемент КМП описывает одну или несколько детерминант. Ниже (табл. 1) приведен пример элемента контекстуальной модели перевода.

Контекстуальная модель перевода (КМП) представляет собой сумму элементов контекстуальной модели перевода. КМП описывает все детерминанты для данного многозначного слова. КМП глаголов изучались на материале первых 70 глаголов исходной выборки. В обследованных алгоритмах можно выделить элементы КМП трех типов.

Элементы КМП первого типа имеют в качестве детерминант существительные. При поиске пропускаются слова, которым приписан признак СФ «определение». Чаще всего это прилагательные и причастия. Для всех элементов этого типа определена трансформация пассива. Перевод производится либо одним русским глаголом, либо сочетанием глагола с предлогом, либо фразеологическим сочетанием. Детерминанта в большинстве случаев не переводится. Элементы КМП первого типа приводятся в табл. 2.

Таблица 2

Элементы КМП первого типа

Лексико-грамматический класс детерминанты. Обозначение элемента КМП	Промежуточные слова	Участие в трансформациях	Чем переводится	Переводится ли детерминанта	Примеры
Сущ. ЭС 1	прилаг. или прич.	пассив	одним русским глаголом	нет	guide story 'инспирировать'
Сущ. ЭС 2	То же	»	сочетанием слов	нет	injure interest 'нанести ущерб'
Сущ. ЭС 3	» »	»	одним глаголом с предлогом	нет	grant point 'согласиться с...'
Сущ. ЭС 4	» »	»	одним глаголом или идиомой	да	gain ground 'укрепиться'

Элементы КМП второго типа характеризуются наречием или предлогом в качестве лексической детерминанты (послелог). Правила поиска разнообразны. Детерминанта не участвует в трансформациях. Перевод производится либо одним глаголом, либо сочетанием глагола с предлогом. Детерминанта переводится вместе с основным глаголом. Элементы КМП второго типа представлены в табл. 3.

* Артикли, порядковые числительные и притяжательные местоимения входят в класс прилагательных, личные местоимения — в класс существительных.

Таблица 3

Элементы КМП второго типа

Обозначение элемента КМП	Промежуточные слова	Участие в трансформациях	Чем переводится	Переводится ли детерминанта	Примеры
ЭП 1	нет	нет	одним глаголом	да	hand over 'передать'
ЭП 2	сущ. и прил. или прич.	»	одним глаголом с предлогом	»	identify with 'отождествить с...'

Элементы КМП третьего типа, в отличие от первых двух, имеют составную детерминанту из существительного и послелога. Правила поиска варьируются. Некоторые виды элементов предусматривают трансформацию пассива. Детерминанты переводятся вместе со словом. Как видно из табл. 4, в большинстве случаев эти детерминанты описывают устойчивые фразеологические словосочетания.

Таблица 4

Элементы КМП третьего типа

Обозначение элемента КМП	Промежуточные слова	Участие в трансформациях	Чем переводится	Переводится ли детерминанта	Примеры
ЭПС 1 послелог + сущ.	нет	нет	идиомой	да	jump to conclusions 'делать поспешные выводы'
ЭПС 2 послелог + сущ.	»	»	одним глаголом	»	kick over traces 'взбунтоваться'
ЭСП 1 сущ. + послелог	прил.	пассив	словосочетанием	»	lose sight 'упустить из виду'

Распределение детерминант внутри КМП показывается величиной, которую будем называть индексом КМП. Индекс КМП представляет собой набор чисел, составленный следующим образом. Все детерминанты в таблице детерминант для данного многозначного слова перенумерованы. В строке таблицы каждой детерминанте соответствует номер перевода. Первое число в индексе КМП показывает, какому номеру строки таблицы детерми-

нант-переводов соответствует последний элемент КМП. Следующее число, записываемое через дефис, является номером последней строки таблицы, обрабатываемой данным элементом КМП. Нумерация строк в индексе КМП производится для удобства с конца, так как последние в КМП элементы чаще всего относятся к последним детерминантам в алгоритме перевода. Если данный элемент КМП приходится на общий выход алгоритма, то его номер в индексе КМП — 0. Следующие числа в индексе таким же порядком нумеруют следующие в модели справа налево элементы КМП и распределяют их по строкам таблицы детерминант-переводов.

Контекстуальные модели перевода глаголов обследованной выборки состоят из небольшого числа элементов. В табл. 5 сведены данные о количестве моделей по элементам и числу глаголов на каждую модель.

Таблица 5

Распределение глаголов по числу элементов в модели

Модель	Число глаголов
Одноэлементная	36
Двухэлементная	28
Трехэлементная	5

Лишь один глагол выборки, а именно to keep, имеет модель с большим числом элементов (5). В табл. 6, 7, 8 приведем данные о распределении глаголов выборки по КМП и виды КМП.

Из приведенных табл. видно, что наибольшее количество глаголов приходится на модели с небольшим числом элементов, а именно с одним или двумя элементами. Это дает основание сделать вывод о том, что в большинстве случаев процедура нахождения значения слова по контексту допускает представление в виде небольшого набора простых операций.

Таблица 6

Одноэлементные модели перевода

Модель	Составные элементы	Глаголы
КМП 1	ЭС 1	arrest, assess, etc.,
КМП 2	ЭС 2	injure
КМП 3	ЭС 3	hit
КМП 4	ЭП 1	guard, lack, link
КМП 5	ЭП 2	lull

Двухэлементные модели перевода

Таблица 7

Модель	Составные элементы	Глаголы
КМП 6	ЭС 1+ЭС 2	finalize, impair
КМП 7	ЭС 1+ЭС 3	initiate, grant, etc.,
КМП 8	ЭС 1+ЭС 4	gain
КМП 9	ЭС 1+ЭП 1	consume, furnish, etc.,
КМП 10	ЭС 1+ЭП 2	implicate, judge, etc.,
КМП 11	ЭС 1+ЭПС 1	jump
КМП 12	ЭП 1+ЭС 2	balance, back, etc.,
КМП 13	ЭП 1+ЭПС 2	kick

Таблица 8

Трехэлементные модели перевода

Модель	Составные элементы	Глаголы
КМП 14	ЭС 1+ЭС 2+ЭСП 1	lose
КМП 15	ЭС 1+ЭС 3+ЭС 4	handle
КМП 16	ЭС 1+ЭС 4+ЭП 1	lay
КМП 17	ЭС 1+ЭП 1+ЭП 2	guide
КМП 18	ЭС 1+ЭП 2+ЭС 2	invite

§ 4. СТРУКТУРА И СИСТЕМАТИЗАЦИЯ АЛГОРИТМОВ ПЕРЕВОДА МНОГОЗНАЧНЫХ ГЛАГОЛОВ

Для того чтобы задать конкретный алгоритм перевода слова, нужно лишь получить таблицу детерминант-переводов. Такая организация многозначного словаря облегчает пополнение и изменение отдельных алгоритмов перевода слов: точнее, алгоритм не меняется как таковой, меняются лишь переводы и детерминанты в таблице.

Правила поиска детерминант относятся к числу средств, формирующих лингвистическую логику алгоритма.

Алгоритм перевода каждого многозначного слова индивидуален в том смысле, что он состоит из отдельного набора детерминант и переводов для этого слова. Однако введенные понятия позволяют сделать некоторые обобщения. Введем понятие «структурный тип» (СТ). Назовем так конструкцию, изображенную ниже:



Здесь $d_{(1,2,\dots,n)}$ — обобщенная правая лексическая детерминанта, $x_{(1,2,\dots,n)}$ — обобщенный перевод, x_m — общий выход

алгоритма. «Обобщенный» означает, что не указывается конкретно, сколько детерминант: структурный тип показывает только, что в алгоритме имеется n детерминант, что все они лексические, правого направления, и каждой из них соответствует один перевод из n переводов, а также есть один общий выход. Так, d_1 соответствует перевод x_1 , d_2 — x_2 и т. д. Стрелка вправо означает отход по «да» при наличии детерминанты, стрелка вниз — отход по «нет» при ее отсутствии. Структурный тип отражает, таким образом, лишь качественный состав алгоритма.

Назовем СТ одноэлементным, если в графическом изображении у него есть только одна вертикальная стрелка, двухэлементным, если две, трехэлементным, если три, и т. д. Составность структурного типа означает, таким образом, сколько разных видов детерминант в нем имеется.

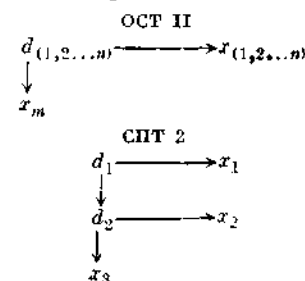
Рассмотрение алгоритмов перевода многозначных слов показало следующее.

1) Множество алгоритмов перевода многозначных слов каждого лексико-грамматического класса содержит алгоритмы, структура которых характеризуется одним СТ. Все множество алгоритмов перевода слов одного класса описывается конечным набором структурных типов. Структурные типы алгоритмов для слов одного лексико-грамматического класса назовем частными структурными типами (ЧСТ).

2) Имеются СТ, общие для алгоритмов перевода слов, принадлежащих к разным лексико-грамматическим классам. Такие СТ называются общими структурными типами (ОСТ).

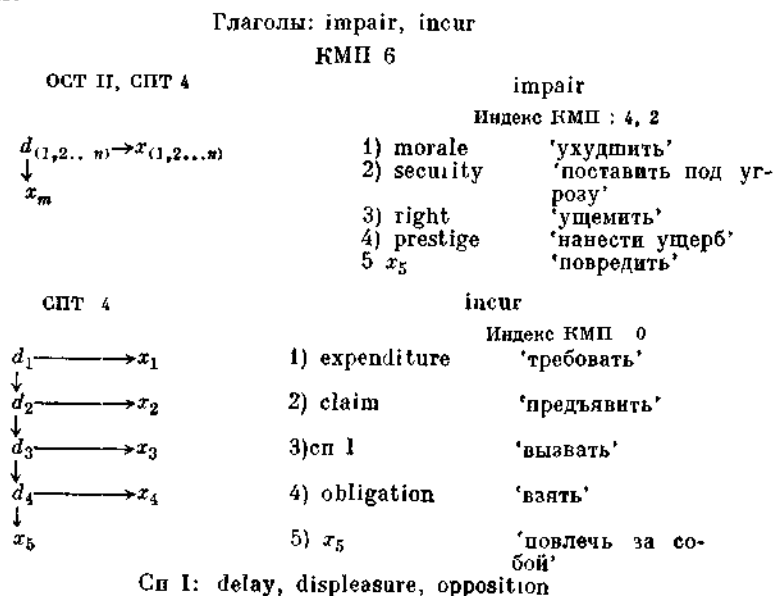
Введем понятие «структурный подтип» (СПТ) структурного типа. В СПТ обозначена каждая детерминанта и каждый перевод. Число детерминант и переводов соответствует их числу в индивидуальном алгоритме. Указанием на СПТ описывается с точностью до одной детерминанты и одного перевода структура алгоритма перевода многозначного слова.

Ниже изображен общий структурный тип (ОСТ II) и структурный подтип (СПТ 2) внутри этого типа. Структуру подтипа 2 имеют алгоритмы перевода многозначных английских глаголов to balance, to consume и др.:



Множество алгоритмов перевода, имеющих один структурный тип, распадается на подмножества, имеющие общий структурный подтип, так что число структурных подтипов значительно меньше числа индивидуальных алгоритмов перевода слова данного лексико-грамматического класса.

Ниже показан алгоритм перевода для некоторых многозначных глаголов.



Чтобы проверить гипотезу о наличии структурных типов среди алгоритмов перевода и об общности этих типов для разных лексикограмматических классов, была сделана выборка по 100 алгоритмов перевода из каждого лексико-грамматического класса. Предлоги и союзы в эту выборку не вошли, поскольку пока не ясно соотношение грамматических и лексических значений этих слов при разрешении лексической многозначности принятым способом. В полученной выборке были выделены типы структур алгоритмов. Были обнаружены три структурных типа, общие для всех лексико-грамматических классов (см. рис. I, в приложении). Кроме общих структурных типов, для каждого лексико-грамматического класса оказалось в наличии несколько частных структурных типов. Некоторое количество алгоритмов не распределилось ни по общим, ни по частным структурным типам. Результаты распределения выборки по типам сведены в табл. II. Из нее следует, что выборка почти полностью распределилась по общим и частным структурным типам. Лучше всего типизировались многозначные глаголы. Хотя прилагательные тоже полностью

распределились, следует учесть, что у них в три раза больше частных структурных типов, чем у глаголов.

Проведенное обследование позволяет считать подтвердившейся гипотезу о том, что алгоритмы перевода имеют общие структурные типы, а число подтипов не является бесконечным, и, во всяком случае, значительно меньше, чем число индивидуальных алгоритмов перевода. Явным образом преобладают ОСТ, на долю которых приходится в среднем 60—80% всех алгоритмов. ОСТ составлены только из лексических детерминант. Два типа односоставны, один двухсоставен. Это говорит о том, что для основных лексико-грамматических классов и основной массы многозначных слов лексическая многозначность разрешается простыми схемами контекстного анализа. Полученные результаты позволяют развить дальше выводы, к которым пришел А. Каплан, исследуя разрешающую силу контекста в рамках своего ограниченного эксперимента. По нашим наблюдениям оказывается, что разрешающая сила контекста, точнее — величина или объем контекста, зависят от лексико-грамматического класса исследуемого слова. Так, из данных табл. II следует, что для глаголов и прилагательных наибольшей разрешающей способностью обладает контекст с правой лексической детерминантой, в то время как для существительных и наречий наибольшей разрешающей силой обладает контекст из одного предшествующего и одного последующего слова.

Отсутствие в ОСТ лексико-грамматических детерминант вовсе не означает их малой роли в различении многозначности. Дело в том, что эти детерминанты чаще всего, по крайней мере на исследованном материале, выступают в совокупности с другими, и поэтому не достигается общность соответствующих СТ для всех лексико-грамматических классов.

В некоторых отдельных случаях приведение алгоритмов к СТ проводилось с использованием свойств эквивалентности детерминант. Однако общее число предпринятых преобразований невелико. Так, для глаголов лишь 5% алгоритмов перевода претерпели отдельные эквивалентные преобразования.

В структуре алгоритмов перевода глаголов, вошедших в выборку, общие и частные СТ оказались представленными следующим образом:

Общие структурные типы глаголов выборки изображены на рис. I в Приложении. На рис. II даны частные структурные типы для глаголов.

Остались нераспределенными по структуре всего 8 глаголов (to throw, to turn, to suggest и др.) с ярко выраженной лексической индивидуальностью, из-за которой предполагаемый подтип всегда окажется единственным. Вообще говоря, допустимым является более широкое использование свойств эквивалентности детерминант, обобщение переводов и т. п., что позволило бы зна-

Таблица 9

Распределение глаголов выборки по типам и подтипам

Общи́и или частные структурные типы	Количество подтипов	Количество глаголов по данному типу
ОСТ I	1	1
» I	10	144
» I	3	9
ЧСТ I	5	5
» II	2	4
» III	13	23
Всего: 6	34	186

чительно сократить число подтипов без ущерба для эффективности алгоритма.

Таким образом, большое количество многозначных глаголов характеризуются общностью в структуре алгоритмов перевода. Эта общность может быть реализована в структурных типах и подтипах без ущерба для индивидуальной лексической сочетаемости.

При введении понятия «структурный тип» мы определили составность структурного типа как характеристику, выраженную числом вертикальных стрелок в типе. Очевидно, что составность характеризует только тип и не относится к подтипу.

При выведении типов из подтипов и конкретных алгоритмов перевода целесообразно ввести иерархию среди лексически-выраженных детерминант и пользоваться ею для типизации. ЛГ-детерминанта считается старше, чем Л-детерминанта. В СТ при наличии старшей детерминанты младшая может быть нулевой. Так, принимаются структурно эквивалентными, т. е. составляющими один СТ, алгоритмы, представленные следующими дизъюнкциями детерминант:

- 1) $DD \vee PD \vee d^9$;
- 2) $DD \vee PD$.

⁹ d — правая Л-детерминанта, a — левая Л-детерминанта, f — правая ЛГ-детерминанта, p — левая ЛГ-детерминанта, γ — М-детерминанта, β — МЛ-детерминанта, α — С-детерминанта. Лексически выраженные детерминанты в составе сложных и составных обозначаются соответствующими заглавными буквами, например: DD — составная Л-детерминанта правого направления, у которой оба компонента являются Л-детерминантами; PD — составная детерминанта, у которой первый компонент — ЛГ-детерминанта, а второй — Л-детерминанта и т. д. Грамматически выраженные детерминанты в составе сложных следуют первыми в написании, если относятся к основному слову, и после той детерминанты, к которой относятся, в противном случае. Например, $D\gamma$ — сложная детерминанта, где М-детерминанта относится к лексическому компоненту — второй детерминанте; αPF — сложная детерминанта, где С-детерминанта относится к основному слову, а справа и слева от него лексические детерминанты и т. д.

Введение факультативной детерминанты целесообразно при типизации. Распространяя понятия иерархии и факультативности на всю систему детерминант, можно получить в конце концов один СТ для любого алгоритма перевода, а конкретные типы будут получаться из него заполнением пустых мест. Однако цель исследования заключается в изучении реально существующих типов и структур, поэтому при рассмотрении составности структуры нецелесообразно пользоваться понятиями факультативной, или нулевой, детерминанты. Все лексически-выраженные детерминанты считаются равноценными и отмечаются в структуре. Структура, отражающая составность, будет в связи с вышеизложенным отличаться от СТ, поэтому соответствующую конструкцию, в отличие от последнего, будем называть просто «структурой».

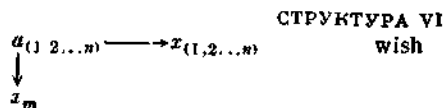
Всего в выборке отмечено 37 структур. По составности они распределяются следующим образом (см. табл. 10).

Из табл. 10 следует, что основное количество структур — двух- и трехсоставные, т. е. варьированность детерминант ограничена.

Приведем наиболее распространенные структуры:

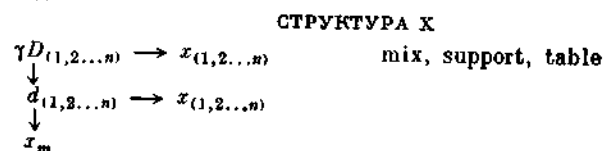
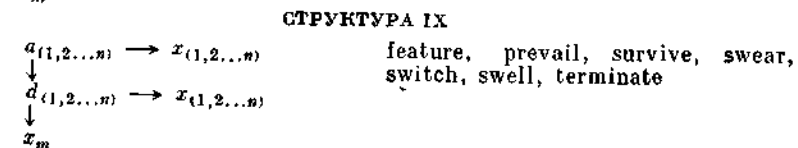
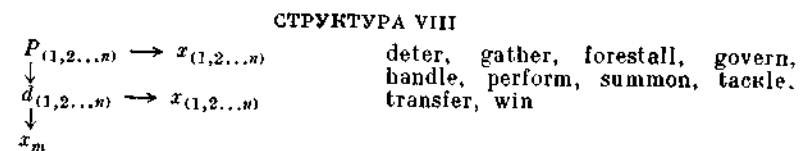
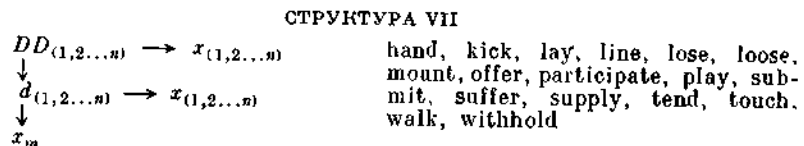
Таблица 10	
Распределение структур выборки по составности	
Составность	Число структур
Односоставные	6
Двухсоставные	14
Трехсоставные	12
Четырехсоставные	3
Прочие (составность выше четырех)	2
Всего:	37

$\begin{array}{c} d_{(1,2,\dots,n)} \longrightarrow x_{(1,2,\dots,n)} \\ \downarrow \\ x_m \end{array}$	<p>СТРУКТУРА I</p> <p>arrest, assess, balance, consume, consolidate, fasten, etc.</p>
$\begin{array}{c} P_{(1,2,\dots,n)} \longrightarrow x_{(1,2,\dots,n)} \\ \downarrow \\ x_m \end{array}$	<p>СТРУКТУРА II</p> <p>pertain, procure, subside, swallow, etc.</p>
$\begin{array}{c} \gamma D_{(1,2,\dots,n)} \longrightarrow x_{(1,2,\dots,n)} \\ \downarrow \\ x_m \end{array}$	<p>СТРУКТУРА III</p> <p>oblige, stumble, surprise, etc.</p>
$\begin{array}{c} DD_{(1,2,\dots,n)} \longrightarrow x_{(1,2,\dots,n)} \\ \downarrow \\ x_m \end{array}$	<p>СТРУКТУРА IV</p> <p>subscribe, vest</p>
$\begin{array}{c} \beta_{(1,2,\dots,n)} \longrightarrow x_{(1,2,\dots,n)} \\ \downarrow \\ x_m \end{array}$	<p>СТРУКТУРА V</p> <p>suggest</p>



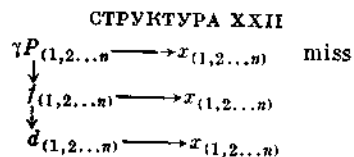
Всего в исходной выборке 128 глаголов, т. е. больше половины, имеют односоставные структуры алгоритмов перевода.

Их двухсоставных структур наиболее частыми оказались следующие:

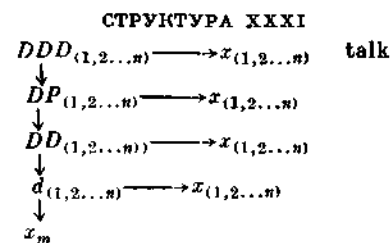


В остальных десяти двухсоставных структурах отмечено по одному глаголу в каждой. Всего на двухсоставные структуры приходится 47 глаголов выборки.

Трёхсоставные структуры многочисленны, но в выборке каждая представлена лишь одним глаголом. Пример трёхсоставной структуры:



Всего в выборке глаголов с трёхсоставной структурой 12. Четырёхсоставных структур отмечено всего три, каждая из которых представлена одним глаголом. Пример четырёхсоставной структуры:



В выборке имеется всего два глагола, у которых разнообразие детерминант не входит в перечисленные структуры. Это to turn, у которого шесть разных видов детерминант, и to keep, имеющий семь видов. Оба глагола находятся на верхней статистической границе выборки и, вообще говоря, могут считаться принадлежащими к наиболее многозначным и наиболее частым глаголам, группа которых не рассматривается здесь специально.

Число глаголов выборки, распределённых по одно- и двухсоставным структурам, составляет 175, т. е. около 85% всех многозначных глаголов выборки. Это позволяет сделать следующий вывод: в большинстве случаев зависимость лексического значения от контекста для глаголов средней многозначности и средней частоты, т. е. для основной массы многозначных глаголов, качественно однородна. Эта качественная однородность зависимости от контекста проявляется в том, что для большинства многозначных глаголов контекстные значения определяются детерминантами одного или двух видов.

Представляет интерес установить, какова составность наиболее часто используемых алгоритмов, т. е. алгоритмов перевода

Таблица 11

Структура алгоритмов наиболее частых глаголов выборки

№№	Глаголы	Абсолютная частота	Число переводов	Структура
1	improve	108	3	односоставная
2	increase	320	8	»
3	invite	102	11	двухсоставная
4	issue	450	11	односоставная
5	keep	180	41	многосоставная
6	lead	120	4	двухсоставная
7	obtain	290	8	односоставная
8	observe	110	7	двухсоставная
9	offer	150	11	»
10	participate	170	5	»
11	press	450	8	трёхсоставная
12	support	260	7	двухсоставная
13	suggest	582	2	односоставная
14	talk	120	8	четырёхсоставная
15	wish	340	2	односоставная

глаголов с наибольшей частотой встречаемости. Из выборки было выделено 15 многозначных глаголов, имеющих абсолютную частоту выше 100. Данные о составности структур алгоритмов перевода этих глаголов сведены в табл. 11.

Из данных этой таблицы следуют выводы: а) наиболее частые глаголы выборки разрешают многозначность в контексте с помощью алгоритмов, устанавливающих всего один-два (редко три) вида связи, и обратно: простая схема запроса контекста позволяет разрешить многозначность значительного числа глаголов в большинстве случаев их употребления; б) на представленном материале не прослеживается сколько-нибудь явной зависимости между числом переводов, составностью структуры и частотой глагола.

§ 5. КОМПЛЕМЕНТАРНЫЕ МЕТОДЫ

При построении алгоритмов распознавания значений полисемантических слов возник ряд случаев, не разрешимых в рамках метода детерминант, описанного нами в § 3. Их можно систематизировать в три типа. Первый можно назвать проблемой о б щ е о в ы х о д а а л г о р и т м а, второй — проблемой ф р а з е о л о г и ч е с к и х е д и н с т в и третий — проблемой м е т а т а к с н ы х п р е о б р а з о в а н и й.¹⁰

Посмотрим, что это за случаи и с помощью каких методов, дополняющих метод детерминант, они могут быть подведены под общий случай снятия омонимии посредством диагностических форм, о котором мы говорили в § 1. Начнем с примера.

Словарная статья при глаголе *to absorb* в словаре В. К. Мюллера имеет следующий вид:

absorb, *v* всасывать, впитывать, абсорбировать, амортизировать (толчки), поглощать (*тж. физ.*).

~ *ed in thoughts* погруженный в мысли.

Дистрибутивные модели Ю. Д. Апресяна для этого же глагола [4]:

1) впитывать: C_1 *absorb*+ C_2 *Blotting paper absorbs ink.*

2) вбирать, поглощать: P *absorb*+ C *The boy absorbs knowledge.*

3) погружаться: P +*to be absorb*+*ed in*+ C . *He is absorbed in reading.*

Будем считать приведенные значения глагола *to absorb* его возможными переводами и рассмотрим эти два описания значений одного английского глагола с точки зрения использования их для автоматического выбора одного из переводов. Список переводов в словаре Мюллера больше списка Апресяна вследствие того, что первый составлен без определения границ применимости в зависимости от характера текстов; сюда вошли и

могущие считаться терминологическими значения 'абсорбировать', 'амортизировать'. Второй список меньше, хотя можно считать, что значения 'абсорбировать', 'амортизировать' удовлетворительно заменяются переводом 'поглощать'. Очевидно, что для механического применения модели значений Апресяна гораздо удобнее, чем традиционное описание слова с его переводами в обычном переводном словаре. Однако даже если отвлечься от принципиальной невозможности использования этих моделей в алгоритмических целях, вследствие того что они предполагают известную синтаксическую модель словосочетания, возможную лишь после полного синтаксического анализа, то остаются еще некоторые особенности, требующие специального рассмотрения. Во-первых, следует учесть трансформационные характеристики моделей. Так, для пассивной конструкции модель 1 изменит структуру. Во-вторых, имеет смысл строить модели для определенной области языка; тогда список значений может оказаться либо неполным, либо некоторые значения будут лишними. В-третьих, должны быть указания о том, как поступить, если в тексте не будет распознана ни одна дистрибутивная модель. В частности, это может произойти при случайном сбое. В принципиальном отношении речь идет о том, следует ли выделять у слова его главное значение.

Если последовательно проводить принцип «каждому значению отвечает определенная дистрибуция», то и для главного значения нужно указать тип дистрибуции. Но каким принципом при этом руководствоваться? Например, можно предположить, что формула дистрибуции для такого значения должна быть наименее специализирована и наиболее частотна. Ю. Д. Апресян ставит знак равенства между понятиями «наиболее—наименее частотна» и «наименее—наиболее специализирована». Однако зависимость между специализированностью и частотностью представляется несколько более сложной. Рассмотрим пример. Глагол *to put* в трех своих значениях детерминируется соответственно тремя формулами дистрибуции (если следовать Ю. Д. Апресяну):

1) примириться: P +*put*+*up*+*with*+ C . *He never put up with injustice.*

2) положить: P +*put*+*up*+ C_1 +*on*+ C_2 . *She put up the book on the shelf.*

3) отложить: P +*put*+*off*+ C . *They put off their visit for a week.*

Которая из этих формул дистрибуции наиболее частотна? Какая наиболее специализирована? Если понимать под специализацией усложненность формулы дистрибуции в смысле количества компонентов формулы, то наиболее специализирована формула 2. Однако далеко не очевидно, что она менее частотна, чем формулы 1 или 3. Во всяком случае частотность не является

¹⁰ L. Tesnière. *Éléments de syntaxe structurale*. Paris, 1959, p. 283.

легко определяемым признаком для каждого из возможных переводов глагола, потому что трудно определить масштаб выборки, в которой все частотные характеристики всех входящих в алгоритм переводов были бы достаточно достоверны, так как в разных наборах контекстов разные значения одного и того же глагола непременно имеют разные частоты. Затруднительна также (хотя и не невозможна) частотная характеристика главного значения. Тем не менее, когда имеется в наличии конкорданс большого объема, можно получить удовлетворительное представление о частоте последнего. Мы не касаемся здесь традиционной лексикологической и семантической проблемы «общего» и «частных» значений многозначного слова, главного и второстепенных значений; вопрос рассматривается в чисто алгоритмическом плане: какое из значений можно считать недетерминируемым и достаточно широким, чтобы оно могло служить «общим выходом» алгоритма перевода?

Выбор основного значения по степени «общности» представляет известные трудности; так как «общность» можно понимать как 1) первое, основное значение слова в обычном переводном словаре; 2) основное, наиболее распространенное значение слова в данном конкордансе; 3) основное значение данного слова в генеральной совокупности текстов данного характера, известное составителю алгоритма перевода, но не совпадающее с первыми двумя пунктами. Поэтому в методе детерминант предусматривается выбор для слова одного основного недетерминируемого значения. Все другие значения, отличающиеся от основного, детерминируются и рассматриваются как частные.

Таким образом, принимается, что каждое многозначное слово имеет одно главное значение, которое специально не детерминируется, хотя оно и может иметь, строго говоря, специальную дистрибутивную формулу, которая не подходит для частных значений. Это значение не детерминируется в алгоритмах разрешения лексической многозначности, построенных по методу детерминант, а является так называемым «общим выходом» алгоритма. Детерминируются значения, отличающиеся от общего выхода. И значит, к методу детерминант подключается метод конвенций по вопросу об иерархии значений полисемантического слова и отбору главного значения.

Следующий тип проблем — это фразеологизмы.

Перевод исследуемого слова может детерминироваться другим словом или словами, образующими вместе с исследуемым словосочетанием разной степени устойчивости. Фразеологическая связанность компонентов словосочетания может быть настолько сильной, что само словосочетание целесообразно включить в словарь оборотов, и тогда оно в наше рассмотрение не входит.¹¹

¹¹ Если словосочетание не включено в словарь оборотов и исследуемое слово выступает как многозначное, то тогда с алгоритмической точки зрения

Таким образом, здесь на помощь методу детерминирования, результаты которого представлены в контекстологическом словаре, приходит метод построения словаря оборотов.

И наконец последнее: вопрос о метатаксных преобразованиях.

Активная роль глагола в синтаксисе предложения и специфика рассмотрения проблемы многозначности в структурном плане приводят к одному известному затруднению. Перевод некоторых глаголов может быть дан только после выяснения того, в каких синтаксических конструкциях, переводимой и переводящей, этот глагол должен участвовать; обратно, применение переводящей синтаксической конструкции зависит от того, какой перевод получит входящий в эту конструкцию глагол. Метод детерминант допускает анализ по существу синтаксических явлений (с помощью грамматически-выраженных детерминант) и включение соответствующих правил в алгоритмы перевода многозначных слов; однако ясно, что сколько-нибудь сложный синтаксический анализ в рамках алгоритмов перевода слов является крайне нежелательным, так как усложняет эти алгоритмы. Практически гораздо выгодней признать неприменимость метода для ряда случаев и очертить реальный круг явлений в исходном конкордансе, выходящих за рамки метода. Перечисляя факты, лежащие вне возможностей метода, мы будем идти от отдельных глаголов к явлениям общего характера.

1. а) В исходном списке отмечено пять глаголов, а именно *to like, to lack, to need, to preface, to seem*, которые во всех своих переводах, зафиксированных в исходном конкордансе, требуют изменения синтаксической структуры предложения;

б) отмечено четыре глагола, которые в пассиве требуют для перевода особых конструкций, неконгруэнтных английским конструкциям; это глаголы *to quote, to say, to report, to know*;

в) глагол *to know* в форме причастия прош. вр. участвует в синтаксических построениях, подбор эквивалентов которым в русском языке затруднителен.

Примеры к пунктам «б» и «в»:

He is quoted as saying. . . ; He was said to be at home; The king was reported to have said. . . ; They were known to be dead; They are known to have been living. . . ; If known. . . when known. . . ; Still known to exist. . .

2. В исходном списке имеется девять глаголов, многозначность которых не разрешается методом детерминант в исходном конкордансе. По-видимому, здесь требуется детерминирование в категориях семантики всего предложения или текста в целом. Это следующие глаголы:

степень фразеологической связанности компонентов словосочетания не имеет значения, она нерелевантна.

to sustain	'выдерживать'/'подтверждать';
to top	'обезглавить'/'завершить';
to trail	'тащиться'/'следить';
to leave	'оставить'/'выехать';
to lie	'лежать'/'лгать';
to make way	'двигаться вперед'/'уступить дорогу';
to walk out	'выйти'/'забастовать';
to loan	'дать займы'/'взять займы';
to govern	'руководить'/'руководствоваться'.

То обстоятельство, что таких случаев собственно лексической многозначности относительно немного — всего девять из общего числа приблизительно в две тысячи глаголов — и что это глаголы сравнительно невысокой частотности, служит лишним доказательством приемлемости метода детерминант.

3. Требуют специального анализа конструкции со вторичной предикацией. П р и м е р ы: I saw him going. . . ; They have let it be known. . . ; I should like you to decide. . .

4. Лексически выраженная детерминанта может находиться так далеко от исследуемого слова, что с помощью существующих правил пропуска ее обнаружить или не удастся, или это будет связано с чрезвычайно громоздкими и малоэффективными процедурами. П р и м е р (правая лексическая детерминанта глагола to take находится в придаточном предложении, в то время как сам глагол в главном): . . . positions they will take. Надо отметить, что такие случаи, хотя они кажутся довольно естественными, достаточно редки; по нашим подсчетам, на 700 употреблений глагола to take в английских газетных текстах таких случаев отмечено всего 4.

5. Для пассивной конструкции, в которой может участвовать каждый английский глагол, правая лексически-выраженная детерминанта становится левой. При пассивной конструкции может быть осуществлена несложная операция повторения поиска правой лексически-выраженной детерминанты влево: сигналом к проведению такой операции будет наличие признака «пассив», который сообщается глаголу специальной схемой анализа, работающей до перевода.

Чтобы разрешить ситуации, перечисленные в п. 1 (стр. 209), требуется полный синтаксический анализ предложения. Список таких глаголов, вероятно, пополнится в новом конкордансе. Таким образом, перечисленные глаголы конкретно, списком, задают границу досинтаксического метода, метода детерминант, и синтаксических методов. Наиболее целесообразным представляется также синтаксический подход и для конструкций, указанных в п. 3. Для разрешения многозначности глаголов, перечисленных в п. 2, синтаксический анализ не будет эффективным, он в данном случае нерелевантен. Проблема здесь должна ре-

шаться в плане семантики всего высказывания в целом. Возможно, что там же следует искать решения ситуации п. 4.

Алгоритмы перевода многозначных глаголов составлялись с учетом описанных выше ограничений.

§ 6. УЛУЧШЕНИЕ КОНКОРДАНСА: ТИПИЗАЦИЯ ЛЕКСИЧЕСКИХ ДЕТЕРМИНАНТ

Рассмотрим лексически-выраженное детерминирование в алгоритмах перевода глаголов с тем, чтобы установить:

1) существует ли возможность преобразования лексических детерминант, заданных одним словом, в списки слов и списков — в лексико-грамматические классы;

2) имеет ли детерминирование с помощью лексико-грамматических детерминант какие-нибудь особенности по сравнению с другими видами детерминирования.

Ответ на эти вопросы может помочь выяснить отношения между методом детерминант и гипотетическим способом определения значений посредством разделения лексики на непересекающиеся словарные классы [7], а также перспективы метода и зависимость его разрешающей способности от величины и полноты исходного конкорданса.

Таблица 12

Полуоператорная запись алгоритма перевода

1	→	сп 1	23	3	
2	п				'задержать'
3	→	attention	23	5	
4	п				'привлечь'
5	→	drift	23	7	
6	п				'остановить'
7	→	policy	23	9	
8	п				'прекратить'
9	п				'арестовать'

Сп 1

- 1) development
- 2) movement

Начнем с конкретного примера. В табл. 12 дана полуоператорная запись алгоритма перевода глагола to arrest, составленного по конкордансу в 49 употреблений этого глагола (пояснение к записи см. на стр. 223). В словаре В. К. Мюллера глаголу to arrest придана следующая словарная статья:

arrest *v* 1) задерживать, останавливать; to a. development задерживать развитие; 2) арестовывать; 3) приковывать (взоры, внимание); 4) выключать (машину, прибор); тормозить.

Если говорить о «значениях» в собственном смысле, то в алгоритме (табл. 12) отражены все эти значения, кроме технического

(‘выключать машину, прибор’). Переходя от значений к переводам, устанавливаем, что отдельные переводы детерминированы следующим образом.

1) ‘З а д е р ж а т ь’ — в словаре В. К. Мюллера не отделяется от ‘останавливать’ и иллюстрируется сочетанием с development. В алгоритме перевода этот перевод детерминируется списком, в который, помимо development, входит еще и слово movement. Видимо, в определяющем контексте сочетание to arrest movement было переведено с глаголом ‘задержать’, а не ‘остановить’, что может быть объяснено реальной ситуацией, а также тем, какое движение имеется в виду (например, политическое движение).

2) ‘О с т а н а в л и в а т ь’ — детерминируется в алгоритме отдельно с помощью существительного drift.

3) ‘А р е с т о в ы в а т ь’ — принят в качестве общего выхода.

4) ‘П р и в л е ч ь’ — детерминируется существительным attention, хотя, как видно из пояснений в словаре Мюллера, возможны и другие слова в качестве детерминант для этого перевода.

Поставим вопрос о том, возможна ли в данном алгоритме перевода замена лексических детерминант-слов детерминантами-списками или лексико-грамматическими детерминантами без дополнительного конкорданса. Попробуем определить пригодность переводов в этом алгоритме для изменяющихся детерминант.

Расширим список 1, включив туда слова-синонимы process, march, advance. Будут ли эти слова также детерминировать перевод ‘задержать’? В отношении существительных process, advance положительный ответ представляется переводчику вероятным; что же касается слова march, то его сочетаемость с to arrest представляется маловероятной; во всяком случае желательно иметь подтверждение возможности такого сочетания в виде примера из реального текста. При пополнении списка приходится, таким образом, решать две проблемы: 1) какие слова одинакового с детерминантой смыслового содержания можно включать в список, предварительно определив общность семантического содержания; 2) сочетаются ли найденные таким образом слова в реальных текстах. Составление алгоритма по конкордансу имеет в этом отношении неоспоримые преимущества. Формулирование семантического содержания детерминант представляет собой нелегкую задачу, как это видно даже из поверхностного знакомства со списками в алгоритмах перевода. Если мы попробуем обобщить список 1, представив его как подкласс существительных, обозначающих «движение, прогресс, развитие», то, кроме двух названных выше проблем, возникает еще и третья, а именно: 3) в каких алгоритмах перевода других многозначных слов можно будет исполь-

зовать выделенный подкласс. Очевидно, что назначение специального подкласса оправдано только для его множественного использования.

Вышеизложенное позволяет сделать вывод, что путь расширения алгоритма начинается, по-видимому, от рассмотрения конкорданса для каждого отдельного многозначного слова, а замена и обобщение детерминант возможны лишь после обследования всех алгоритмов перевода, для которых такая замена предполагается. Предпочтительна, по-видимому индуктивная методика: от текста и от реально наблюдаемой лексической и лексико-грамматической сочетаемости к составлению списков слов определенной тематики, которые в конце концов могли бы выступать в качестве объединенных детерминант.

Лексико-грамматические детерминанты в алгоритмах перевода глаголов встречаются достаточно часто. В исходной выборке в 41 алгоритме имеются лексико-грамматические детерминанты. Это алгоритмы перевода глаголов to back, to deter, to forestall, to gather и др. Среди этих глаголов у четырех различие значений происходит только с помощью лексико-грамматических детерминант. Это to pertain, to procure, to subside, to swallow. Все они односоставны. Рассмотрим некоторые значения, детерминируемые лексико-грамматически. В алгоритме перевода глагола to back имеются значения ‘отступить’, ‘уклониться’, детерминируемые послелогоми down, out, away. Перевод ‘поддержать’ детерминируется последующим существительным и послелогом. Как видно, приведенные выше переводы могут быть сгруппированы по общности семантического содержания: одна группа состоит из переводов с единым значением ‘отказаться’, другая — со значением ‘поддержать’. Второе значение детерминируется лексико-грамматической детерминантой. В алгоритме перевода глагола to deter имеются всего три перевода: ‘запугать’, ‘сдерживать’ и ‘удержать от’. Перевод ‘запугать’ детерминируется подклассом «одушевленные существительные». Заметим, что ‘запугать’ по смыслу дальше от ‘сдерживать’, детерминируемого лексически (aggression), которое приближается к значению, взятому как общий выход (‘удержать от’). И здесь лексико-грамматическая детерминанта определяет значение, наиболее отстоящее по смыслу от остальных. В алгоритме перевода глагола to forestall лексико-грамматическая детерминанта «одушевленное существительное» детерминирует значение ‘помешать’; другие значения этого глагола — ‘перехватить’, ‘задержать’, ‘предотвратить’. Два первых детерминируются лексически, третье — общий выход. Отметим, что у этого глагола, как и у предыдущего, лексико-грамматической детерминантой является, в отличие от глагола to back, не класс, а подкласс слов, соответственно с чем и различие в детерминируемых значениях для двух последних глаголов не так велико. Лексико-грамматическая детерминанта «глагол» в алго-

ритме перевода глагола to lead детерминирует значение 'заставить', существенно отличающееся от остальных ('возглавить', 'завести', 'привести'). Глагол to maneuver имеет составную детерминанту, где главную роль играет лексико-грамматическая детерминанта «одушевленное существительное». Эта детерминанта определяет перевод 'заставить занять'; другие переводы: 'распустить', 'создать', 'обмануть', 'маневрировать'. Имеются и другие примеры такого же рода. На этой основе представляется возможным высказать следующее предположение относительно лексико-грамматических детерминант.

Лексико-грамматические детерминанты определяют выбор значения, которое по смыслу в наибольшей степени отличается от других имеющихся значений глагола. Существует также прямая зависимость характера лексико-грамматической детерминанты (класс или подкласс слов) и отдаленности детерминируемого ею значения от остальных. Класс детерминирует наиболее отдаленные значения, подкласс — менее отдаленные. Различие в значении коррелирует со степенью обобщенности детерминанты.

§ 7. УЛУЧШЕНИЕ КОНКОРДАНСА: РЕЗУЛЬТАТЫ РЕАЛИЗАЦИИ АЛГОРИТМОВ НА ЭВМ

Машинный эксперимент по переводу многозначных слов был проведен как часть проверки всей системы машинного перевода английских публицистических текстов [11]. Переводились тексты длиной в 1000 — 2000 слов. В одной серии испытаний на машине было переведено около 30 тыс. словоупотреблений; результаты этого прогона и рассматриваются в настоящей работе. В ходе машинного эксперимента были проверены алгоритмы перевода всех многозначных слов, всех однозначных слов и прочие алгоритмы, входящие в систему машинного перевода, частью которой является рассматриваемый контекстологический словарь. Переводились тексты, не использовавшиеся при составлении исходного конкорданса.

Общая эффективность перевода многозначных слов иллюстрируется на примере текста № 2, данного в приложении. В этом тексте имеется 42 употребления многозначных глаголов; из них неправильно перевелось 7. Эффективность перевода, таким образом, составляет 84%. Несмотря на значительный разброс в качестве перевода отдельных текстов, общая результативность перевода многозначных слов довольно устойчиво остается в пределах 80%. Машинный эксперимент позволяет проверить алгоритмы перевода многозначных слов «в ширину», т. е. работу их совместно с переводом однозначных слов, с другими многозначными словами, зависимость правильности перевода от полученной на входе грамматической информации, количество случайных сбоях и пр.

В то же время естественным образом срабатывают в основном алгоритмы перевода наиболее часто встречающихся многозначных слов.

В каждом конкретном случае многозначности можно всегда установить границы, за которыми выбор переводного эквивалента может быть осуществлен с помощью средств, не входящих в первоначальный арсенал. Однако неизвестно, сколь часты случаи сложной контекстной зависимости, вызывающие к жизни особые диагностические формы, и как распределяются сложные ситуации среди основного количества многозначных слов. Машинный эксперимент и анализ неправильных переводов помогают ответить на вопрос о том, каковы возможности метода детерминант и какими затратами достигается улучшение конкорданса.

Составитель алгоритма стремился объединить в общем выходе, во-первых, наиболее частые и семантически близкие значения и, во-вторых, те, которые трудно детерминировать. Вследствие такой разнородности требований к общему выходу назначение его для ряда многозначных слов не является простой задачей. Частота отдельных значений может быть хорошо определена лишь в весьма большом конкордансе, практически почти не реализуемом. Имея перед собой исходный конкорданс, составитель отмечал наиболее частый перевод, оценивал его с точки зрения общности, намечал общий выход, детерминировал остальные переводы и сводил детерминанты и переводы в алгоритм. Естественно ожидать, что в большинстве случаев алгоритм будет переводить слово по общему выходу. Машинный эксперимент показывает, что это именно так и происходит.

Объем исходного конкорданса и число отдельных словоупотреблений глаголов исходной выборки иллюстрируются данными, приведенными в табл. III Приложения. Рассматривая данные этой таблицы, можно прийти к заключению, что для глаголов, перечисленных в ней, переводы достоверны, поскольку они получены с учетом большого количества употреблений, за исключением некоторых, не обладающих высокой частотностью. Между тем интересно проверить алгоритмы на как можно большем числе употреблений для каждого из них, т. е. «в глубину». Поэтому результаты чисто машинного прогона были дополнены специальным отдельным просмотром эффективности алгоритмов перевода глаголов выборки, не охваченных машинным экспериментом, на дополнительном материале. Объем этих дополнительных материалов также составил 30 тыс. словоупотреблений.

Для проверки были взяты 194 алгоритма перевода многозначных глаголов, восемь из которых в Приложении. В качестве материала для проверки были использованы тексты публицистической тематики из разных источников, но не из тех, по которым составлялся исходный конкорданс (для проверки «на прочность»). Работа производилась в следующем порядке:

- 1) отмечались в текстах подряд все глаголы;
- 2) среди глаголов отмечались те, для которых в Приложении имеются алгоритмы перевода;
- 3) выписывался в специальную таблицу контекст, в котором встретился глагол, и производился его перевод с помощью алгоритма.¹²

Таким способом было проверено 250 употреблений глаголов заданной выборки и 84 алгоритма.

В соответствии со сказанным ранее, при проверке учитываются следующие операции, считающиеся выполненными: а) изменение направления поиска детерминанты при пассивных конструкциях, б) отыскание лексической детерминанты в том случае, если она находится не в том предложении, где исследуемое слово. Эти операции выполняются специальными алгоритмами. Перевод глагола, по условию, дается в совершенном виде. Если перевод требует другого вида глагола, то в тех случаях, когда подбор другого глагола видовой пары не вызывает затруднений и совершается по достаточно простым правилам ('завершать' — 'завершить'), которые реализованы в специальном алгоритме, неправильный вид ошибкой не считается. Если глагола видовой пары нет или существует множественный выбор, то неправильный вид глагола фиксируется как ошибка алгоритма — нужно было при составлении алгоритма предусмотреть вид.

Из 250 употреблений глаголов неправильно перевелось 58, т. е. 20% всех случаев. Рассмотрим неправильные переводы.

Наиболее распространенным случаем ошибочного перевода является ошибка № 1. Она характеризуется наличием следующей ситуации. В алгоритме есть нужный перевод и этот перевод может быть легко детерминирован. Детерминанта отсутствует в списке детерминант алгоритма. Она может быть туда включена. Ошибка № 1 является следствием недостаточного объема или представимости исходного конкорданса. **Пример:** глагол *to perform* (в предложении *These objective causes of our weakness must be rapidly removed to enable the working class to perform its vanguard role*) перевелся 'сделать', хотя среди его переводов в алгоритме есть 'выполнить', детерминируемый списком. Достаточно пополнить этот список детерминантой *role*. Всего ошибок № 1 в текстах 19.

Ошибка № 2 происходит при отсутствии в алгоритме необходимого перевода и необходимой детерминанты, хотя перевод может быть детерминирован, а также тогда, когда одного лексического детерминирования недостаточно и требуется (возможно) введение грамматически-выраженной детерминанты. **Примеры:** алгоритм перевода глагола *to gain* не имеет перевода

'получать' или 'приобрести', вследствие чего этот глагол перевелся неправильно (в предложении *The federal rulers believe that the best way to gain possession of nuclear armaments is through a Bonn-Washington nuclear axis*). Нужное значение глагола хорошо детерминируется составной детерминантой *possession of*. Глагол *to press* (в предложении *Clearly, the independence of the trade unions is one of the most pressing problems*) переводится 'нажать на', что неправильно. Правильный перевод обеспечивается введением М- или МЛ-детерминанты. Всего ошибок № 2 в текстах 27.

Ошибка № 3 происходит вследствие неправильного выбора глагола из пары «возвратный — невозвратный». Правильное решение можно получить, детерминируя глагол признаком переходности, однако это требует некоторого изменения исходной организации русского словаря. В проверяемых алгоритмах решение проведено лишь там, где допустимо лексическое детерминирование (детерминирование отсутствием не производится). **Пример:** глагол *to increase* (в предложении: *The number of the American planters in the Republic of Ivory Coast increased more than twofold between 1947 and. . .*) переводится 'увеличить', в то время как требуется перевод 'увеличиться' или 'возрастать'. Всего ошибок такого рода в текстах отмечено 4.

Ошибка № 4 происходит вследствие неправильного общего выхода алгоритма. Она иллюстрируется примером перевода глагола *to win*, данным ниже.

Ошибка № 5 отражает ситуации, когда в контексте нет детерминанты. Именно эти непереуведенные (неправильно переведенные) случаи образуют неформализованный остаток, который

Таблица 13

Классификация неправильных переводов

№№	Причина неправильного перевода	Число ошибок по этой причине
1	В алгоритме отсутствует детерминанта. Перевод есть и может быть детерминирован	19
2	В алгоритме нет детерминанты и перевода. Требуется грамматическое детерминирование. Детерминирование возможно	27
3	Недетерминирован выбор глагола из пары «возвратный — невозвратный». Детерминирование возможно	4
4	Неправильен общий выход алгоритма	1
5	В контексте нет детерминанты. Детерминирование в данном конкордансе невозможно. Требуется расширение конкорданса	7
	Итого:	58

¹² Используя алгоритмы, данные в приложении, каждый может воспроизвести указанный эксперимент.

уменьшается в результате дальнейшего пополнения и улучшения исходного конкорданса.

Сведем в таблицу результаты проверки по типам ошибок.

Рассмотрим случаи отсутствия детерминант в тех ситуациях, когда ближайший контекст не дает возможности определить перевод. Возьмем два наиболее характерных случая.

1) But such healthy antipropaganda propaganda was not to be allowed to win so easily.

Здесь перевод глагола *to win* 'выиграть', полученный по общему выходу, является неудачным (хотя заметим кстати, что при правильном определении другой контекстной информации смысл предложения в целом был бы понятен и при таком переводе; обстоятельство, лежащее в основе точки зрения о возможности машинного перевода с одними однозначными словами). Попытка детерминировать значение 'победить' в данном контексте оказывается неудачной. Слово *propaganda*, с которым по смыслу сочетается глагол *to win*, не может служить детерминантой глагола, вследствие того что связь между ними слишком случайна. Синтаксический анализ не поможет в данном случае детерминировать перевод. Отсутствуют формальные признаки, которые могли бы послужить основой для грамматически-выраженного детерминирования. Правильный перевод может быть сделан после учета семантических связей между предикатом (глаголом) и субъектом на уровне семантического анализа, что выходит за рамки метода. Для данного глагола эффективным может оказаться изменение общего выхода.

2) Lord Atlee was slightly improved today from his ulcer operation according to a hospital bulletin.

Перевод глагола *to improve* не может быть сделан без учета семантики всего высказывания. В качестве условной детерминанты может быть предложена следующая: «пассив глагола *to improve*» и левая лексико-грамматическая детерминанта «одушевленное существительное». По-видимому, для какого-то числа случаев такая детерминанта может быть эффективной. Существенно то обстоятельство, что синтаксический контекст в данном случае не разрешает многозначности.

Подведем итоги. Больше всего неправильных переводов происходит тогда, когда в алгоритме нет перевода, либо есть перевод, но нет детерминанты, причем и детерминанта, и перевод вполне возможны в рамках метода. Преобладание этого рода неправильных переводов говорит о недостаточности исходного конкорданса для некоторых многозначных глаголов, вошедших в выборку. Грамматически-выраженное детерминирование приводит к решению вопроса в довольно многих случаях. Отсюда следует вывод, что сложное или составное детерминирование могут привести к существенному повышению результативности. Остается определенное количество контекстов, для которых не

удалось найти диагностических форм, разрешающих многозначность. Для этих случаев характерно также отсутствие связи между многозначностью и синтаксической структурой. Всего таких контекстов 7, что составляет примерно 12% от общего количества неправильных переводов, или около 3% от общего количества сделанных при проверке переводов. 88% неправильно переведенных глаголов допускают удовлетворительное разрешение многозначности в предложенной теории.

Таким образом, по полученным данным верхняя граница эффективности метода в принципе может быть поднята до 97%. Однако достижение такой эффективности на практике представляется маловероятным, поскольку для этого требуется, чтобы в конкордансе были исчерпывающим образом представлены все значения всех глаголов, определенных как многозначные, а такой конкорданс практически очень трудно, если не невозможно, реализовать. Частотность отдельных значений глаголов слабо зависит от частотности самих глаголов, тем более, что плохо подтверждается зависимость количества значений слова от частоты.

§ 8. ИТОГИ ИССЛЕДОВАНИЯ

Принципиально возможно устранение лексической многозначности слов исследуемого языка в переводе на эталонный с помощью правил анализа контекста. Разные лексико-грамматические классы обладают разными уровнями многозначности. Из знаменательных лексико-грамматических классов наиболее многозначны глаголы, наименее — существительные. Метод детерминирования значений с помощью контекстных признаков позволяет строить специальные алгоритмы разрешения лексической многозначности. Детерминирование лексических значений (переводов) слов в контексте можно осуществить с помощью пяти основных детерминант: лексической, лексико-грамматической, морфемной, морфологической и синтаксической. Разнообразие контекстных связей слова, приводящее к различию лексических значений, можно учитывать при построении алгоритмов перевода различными способами детерминирования, которые отражаются в различных объединениях детерминант.

Определяющий контекст — контекст, в котором разрешается многозначность, — различен для разных лексико-грамматических классов. Для существительных и наречий этот контекст простирается вправо и влево от слова, для глаголов и прилагательных разрешающий контекст в основном находится справа. В алгоритмах перевода многозначных глаголов преобладают правые лексически-выраженные детерминанты.

Множество алгоритмов перевода многозначных слов образует систему. Система строится по следующим параметрам: структуре

алгоритмов как совокупности детерминант и лингвистическому содержанию операций по переводу слова. Имеется небольшое число структурных типов, общих для всех многозначных слов знаменательных лексико-грамматических классов. Эти общие структурные типы охватывают большую часть многозначных слов и характеризуются несложной структурой с однотипными детерминантами. Основная масса многозначных слов связана с контекстом однородными зависимостями. Лингвистическое содержание операций по переводу отражено в контекстуальной модели перевода. Эта модель состоит из элементов, число которых может быть различным. Для большинства алгоритмов перевода глаголов контекстуальные модели перевода одно-, двух- и трехэлементны. Структурный тип алгоритма перевода, детализируемый структурным подтипом, контекстуальная модель перевода, таблица детерминант, таблица переводов и индекс модели задают полностью отдельный алгоритм перевода многозначного слова. Алгоритм перевода становится типовым, он отделяется от таблиц детерминант и переводов, реализуется отдельной программой и облегчает автоматическое использование индивидуальных контекстных связей слова, учет и пополнение их. Механизм разрешения многозначности отделяется от таблиц.

Экспериментальное исследование показало, что алгоритмы разрешения лексической многозначности, не подвергавшиеся корректировке, обеспечивают правильный перевод многозначных слов в 80% случаев. Анализ неправильных переводов показал, что в 80% случаев неправильных переводов устранение неправильного перевода происходит либо добавлением в алгоритм детерминанты, либо детерминанты и перевода в полном соответствии с предложенным методом. Метод детерминант теоретически обеспечивает, таким образом, правильность перевода многозначных слов на 90—95%. Такая эффективность позволяет использовать метод для решения большинства известных в настоящее время прикладных задач, где требуется разрешение лексической многозначности.

Алгоритмы перевода многозначных слов, построенные для каждого многозначного слова и организованные в систему по найденным параметрам этой системы, образуют автоматически действующий контекстологический словарь для перевода многозначных слов. Он может быть использован для перевода многозначных слов в системах машинного перевода, в системах автоматического реферирования и индексирования при разрешении многозначности ключевых слов, в программированном обучении иностранным языкам, как механическое пособие для человека-переводчика и пр.

ПРИЛОЖЕНИЕ

Таблица I

Распределение многозначных слов во входном словаре

Лексико-грамматический класс	Количество многозначных слов в процентах к общему количеству многозначных слов словаря
Существительные	45.0
Глаголы	25.0
Прилагательные	22.6
Наречия	3.7
Предлоги	2.5
Союзы	1.2

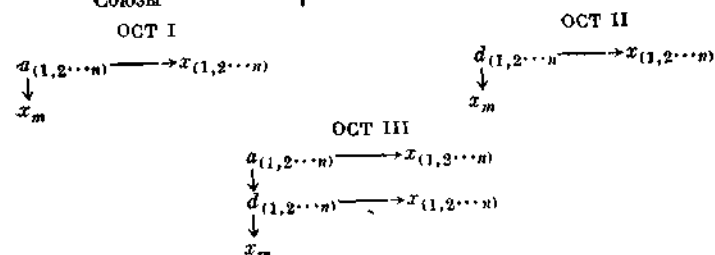


Рис. I. Общие структурные типы (ОСТ).

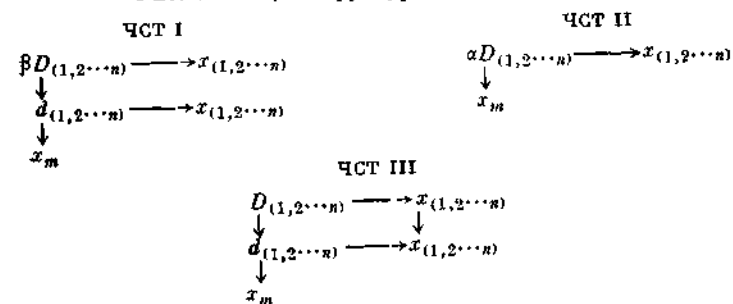


Рис. II. Частные структурные типы глаголов.

Таблица II

Распределение алгоритмов ПМС по общим и частным структурным типам

Лексико-грамматические классы	Число алгоритмов ПМС, распределенных по общим типам			Число алгоритмов ПМС, распределенных по частным типам						Всего алгоритмов, поданных типизации
	I	II	III	I	II	III	IV	V	VI	
Существительные	26	23	35	5	3	2	3	1	10	98
Прилагательные	9	41	17	4	5	5	4	5	—	100
Глаголы	—	66	14	14	6	—	—	—	—	100
Наречия	14	27	30	3	3	3	—	—	—	85

Таблица III

Глаголы выборки, прошедшие проверку

№№	Глаголы	Частоты		Число неправильных переводов
		в тексте	в конкордансе	
1	arrest	3	49	
2	assess	1	13	
3	back	3	37	
4	balance	1	31	
5	consolidate	1	8	
6	consume	2	6	1
7	gain	5	48	1
8	gather	3	14	2
9	govern	1	28	
10	grant	1	91	
11	guard	2	33	
12	guide	2	31	
13	halt	1	9	
14	hand	1	41	
15	handle	1	90	
16	heighten	1	41	1
17	hit	2	13	2
18	identify	1	38	
19	ignore	2	22	
20	improve	1	108	1
21	incur	1	8	
22	injure	4	16	3
23	increase	9	320	1
24	initiate	3	35	2
25	institute	1	33	
26	invite	2	102	
27	issue	6	450	1
28	judge	1	21	
29	jump	1	5	
30	keep	8	180	2
31	lay	7	21	5
32	lead	12	120	
33	link	6	25	
34	lose	9	41	1
35	mix	1	15	
36	miss	2	28	
37	manage	1	7	
38	mount	1	15	
39	match	1	7	1
40	name	4	64	
41	notify	4	52	1
42	obtain	1	290	
43	observe	1	110	1
44	offer	7	150	
45	outline	3	90	
46	overcome	3	12	1
47	overthrow	2	3	2
48	pave	1	5	
49	participate	1	170	
50	perform	1	15	1
51	pile	1	2	

Таблица III (продолжение)

№№	Глаголы	Частоты		Число неправильных переводов
		в тексте	в конкордансе	
52	play	15	90	
53	pocket	1	2	
54	pose	3	21	
55	press	4	450	3
56	qualify	3	33	
57	stumble	1	2	
58	suffer	4	23	1
59	succeed	4	40	
60	suspend	4	32	2
61	sum	3	8	1
62	support	18	260	
63	survive	2	2	1
64	suppress	1	8	
65	submit	1	56	
66	suggest	2	582	1
67	substitute	2	18	1
68	sweep	2	5	2
69	swell	1	2	
70	switch	1	5	
71	tailor	1	2	
72	talk	5	120	
73	tend	2	33	
74	test	3	28	
75	throw	2	27	
76	transfer	2	98	
77	turn	8	61	
78	vote	9	61	1
79	undertake	4	87	
80	watch	3	20	2
81	win	13	27	7
82	witness	4	7	1
83	wish	2	340	
84	wipe	1	2	

Алгоритмы перевода многозначных глаголов (выборка)

В столбцах полуоператорной записи следующие данные:

Первый столбец — номер строки (оператора) по порядку;

Второй — код оператора; \rightarrow — проверка вправо (влево) на слово или список слов; \leftarrow — проверка вправо (влево) на лексико-грамматический класс (подкласс); Δ — проверка слова на признак; π — перевод; π^1 — перевод слова вместе с детерминантой;

Третий — номер оператора слова, проверяемого на признак;

Четвертый — детерминанта;

Пятый — номер подпрограммы пропусков (см. ниже);

Шестой — номер оператора отхода по «нет»;

Седьмой — номер оператора слова, переводимого вместе с основным;

Восьмой — перевод.

Подпрограммы пропусков:

- 0 — ничего не пропускать;
- 2 — пропускать прилагательные;
- 10 — » существительные и прилагательные;
- 15 — » наречия;
- 16 — » все слова в пределах фразы;
- 19 — » » » предложения;
- 23 — » слова, имеющие СФ «определение».

back								
1	→	down	0	3	1	'отступиться'		
2	п ¹							
3	→	out	0	6				
4	→	of	0	6				
5	п ¹						3.4	'уклониться от'
6	→	away	15	8				
7	п							'отказаться'
8	→	up	0	10				
9	п ¹						8	'поддержать'
10	→	сущ.	23	13				
11	→	up	0	13				
12	п ¹						11	'поддержать'
13	п							

consume								
1	→	with	0	3	1	'поглотить'		
2	п							
3	→	Сп. 1	23	5				
4	п						1	'потреблять'
5	п							

- Сп. 1**
 1) food
 2) product
 3) butter

deter								
1	→	одушевлени. сущ.	23	3	1	'запугать'		
2	п							
3	→	aggression	23	5				
4	п						1	'сдержань'
5	п							

fasten

1	→	attention	23	3	1	'сосредоточить'
2	п					
3	п					

feature

1	→	Сп. 1	15	3	1	'изобразить'		
2	п							
3	→	protest	23	5				
4	п						1	'выразить'
5	п							

Сп. 2

- 1) exhibit
 2) illustration

gain								
1	→	сп. 1	23	3	1	'получить'		
2	п							
3	→	сп. 2	23	5			1	'достигнуть'
4	п							
5	→	control	23	7			1	'установить'
6	п							
7	→	credence	23	9			1	'снискать'
8	п							
9	→	сп. 3	23	11			1	'добиться'
10	п							
11	→	prestige	23	13			1	'повысить'
12	п							
13	→	сп. 4	23	15			1	'обеспечить'
14	п							
15	→	initiative	23	17			1	'захватить'
16	п							
17	→	сп. 5	23	19			17	'укрепиться'
18	п ¹							
19	→	stature	2	21			19	'занять свое место'
20	п ¹							
21	п							'выиграть'

Сп. 1

- 1) monopoly
 2) impression
 3) concession
 4) advantage
 5) vote
 6) freedom
 7) benefit
 8) support
 9) knowledge
 10) concurrence

Сп. 2

- 1) objective
 2) aim

Сп. 3

- 1) independence
 2) acceptance

Сп. 4

- 1) exception
 2) access

Сп. 5

- 1) ground
 2) strength

rule								
1	→	сущ.	23	10	1	'исключить'		
2	п ¹	out		4				
3	→						2	
4	п ¹							
5	→	roost		6			4	'командовать'
6	п ¹							
7	→	as		9			23	
8	п	сп. 1		9				
9	→						10	'использовать'
10	п	out		12				
11	→						10	'исключить'
12	п ¹	as		14				
13	→						16	'править'
14	п	that		16				
15	→						16	'поставить'
16	п							

Сп. 1

- 1) tool
 2) means

safeguard				
1	△	«определение»	3	
2	п			
3	→	customer	23	5
4	п			
5	→	сп. 1	23	7
6	п			
7	→	against	15	9
8	п			
9	п			

Сп. 1

- 1) quantity
2) deficit

ОБРАЗЦЫ МАШИННОГО ПЕРЕВОДА

Текст № 1

The exchange of letters between president Johnson and Hochiminh, recently disclosed by Hanoi, helps to clear the air on certain questions which the Johnson administration had deliberately befogged. And in doing so, it reveals once again the utter hypocrisy of Johnson's protestations of his willingness to negotiate and his desire for peace.

A year ago, Johnson was loudly proclaiming his readiness to halt the bombing if only Hanoi would indicate its willingness to negotiate. It was their refusal to do so, he insisted, which was the only roadblock to de-escalation of the war. But then the government of North Vietnam proceeded to give distinct indications that no such roadblock existed. These indications came through various channels.

Информационные ячейки к тексту № 1

Цифровой эквивалент	Грамматическая информация	Номер перевода	Слово в тексте	Перевод
19 24 16	0000000010000	00 03 26	.	(точка)
19 24 17	0000000010001	17 30 01	.	(абзац)
12 00 02	0000000000110	00 00 01	the	
22 10 25	00000010010011	15 24 25	exchange	'обмен'
26 00 00	0000000001010	00 00 01	of	
22 20 02	00001110101102	06 05 17	letters	'писем'
26 00 22	0000000001011	08 10 23	between	'между'
20 12 28	1000011011101	16 01 03	president	'президентом'
11 30 31	0000000000000	00 00 02	Johnson	
19 09 02	0000000000000	00 04 30	and	'и'
06 04 31	1000011010010	00 00 01	Hochiminh	
19 24 00	0000000000000	00 03 06	,	(зпт)
28 02 28	0000000001000	01 07 14	recently	'недавно'
26 17 02	01010011010111	10 20 16	disclosed	'раскрытым'
26 00 24	0000000001010	00 00 01	by	
11 30 31	0000000000000	00 00 02	Hanoi	
12 24 00	0000000000000	00 03 06	,	(зпт)
26 13 19	00010000010100	29 15 19	helps	'помогает'
18 00 00	0000000001000	00 00 01	to	
26 13 08	00010111000100	28 00 08	clear	'разрядить'

Продолжение

Цифровой эквивалент	Грамматическая информация	Номер перевода	Слово в тексте	Перевод
12 00 02	00000000000110	00 00 01	the	
21 31 29	00000010101100	00 00 01	air	'атмосферу'
26 01 07	00000000001011	11 20 02	on	'на'
19 03 02	0000000001000	01 10 11	certain	'несомненно'
10 06 14	00001010101101	14 15 15	questions	'вопросов'
19 24 14	00000000000000	00 00 01	,	(зпт)
00 00 09	00001010010010	06 07 20	which	'которые'
12 00 02	00000000000110	00 00 01	the	
11 30 31	00000000000000	00 00 02	Johnson	
21 31 18	00000100101101	25 04 10	administration	'администрации'
27 24 27	0001111100100	00 00 01	had	
18 04 03	00000000001000	02 05 11	deliberately	'умышленно'
17 31 31	00000000000000	00 00 02	befogged	
19 24 16	00000000000000	00 03 26	.	(точка)
19 09 02	00000000000000	00 04 30	and	'и'
26 01 14	00000000001011	11 20 00	in	'в'
26 02 12	00010110010101	09 31 06	doing	'делает'
28 05 05	00000000001000	02 01 29	so	'таким образом'
19 24 00	00000000000000	00 03 06	,	(зпт)
20 00 15	00000010010010	09 03 05	it	'этот'
27 02 24	00010110010101	10 20 16	reveals	'раскрывает'
19 08 30	00000000001000	00 28 14	,	'еще раз'
12 00 02	010100000000110	00 00 01	the	
25 17 17	00000010110110	13 03 19	utter	'полному'
09 18 18	00000010111100	00 00 01	hypocrisy	
26 00 00	00000000001010	00 00 01	of	
10 05 20	00001011001101	16 14 18	protestations	'протесты'
26 00 00	00000000001010	00 00 01	of	
11 30 31	00000000000000	00 00 02	Johnson	
19 08 30	01000000000110	00 00 01	,	
19 08 30	00000000000110	00 00 01	of	
26 00 00	00000000001010	00 00 01	his	'его'
12 00 13	00000100100113	07 15 27	willingness	'готовности'
10 24 17	00000100101100	25 13 13	to	
18 00 00	00000000001000	00 00 01	negotiate	'вести переговоры'
26 27 28	00010111000100	23 14 07	and	'и'
19 09 02	00000000000000	00 04 30	desire	'желания'
09 06 16	00000101101102	05 10 16	for	
26 00 31	00000000001010	00 00 01	peace	'мира'
10 00 16	00000010101101	15 18 23	,	(точка)
19 24 16	00000000001000	00 03 26	.	

Перевод текста № 1

Обмен писем между президентом Johnson и Hochiminh, недавно раскрытым Ханой помогает разрядить атмосферу на несомненно вопросов которые Johnson администрация умышленно befogged и в делает таким образом, этот раскрывает еще раз полностью hypocrisy протесты Johnson его готовности вести переговоры и его желания мира.

Год тому назад, Johnson громко провозглашала его готовность прекратить бомбардировку, если единственный Ханой указал его готовность вести переговоры. Этот был их отказом сделать таким образом, он настоял, который был

единственным roadblock de=расширение войны. Но тогда правительство Северного Вьетнама продолжило действовать проявить симптомы ясных указаний что никакой такой roadblock не существует. Эти указания присоединились через посредство различных каналов.

Пояснения. 1) В массиве информационных ячеек многозначные слова имеют цифровой эквиваленты в диапазоне 20 00 00 — 28 31 31. 2) Массив информационных ячеек приводится только для первого абзаца текста. 3) Часть номера перевода записывается в грамматических разрядах.

Текст № 2

The French people will go to the polls next Sunday in a general election that is likely to be historical.

After nine years of gaullist rule that has maintained a regime of personal power and has ensured the unhindered growth of the power of the big monopolies the coming elections may well mark a new democratic beginning.

Over this period although industrial production has grown by six zero per cent and the productivity of labor by five zero percent, there are still over four million salaried employees who receive less than a living wage.

The purchasing power of the majority of the people has been cut by the high prices and the increase in taxes both direct and indirect. In spite of the fact that there is a rising unemployment figure France continues to have the longest working-week in Europe.

For the first time since the victory of the popular front before the war and the democratic elections immediately after the war when all the anti-fascist parties had gone to the electorate together, are the French Communist Party and the other democratic and left parties fighting the elections together on the basis of an election alliance.

Информационные ячейки к тексту № 2

Цифровой эквивалент	Грамматическая информация	Номер перевода	Слово в тексте	Перевод
19 24 16	0000000010000	00 03 26		
12 00 02	00000000000110	00 00 01	the	
24 01 07	00000010010110	08 26 15	french	'французский'
22 24 11	00000010010011	15 21 08	people	'народ'
16 02 10	00010111140100	00 00 01	will	
26 22 27	00010111110100	19 14 02	go	'направится в'
26 01 19	00000000001010	29 06 06	to	'к'
12 00 02	00000000000110	00 00 01	the	
10 02 20	00001010111101	20 24 21	polls	'избирательным участкам'
20 02 02	00000110111102	05 03 15	next Sunday	'будущему'
26 01 14	00000000001011	11 20 02	in	'воскресенью'
12 00 01	00000000000110	00 00 01	a	'на'
25 07 21	00010111001110	06 30 25	general	'общих'
22 09 09	00010111011102	11 18 11	election	'выборах'
19 24 14	00000000010000	00 00 01	,	
20 00 00	00000010010010	09 03 05	that	'этот'
26 02 15	00010111001010	00 21 06	is	'должен'
18 09 31	00000000001000	00 15 30	likely	'вероятно'
18 00 00	00000000001000	00 00 01	to	
26 02 10	00010111000100	19 25 08	be	'явится'
12 19 19	00000011010100	09 05 13	historical	'историческим'
19 24 16	00000000010000	00 03 26	.	(тчк)
19 24 17	00000000001010	01 21 04	after	'после'

Цифровой эквивалент	Грамматическая информация	Номер перевода	Слово в тексте	Перевод
00 02 12	00000010010111	14 21 23	nine	'9'
20 02 08	00001010101101	14 19 00	years	'годов'
26 00 00	00000000001010	00 00 01	of	
11 30 31	00000000000000	00 00 02	gaullist	
10 15 11	00000110101102	06 10 31	rule	'правила'
19 24 14	00000000010000	00 00 01	,	(энт)
20 00 00	0000010010010	09 03 05	that	'этот'
26 02 05	00010111100100	00 00 01	has	
26 26 26	00010111100101	10 23 29	maintained	'сохранил'

Перевод текста № 2

Французский народ направится в избирательным участкам будущему воскресенью на общих выборах этот должен вероятно явиться историческим. После 9 годов Gaullist правила этот сохранил режиму личную власть и обеспечил не мешавшего увеличения власти больших монополий будущих выборов может вполне отметить новое демократическое начало.

В течение этого периода несмотря на то, что промышленное производство росло 60 процентов и производительность труда 50 процентов, будут все еще свыше 4 миллионов плаченных жалованье служащих кто получает менее чем жизненная заработная плата.

Закупающая власть большинства народа была урезана высокими ценами и увеличением налогов как прямых и косвенных. Несмотря на то обстоятельство что будет растущей безработицы цифры Франция продолжается иметь длинная рабочая-неделя в Европе.

Впервые с победой Народного фронта перед войной и демократическими выборами немедленно после войны, когда все антифашистские партии шли к избирательно вместе, будут Французской Коммунистической партии и других демократических и левых партий боевых выборов вместе союза выборы.

ЛИТЕРАТУРА

1. Алексеев П. М. Частотный словарь английского подъязыка электроники. Стр. Л., 1968, стр. 151—161.
2. Алексеев П. М. Лексическая и морфологическая статистика английского подъязыка электроники. Там же, стр. 120—131.
3. Амосова Н. Н. О синтаксическом контексте. Лексикографический сб. АН СССР, вып. V, 1962, стр. 36—45.
4. Апресян Ю. Д. Дистрибутивный анализ значений и структурные семантические поля. Там же, стр. 52—72.
5. Бельская И. К. О принципах построения словаря для Машинного перевода. ВЯ, 1959, № 3, стр. 89—94.
6. Головин В. Н. Введение в языкознание. М., 1966.
7. Ингве В. Синтаксис и проблема многозначности. В сб.: Машинный перевод, М., 1957, стр. 280—304.
8. Марчук Ю. Н. Определение роли существительного и глагола в английском тексте при грамматическом анализе для машинного перевода. НТИ, ВИНТИ АН СССР, 1965, № 3, стр. 36—40.
9. Марчук Ю. Н. О соотношении лексической и синтаксической информации при определении лексических значений многозначного слова. Труды III Всесоюзной конференции по информационно-поисковым системам и семантическим проблемам автоматической обработки текстов, т. II, ВИНТИ АН СССР, 1967, стр. 451—459.

10. М е л ь ч у к И. А. О терминах «устойчивость» и «идиоматичность». ВЯ, 1960, № 4, стр. 73—80.
11. М о т о р и н Ю. А., М а р ч у к Ю. Н. Реализация автоматического перевода на современных серийных ЭВМ общего назначения. Вопросы радиоэлектроники, серия электронно-вычислительной техники, вып. 7, 1970, стр. 20—29.
12. С о с н о в с к а я В. В. Опыт установления лексических классов глаголов по диагностирующим синтаксическим конструкциям. В сб.: Проблемы формализации семантики языка, I Московский ГПИ иностранных языков, 1964, стр. 143—147.
13. К а р л а н А. An Experimental Study of Ambiguity and Context. Mechanical Translation, vol. 2, № 2, 1955, pp. 39—45.
14. K a t z J., F o d o r J. The Structure of a Semantic Theory. In: The Structure of Language. Prentice-Hall (New Jersey), 1965, pp. 479—519.
15. K o u t s o u d a s A. Defining Linear Context to Resolve Lexical Ambiguity. Language and Speech, 1959, № 2, pp. 241—235.
16. M c C o n l o g u e K., S i m m o n s R. F. Analysing English Syntax with a Pattern-Learning Parser. Communications of the Association for Computing Machinery, 1965, nov., vol. 8, № 1, pp. 687—698.
17. L u k j a n o w a A. Semantic Classification. Proceeding of the National Symposium on Machine Translation. Englewood Cliffs, New York, 1961, pp. 394—397.
18. T e s n i è r e L. Élément de syntaxe structurale. Paris, 1959, p. 283.

В. Я. Шабес, В. В. Рожков

УСТРАНЕНИЕ ПРЕДЛОЖНО-ПАДЕЖНОЙ МНОГОЗНАЧНОСТИ НА БАЗЕ ТЕЗАУРУСА

Наряду с методами исследования окружающего контекста (см. здесь статьи Ю. Н. Марчука и Л. И. Тривис), а также с иконическим подходом, представленным в статье К. Б. Бектаева, Л. И. Белоцерковской и других, для различения многозначности при автоматической переработке текста возможно использование моделей, описывающих определенные микроучастки семантического пространства.

При таком подходе некоторое синтаксическое явление можно представить в виде знака со специфическим дескриптом, развернутым в синтагматике на отдельные сегменты, но обладающим семантической цельностью на денотативном уровне. Элементы синтаксического знака в большинстве случаев асимметричны. Именно это обстоятельство приводит некоторых исследователей к пессимистическим выводам относительно возможности алгоритмического устранения синтаксической неоднозначности.¹

Рассматривая синтаксическое явление как знак, мы тем самым хотим подчеркнуть возможность его алгоритмического описания как на формально-грамматическом, так и на содержательном уровнях. Учитывая специфичность и многофункциональность данного знака, более рациональным представляется описание некоторого участка семантического пространства с последующей экспликацией полученных единиц на уровне формы.

Описываемым множеством в данном случае явится набор семантико-синтаксических инвариантов, соответствующих реляционным значениям. Вслед за С. К. Шаумяном будем называть такие инвариантные значения «реляторами» [6, стр. 151].

В лингвистической литературе известны попытки системного описания грамматических значений, причем обычно исследуется

¹ С. Куно и А. Эггингер, например, считают, что проблема синтаксической неоднозначности алгоритмически неразрешима даже для специальных контекстно-свободных языков [2, стр. 125].

какой-либо один формально-грамматический класс, хотя с содержательной точки зрения такое ограничение не всегда оправдано (см., например, [9, 8]).

Выявление границ, общей схемы исследуемого семантического пространства, обнаружение собственно реляторов (R) предполагает использование некоторого приема, который может носить как индуктивный, так и дедуктивный характер. Одним из индуктивных приемов обнаружения искомым универсалий является анализ знаков иной, более абстрактной (мощной) языковой категории, семантически объединяющей изучаемые частные категории.

В качестве такой категории в нашем случае может служить набор вопросительных словоформ. С семантической точки зрения вопросительные словоформы являются особым разрядом субститутов, имеющих соответствия во всех неслужебных частях речи данного языка и не сводимых к какой-либо одной части речи. Для анализа были привлечены вопросительные словоформы 45 языков. При отборе учитывались как типологические, так и генеалогические характеристики.

Вопросительные словоформы при попарном сравнении дают возможность строить корреляции, выявляя их парадигматические связи, причем одновременно эксплицируется не только тип связи, но и «содержание» ее, или дифференциальный семантический признак (семантический множитель).

В работах по структурной семантике можно найти описания типов парадигматических отношений на уровне смыслов лексических единиц. Так, В. А. Москович [3, стр. 108] выделяет отношения синонимии, антонимии, отношение «род»—«вид», отношение «род»—«вид» с соподчинением между видами, отношения типа «причина»—«следствие», «часть»—«целое», «уточняющая связь», а также связь по функциональному сходству.

У Дж. Лайонза [10, стр. 72] классификация типов парадигматических связей отличается от вышеизложенной. Он отмечает следующие типы отношений: различие, несовместимость, антонимия, синонимия, гипонимия, конверсивность, следствие, деривация.

Если же попытаться из приводимых обоими исследователями типов парадигматических связей выделить наиболее абстрактные признаки и систематизировать их,² то полученная классификация может быть использована для частичной формализации логико-интуитивной деятельности лингвиста при обнаружении типа и содержания парадигматической связи (r), причем анализ производится на уровне дифференциального семантического признака.

² В. А. Виноградов предлагает следующее дефиниционное утверждение: «Совокупность объектов образует систему, если набор признаков, постулируемых для описания этих объектов, образует систему» [4, стр. 175].

На рис. 1 представлена возможная схема такой классификационной модели. На входе имеется пара объектов A и B (в нашем случае это корреляция вопросительных словоформ, связанных некоторым парадигматическим отношением r). На первом шаге классификация предполагает

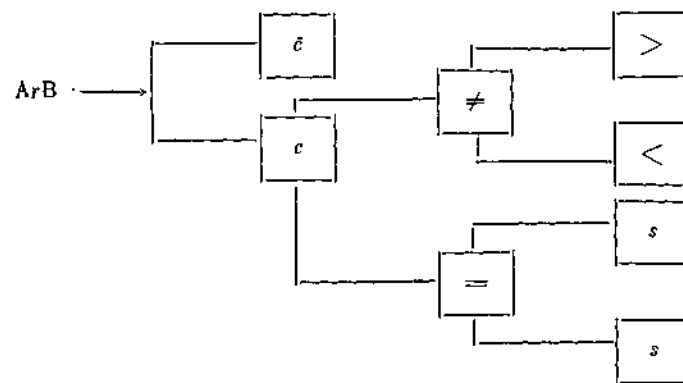


Рис. 1.

установление семантической несравнимости (ϵ), например, «сколько» и «где» несравнимы по заданному (выявленному ранее) признаку «локальность», или сравнимости (c) с установлением по корреляции содержания связи (те же словоформы сравнимы по признаку «воспросительность», так как имеется множество коррелирующих пар в разряде вопросительных словоформ).

Второй шаг требует дифференциации типа связи. Равенство ($=$) предполагает, что обнаруженный признак «равновелик» в объекте A и объекте B . В этом случае возможна синонимия (s) или антонимия (z). Неравенство (\neq) устанавливает иерархическое подчинение по степени абстрактности найденного признака в A и B . Имеются в виду случаи типа: «часть» < «целое», «вид» < «род», «где» > «над чем», «из чего» < «откуда» и т. д.³ Другими словами, здесь определяется общее и частное на уровне дифференциального семантического признака.

В табл. 1 приводятся примеры типов парадигматической связи с указанием содержания на материале вопросительных словоформ.

Обработав подобным образом вопросительные словоформы, мы определили набор основных дифференциальных семантических признаков, характеризующих как общую схему семантического пространства, так и отдельные его участки [5, стр. 84—89]. Одновременно получены данные для построения парадигматической модели, формально выраженной в виде класси-

³ Русские аналитические конструкции с предлогом используются в статье лишь в том случае, когда в некотором языке есть моноформа-синоним.

Таблица 1

A	Тип _r	B	Содержание связи
Отчего?	$c=s$	зачем?	«каузальность»
Откуда?	$c=s$	зачем?	«финальность»
Откуда?	$c=s$	куда?	«финальность»
На что?	$c \neq <$	куда?	«локальность»
Во что?	$c=s$	в каком направлении?	«предельность»
Какой?	δ	как?	«локальность»
Какой?	$c=s$	как?	«предметность»
Какой?	$c=s$	как?	«атрибутивность»
На чем?	$c=s$	куда?	«динамика»

фикационного дерева — иерархии, или тезауруса. Тезаурус построен на бинарном принципе, т. е. каждая пара семантических признаков, обозначенная на ребрах графа, связана отношением $c=s$ между собой и отношением $c \neq < c$ признаком, отмеченным на вершине графа.

Тезаурус локальных реляторов представлен на рис. 2. Дерево состоит из четырех кустов, причем каждый последующий куст является продолжением любого из окончаний предыдущего. Выходы из окончаний куста обозначены стрелками. Символы на узлах дерева обозначают следующие дифференциальные семантические признаки:

- 1) +A «локальность»;
- 2) +B/-B «точка»/«сложный объект»;
- 3) +C/-C «между»/«немежду»;
- 4) +D/-D «два»/«много»;
- 5) +E/-E «внутри»/«снаружи»;
- 6) +F/-F «поле»/«объем»;
- 7) +G/-G «дифференцировано в пространстве»/«недифференцировано в пространстве»;
- 8) +H/-H «горизонталь»/«вертикаль»;
- 9) +Z/-Z «координата I»/«координата II»;
- 10) +J/-J «А-конец координаты»/«Б-конец координаты»;
- 11) +K/-K «дифференцировано по расстоянию»/«недифференцировано по расстоянию»;
- 12) +L/-L «неконтакт»/«контакт»;
- 13) +N/-N «фиксированность»/«нефиксированность»;
- 14) +Q/-Q «исчисляемость»/«неисчисляемость»;
- 15) +U/-U «невязную»/«вязную»;
- 16) +R/-R «далеко»/«близко»;
- 17) +M/-M «статика»/«динамика»;
- 18) +P/-P «направленность»/«ненаправленность»;
- 19) +T/-T «неопределенность»/«определенность»;
- 20) +V/-V «транзитность»/«нетранзитность»;
- 21) +W/-W «финальность»/«инициальность»;
- 22) +O/-O «реальная локальность»/«псевдолокальность»;
- 23) +S/-S «активность»/«предметность»;
- 24) +X «конец ветви».

Дерево порождает 2352 комбинации семантических признаков, или инвариантных (речевых) синтаксических значений локальной связи. Релятором считается цепочка (кортеж) символов от +A до +X. Небольшое количество «сухих» ветвей устраняется специальным блоком алгоритма.

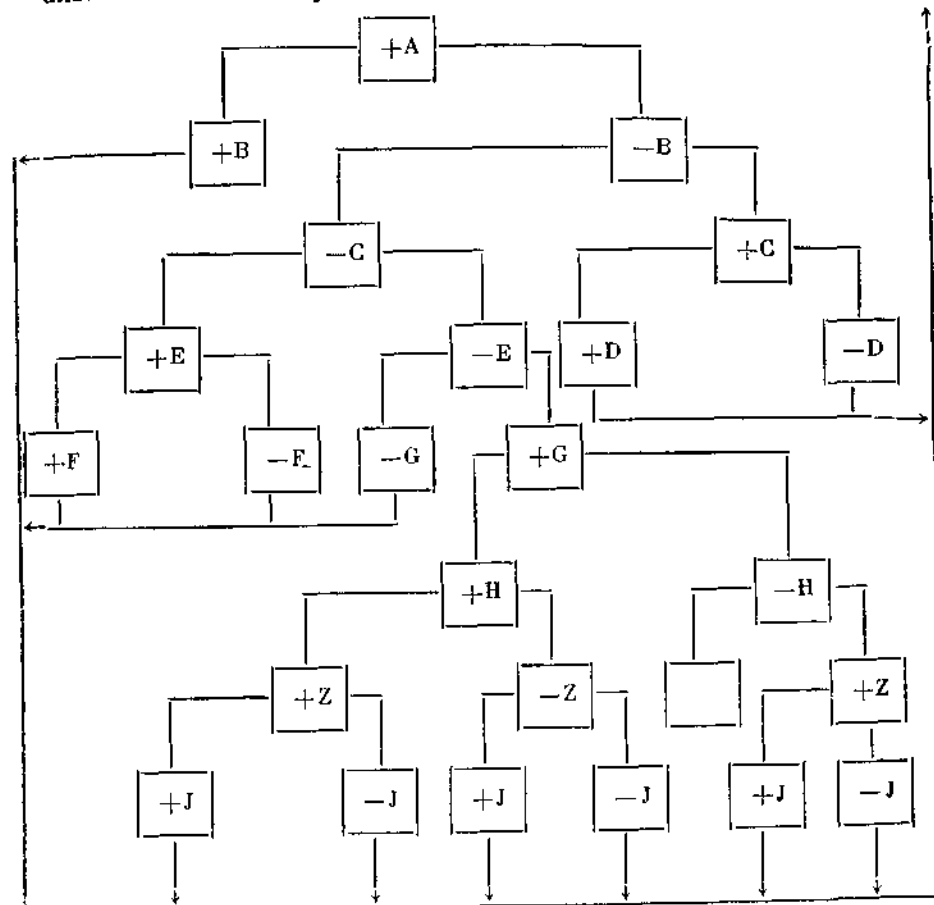


Рис. 2.

Чтобы описать в терминах тезауруса некоторый речевой сегмент, значение этого сегмента сопоставляется с тезаурусом. В ходе этого сравнения последовательно ищутся необходимые дифференциальные признаки начиная с +A. Например, в ходе такого анализа английский сегмент was coupled to it приобретает вид [+A-B-C-E+K-L-M+P-T+W+O+S], а сегмент arrival from England получит запись [+A-B-C+E+F-K-M+P-T-W+O-S].

Эта процедура была реализована на 5000 примеров, иллюстрирующих локальные связи. Примеры отбирались из английских текстов различной тематики. Опорной формой считался предлог,

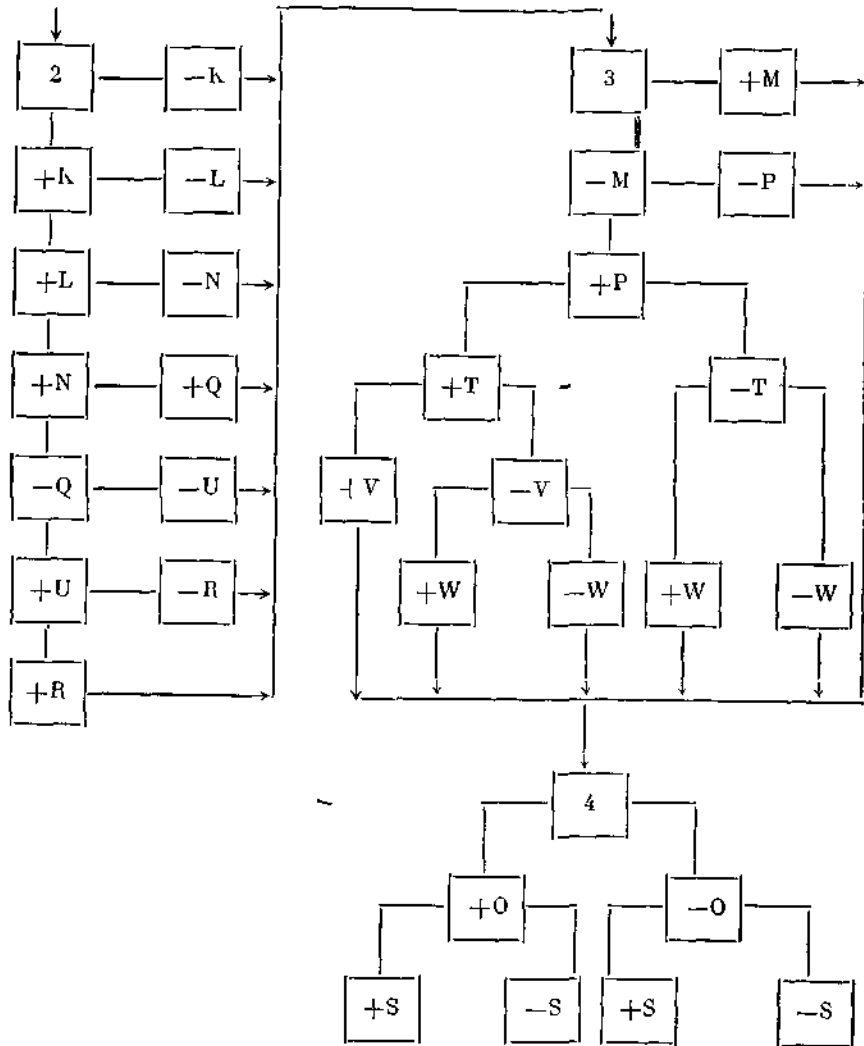


Рис. 2 (продолжение).

управляющая и управляемая словоформы рассматривались в качестве компонентов десигната синтаксического знака.⁴ В таком виде синтаксический знак, естественно, симметричен.

⁴ В русском варианте опорными формами, по-видимому, следует считать морфемы и падежа и предлога; ср. *под чем* и *под(о) что*.

Полученные реляторы могут быть сгруппированы по признаку любого из компонентов десигната синтаксического знака. При этом в подавляющем большинстве случаев компоненты десигната описываются более чем одним релятором, или, другими словами, компоненты знака, также являющиеся знаками (k -знаки), асимметричны. Так, основными реляторами k -знака *from* являются:

- 1) $+A+B+K+L+N+Q+M+O+S$
 - 2) $+A+B+K+L+N+Q+M+O-S$
 - 3) $+A+B+K+L+N+Q-M-P+O+S$
 - 4) $+A+B+K+L+N+Q-M-P+O-S$
 - 5) $+A+B+K+L+N-Q+U+R+M+O+S$
 - 6) $+A+B+K+L+N-Q+U+R+M+O-S$
 - 7) $+A+B-K-M+P+T-V-W+O+S$
 - 8) $+A+B-K-M+P+T-V-W+O-S$
 - 9) $+A+B+K+L+N-Q+U-R+M+O+S$
 - 10) $+A+B+K+L+N-Q+U-R+M+O-S$
 - 11) $+A+B-K-M+P-T-W+O+S$
 - 12) $+A+B-K-M+P-T-W+O-S$
 - 13) $+A-B-C+E-F-K-M+P-T-W+O+S$
 - 14) $+A-B-C+E-F-K-M+P-T-W+O-S$
 - 15) $+A-B-C+E+F-K-M+P-T-W+O+S$
 - 16) $+A-B-C+E+F-K-M+P-T-W+O-S$
 - 17) $+A-B-C-E+G-H+Z+J+K-L-M+P-T-W+O+S$
 - 18) $+A-B-C-E+G-H+Z+J+K-L-M+P-T-W+O-S$
 - 19) $+A-B-C-E-G+K-L-M+P-T-W+O+S$
 - 20) $+A-B-C-E-G+K-L-M+P-T-W+O-S$
 - 21) $+A+B-K-M+P+T-V-W+O+S$
 - 22) $+A+B-K-M+P+T-V-W+O-S$
- и т. д.

Рассматривая набор реляторов, соотносимый с предлогом *from*, можно убедиться в том, что некоторые дифференциальные признаки входят в состав всех реляторов данной группировки. Отсюда можно заключить, что именно эти признаки принадлежат *from* и составляют его собственную, регулярно воспроизводимую языковую потенцию. Поскольку группировки реляторов возможны по десигнату (вернее, по экспоненту) любого из k -знаков, ясно, что дифференциальные семантические признаки могут быть «рассортированы» между ними и каждому из k -знаков приписано его языковое значение.

Для пояснения метода получения языковых значений k -знаков из реляторов и обратного пути удобно воспользоваться теоретико-множественным аппаратом [7].

Тезаурус (T) является множеством с отношением строгого порядка $<$.

Множество T конечно и имеет максимальный элемент $+A$, причем если некоторый элемент множества $x < y$ и $x < z$, то y и z сравнимы (равны и антонимичны), т. е. отношение на T является древесным порядком. Каждая ветвь тезауруса представляет собой кортеж, или упорядоченный набор элементов (релятор). Обозначим кортежи $R_1, R_2, R_3, \dots, R_n$. Таким образом, $R_1 \subseteq T, R_2 \subseteq T, R_3 \subseteq T, \dots, R_n \subseteq T$.

Каждый кортеж R_x представляет собой инвариантное значение (денотат), реализуемый в тексте некоторым синтаксическим экспонентом, или, при определенных лингвистических допущениях, $R_x = F_x$, т. е. $\langle R_x F_x \rangle$ есть симметричный синтаксический знак. В свою очередь экспонент F_x является сложным образованием и может состоять из множества элементов: $K_1, K_2, K_3, \dots, K_n$, т. е. $K_1 \in F_x; K_2 \in F_x; K_3 \in F_x, \dots, K_n \in F_x$.

Если мы сгруппируем кортежи-денотаты по признаку одного элемента K_i , то принципиально возможна асимметрия, т. е. $K_x \equiv \equiv R_1^x; R_2^x; R_3^x; \dots, R_n^x$.

Для устранения асимметрии и нахождения языкового значения K_i , которое обозначим L_{K_i} , возможно использовать операцию объединения множества R . Результат объединения обозначим Y_{K_i} . Тогда:

$$Y_{K_i} = R_1^x \cup R_2^x \cup R_3^x \cup \dots \cup R_n^x.$$

Но полученное множество содержит пары элементов с противоположным знаком (антонимичные элементы), которые не дают никакой информации о содержании исследуемого K_i . Для устранения антонимичных элементов в полученном множестве достаточно получить алгебраическую сумму элементов (a_i) указанного множества. Тогда языковое значение некоторого K_i будет равно:

$$L_{K_i} = \sum_{i=1}^n (a_i).$$

Знание языковых $L_{K_1}, L_{K_2}, L_{K_3}, \dots, L_{K_n}$ дает возможность найти речевой кортеж R_x путем пересечения объединения известных L_K со множеством кортежей. Поскольку все возможные кортежи порождаются тезаурусом, возможно пересечение непосредственно с тезаурусом:

$$R_x = T \cap (L_{K_1} \cup L_{K_2} \cup L_{K_3} \cup \dots \cup L_{K_n}).$$

Таким образом:

$$R_x = f(T, L_{K_1}, L_{K_2}, L_{K_3}, \dots, L_{K_n}).$$

Произведя несложные вычисления с реляторами, приведенными на стр. 237, мы можем отыскать языковое (локальное) значение предлога from:

$$Y_{K(\text{from})} = [+A + B - B - C + E - E + F - F + G - G - H + + Z + J + K - K + L - L + N + Q - Q + U + R - R + M - - M + P - P + T - T - V - W + O - O + S - S].$$

Следовательно:

$$L_{K(\text{from})} = [+A - C - H + Z + J + N + U - V - W].$$

Последнюю формулу можно прочесть так: «локальность», «не- между», «вертикаль», «координата I», «А-конец координаты» (комбинация $[-H+Z+J]$ обозначает «верх»), «фиксированность», «нетранзитивность», «инициальность». Данные семантические признаки дают возможность равновероятно предполагать любое из значений from, т. е. выражают все его потенциальные локальные значения.

Предположим, что нам известно выявленное аналогичным образом локальное синтаксическое значение k -знака England справа от k -знака from: $[-B + E + F - K + M + O - S]$, расшифровывающееся следующим образом: «сложный объект», «внутри», «поле», «недифференцировано по расстоянию», «статика», «реальная локальность», «предметность». Левый от from k -знак arrival имеет значение $[+B - K - M + P - T + O - S]$, т. е. «точка», «недифференцировано по расстоянию», «динамика», «направленность», «предельность», «реальная локальность», «предметность». Тогда искомый релятор может быть найден по тезаурусу (см. рис. 3).

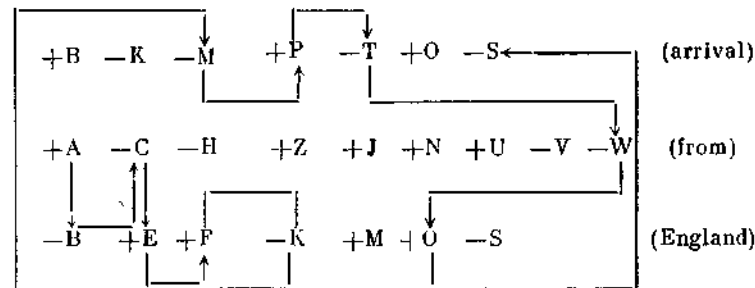


Рис. 3

На данном рисунке представлена схема синтеза тезаурсного релятора из языковых значений (L_K).

Необходимо отметить, что обращение к тому или иному L_K происходит в определенной последовательности в зависимости от позиции соответствующего K в F (т. е. в зависимости от положения по отношению к from как England, так и arrival).

В настоящее время на ЭВМ «Минск-22» осуществляется реализация программы различения многозначности предлогов, падежей и предложно-падежных конструкций на основе локального тезаурса.

Блок-схема алгоритма представлена на рис. 4. Поясним работу некоторых блоков, которые, как нам представляется, требуют более детального описания.

Б л о к 1. С перфоленты в оперативную память ЭВМ вводится массив текста объемом в 1000 словоформ. В конце массива записывается команда останова.

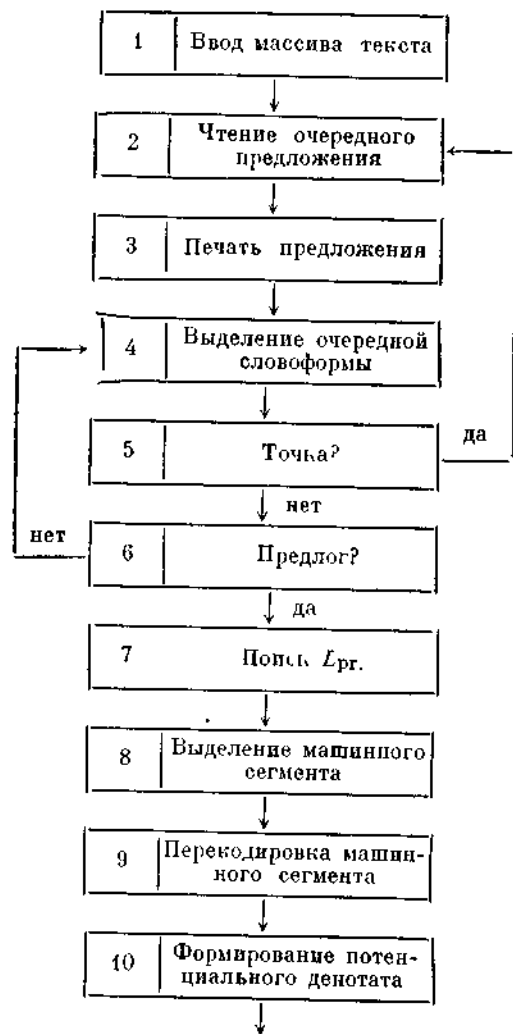


Рис 4.

Б л о к 2. Из массива выделяется очередное предложение с одновременной перекодировкой во внутренний код машины. После обработки последнего предложения происходит останов машины по команде, полученной из блока 1.

Б л о к 6. Производится поиск опорной словоформы путем сравнения каждой из словоформ со списком опорных k -форм (предлогов). Определяется номер предлога в списке.

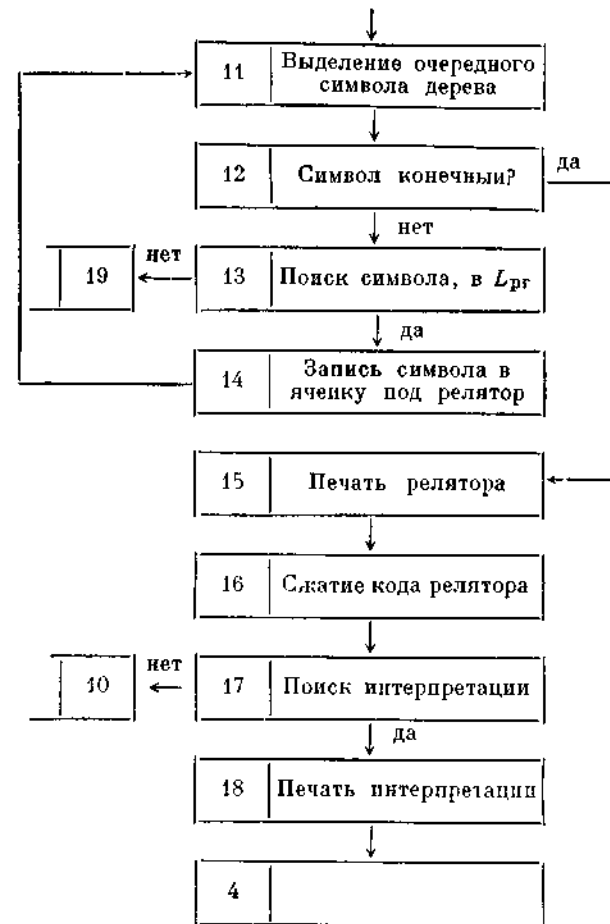


Рис. 4 (продолжение).

Б л о к 7. Определяется языковое значение опорной формы, в английском варианте — предлога (L_{pr}) по таблице, находящейся в машинной памяти.

Б л о к 8. Выделяется машинный сегмент, включающий в себя F_s , на глубину, определяемую особенностями данного реального языка. Для английского языка принята глубина сегмента в 5 словоформ влево и 5 вправо от опорной k -формы.

Б л о к 9. Сегмент перекодировается при помощи машинного (микромашинного) словаря; k -формам приписываются кодовые эквиваленты (номера L_k).

Б л о к 10. Данный блок предназначен для формирования триад из элементов L_k , которые условно можно считать потенци-

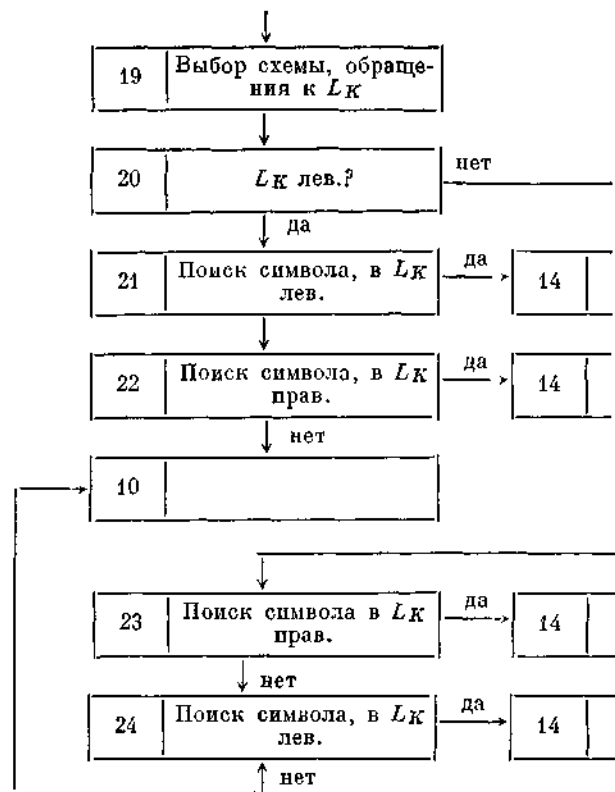


Рис. 4 (продолжение).

альными компонентами R_c (денотата синтаксического знака). Возможность синтеза релятора из каждой сформированной триады проверяется по тезаурусу (см. блоки 11 — 14, 19 — 24). Таким образом, тезаурус является своеобразным фильтром, отбрасывающим недостоверные (нелогичные) комбинации L_k . При каждом обращении к блоку 10 формируется новая триада, отличающаяся от предыдущей только одним левым или правым L_k .

Блоки 11, 12, 13, 14, 19, 20, 21, 22, 23, 24 составляют в совокупности механизм синтеза денотата (релятора). Схема работы этих блоков представлена на рис. 4.

Алгоритм предусматривает обработку неиндексированного текста с применением машинного словаря. Поскольку работы по созданию семантико-синтаксических кодов для машинного словаря, разработанного в группе «Статистика речи», не завершены, эксперименты проводятся как на базе микромашиного словаря (500 словоформ), так и с предварительной индексацией неопорных k -форм. Результаты машинного эксперимента по построению релятора

на основе двух переменных (предлог и индексы неопорных L_k), проведенного на основе локального тезауруса, представлены на рис. 5.

Полученному по тезаурусу релятору может быть приписан эквивалент на заданном реальном языке (интерпретация релятора).

○	675	HE TOLD IT THE CONGRESS +A-B-C-E-F-K-M-O-S /CP/ (НАХОЖИ.) В . НА /СДП/
○	676	CONCLUSIONS DEDUCED FROM MEASUREMENTS +A-B-C-E-F-K-M-P-T-W-O-S /C/ (УДАЛЕИ.) ИЗ /СДП/ (ПОЛЕ)
○	677	TO CONFESS TO THEM +A-B-K-M-P-T-W-O-S /CP/ (ПРОИЗВОД.) /СДП/
○	678	A VISIT TO MOSCOW +A-B-C-E-F-K-M-P-T-W-O-S /C/ (ПРИЕХИ.) В . НА /СДП/
○	679	A BASE CAMP AT SHU CHI +A-B-C-E-F-T-W-O-S /C/ (НАХОЖИ.) В . НА /СДП/
○	680	WERE SOLD GIRD TO THE ENDS +A-B-C-E-G-K-L-M-P-T-W-O-S /CP/ (ПОМЕЧ.) НА /СДП/
○	681	THE HIGHWAY LEADING TO SAIGON +A-B-K-M-P-T-V-W-O-S /C/ (ПРИБЛИЖ.) ПО НАПРАВЛЕНИЮ К
○	682	IT LED TO THE STARS +A-B-K-M-P-T-V-W-O-S /CP/ (ПРИБЛИЖ.) ПО НАПРАВЛЕНИЮ К
○	683	THE ARTICLE FROM OUR CORRESPONDENT +A-B-K-M-P-T-W-O-S /C/ (УДАЛЕИ.) ОТ /СДП/
○	684	WERE POSITIONED ON DR NEAR IT +A-B-C-E-G-K-L-M-O-S /CP/ (НАХОЖИ.) НА /СДП/ (ПОВЕРХИ.) ****

Рис. 5. Машинный результат.

Так, например, возможно: [arrival from England] → [+A—B—C—E—F—K—M+P—T—W+O—S] → [N (движение) из N (поле) G], где N — сущ., а G — флексия род. пад. Можно предположить, что по заданному релятору возможен синтез необходимых конструкций по морфологическим правилам данного языка. Теоретически возможным является и целенаправленный поиск по ма-

шинному словарю в зависимости от заданной семантической сочетаемости некоторого K_i , а так же нахождение R_i по неполному F_i (на основе избыточной семантической информации, ср. рис. 3).

Думается, что семантико-синтаксические исследования на базе тезаурусов, моделирующих определенные микроучастки семантического пространства, могут представлять не только прикладной, но и теоретический интерес, так как позволяют моделировать на семантическом уровне с известной степенью достоверности как язык и речь, так и процесс речевой деятельности. Возможны также компаративно-семасиологические исследования. Использование метаязыка тезауруса позволяет сравнивать структуры языковых значений k -форм как в одном, так и в нескольких языках (см., например, табл. 2).

Таблица 2

Языки	Формы	Языковое значение L_K
Русский	над+тв. пад.	+A-E+G-H+Z+J+K+ +L+Q+U-R+T+V
Английский	above	+A-E+G-H+Z+J+K+ +L-N
»	over	+A+F+G+Z+L-N+T+V
Испанский	sobre	+A-E+G+H+Z+J+K- -N-W

Л И Т Е Р А Т У Р А

1. Виноградов В. А. Всегда ли система системна? В сб.: Система и уровни языка, М., 1969.
2. Куно С. и Эттингер А. Синтаксические структуры фраз и синтаксическая неоднозначность в английском языке. АП.
3. Москович В. А. Информационные языки. М., 1971.
4. Петровский Р. Г. Отраслевой вероятностный машинный перевод. СТ II.
5. Шабес В. Я. Тезаурус реляционных значений. В сб.: Принципы и методы лексико-грамматических исследований, Л., 1972.
6. Шаумян С. К. Моделирование синтаксических связей в аппликативной грамматике. В сб.: Инвариантные синтаксические значения и структура предложения, М., 1969.
7. Шрейдер Ю. А. Равенство, сходство, порядок. М., 1971.
8. Hjelmslev L. La catégorie des cas. Partie II. Acta Jutlandica, Aarhus, 1937, v. IX.
9. Kuryłowicz J. Le problème du classement des cas. Biuletyn polskiego towarzystwa językoznawczego, IX, Kraków, 1949.
10. Lyons J. Structural Semantics. Oxford, 1963.

М. Г. Зорев

МАШИННАЯ МОРФОЛОГИЯ И ЕЕ ПРИМЕНЕНИЕ
В НЕМЕЦКО-РУССКОМ АВТОМАТИЧЕСКОМ СЛОВАРЕ

§ 1. МАШИННАЯ И ТРАДИЦИОННАЯ МОРФОЛОГИЯ
НЕМЕЦКОГО ЯЗЫКА

Опыт моделирования лексики и морфологии некоторых пидевропейских языков для нужд инженерной лингвистики показал, что построение двуязычных автоматических словарей (АС) на базе лексем-основ или лексем — исходных форм слова, с одной стороны, и словоформ, с другой, имеет свои неудобства. В первом случае мы вынуждены строить сложную морфологическую программу, связанную с многократным обращением ко внешней памяти на магнитных лентах, что быстро приводит к сбоям последних в ЭВМ, не имеющей промежуточной памяти в виде магнитных барабанов и дисков. Во втором случае при малой морфологической программе мы имеем очень громоздкий словарь (особенно для флективных языков).

В связи с этим при машинной переработке текстов, написанных на флективных языках, возникает необходимость искать компромиссное решение, состоящее в том, чтобы не очень сложная морфологическая программа сочеталась с не очень громоздким автоматическим словарем. Поискам такого компромиссного решения на материале немецкого языка и была посвящена настоящая работа.

Существо предлагаемого решения состоит в создании новой машинной морфологии немецкого языка, опирающейся на нетрадиционное выделение так называемых машинных флексий и создание словаря машинных основ (квазиснов).

Возникает вопрос, какие преимущества представляет автоматический словарь квазиснов (АСК) по сравнению с существующими автоматическими словарями словоформ и основ?

Как показывают данные в табл. 1, преобразование частотного словаря словоформ (ЧСС) в частотный словарь квазиснов (ЧСК)

сокращает объем списка на 4120 словарных единиц, или на 20.25%, а преобразование ЧСС в частотный словарь основ (ЧСО) дает сокращение на 5827 словарных единиц, или на 28.56%. Одновременно ЧСО и ЧСК показывают более быстрый рост покрываемости: первые 500 единиц ЧСО покрывают на 22% больше текста, чем 500 первых единиц ЧСС, для ЧСК эта покрываемость выше на 16%. Первые 1000 лексических единиц в ЧСО покрывают на 18%, а в ЧСК на 15% больше текста, чем первые 1000 словоформ ЧСС (покрываемость ЧС принимается за 100%).

Таблица 1

№№	Словоформы		Основы		Квазисносы	
	количество Σ_1	% покрываемости	количество Σ_2	% покрываемости	количество Σ_3	% покрываемости
1	100	48.3	100	56.8	100	54.1
2	500	61.7	500	75.1	500	71.8
33	1 000	69.7	1 000	82.8	1 000	80.5
44	20 405	94.8	14 578	97.1	16 285	96.3

Все это дает нам право отказаться от использования на входе нашего АС словоформ и обратиться к использованию либо основ, либо квазисносов. Чтобы решить, какая из этих лексических единиц должна быть использована в АС, необходимо учитывать объем той морфологической программы, которая будет придана тому или иному типу АС.

Для этого дадим оценку общего числа дополнительных операций по морфологическому анализу текстовых словоупотреблений при использовании входного словника АС, построенного на традиционных основах (будем именовать его сокращенно АСО), и словника, использующего квазисносы (АСК).

Обращаем внимание читателя на то, что мы ограничиваем максимальное число отсечений в правой части словоформы двумя буквами на том основании, что большинство немецких окончаний содержит одну или две буквы. Одновременно малое число отсечений сильно сокращает время обращения к словарю квазисносов при машинном морфологическом анализе. Наконец, ограничение, по которому машинные флексии не могут превышать двух букв, хотя и несколько увеличивает количество квазисносов по отношению к основам, но зато сильно сокращает число разных окончаний. Так, например, традиционная парадигма глагола folgen включает 10 различных окончаний (в том числе до четырехбуквенных), сочетающихся с основой folg-, ср. наст. вр. изъявительного наклонения: -e, -st, -t, -en, претеритум изъявительного наклонения: -te, -test, -ten, -tet, наст. вр. сослагательного наклонения: -est, -et. Преобразуя традиционные окончания в одно- и двухбуквенные, мы получаем только 5 различных машинных флексий, ср.

folg-e, -st, -t, -en — для наст. вр. изъявительного наклонения;
 folge-n — для наст. вр. сослагательного наклонения;
 folgte- (повторение вышеуказанных машинных флексий) — для претеритума изъявительного и сослагательного наклонения.

Преимущество квазисносы по сравнению с основой заключается в существенном сокращении количества отсечений флексий. По сравнению со словоформой это преимущество состоит в существенном увеличении эффективности: для покрытия парадигмы спряжения в указанных формах требуются 8 различных словоформ и всего 3 квазисносы.

Дополнительные операции связаны с преобразованием входного словника, построенного на словоформах, во входной словник основ (АСО), это показано в табл. 2. Для того чтобы разъяснить построение табл. 2, рассмотрим расчет в первой подзоне. Первая подзона включает три изменяемые словоформы:

eine (F = 2187)
 werden (F = 1838)
 wird (F = 1804)

При использовании АСО у eine будет отсекается одна буква, у werden — две конечных буквы с тем, чтобы привести эти словоформы к их основам ein- и werd-. Словоупотребления wird отсечениям окончаний подвергаться не будут, зато здесь нужно будет ликвидировать аблаут (wird-, werd-). Иными словами, сумма отсечений окончаний у изменяемых словоупотреблений первой подзоны будет равна $2187 + 1838 \cdot 2 = 5863$. Если к этому прибавить 1.804 случаев аблаута, то общее число морфологических операций по опознанию трех указанных словоформ в массиве из 200 тысяч словоупотреблений будет равно 7667.

При использовании квазисносов eine-, werde-, wird- число морфологических операций сократится до 1838. Морфологические операции в АСК включают только отсечения конечных букв; устранений префиксов, инфиксов и чередований в словоупотреблениях здесь не требуется.

Как показывают итоговые данные табл. 2 (см. столбцы 15 и 19, или 16 и 20), использование АСК значительно сокращает число машинных операций по морфологическому анализу и опознанию текстовых словоупотреблений. Коэффициент словоизменения для словаря основ в три раза выше коэффициента словоизменения для словаря квазисносов (1.03 : 0.34). Хотя все наши расчеты относятся к первым шестистам словоформам ЧСС, выводы могут быть распространены и на остальную часть частотного списка, поскольку большинство редких слов представляет собой изменяемые словоформы.

Использование АСК освобождает нас от необходимости введения в систему немецко-русского МП большого числа подпрограмм морфологического анализа отсечения префиксов, а также обработки

Таблица 2

№М	Зона в ЧСС	Количество изменений в ЧСС	Общее количество изменений в ЧСС	Традиционные основы				Квазиосновы											
				суммарная частота грамматической аффиксации и альтернативная словоупотребления		чередование в основе		суммарная частота традиционных флексий		общее количество морфологических операций при использовании АСО		суммарная частота традиционных флексий в словоупотреблении		общее количество морфологических операций при использовании АСО					
				префиксы	суффиксы	умлаут	аблаут	однобуквенные	двухбуквенные	трехбуквенные	общее количество	флексии	флексии	флексии	флексии				
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1-20	3	5829	—	—	—	—	1804	—	2187	1838	—	5863	7667	—	1838	—	1838	—
2	21-50	6	5288	—	—	—	825	—	—	1373	3803	—	8979	9804	—	3803	—	3803	—
3	51-100	11	4502	—	—	—	337	515	—	1530	1359	—	4248	5100	—	1200	426	2052	—
4	101-200	38	8153	—	—	—	511	135	—	2511	1949	—	6409	7055	—	1269	231	1731	—
5	201-300	33	6175	—	155	—	77	280	—	1174	1320	—	3814	4326	—	1460	296	2052	—
6	301-400	31	4676	276	417	276	—	—	48	1004	581	—	2166	2835	—	823	—	823	—
7	401-500	30	3524	97	192	338	—	112	79	568	478	—	1524	2341	—	472	49	570	—
8	501-600	37	2493	41	121	159	159	79	—	654	742	39	2255	2655	—	618	230	1078	—
	Всего:		40700	414	464	735	1084	3780	48	11001	12070	39	35258	41783	1.03	11483	1232	13947	0.34

Примечание. При выделении традиционных основ допускаются в основном следующие операции: а) отбрасывание префиксов; б) отбрасывание инфиксов; в) «устранение» умлаута, аблаута, чередований согласных и грамматических суффиксов. Неотделяемые и отделяемые приставки, а также словообразовательные суффиксы считаются частью основы. Суфлитивные формы считаются самостоятельными основами.

инфиксов и чередований основы. В исследованной нами зоне достоверных частот нашего ЧСС число таких подпрограмм достигло бы шести.

Из всего сказанного как будто следует, что при построении входного словника немецко-русского АС целесообразно пользоваться не традиционными, но машинными основами (квазиосновами).

Но не приведет ли это использование квазиоснов в АСК к сильному усложнению машинной грамматики — усложнению, которое может потребовать введение дополнительных подпрограмм, увеличить объем кодов? Чтобы ответить на этот вопрос, сравним статистико-грамматическую структуру немецких текстов по электронике со статистической картиной машинной грамматики этих текстов.

Таблица 3

№	Части речи	Число разных словоформ (i)			Число разных словоупотреблений (F)		
		Σi	%	δ	F	%	δ
1	Имена существительные, в том числе имена собственные	13.180	64.67	0.02	60 686	30.34	0.78
2	Глаголы, в том числе:						
	вспомогательные	3.832	18.77	0.03	27 750	13.88	0.01
	модальные	54	0.27	0.30	11 241	5.62	0.01
	причастие I	308	1.50	0.11	1166	0.58	0.05
	причастие II	1214	5.92	0.06	4771	2.38	0.04
3	Прилагательные	2536	12.42	0.04	15 320	7.66	0.01
4	Наречия	438	2.14	0.08	9399	4.70	0.02
5	Местоимения	117	0.57	0.18	9097	4.55	0.02
6	Числительные	76	0.37	0.23	1161	0.58	0.05
7	Предлоги	49	0.22	0.27	25 834	12.94	0.01
8	Союзы	41	0.20	0.30	12 380	6.19	0.01
9	Артикли	12	0.05	0.57	32 769	16.38	0.01
10	Частицы	10	0.04	0.62	2394	1.20	0.01
11	Сокращения	85	0.41	0.21	1556	0.78	0.05
12	Прочие	29	0.14	0.36	1604	0.80	0.05
	Итого:	20.405	100.00	—	200.000	100.00	—

В таблице: δ — относительная ошибка наблюдения, вычисляемая по формуле: $\delta = \frac{Z_p \sqrt{1-f}}{\sqrt{fN}}$, где f — частотность, а N — объем выборки.

§ 2. СТАТИСТИЧЕСКОЕ РАСПРЕДЕЛЕНИЕ ТРАДИЦИОННЫХ ГРАММАТИЧЕСКИХ КЛАССОВ В СЛОВАРЕ И ТЕКСТЕ НЕМЕЦКОГО ПОДЪЯЗЫКА ЭЛЕКТРОНИКИ И КОМПРЕССИЯ ЛЕКСИЧЕСКИХ ЕДИНИЦ ПРИ ПЕРЕХОДЕ ОТ ЧСС К ЧСК

Общие сведения о распределении грамматических классов в частотном списке словоформ и в исследуемых текстах даны в табл. 3.

Нетрудно заметить, что ведущее место в нашем подъязыке занимают существительные и имеющие аналогичные с ними парадигмы именные слова (Gegenstandserwägungen), прилагательные и адъективированные причастия в роли определения и местоимения. Эта группа словоформ составляет в наших текстах 64.19%, что вдвое превышает долю в общеязыковых текстах.¹ Это объясняется, вероятно, номинативным характером научно-технического функционального стиля, к которому относится подъязык электроники. При составлении оптимальных систем машинной переработки, накопления и передачи информации подпрограммам анализа и синтеза этих именных форм следует уделить наибольшее внимание.²

В табл. 4 дано сопоставление статистических данных о частях речи в ЧС словоформ, с одной стороны, со статистикой частей речи в ЧСК и ЧСО, с другой.

Преобразование ЧСС в ЧСК дает следующее сокращение числа лексических единиц:

1) в классе существительных	— на 1899 единиц
2) в классе глаголов	— » 421 »
2) в классе детерминативов	— » 1800 » ³
всего	— » 4120 »

Основное сокращение объема ЧСС в пользу ЧСК осуществляется за счет существительных и прилагательных. Это объясняется предметным характером стиля сообщения и тем, что в немецком языке прилагательные, в отличие от других флективных частей речи, имеют двойную парадигму склонения в зависимости от членного и нечленного употребления именной группы, ср. der hohe Widerstand и hoher Widerstand ('высокое сопротивление'), а также тем, что словарные единицы этой группы имеют больше отсекаемых флексий, чем глаголы.

Сокращение ЧСС при переходе к ЧСК в глагольных словах относится в основном к адъективированному причастию I ($\sum i = 206$) и причастию II ($\sum i = 398$).

¹ По данным Х. Мейера, в немецких текстах разных подъязыков и стилей именные слова (Gegenstandserwägungen) занимают в среднем 30%; см.: [4], S. 196—198.

² См.: [1], стр. 128.

³ У детерминативов имеем следующую компрессию (в единицах): у прилагательных — на 1132, у причастию I — на 205, у причастию II — на 398, у местоимений — на 49, у числительных на — 12, в классе артикля на 4; всего на 1800 единиц.

№№	Части речи	В ЧС словоформ			В ЧС квазиоснов			В ЧС основ		
		i	%	δ	i	%	δ	i	%	δ
1	Имена существительные	13 180	64.67	0.02	11 281	71.09	0.02	10 392	74.71	0.02
2	Глаголы, в том числе вспомогательные и модальные	3832	18.77	0.03	2808	16.40	0.4	2025	13.89	0.5
	причастие I	54	0.27	0.30	46	0.24	0.27	44	0.32	0.27
	причастие II	408	1.50	0.11	203	1.16	0.13	122	0.83	0.17
3	Прилагательное	1114	5.42	0.06	716	4.14	0.07	526	3.60	0.08
4	Местоимения	2536	12.42	0.04	1404	8.11	0.05	884	0.07	0.06
5	Числительные	117	0.57	0.16	68	0.39	0.23	62	0.42	0.25
6	Артикли	67	0.37	0.23	64	0.37	0.23	56	0.38	0.26
	Итого флективных частей речи	12	0.05	0.57	8	0.05	0.69	7	0.05	0.74
7	Неизменяемые части речи*	19 753	96.85	—	15 633	96.41	—	13 926	95.52	—
	Всего (Σi)	20 405	100.0	—	16 285	100.0	—	14 578	100.0	—

* Неизменяемые части речи не расшифровываются, так как их статистический вес не меняется при преобразовании ЧСС в ЧСК и ЧСО.

Сокращение за счет личных форм глагола несущественно ($\sum i = 390$). Это объясняется почти полным отсутствием 1-го и 2-го лица ед. ч. и 2-го лица мн. ч. всех временных форм в исследованных текстах.

Сжатие ЧСС за счет местоимений касается указательных, относительных и притяжательных местоимений. Компрессию дают также порядковые числительные и неопределенный артикль. Сходную картину мы получаем при переходе от ЧСС к ЧСО.

§ 3. МАШИННАЯ МОРФОЛОГИЯ НЕМЕЦКОГО ПОДЪЯЗЫКА ЭЛЕКТРОНИКИ

Возникающая в связи с использованием квазиоснов машинная морфология должна приниматься читателем с двумя оговорками:

а) будучи продиктованной результатами исследований базового языка в условиях немецко-русской двуязычной ситуации,

машинная морфология не может претендовать на какую-либо универсальность;

б) она может менять свое построение в зависимости от направленности и построения автоматической переработки текста.

Функционирование машинной морфологии осуществляется в три этапа:

1) отсечение машинных окончаний у изменяемых словоформ текста; отождествление полученных из текстовых единиц квазиоснов с квазиосновами словаря (словарными единицами);

2) расцифровка грамматических и лексических кодов словарной единицы и отождествление отсеченного окончания текстового словоупотребления с определенным окончанием парадигмы;

3) выбор русского эквивалента немецкой квазиосновы и русского эквивалента немецкой флексии; синтез русской основы и окончания.

Рассмотрим подробно каждый из этих этапов.

1) Изменяемые словоупотребления объединяются с точки зрения машинной грамматики в три класса: а) существительные (формальной примером которых является начальная заглавная буква); б) глаголы (их формальными признаками являются окончания, входящие в парадигмы, представленные в табл. 7); в) детерминативы, имеющие окончания, отраженные в табл. 8.

Каждый класс имеет группу парадигм, каждая из которых имеет свой код. Так, парадигмы существительных обозначены кодами от 11 до 17, парадигмы глаголов имеют коды от 22 до 26, а детерминативы — от 30 до 38 (см. табл. 6—8). Коды 18—21 являются запасными для парадигм существительных, 27—29 — для глаголов, 39—40 — для детерминативов.

Все лексемы с чередованием в корне, в частности все немецкие глаголы с умлаутом, аблаутом, чередованием согласных (fahren — fährt, sehen — sieht, bringen — brachte) рассматриваются как отдельные квазиосновы. В качестве разных квазиоснов выступают и орфографические дубликаты (Telephon — Telefon, Photo — Foto, Phono — Fono, Telegraph — Telegraf).

У существительных формы ед. ч. группируются обычно вокруг одной квазиосновы, а формы мн. ч. — вокруг другой⁴ (исключение составляют существительные типа Spannung, Mensch, см. стр. 254—255).

Неизменяемые словоформы не требуют каких-либо машинных трансформаций и полностью совпадают с машинными квазиосно-

⁴ Приведение парадигмы каждого существительного к одной машинной основе существенно усложнило бы программу морфологического анализа. Так, например, если в немецкой именной парадигме Rat — Rates — Rate — Rat — Ratschläge — Ratschläge — Ratschlägen — Ratschläge в качестве квазиосновы взять цепочку Rat, то система машинных флексий будет иметь вид: нулевая флексия — es- e — schläge — schlägen. При машинном анализе придется производить до восьми отсечений букв в правой части словоформы.

вами. К этим словоупотреблениям относятся следующие традиционные классы: союзы, предлоги, частицы, наречия (степени сравнения даются как отдельные неизменяемые словоформы), числительные (за исключением склоняющихся числительных), местоимения (все супплетивные и несклоняющиеся местоимения), определенные артикли. Кроме того, в этот класс должны включаться неизменяемые существительные типа Relais.

К изменяемым словоупотреблениям относятся также все формы глагола, кроме презенса и имперфекта действительного залога, а также глаголы, подвергающиеся в парадигме спряжения чередованиям типа gebe — gibst и прилагательные в ролипре диактивов.⁵

В тех случаях, когда речь идет о неизменяемых словоупотреблениях, отождествление текстового словоупотребления со словарной квазиосновой достигается при первом обращении к АСК. В том случае, когда мы имеем дело с изменяемыми словоупотреблениями, приходится после первого обращения к АСК производить отсечение сначала одной, а потом второй конечной букв.⁶ После каждого отсечения оставшая цепочка букв сразу сравнивается с квазиосновой АСК до тех пор, пока не будет достигнуто их полное отождествление. Если после отсечения двух букв отождествления все-таки не достигается, значит искомая квазиоснова в словаре отсутствует.

2) Если квазиоснова, полученная из текста, оказалась отождествленной со словарной единицей, то первой приписывается лексическая и грамматическая информация последней. Как уже говорилось, грамматическая информация, так называемый сжатый код (СК)⁷ представляет собой отсылку к той или иной строке определенной таблицы — строке, содержащей парадигму данной машинной основы и другие ее грамматические характеристики. Отождествляя окончание текстового словоупотребления с окончаниями в табл. 6—8, мы получаем соответствующую грамматическую информацию к текстовому словоупотреблению.

Подчеркиваем, что на этом этапе — после отсечения окончания текстового словоупотребления и извлечения отсюда грамматической информации — квазиоснов уже нет и СК относится исключительно к текстовому словоупотреблению.

3) Лексическая информация текстового словоупотребления, полученная в результате реализации второго этапа, представляет

⁵ Ср. аналогичный прием при составлении русско-немецкого автоматического словаря основ в некоторых немецких коллективах МП; см.: [5], S. 41 ff.

⁶ Флексии существительных и глаголов могут иметь до двух букв, подлежащих отсечению, а машинные флексии детерминативов — не более одной буквы.

⁷ А. Д. Борисевич, В. В. Гончаренко... и др. См.: [2], стр. 331—614.

собой один или несколько кодов, отсылающих к соответствующим основам выходного русского словника. Грамматическая информация входного языка, полученная на втором этапе, имеет соответствие в виде кодов русских окончаний. Присоединяя эти окончания к основам, взятым из русского словника, мы получаем набор выходных словоупотреблений, соответствующих входному словоупотреблению. Специальные подпрограммы выбирают из множества выходных словоупотреблений то словоупотребление, которое уместно в данном тексте.

§ 4. РАСПРЕДЕЛЕНИЕ ЧАСТЕЙ РЕЧИ В АСК

Распределение в АСК изменяемых и неизменяемых частей речи машинной грамматики показано в табл. 5.

Как видно из табл. 5, словоформы, принадлежащие к неизменяемым частям речи, составляют большинство словарных единиц АСК (66.14%).

Теперь рассмотрим каждую из указанных трех изменяемых частей речи в отдельности.

С у щ е с т в и т е л ь н ы е. При значительном сокращении количества существительных в ЧСК увеличение объема грамматической программы, связанное с распознаванием именной флексии, невелико. Как показано в табл. 6, машинная грамматика имени существительного состоит из семи типов склонений, использующих только семь различных флексий (нулевая, -е, -en, -es, -n, -ns, -s).

Флексии, показанные в табл. 6, выбраны из задаваемых традиционной грамматикой парадигм склонения существительных (см. [3] §§ 261—324).

Существительные ж. р., имеющие нулевые падежные окончания в ед. и мн. ч. (типа Spannung — Spannungen),⁸ могут быть

Таблица 5

№ №	Части речи	Квазисносы		
		а	% от всех словарных единиц	б
1	Имя существительное (без неизменяемых существительных)	3523	20.35	0.03
2	Глаголы	706	4.08	0.07
3	Детерминативы	1285	7.42	0.06
4	Всего изменяемых словоформ:	5514	33.86	
	Неизменяемые словоформы (включая неизменяемые существительные)	10 771	66.14	0.02
	Всего квазисносов:	16 285	100	

⁸ Эти существительные составляют 47.11% от числа всех существительных нашего ЧСЭ по электронике.

включены в АСК только в ед. ч. Отождествление мн. ч. этих текстовых словоупотреблений со словарными единицами АСК осуществляется путем отсечения окончания -en в соответствии с парадигмой, код 17 (табл. 6).⁹

Таблица 6

№ №	Код парадигмы	Падежи				Примеры
		N.	G.	D.	A.	
1	11	—	s	—	—	Tag Peter (м. р. ед. ч.)
2	12	—	—	n	—	Tage (мн. ч.)
3	13	—	es	e	—	Bild Tag (ср. р. ед. ч.) (м. р. ед. ч.)
4	14	—	en	en	en	Mensch (м. р. ед. ч.)
5	15	—	es	n	n	Name (м. р. ед. ч.)
6	16	—	n	n	n	Abgeordnete (м. р. ед. ч.)
7	17	en	en	en	en	Spannungen (мн. ч.)

Нетрудно также заметить, что необходимость отсечения падежных флексий существительных слабого склонения также отпадает ввиду того, что парадигма склонения всех падежей ед. ч., кроме номинатива, совпадает с формами квазисносов мн. ч. (ср. N. der Mensch; G. des Menschen; D. dem Menschen; A. den Menschen; мн. ч. всех падежей: die Menschen). Здесь мы снова имеем две неизменяемые основы: Mensch и Menschen. Таким образом, морфологический анализ флексий существительных осуществляется только относительно существительных м. и ср. р. сильного склонения.

Имена собственные м. р. отнесены к парадигме склонения нарицательных существительных (табл. 6, парадигма 11).

Существительные сильного склонения, имеющие два окончания в род. пад. ед. ч. -es, -s, относятся по лингвистической традиции к одному и тому же типу склонения. В машинной же грамматике эти существительные дают две парадигмы. Так, например, в табл. 6 существительное Tag показано дважды: в парадигме 11 (Tags) и в парадигме 13 (Tages).

Падежами существительных, которые получают в машинной грамматике особые флексии, являются: род. ед. ч. с окончаниями -es или -s дат. ед. ч. на -e, дат. мн. ч. на -en. Таким образом, только три указанные падежные формы существительных м. и ср. р. подвергаются в ходе автоматического поиска однобуквенному и максимум двухбуквенному усечению.

Глаголы. Машинная парадигма немецких глаголов-квазисносов включает пять типов спряжения и шесть потенциаль-

⁹ О возможности сокращения объема АСК за счет этой группы существительных упоминается в настоящей статье впервые. Этим сокращение объема словника словоформ в пользу словника квазисносов увеличивается от ранее упомянутых 15.17 до 20.25%.

ных флексий (нулевая, -e, -st, -t, -en, -n; ср. табл. 7). Отсутствие в текстах немецкого подязыка электроники 1-го и 2-го лица ед. ч. и 2-го лица мн. ч. наст. вр. сокращает количество потенциальных флексий до четырех (нулевая, -t, -n, -en).

Таблица 7

№ №	Код парадигмы	Лицо и число						Примеры
		1-е	2-е	3-е	1-е	2-е	3-е	
		ед.			мн.			
1	22	e	st	t	en	t	en	folg- — наст. вр. изъяв. накл. folgte- — претеритум изъяв. накл.
2	23	—	st	—	n	t	n	folge- — наст. вр. сослагательн. накл.
3	24	—	e	—	en	t	—	folg- — повелительн. накл.
4	25	—	st	—	en	t	en	kam- — претеритум изъявит. накл.
5	26	—	st	t	n	t	n	zeichne- — наст. вр. изъявительн. и сослагательн. накл.

Выше указывалось, что максимальное количество букв в машинной флексии не должно превышать двух, поэтому в тех случаях, когда окончания образуются из трех букв (ср. *zeichn-est*, тип парадигмы 26), третья с конца буква (e) включается в основу. См., например, в этом смысле спряжение глагола *zeichnen* в наст. вр.:

1-е л. ед. ч.	<i>zeichne</i>
2-е » » »	<i>zeichne-st</i>
3-е » » »	<i>zeichne-t</i>
1-е лицо мн. ч.	<i>zeichne-n</i>
2-е » » »	<i>zeichne-t</i>
3-е » » »	<i>zeichne-n</i>

Остальные временные формы квазисосновы типа *zeichnen* совпадают с глаголами упомянутого типа *folg-*. Поиск глагольных флексий ограничивается окончаниями в презенсе и претеритуме действительного залога. Для остальных глагольных временных форм (инфинитивы, вспомогательные глаголы, входящие в состав сложных временных форм и др.) поиск флексии не проводится, поскольку эти словоупотребления совпадают со словарными формами в АСК.

Детерминативы. Выше уже говорилось, что все изменяемые слова, за исключением глаголов и существительных, объединяются в машинной грамматике в класс детерминативов. Это объясняется тем, что традиционные прилагательные, прича-

стия, неопределенный артикль, ряд местоимений и числительных, входящих в этот класс, используют общие парадигмы, причем все эти парадигмы имеют окончания, состоящие максимум из одной буквы (см. табл. 8).

Таблица 8

№ №	Код парадигмы	Падежи				Наличие артикля в именной группе	Примеры
		N.	G.	D.	A.		
1	30	—	n	n	n	с определен. артиклем	<i>der gute Schalter</i>
2	31	r	n	m	n	без артикля	<i>guter Schalter</i>
3	32	n	n	n	n	с определен. артиклем	<i>die guten Schalter</i>
4	33	—	n	n	—	без артикля	<i>gute Schalter</i>
5	34	—	n	n	—	с определен. артиклем	<i>die gute Arbeit</i> <i>das gute Werk</i>
6	35	—	r	r	—	без артикля	<i>gute Arbeit</i>
7	36	s	n	m	s	» »	<i>gutes Werk</i>
8	37	s	n	n	s	с определен. артиклем	<i>ein grosses Land</i>
9	38	r	n	n	n	То же	<i>ein grosser Widerstand</i>

Разъясним это на следующем примере.

В АСК причастие указывается дважды: как неизменяемая словарная единица, выступающая в роли составной части сложных временных форм глагола (типа *ist gemacht*), и как изменяемая словарная единица, выступающая в роли адъективированного причастия, ср.: *gemacht-e Arbeit*, *gemacht-er Arbeit*, *gemacht-er Arbeit*, *gemacht-e Arbeit*.

Нетрудно видеть, что окончание -e проходит по всей машинной парадигме склонения адъективированного причастия. Включение этой буквы в квазисоснову (*gemachte-*) ограничивает флексию одной буквой г.

Машинные флексии детерминативов всех родов показаны в табл. 8.

Машинная грамматика детерминативов состоит из девяти типов склонения, включающих пять окончаний (нулевое, -n, -r, -m, -s).

Дальнейший автоматический анализ упрощается тем, что детерминативы в немецком языке, как известно, всегда стоят в позиции. Они согласуются в роде, числе и падеже со следующими за ними именными словоформами, поэтому для выбора флексий детерминативов достаточно обратиться к СК следующих за ними именных слов.

Неизменяемые словоформы. В машинный грамматический класс неизменяемых словоформ входят числительные, наречия, определенный артикль, предлоги, союзы, аб-

формировать эти описания, превратив их в машинную морфологию, которая, с одной стороны, была бы ориентирована на лингвистические особенности перерабатываемых на ЭВМ научно-технических и деловых текстов, а с другой — учитывала бы специфику организации памяти ЭВМ и особенности программирования.

§ 2. ПОСТРОЕНИЕ МАШИННОЙ МОРФОЛОГИИ РУССКОГО ГЛАГОЛА

А. А. Пиотровская

МАШИННАЯ МОРФОЛОГИЯ РУССКОГО ГЛАГОЛА

§ 1. ЗАДАЧИ МАШИННОЙ ГРАММАТИКИ

Лингвистическое построение грамматики и словаря для автоматического анализа и синтеза текста зависит не только от строя языка, но и от технических характеристик той машины, на которой предполагается реализовать грамматику и словарь. При использовании малых и средних ЭВМ, обладающих небольшим ОЗУ и достаточно большой внешней памятью, целесообразно использовать АС словоформ, сопряженные с упрощенной грамматикой [5]. Напротив, работая с современными ЭВМ, обладающими достаточно емким ОЗУ, целесообразно использовать АС основ и исходных форм, сочетающийся с развернутым алгоритмом анализа и порождения именных и глагольных парадигм или их части.

Построение такого морфологического алгоритма и его реализация на ЭВМ интересна не только с практической — инженерно-лингвистической точки зрения, но имеет также и теоретическую перспективу. Структура и работа такого алгоритма может рассматриваться в плане формальной экспликации и моделирования лингвистического поведения человека при анализе и порождении текста.

Настоящая работа представляет собой попытку такой формальной экспликации морфологии русского глагола.

Все множество морфологических форм русского глагола представляет собой причудливое переплетение регулярных схем, порожденных современной системой с архаическими рудиментами и нормализовавшимися диалектизмами [1; 3; 6; 8; 10; 11]. Поэтому традиционно-лингвистическое описание русской морфологии не образует последовательную и непротиворечивую систему правил, легко поддающуюся формализации. Знакомство с материалом традиционных и даже структурных построений русской морфологии [7; 16; 19; 20] убеждает нас, что для решения инженерно-лингвистических задач целесообразно несколько транс-

Обычно называют следующие категории, характеризующие русский глагол: лицо, вид, время, залог, наклонение, род, число [6, стр. 452 и сл.; 8, стр. 409]. Кроме того, некоторые исследователи выделяют категорию репрезентации, обнаруживающуюся в противопоставлении личных форм инфинитиву, дееспричастию и причастию [7, стр. 118; 15, стр. 245—248]. Наконец, при рассмотрении причастий следует учитывать и категорию падежа.

Статистическое обследование западноевропейских и русских научно-технических и публицистических текстов (см. материалы сборников СтР, СТ, СтРААТ) показывают, что первые два лица единственного числа, а также формы повелительного наклонения в этих текстах практически не употребляются. В связи с этим указанные грамматические формы в нашей машинной грамматике не учитываются.

Исходной формой глагольной парадигмы считается инфинитив, причем инфинитивы несовершенного и совершенного вида вместе с порождаемыми парадигмами рассматриваются как две разные лексемы. Таким образом, категория вида из машинной морфологии выводится и рассматривается в качестве лексической категории, отражаемой в АС.

В качестве особых лексем рассматриваются также причастия и дееспричастия. И те и другие не включаются в машинную морфологию глагола. В связи с этим в нашей грамматике отсутствует категория падежа и не учтены формы страдательного залога, выраженного с помощью кратких форм причастий (эти формы могут быть синтезированы через АС).

Категория репрезентации сокращается и ограничивается противопоставлением инфинитива и личных форм спряжения.

Итак, описываемая машинная морфология включает категории времени, рода, числа, лица, залога, наклонения и репрезентации. Последние три категории заданы в сокращенном виде.

Грамматические категории выявляются в машинной грамматике либо через словоизменение, либо, как это имеет место с сослагательным наклонением и будущим временем глаголов несовершенного вида изъявительного наклонения, с помощью средств аналитической морфологии.

Таблица 1

п ₂	Лексическая основа						Грамматические постфиксы			
	п ₁		к		а ₁		а ₂	а ₃	а ₄	а ₅
	п ₁₁	п ₁₂	к ₁	к ₂	а ₁₁	а ₁₂				
<i>пре</i>	<i>об</i>	—	<i>раз</i>	<i>ов</i>	<i>а</i>	<i>ть</i>	—	—	—	<i>ся</i>
<i>пре</i>	<i>об</i>	—	<i>раз</i>	<i>ов</i>	—	—	<i>а</i>	<i>и</i>	—	<i>сь</i>
<i>пре</i>	<i>об</i>	—	<i>раз</i>	<i>у</i>	—	—	—	—	<i>ет</i>	<i>ся</i>
—	<i>об</i>	<i>о</i>	<i>ги</i>	—	<i>а</i>	<i>ть</i>	—	—	—	—
—	<i>об</i>	—	<i>гон</i>	—	—	—	—	—	<i>ят</i>	—

Фузионный характер русской морфологии затрудняет и усложняет алгоритмизацию и программирование. Для упрощения задачи мы построили квазиагглютинативную схему анализа и синтеза русской глагольной словоформы. Эта схема включает: префиксы (п₁, п₂, п₃) и корень (к), которые вместе образуют лексическую основу и грамматические постфиксы (а₁, а₂, а₃, а₄, а₅). Префикс (п₁), корень (к) и постфикс (а₁) в свою очередь распадаются каждый на основной компонент (п₁₁, к₁, а₁₁) и его дополнение (п₁₂, к₂, а₁₂). Полная схема русского глагола принимает

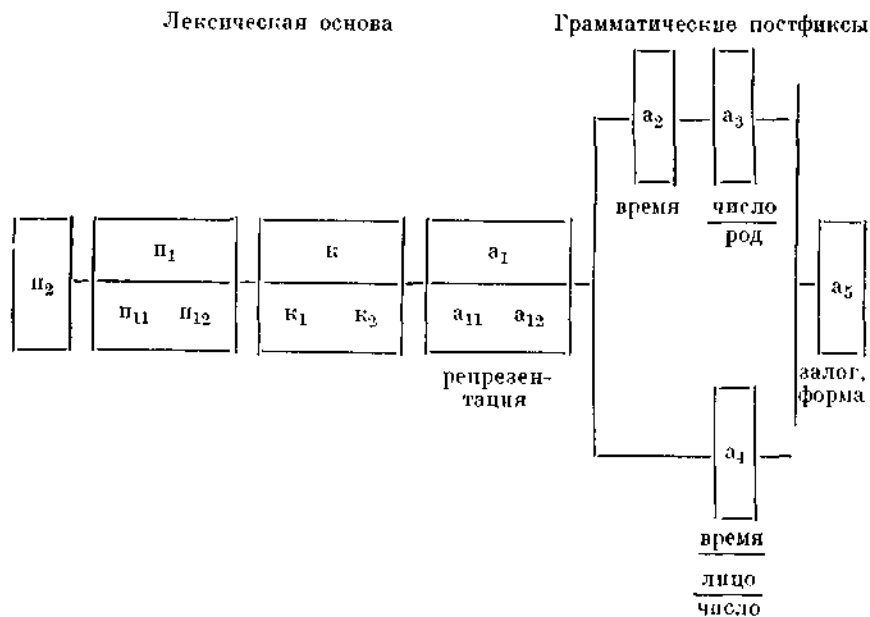


Рис. 1. Схема построения русской глагольной словоформы.

в этом случае вид агглютинативной цепочки машинных морфем, как это показано на рис. 1 и в табл. 1. Этот подход позволяет выделить 90 типов формообразования глагола. Классификация

этих типов и механизм машинного образования настоящего, простого будущего и прошедшего времени показаны в табл. 2 и 3. При составлении этой классификационной таблицы использовался обратный словарь русского языка Г. Вильфельдта [17], с учетом современных лексических норм русского литературного языка [13].

§ 3. МАШИННЫЙ АНАЛИЗ И СИНТЕЗ РУССКИХ ГЛАГОЛЬНЫХ ФОРМ

Описанное в предыдущем параграфе квазиагглютинативное построение русской глагольной формы позволяет унифицировать и упростить составление алгоритма и машинной программы ана-

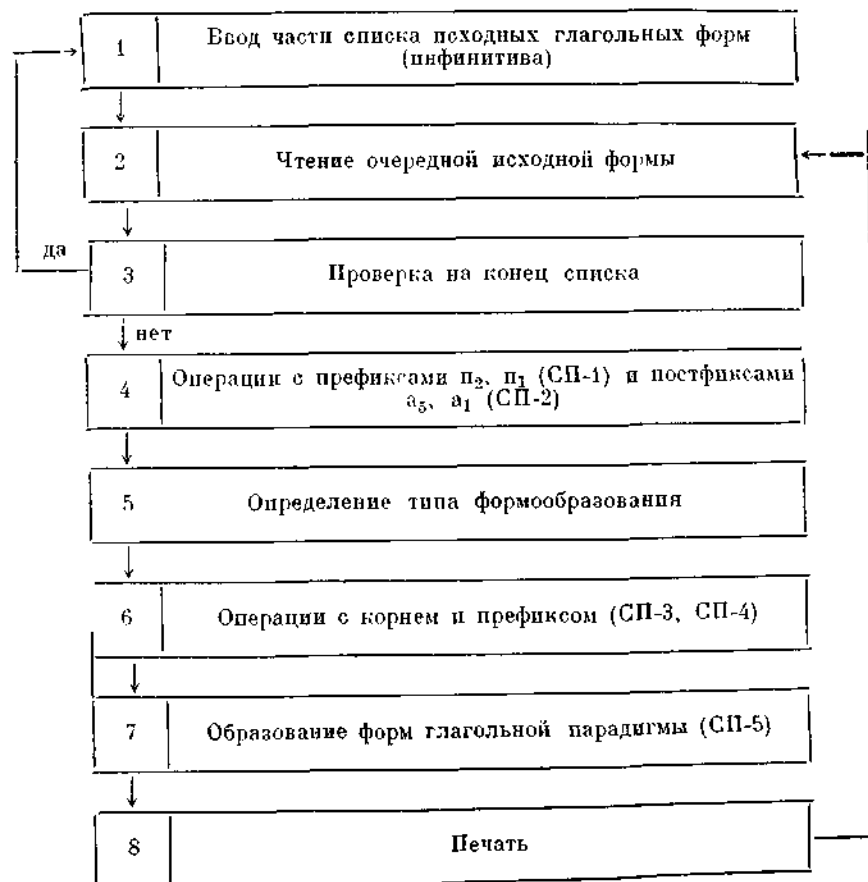


Рис. 2. Принципиальная блок-схема порождения личных глагольных форм.

№ №	Типы формообразования глаголов	Исходная форма (инф.)						операция с преф. п.	Пример	
		преф. п.		корень		постф. а ₁	основной комп. п. ₁₁			доп. п. ₁₂
		основной комп. п. ₁₁	доп. п. ₁₂	основной комп. к ₁	доп. к ₂	постф. а ₁				
1	a-1	все		-бе	ж-	ать	не пзм.			
2	a-2	все, кр. сп. 1		-б	р-	ать	» »			
3	a-3	сп. 1	о-	-б	р-	ать	» »	отбр. о-		
4	a-4	все, кр. сп. 1, 2		-г	к-	ать	» »			
5	a-5	сп. 1, 2	о-	-г	к-	ать	» »	то же		
6	a-6	все		-д		ать	» »			
7	a-7	все, кр. сп. 1, 2		-д	р-	ать	» »			
8	a-8	сп. 1, 2	о-	-д	р-	ать	» »	отбр. о-		
9	a-9	все		-е	х-	ать	» »			
10	a-10	все, кр. сп. 3		-ж		ать	» »			
11	a-11	сп. 3	о-	-ж		ать	» »	приб. о-		
12	a-12	все, кр. сп. 1		-з	в-	ать	» »			
13	a-13	сп. 1	о-	-з	в-	ать	» »	отбр. о-		
14	a-14	все		-л	г-	ать	» »			
15	a-15	»		-сп		ать	» »			
16	a-16	»		-с	л-	ать	» »			
17	a-17	все, кр. сп. 1, 2		-ст	л-	ать	» »			
18	a-18	сп. 1, 2	о-	-ст	л-	ать	» »	то же		
19	a-19	все		сп. 1		ать	» »			
20	a-20	»		сп. 2	т-	ать	» »			
21	a-21	»		сп. 3	т-, к-	ать	» »			
22	a-22	»		сп. 4	в-	ать	» »			
23	a-23	»		сп. 5		ать	» »			
24	a-24	»		сп. 6	з-	ать	» »			
25	a-25	»		сп. 7		ать	» »			
26	a-26	»		сп. 8	с-, х-	ать	» »			
27	a-27	»		сп. 9	ев-	ать	» »			
28	a-28	»		сп. 10	ев-	ать	» »			

Образование наст. и буд. простого вр.								Пример
операция с корнем		операция с постф. а ₁		присоединяемые личные окончания (постф. а ₁)				
основной комп. к ₁	доп. к ₂	доп. а ₁₁	основной комп. а ₁₂	ед. ч.		мн. ч.		
				3-е	1-е	2-е	3-е	
не пзм.	для 3-го л. мп. ч. зам. ж- на г-	отбр.	отбр.	ит	им	ите	ут	(по)бежать
зам. -б	не пзм.	»	»	ет	ем	ете	ут	(со)братъ
на -бе	» »	»	»	ет	ем	ете	ут	(разо)братъ
то же	» »	»	»	ит	им	ите	ят	(на)гнать
зам. -г	» »	»	»	ит	им	ите	ят	(обо)гнать
на -го	» »	»	»	ст	дим	дите	дут	(с)дать
то же	приб. а-	»	»	ет	ем	ете	ут	(у)драть
не пзм.	не пзм.	»	»	ет	ем	ете	ут	(со)драть
зам. -д	» »	»	»	ет	ем	ете	ут	(у)ехать
на -де	» »	»	»	ет	ем	ете	ут	(на)жать
то же	зам. х- на д-	»	»	ет	ем	ете	ут	(с)жать
не пзм.	приб. м-	»	»	ет	ем	ете	ут	(на)звать
»	» »	»	»	ет	ем	ете	ут	(ото)звать
»	то же	»	»	ет	ем	ете	ут	(обо)лгать
зам. -з	не пзм.	»	»	ет	ем	ете	ут	спать
на -зо	» »	»	»	ет	ем	ете	ут	(по)слать
то же	» »	»	»	ет	ем	ете	ут	(за)стлать
не пзм.	для 3-го л. ед. ч. и 1-го, 2-го л. мн. ч.	»	»	ет	ем	ете	ут	(разо)стлать
» »	зам. с-	»	»	ит	им	ите	ят	(подо)ждать
» »	на ж-	»	»	ет	ем	ете	ют	(за)скрежетать
зам. -с	не пзм.	»	»	ет	ем	ете	ют	шептать
на -ш	» »	»	»	ет	ем	ете	ют	создавать
зам. -сте	» »	»	»	ит	им	ите	ат	(но)стучать
на -сте	» »	»	»	ет	ем	ете	ют	(рас)сказать
зам. -ст	» »	»	»	ет	ем	ете	ют	(но)колебать
на -сте	зам. т-	»	»	ет	ем	ете	ют	(на)писать
не пзм.	на щ-	»	»	ет	ем	ете	ют	(раз)вальцевать
» »	зам. т-, к- на ч-	»	»	ет	ем	ете	ют	(за)шпаклевать
» »	отбр.	»	»	ет	ем	ете	ют	
» »	не пзм.	»	»	ет	ем	ете	ют	
зам. з-	на ж-	»	»	ет	ем	ете	ют	
на ж-	приб. л-	»	»	ет	ем	ете	ют	
зам. с-, х-	зам. с-, х-	»	»	ет	ем	ете	ют	
» »	на ш-	»	»	ет	ем	ете	ют	
» »	зам. ев-	»	»	ет	ем	ете	ют	
» »	на ч-	»	»	ет	ем	ете	ют	
» »	зам. ев-	»	»	ет	ем	ете	ют	
» »	на ю-	»	»	ет	ем	ете	ют	

* Сокращения к табл. 2, 3

доп. — дополнение; зам. — заменить; пзм. — изменяется; комп. — компо-

нент; кр. — кроме; осн. — основа; ост. — остальной (-ная, -ное, -ные); отбр. — отбросить; постф. — постфикс; преф. — префикс; приб. — прибавить.

№ №	Типы формообразования глаголов	Исходная форма (инф.)						операция с преф. п ₁	Пример	
		преф. п ₁		корень		постф. а ₁	основной комп. п ₁₁			доп. п ₁₂
		основной комп. п ₁₁	доп. п ₁₂	основной комп. к ₁	доп. к ₂					
29	а-29	все		сп. 11	ск-	ать	не изм.			
30	а-30	»		все, кр. <i>основать, одобровать, уповать</i>	ов-	ать	» »			
31	а-31	»		все основы, кр. типов а-1 а-30 -д		ать	» »			
32	е-1	»				еть	» »			
33	е-2	все, кр. сп. 3, 4		-ме, -не, -те	р-	еть	» »			
34	е-3	сп. 3		-ме, -не, -те	р-	еть	» »	приб. о зам. с- на -зо		
35	е-4	сп. 4	с-	-ме, -не, -те	р-	еть	» »			
36	е-5	все		-н		еть	» »			
37	е-6	»		-го	т-	еть	» »			
38	е-7	»		сп. 12		еть	» »			
39	е-8	»		все основы, кр. типов е-1 ÷ е-7		еть	» »			
40	и-1	»		все основы		ить	» »			
41	и-2	»		-бр		ить	» »			
42	и-2	»		-гн		ить	» »			
43	и-3	»		-ж		ить	» »			
44	и-4	»		-шиб		ить	» »			
45	и-5	все, кр. сп. 3, 4		сп. 13		ить	» »			
46	и-6	сп. 3		»		ить	» »	приб. о- зам. с- на во-		
47	и-7	сп. 4	с-	»		ить	» »			
48	и-8	все		все основы на ж-, ч-, ш-, щ-		ить	» »			
49	и-9	»		кр. типов и-3, и-5		ить	» »			
50	о-1	»		все основы, кр. типов и-1 ÷ и-8 -мо	л-	оть	» »			
51	о-2	»		все основы, кр. типа о-1		оть	» »			

Образование наст. и буд. простого вр.										Пример
операция с корнем		операция с постф. а ₁		присоединяемые личные окончания (постф. а ₂)						
основной комп. к ₁	доп. к ₂	доп. а ₁₁	основной комп. а ₁₂	ед. ч. мн. ч.						
				3-е	1-е	2-е	3-е			
не изм.			отбр.	ет	ем	ете	ют	искать		
» »	зам. <i>ск-на щ-</i>		»	ет	ем	ете	ют	(по)требовать		
» »	зам. <i>ов-на щ-</i>		»	ет	ем	ете	ют	(с)дать		
» »	приб. а-		»	ет	ем	ете	ют	(с)дать		
» »	приб. <i>ен-</i>		»	ет	ем	ете	ют	вдеть		
зам. <i>-ме, -не, -те</i>			»	ет	ем	ете	ют	(на)тереть		
на <i>-ж, -н, -т</i>	приб. р-		»	ет	ем	ете	ют	(о)тереть		
то же	» »		»	ет	ем	ете	ют	(ис)тереть		
» »	» »		»	ет	ем	ете	ют	(с)петь		
не изм.	приб. о- для 3 л.		»	ет	им	ите	ят	(за)хотеть		
» »	ед. ч. зам. т- на ч-		»	ет	им	ите	ят	(у)видеть		
» »	не изм.		»	ет	ем	ете	ют	(по)светлеть		
» »	приб. е-		»	ет	ем	ете	ют	(о)лезть		
» »	приб. з-		»	ет	ем	ете	ют	(по)брить		
» »	приб. е-		»	ет	ем	ете	ют	(за)гнить		
» »	приб. и-		»	ет	ем	ете	ют	(про)жить		
приб. <i>ио-</i>	не изм.		»	ет	ем	ете	ют	(и)шибить		
не изм.	» »		»	ет	ем	ете	ют	(на)бить		
приб. б-	» »		»	ет	ем	ете	ют	(о)бить		
то же	» »		»	ет	ем	ете	ют	(рас)шить		
» »	» »		»	ет	ем	ете	ют	(о)ключить		
не изм.	» »		»	ит	им	ите	ап	(о)ключить		
» »	» »		»	ит	им	ите	ап	(у)становить		
зам. <i>-мо</i>	» »		»	ет	ем	ете	ют	(раз)молоть		
на <i>-ме</i>	» »		»	ет	ем	ете	ют	(рас)колоть		
не изм.	» »		»	ет	ем	ете	ют	(рас)колоть		

№ №	Типы формообразования глаголов	Исходная форма (инф.)						Пример
		преф. п ₁		корень		постф. а ₁	операция с преф. п ₁	
		основной комп. п ₁₁	доп. п ₁₂	основной комп. к ₁	доп. к ₁	основной комп. п ₁₁	доп. п ₁₂	
52	с-1	все		-е		сть	не изм.	
53	с-2	»		-с	е-	сть	» »	
54	с-3	все, кр. сп. 3, 4		-ч	е-	сть	» »	
55	с-4	сп. 3		-ч	е-	сть	» »	приб. о-
56	с-5	сп. 4	с-	-ч	е-	сть	» »	зам. с- на зо-
57	с-6	все		все основы, кр. типов с-1 ÷ с-5	—	сть	» »	
58	у-1	»		нулевая основа или -д		уть	» »	
59	у-2	»		сп. 14	н-	уть	» »	
60	у-3	»		все основы, кр. типов у-1, у-2		уть	» »	
61	ы-1	»		-б		ыть	» »	
62	ы-2	»		-са, -па		ыть	» »	
63	ы-3	»		-ст		ыть	» »	
64	ы-4	»		все основы, кр. типов ы-1, ы-2		ыть	» »	
65	я-1	нет		-в	з-	ять	» »	
66	я-2	все, кр. сп. 3		-ж		ять	» »	
67	я-3	сп. 3		-ж		ять	» »	
68	я-4	все, кр. сп. 3		-н		ять	» »	приб. о-
69	я-5	сп. 3		-н		ять	» »	
70	я-6	при		-н		ять	» »	
71	я-7	е		-н		ять	» »	
72	я-8	все		-сто		ять	» »	
73	я-9	»		сп. 15		ять	» »	
74	я-10	»		все основы, кр. типов я-1 ÷ я-9		ять	» »	
75	т-1	все, кр. сп. 2		-бре	с-	ти	» »	
76	т-2	сп. 2		-бре	с-	ти	» »	
77	т-3	все		-блю, -ве	с-	ти	» »	

Образование наст. и буд. простого вр.										Пример
операция с корнем		операция с постф. а ₁		присоединяемые личные окончания (постф. а ₂)						
основной комп. к ₁	доп. к ₁	доп. а ₁₁	основной комп. а ₁₂	р.д.ч.		мн. ч.				
				3-е	1-е	2-е	3-е			
не изм.	» »	»	»	ст	дйм	дите	дят	(съ)есть		
» »	зам. е- на яд-	»	»	ет	ем	ете	ут	(под)есть		
» »	зам. е- на т-	»	»	ет	ем	ете	ут	(у)честь		
» »	зам. е- на т- то же	»	»	ет	ем	ете	ут	(с)честь		
не изм.	приб. д-	»	»	ет	ем	ете	ут	класть		
» »	приб. у-	»	»	ет	ем	ете	ют	(раз)дуть		
» »	не изм.	»	»	ет	ем	ете	ут	(по)гнзнуть		
» »	» »	»	»	ет	ем	ете	ут	(со)гнзнуть		
зам. -б, на -буд	» »	»	»	ет	ем	ете	ут	(от)быть		
не изм.	приб. ы-	»	»	ет	ем	ете	ут	(от)пльть		
не изм.	приб. ын-	»	»	ет	ем	ете	ут	(за)стыть		
не изм.	приб. о-	»	»	ет	ем	ете	ут	(по)крыть		
» »	зам. з- на озм-	»	»	ет	ем	ете	ут	взять		
» »	приб. н-	»	»	ет	ем	ете	ут	(на)мять		
» »	приб. н- не изм.	»	»	ет	ем	ете	ут	(с)мять		
зам. -н на -йм	не изм.	»	»	ет	ем	ете	ут	(за)нять		
не изм.	приб. им-	»	»	ет	ем	ете	ут	(с)нять		
зам. -н на -ж	не изм.	»	»	ет	ем	ете	ут	(при)нять		
не изм.	приб. емл-	»	»	ет	ем	ете	ют	(е)нять		
» »	не изм.	»	»	ит	им	ите	ят	(на)стоять		
» »	» »	»	»	ет	ем	ете	ют	(под)таять		
» »	приб. я-	»	»	ет	ем	ете	ют	(с)клонять		
» »	зам. с- на д-	»	»	ет	ем	ете	ут	(на)брести		
» »	зам. с- на т-	»	»	ет	ем	ете	ут	(изо)брести		
» »	зам. с- на д-	»	»	ет	ем	ете	ут	(при)вести		

Таблица 2 (продолжение)

№ №	Типы формообразования глаголов	Исходная форма (инф.)						операция с преф. п ₁	
		преф. п ₁		корень		постф. а ₁	основной комп. п ₁₁		доп. п ₁₂
		основной комп. п ₁₁	доп. п ₁₂	основной комп. к ₁	доп. к ₂				
78	т-4	все		-гре, -скре	с-	ти	не изм.		
79	т-5	нет		-ид		ти	» »		
80	т-6	все		-й		ти	» »		
81	т-7			-ме, -плас, -цае	с-	ти	» »		
82	т-8	»		-ра	с-	ти	» »		
83	т-9	»		все основы, кр. типов т-1 ÷ т-8		ти	» »		
84	ч-1	все, кр. сп. 3		-ж	с-	чь	» »		
85	ч-2	сп. 3		-ж	с-	чь	» »	приб. о	
86	ч-3	все		-сти	с-	чь	» »		
87	ч-4	»		-тол	с-	чь	» »		
88	ч-5	»		-а	с-	чь	» »		
89	ч-6	»		сп. 16		чь	» »		
90	ч-7	»		все основы, кр. типов ч-1 ÷ ч-6		чь	» »		

Образование наст. и бул. простого вр.								Пример	
операция с корнем		операция с постф. а ₁		присоединяемые личные окончания (постф. а ₁)					
основной комп. к ₁	доп. к ₂	доп. а ₁₁	основной комп. а ₁₂	ед. ч.		мн. ч.			
				3-е	1-е	2-е	3-е		
не изм.			отбр.	отбр.	ет	ем	ете	ут	(на)скрести
» »			»	»	ет	ем	ете	ут	идти
» »			»	»	ет	ем	ете	ут	(со)йти
» »			»	»	ет	ем	ете	ут	(под)мести
» »			»	»	ет	ем	ете	ут	(про)расти
» »			»	»	ет	ем	ете	ут	(при)ползти
» »			»	»	ет	ем	ете	ут	(при)жечь
» »			»	»	ет	ем	ете	ут	(с)течь
» »			»	»	ет	ем	ете	ут	(до)стичь
» »			»	»	ет	ем	ете	ут	(на)толочь
» »			»	»	ет	ем	ете	ут	(с)лечь
» »			»	»	ет	ем	ете	ут	(у)беречь
» »			»	»	ет	ем	ете	ут	(с)течь

№	Тип формообразования глагола	Образование прош. вр.			
		Операция с преф. п ₁		Операция с корнем	
		основной комп. п ₁₁	доп. п ₁₂	основной комп. к ₁	доп. к ₂
1	a-1 ÷ a-31	не изм.		не изм.	приб. букву, совпадающую с a ₁₁
2	e-1, e-5 ÷ e-8	» »		» »	то же
3	з-1	» »		» »	» »
4	u-1 ÷ u-9	» »		» »	» »
5	o-1 ÷ o-2	» »		» »	» »
6	y-1, y-3	» »		» »	» »
7	ы-1 ÷ ы-4	» »		» »	» »
8	я-1 ÷ я-10	» »		» »	» »
9	e-2 ÷ e-4 (корни -мер, -пер, -тер)	» »		» »	» »
10	c-1, c-2, c-6	» »		» »	не изм.
11	c-3 (корень -че)	» »		» »	для м. р. ед. ч. не изм., в остальных формах отбр. то же
12	c-4 (преф. сп. 3 и и корень -че)	» »	приб. о-кр. м. р. зам. с- на зо-	» »	» »
13	c-5 (преф. сп. 4 и корень -че)	» »	» »	» »	» »
14	y-2	» »		» »	отбр. к-
15	m-1 ÷ m-3, m-7 (корни типа -брес, -вес; -мес, -плес)	» »		» »	отбр. с-
16	m-4 (корни типа -скрес)	» »		» »	зам. с- на б-
17	m-5, m-6 (корни типа -ид, -й)	» »		зам. на -ше	не изм.
18	m-8 (корень -рас)	» »		зам. -ра на -ро	» »
19	m-9 (корень -лола)	» »		не изм.	» »
20	ч-1 (корень -же)	» »		» »	для м. р. ед. ч. приб. г-, в ост. формах зам. г- на з-
21	ч-2 (корень -же)	» »		» »	то же
22	ч-3, ч-5, ч-6 (корень -сти, -ле)	» »		» »	приб. г-
23	ч-4 (корень -толо)	» »		» »	для м. р. ед. ч. приб. к-, в ост. формах зам. о- на к-
24	ч-7	» »		» »	приб. к-

Операция с преф. а ₁		Присоединение постф.			Пример	
доп. а ₁₁	основной комп. а ₁₂	ед. чис.				мн. ч.
		м. р.	ж. р.	с. р.		
отбр.	отбр.	л	ла	ло	ли	(по)бежать
»	»	л	ла	ло	ли	(о)деть
»	»	л	ла	ло	ли	(о)лечь
»	»	л	ла	ло	ли	(у)становить
»	»	л	ла	ло	ли	(раз)молоть
»	»	л	ла	ло	ли	(раз)дуть
»	»	л	ла	ло	ли	(от)быть
»	»	л	ла	ло	ли	(с)нять
»	»	л	ла	ло	ли	(от)мереть
»	»	л	ла	ло	ли	(съ)есть
»	»	л	ла	ло	ли	(у)честь
»	»	л	ла	ло	ли	(с)честь
»	»	л	ла	ло	ли	(рас)честь
»	»	л	ла	ло	ли	(по)гибнуть
»	»	л	ла	ло	ли	(изо)брести
»	»	л	ла	ло	ли	(на)скрести
»	»	л	ла	ло	ли	идти, (на)йти
»	»	л	ла	ло	ли	(про)расти
»	»	л	ла	ло	ли	(при)ползти
»	»	л	ла	ло	ли	(при)лечь
»	»	л	ла	ло	ли	(об)лечь
»	»	л	ла	ло	ли	(с)лечь
»	»	л	ла	ло	ли	(на)толочь
»	»	л	ла	ло	ли	(с)течь

лиза и синтеза этих форм. Этого упрощения и унификации удастся достичь путем применения следующих пяти стандартных подпрограмм:

- а) СП-1 — «Выбор и отделение *n* букв, стоящих в начале слова» (ср. префиксы P_1, P_2);
- б) СП-2 — «Последовательный выбор *n* букв, начиная с конца слова» (ср. постфиксы $a_1 - a_3$ и дополнение k_2); иначе эта подпрограмма называется ЭЛРМ-3 [см. 9, стр. 403];

ИСХОДНЫЕ ФОРМЫ, ВХОДЯЩИЕ В ЭВМ	ПАРАДИГМЫ, ПОЛУЧЕННЫЕ ПО У				
ХОТЕТЬ	ХОТЕА ХОЧЕТ БУДЕТ	ХОТЕРА ХОТИМ БУДЕМ	ХОТЕРО ХОТИТЕ БУДЕТЕ	ХОТЕЛИ ХОТЯТ БУДУТ	ХОТЕТЬ
ВЕРТЕТЬ	ВЕРТЕА ВЕРТИТ БУДЕТ	ВЕРТЕРА ВЕРТИМ БУДЕМ	ВЕРТЕРО ВЕРТИТЕ БУДЕТЕ	ВЕРТЕЛИ ВЕРТЯТ БУДУТ	ВЕРТЕТЬ
ТЕЧЬ	ТЕЧА ТЕЧЕТ БУДЕТ	ТЕЧЕА ТЕЧЕМ БУДЕМ	ТЕЧЕО ТЕЧЕТЕ БУДЕТЕ	ТЕЧЕИ ТЕЧУТ БУДУТ	ТЕЧЬ
ПРИЖИТЬСЯ	ПРИЖИ ПРИЖИЕТ	ПРИЖИА ПРИЖИМ ПРИЖИТЕ	ПРИЖИО ПРИЖИТЕ ПРИЖИТЕ	ПРИЖИИ ПРИЖИУТ	ПРИЖИТЬСЯ
ДАВАТЬ	ДАВАА ДАЕТ БУДЕТ	ДАВААА ДАЕМ БУДЕМ	ДАВААО ДАЕТЕ БУДЕТЕ	ДАВААИ ДАДУТ	ДАВАТЬ
ИЗДАВАТЬ	ИЗДАВАА ИЗДАЕТ БУДЕТ	ИЗДАВААА ИЗДАЕМ БУДЕМ	ИЗДАВААО ИЗДАЕТЕ БУДЕТЕ	ИЗДАВААИ ИЗДАДУТ	ИЗДАВАТЬ

Рис. 3. Машинный результат.

- в) СП-3 — «Уничтожение *n* букв с конца словоформы или ее части»;
- г) СП-4 — «Присоединение справа *n* букв к словоформе или к ее части»;
- д) СП-5 — «Формирование слова путем объединения *n*-буквенных цепочек» (машинных морфем).

Принципиальная блок-схема алгоритма для порождения личных глагольных форм изъявительного и сослагательного наклонений по описанной выше машинной грамматике показана на рис. 2.

Этот алгоритм был реализован в виде машинной программы для ЭВМ «Минск-22». Объем программы — около трех тысяч команд. Выборочные результаты программы показаны на рис. 3.

Следует иметь в виду, что реализация нашего морфологического построения на малой ЭВМ «Минск-22» имеет экспери-

ментальный характер. Описанная машинная морфология и алгоритм предназначены для синтеза русских глагольных словоформ, осуществляемого в ходе реализации бинарного вероятностного МП на средних и больших ЭВМ, например: «Минск-32», ЕС 1030 «БЭСМ-6».

Машинная морфология русского глагола может быть использована также и для нужд оптимизации преподавания русского языка иностранцам.

СПИСКИ ПРЕФИКСОВ

Сп. 1	Сп. 2	2) вз	9) с
1) во	1) изо	3) из	
2) обо	2) о	4) над	Сп. 4
3) ото	3) со	5) об	1) ис
4) подо	Сп. 3	6) от	2) рас
5) разо	1) в	7) под	
		8) раз	

СПИСКИ КОРНЕЙ

Сп. 1*	6) -мет-	7) -дребезж-	6) -обяз-
1) -вр-	7) -плак-	8) -дрожж-	7) -рез-
2) -жажд-	8) -прят-	9) -дыш-	8) -сказ-
3) -жд-	9) -рокоп-	10) -жужж-	
4) -жр-	10) -скак-	11) -журч-	Сп. 7
5) -ор-	11) -гопот-	12) -крчч-	1) -дрем-
6) -полз-	12) -гопт-	13) -леж-	2) -колеб-
7) -сос-	13) -хлопот-	14) -молч-	3) -сып-
8) -стон-	14) -холот-	15) -мч-	4) -щип-
9) -рв-	15) -шепт-	16) -мыч-	
10) -рж-	16) -щебет-	17) -огорч-	Сп. 8
11) -тк-	17) -щекот-	18) -пшч-	1) -брех-
12) -тряс-		19) -рыч-	2) -мах-
	Сп. 4	20) -слыш-	3) -пояс-
Сп. 2	1) -дав-	21) -стуч-	4) -пах-
1) -клевет-	2) -здав-	22) -торч-	5) -пис-
2) -скрежет-	3) -зпав-	23) -трещ-	6) -пляс-
3) -трепет-	4) -став-	24) -уничтож-	7) -тсс-
4) -ропт-		25) -урч-	8) -чес-
	Сп. 5	26) -фырч-	
Сп. 3	1) -брепч-	Сп. 6	Сп. 9
1) -бормот-	2) -брюзж-	1) -вяз-	1) -бичев-
2) -грохот-	3) -бурч-	2) -каз-	2) -бушев-
3) -клокот-	4) -визж-	3) -лпз-	3) -вальцев-
4) -лепет-	5) -ворч-	4) -маз-	4) -врачев-
5) -лопот-	6) -держ-	5) -низ-	5) -гарцев-

* Следует иметь в виду, что приводимые списки корней не всегда идентичны спискам, используемым в машинной программе («машинным» спискам корней). Их различие состоит в том, что машинные списки содержат также ложные корни, получаемые в том случае, когда первые буквы корня графически совпадают с приставкой, ср. -р- из -вр- и -ор-, -тон- из -стон- (см. сп. 1).

6) -глянцев-	6) -ыск-	Сп. 13	23) -тихп-
7) -жев-		1) -б-	24) -тухи-
8) -квасцев-	Сп. 12	2) -в-	Сп. 15
9) -корчев-		3) -л-	
10) -кочев-	1) -блест-	4) -п-	1) -ба-
11) -линчев-	2) -верт-	5) -ш-	2) -бле-
12) -лицев-	3) -вел-	Сп. 14	3) -ве-
13) -лушчев-	4) -вид-		4) -де-
14) -ночев-	5) -вис-	1) -блекн-	5) -ла-
15) -потчев-	6) -галд-	2) -брюзгн-	6) -леле-
16) -свинцев-	7) -гляд-	3) -висн-	7) -ма-
17) -спринцев-	8) -гор-	4) -выкн-	8) -па-
18) -танцев-	9) -дуд-	5) -вязн-	9) -ре-
19) -торцев-	10) -звеч-	6) -гасн-	10) -се-
20) -тушев-	11) -кип-	7) -гибн-	11) -сме-
21) -шприцев-	12) -корп-	8) -глохи-	12) -та-
22) -фальцев-	13) -кряхт-	9) -грязн-	13) -те-
Сп. 10	14) -лет-	10) -дохп-	14) -ха-
1) -горев-	15) -пыхт-	11) -дрябн-	15) -ча-
2) -клев-	16) -свист-	12) -зябн-	16) -ну-
3) -шалев-	17) -сид-	13) -кисн-	Сп. 16
4) -плев-	18) -сип-	14) -липпн-	1) -бере-
Сп. 11	19) -скрип-	15) -мерзн-	2) -мо-
1) -иск-	20) -смотр-	16) -мозгн-	3) -стере-
2) -плеск-	21) -соп-	17) -молки-	4) -стри-
3) -полоск-	22) -тарахт-	18) -мякн-	5) -пра-
4) -прыск-	23) -терп-	19) -никн-	
5) -рыск-	24) -храп-	20) -сохн-	
	25) -хруст-	21) -слабн-	
	26) -шелест-	22) -слепн-	
	27) -шип-		

12. Кулагина О. С. О машинном переводе с французского языка на русский. ПЖ, вып. 3, 1960.
13. Ожегов С. П. Словарь русского языка. Изд. 3-е, М., 1953.
14. Пиотровская А. А., Рихтер Н. А. Синтез видовых форм русского глагола. СТ II.
15. Смирницкий А. И. Морфология английского глагола. М., 1959.
16. Халле М. О правилах русского спряжения. In: American Contributions to the V-th International Congress of Slavists, The Hague, 1963.
17. Bielfeldt H. H. Rückläufiges Wörterbuch der russischen Sprache der Gegenwart. 2 Aufl. Berlin, 1965.
18. Āurovič L. Paradigmatika spio-vnej ruštiny. Bratislava, 1964.
19. Jakobson R. Russian Conjugation. Word, 4, № 3, 1948.
20. Trubetzkov N. S. Das morphologische System der russischen Sprache. Travaux du cercle linguistique de Prague, 5, № 2, 1934.

ЛИТЕРАТУРА

1. Аванесов Р. И. Русская диалектология. М., 1960.
2. Белоногов Г. Г. Определение грамматических признаков «новых» слов с помощью словаря. ИЛ.
3. Борковский В. И. и Кузнецов П. С. Историческая грамматика русского языка. М., 1963.
4. Василевский А. Л., Кравец Л. Г., Москович В. А., Поволоцкая С. К., Самойлович М. В., Тарасова Г. А., Эмдина Ю. М. Англо-русский автоматический перевод. Труды ЦНИИПИ, серия 3, М., 1967.
5. Вертель В. А., Вертель Е. В., Крисевиц В. С., Пиотровский Р. Г., Трибис Л. И. Автоматические словари в системе бинарного вероятностного МП. ИЛ.
6. Виноградов В. В. Русский язык. Грамматическое учение о слове. М.—Л., 1947.
7. Волоцкая З. М., Молошная Т. Н., Николаева Т. М. Опыт описания русского языка в его письменной форме. М., 1964.
8. Грамматика русского языка. I, М., 1952.
9. Зубов А. В. Переработка текста естественного языка в системе «человек — машина». СтРААТ.
10. Исаченко А. В. Грамматический строй русского языка в сопоставлении со словацким. Морфология, I. Братислава, 1954.
11. Кузнецов П. С. Историческая грамматика русского языка. Морфология. Изд. Московского ун-ва, 1953.

Часть IV. СИНТАКСИЧЕСКИЙ МАШИННЫЙ ПЕРЕВОД

Л. З. Соса

ИНВАРИАНТ СТРУКТУРНО-СИНТАГМАТИЧЕСКИХ ОТНОШЕНИЙ

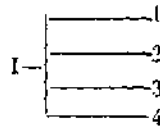
1. Вопрос о соотношении формы и значения сейчас является одним из самых актуальных в лингвистике. От его решения зависит конструирование различных моделей речевой деятельности, поиск наиболее эффективных методов обучения иностранным языкам, создание автоматического перевода и много других, больших и малых задач.

Особенную важность приобретает этот вопрос при построении универсальных грамматик и выяснении соотношения языка и мышления. Создание универсальных грамматик связано с проблемой поиска инвариантов в парадигматике языка и в его синтагматике. Инвариантам второго типа как раз и посвящается данное исследование.

2. Возьмем четыре предложения:

- | | |
|---------------------|-------------------|
| 1) I was a teacher | (английск. язык); |
| 2) Я был учителем | (русск. »); |
| 3) Я учительствовал | (» »); |
| 4) Ngomfundisi | (язык зулу). |

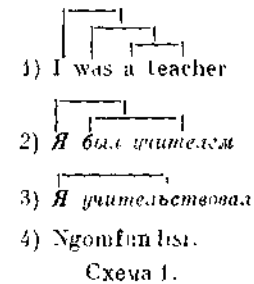
При установлении межъязыковых или внутриязыковых соответствий приведенные предложения рассматриваются как переводы (перифразы) друг друга. Это означает, что в каждом из указанных предложений выделяется некий смысловой инвариант (I), а сами предложения анализируются как различные формы реализации инварианта:



Займемся изучением этого инварианта. Предварительно введем два термина: синтагматический и парадигматический. Синтагматическим будем называть любое множество, упорядоченное во времени (т. е. любую цепочку), парадигматическим —

множество, не удовлетворяющее этому требованию (т. е. любой ряд). Обращаясь к характеристике инварианта анализируемых предложений, нельзя не заметить парадигматическую составляющую этого инварианта (см. ряды: а) I — я — ng; б) teacher — учитель — umfundisi). Синтагматическая составляющая менее заметна, и на ней мы остановимся подробнее.

Начнем с того, что попробуем зафиксировать синтагматическую структуру каждого предложения с помощью схемы. Для этого воспользуемся инструментом, разработанным в классическом анализе и известным нам из школы, — построим для каждого предложения схему его синтаксической структуры. Например:

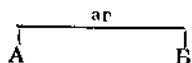


Сопоставим полученные схемы. Результат сопоставления может быть один — констатация полного различия всех синтаксических структур (так, в первом предложении синтаксическая структура выступает как совокупность трех синтаксических связей, во втором — двух, в третьем — одной, в четвертом вообще нет синтаксических связей). И значит, ни о каком инварианте структурно-синтагматического плана при таком способе экспликации синтагматической структуры анализируемых предложений не может быть и речи. Другими словами, указанный способ экспликации синтагматической структуры исследуемых предложений заставляет утверждать одно: все четыре предложения имеют совершенно различную синтагматическую структуру, при переходе от языка к языку никакой структурно-синтагматический инвариант не сохраняется, а происходит полное уничтожение синтагматической структуры исходного предложения и замена ее синтагматической структурой предложения-эквивалента.

Теперь отвлечемся от проиллюстрированного выше способа экспликации синтагматической структуры и попробуем разобраться в том, почему все четыре предложения являются переводами друг друга. На каком основании мы считаем их взаимозаменяемыми эквивалентами? На этот вопрос можно ответить так. По-видимому, критерием эквивалентности всех четырех предложений является их соотносительность с одним и тем же денотатом: тождество картинок, которые можно нарисовать в соответствии

с этими предложениями, дает право говорить о переводимости предложений друг в друга.

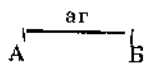
Очевидно, что, в соответствии с нашим определением синтагматических и парадигматических явлений, в каждой картинке можно выделить элементы обоих этих типов. При этом не менее очевидно, что парадигматические явления картинок находят выражение посредством уже выделенных нами выше парадигматических языковых рядов. А о каких синтагматических элементах исследуемых картинок может идти речь? Нам представляется, что в качестве таковых можно рассматривать три элемента: два узла (я — узел А и «действие — учительствование» — узел Б) и связь между ними (агентивную — аг). Таким образом, синтагматическую структуру анализируемых картинок можно представить в виде схемы:



одной и той же для всех картинок, а через их посредство — одной и той же для всех исследуемых предложений. Следовательно, указанная синтагматическая структура картинок и является тем синтагматическим инвариантом, который позволяет рассматривать все четыре предложения как переводы друг друга. И значит, в синтагматической структуре каждого предложения можно выделить два компонента: синтагматический инвариант, представляющий собой синтагматическую структуру соответствующего денотата, и синтагматический спецификатор, изменяющийся от языка к языку. Назовем первый из выделенных компонентов смысловой структурой предложения, а второй — его формальной структурой.

Нарисуем смысловые структуры всех четырех предложений, обращая внимание на то, посредством каких конкретных слов передаются компоненты картинок:

Синтагматическая структура картинок



Смысловые синтагматические структуры предложений

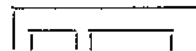
- 1) I (A) was a teacher (B)
- 2) Я (A) был учителем (B)
- 3) Я (A) учительствовала (B)
- 4) Mina (A) ngomfundisi (B)

Схема 2.

Отметим, что один синтагматический элемент картинки может быть выражен одним словом (узел Б в предложении 3), несколькими словами (узел Б в предложениях 1 и 2) или морфемой (т. е. в виде отдельного слова в предложении он может не выражаться;

см. узел А в предложении 4: *mina* 'я', восстанавливаемый на картинке по морфеме *nga*, но в соответствующем предложении отдельным словом не представленный).

Перейдем к формальным синтагматическим структурам исследуемых предложений. Для этого напишем, например, предложение 2 и, опираясь на школьный анализ, зафиксируем в нем



следующие связи: Я был учителем. Связь между я и был назовем, как принято, согласованием (ср. я был, но мы были), между был и учителем — управлением (ср. был учителем, но не был в учителя, тогда как превратился в учителя, но не превратился учителем), между я и учителем — согласованием (ср. я учителем, но мы учителями). Эти связи, как и те, анализом которых мы занимались выше, безусловно, относятся к связям синтагматического типа. Однако к картинке они никакого отношения не имеют: областью действия этих связей являются только языковые формы, вне какого-либо отношения к денотатному плану (согласование — это взаимоуподобление форм обоих членов пары друг другу, управление — обуславливание формы одного члена пары другим ее членом).

В соответствии с такой трактовкой связей в предложении 2 попробуем и для остальных предложений построить структуры синтагматических связей, не имеющих отношения к денотату. Нам представляется, что они будут иметь такой вид:

- 1) $\begin{array}{c} \text{С+Т} \\ \text{---} \\ \text{У+Т} \\ \text{---} \\ \text{C+T} \quad \text{T} \quad \text{C+T} \end{array}$
- 3) $\begin{array}{c} \text{С} \\ \text{---} \\ \text{Я} \text{ учительствовала} \end{array}$
- 4) Ngomfundisi

Схема 3.

В предложении 1 между I и was мы фиксируем согласование (ср. I — was, но we — were), между I и teacher — согласование (ср. I — teacher, но we — teachers), между was и teacher — управление (ср. He was he, но не He was him), между a и teacher — согласование (ср. a teacher, но не a teachers); кроме того, между всеми словоформами здесь есть еще формальные связи «таксемы порядка слов». В предложении 3 между я и учительствовала — согласование (ср. я учительствовала, но мы учительствовали); в предложении 4 формальных связей нет.

Назовем полученные схемы схемами формальных синтагматических структур и отметим, что эти структуры зависят только от специфики каждого языка, при

давал возможность адекватно отражать классическое понимание процесса.

Принятие второго «либо» привело нас к построению теории, изложенной выше. В результате вместо одной синтагматической структуры предложения мы получили две синтагматические структуры того же самого предложения. Какие это дает преимущества, мы отметили. Теперь остается осветить еще один вопрос: как соотносятся оба способа представления синтагматической структуры? При этом больше всего нас будет интересовать следующая проблема: одна ли и та же информация фиксируется посредством обоих способов или наличие двух способов фиксации отражает на самом деле наличие двух типов знаний об исследуемых объектах? Другими словами: один и тот же объект эксплицирован посредством двух способов фиксации синтагматической структуры или перед нами — два разных объекта?

Обратимся опять к предложению 3. В соответствии с классическим пониманием (см. схему 1), синтагматическая структура этого предложения состоит из одной (предикативной) связи между двумя узлами (подлежащим и сказуемым). В соответствии с введенным нами описанием (см. схему 4), синтагматическая структура предложения 3 состоит из двух структур: смысловой и формальной. При этом формальная структура состоит из одной формальной связи (согласования) между двумя узлами (словоформами *я* и *учительствовал*). Смысловая структура состоит так же из одной связи, но уже смыслового плана: агентивной. Аналогично зафиксированы и узлы: в смысловой структуре *я* выступает не как словоформа, а как агент действия *учительствовал*. Иначе говоря, при выделении формальной структуры слов интересуют нас в плане номинации, а при построении смысловой структуры — в плане денотации.

Поэтому, если предположить, что наши схемы эксплицируют тот же объект, что и классические, то одна классическая синтагматическая схема должна быть «равна» двум нашим. Другими словами, если мы действительно исследуем те же языковые явления, что и «классики», то выделенные нами формальные и смысловые структуры — это не просто формальные и смысловые структуры, существующие сами по себе, безотносительно одна к другой, но такие, которые являются означаемым и означающим по отношению друг к другу. Аналогичное рассуждение должно быть справедливо и в отношении узлов. Таким образом, мы приходим к необходимости зафиксировать два метода: 1) метод «сборки» некоторых конструкций из означаемых и означающих (синтез) и 2) метод «разборки» конструкций на означаемые и означающие (дуализацию).

Называя соответствующие конструкты синтагматическим узлом,

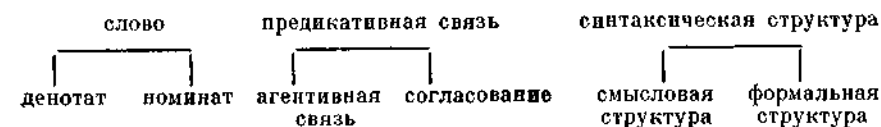
синтагматической связью и синтагматической структурой, для предложения 3 мы получаем:



Сопоставляя полученные конструкты с соответствующими элементами схемы классического анализа, можно отметить, что для предложения 3 имеет место следующее соотношение:

элементы схемы 4	=	элементы схемы 1
синтагматический узел	=	слово
синтагматическая связь	=	предикативная связь
синтагматическая структура	=	синтаксическая структура

Отсюда: если мы изучаем тот же объект, что и «классики», то структура элементов справа от знака равенства должна быть той же, что структура элементов слева. И значит, применив методы дуализации и синтеза к «классическим элементам» (слову, предикативной связи, синтаксической структуре), мы должны будем получить те же компоненты, что и при оперировании с нашими конструктами:



Следовательно, анализ предложения 3 показал, что означаемое и означающее можно выделить и при «классическом разборе» не только у материальных компонентов языка (слов), но и у идеальных, например, таких, как предикативная связь и синтаксическая структура.²

² Синтагматические связи и синтагматические структуры — это элементы того типа, которые обычно относят к синтактике. Иногда говорят, что знак не билатерален, а допускает параметризацию по четырем составляющим: 1) план означаемых, 2) план означающих, 3) синтактика, 4) прагматика. Материалы данной статьи показывают, что синтактика — это не еще один из параметров знака, а, образно говоря, — тоже знак. Более точно: синтактика — область элементов, каждый из которых может быть представлен как знак, т. е. разложен на означаемое и означающее. Таким образом, к сторонам знака синтактика отношения не имеет. С прагматикой дело обстоит, по-видимому, так же. Синтактика и прагматика языковых элементов — это то, что дано исследователю в самом языковом материале. «Латеральность» — то, что дает исследователь, обрабатывая материал: один ученый может представить наблюдаемые им явления в виде монад, другой — в виде дуалов (знаков), третий — в виде триад и т. д.

Теперь посмотрим, как обстоит дело с элементами других предложений.

Начнем с предложения 2 (см. схемы 1 и 4). При классическом анализе в предложении 2 устанавливается синтаксическая структура, состоящая из двух связей: предикативной (между подлежащим *я* и сказуемым *был учителем*) и полупредикативной (между компонентами сказуемого *был* и *учителем*). В соответствии с предложением нами анализом, на схемах 4 им должны соответствовать две структуры, состоящие: одна — из двух формальных, вторая — из двух смысловых связей. Однако в действительности дело обстоит иначе. Полупредикативная связь представлена только одной связью (управлением), а предикативная связь — не двумя, но тремя связями (одной смысловой — агентивной, и двумя формальными — согласованием). И значит, принцип двухкомпонентности сохраняется только для всей структуры в целом, но не для отдельных ее частей: каждый элемент синтаксической структуры не обязательно соответствует элементу формальной и элементу смысловой структуры. А это означает, что формальный, смысловой и формально-смысловой планы могут быть взаимно независимыми, т. е. единицы одного плана можно выбирать независимо от единиц другого плана, а затем устанавливать соответствие между планами через посредство структур в целом. Например, в качестве единиц формального плана можно принять словоформы и связи между словоформами, в качестве единицы смыслового плана — словосочетания и связи между словосочетаниями, а затем через посредство формальных и смысловых синтагматических структур установить взаимное соответствие обоих планов. Отсюда напрашивается вывод: изоморфизм формы и значения в синтагматике имеет место на уровне *с т р у к т у р в ц е л о м* и не имеет места на уровне *о т д е л ь н ы х к о м п о н е н т о в* этих структур.

Этот тезис, по-видимому, объясняет следующие явления. На схемах классического анализа можно отметить связи двух типов: между словами (например, связь между *был* и *учителем*) и между словосочетаниями (например, связь между *я* и *был учителем*). Разнотипность связей подчеркнута разностильностью их изображения: связи между словами фиксируются в виде скобок между узлами, а связи между словосочетаниями — в виде скобок между скобками. При этом в одной и той же схеме уживаются оба типа связей. Больше того, одна часть связи трактуется как связь между словами (на рисунке скобка упирается в узел, см. схему 1), другая часть связи — как связь между словосочетаниями (скобка упирается в скобку). Иначе говоря, в схемах классического анализа наблюдается нивелирование двух уровней — слова и словосочетания.

Задумываясь над причинами нивелирования уровней, можно предположить, либо что нивелирование является необходимым

условием фиксации информации именно того типа, которую передает классический анализ, либо что оно не связано с типом информации, но обусловлено какими-то другими факторами. Например, таким фактором может быть отсутствие единого принципа при выборе единиц анализа: в случае отсутствия изоморфизма плана формы и значения на уровне единиц синтагматических структур при проведении формально-смыслового анализа в качестве единиц могут выбираться то формальные единицы, которым сопоставляется затем некоторое значение (слова), то смысловые единицы, определенным образом впоследствии оформляющиеся (словосочетания).

Покажем, что первое «либо» в случае классического анализа не имеет места: нивелирование уровней не имплицитно типом передаваемой информации.

Для этого решим следующую задачу. Построим синтагматическую структуру предложения 2 (см. схему 1) при условии, что связи устанавливаются между блоками, каждый из которых обязательно равен слову. Иначе говоря: от классической схемы анализа предложения 2 по словам-и-словосочетаниям перейдем к схеме анализа того же предложения по словам. Для решения этой задачи воспользуемся известной нам аналогией из истории лингвистики.

Отметим, что указанная задача сформировалась в связи с решением ряда проблем прикладного характера (например, в связи с построением алгоритмов автоматического анализа). Ее решение было выработано на базе формалистских теорий. В соответствии с этими теориями синтаксическая структура предложения 2, изображенная на вышеприведенной схеме (стр. 273), стала трактоваться так:

|——| |——|
Я был учителем
Схема 5.

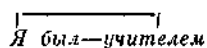
Проанализируем соотношение этой синтаксической структуры со структурой на схеме 1. На первый взгляд может показаться, что перед нами — два разных изображения одной и той же синтаксической структуры: классическая структура (схема 1) состоит из двух связей, формалистская структура (схема 5) — тоже из двух связей. Однако более внимательный анализ показывает, что связи на обеих схемах — разные. При классическом понимании синтаксической структуры предложения речь идет о связи между *я* и *был учителем*, т. е. предполагается, что связь установлена не между *я* и одним из элементов комплекса, но между *я* и всем комплексом (на это указывает проведение стрелки связи ко всему комплексу, а не к одному из его элементов). Последнее можно трактовать как связь между *я* и каждым элементом комплекса.

При формалистском понимании имеется в виду только связь между *я* и *был*. Связь с другим элементом комплекса, имплицитно присутствующая в классическом разборе, в формалистском анализе исчезает. Другими словами, при переходе от классического анализа к формалистскому (в случае рассмотренной нами предикативной связи) на деле происходит утрата информации. А это означает, что на схеме 5 фактически изображена не та же самая синтаксическая структура, что на схеме 1, а только какая-то ее часть. Другая часть утеряна.

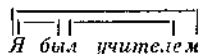
Поэтому схема 5 не может быть принята в качестве решения поставленной выше задачи (зафиксировать в применении к анализу по словам ту же синтаксическую структуру, которую выявляла классическая лингвистика, анализируя предложение по более крупным блокам — словосочетаниям): первая часть задачи (переход от уровня словосочетаний к уровню слов) таким путем может быть решена, вторая часть задачи (не терять при этом переходе информацию о синтаксических отношениях) — нет.

Посмотрим, какое представление интересующей нас синтаксической структуры можно предложить, чтобы справиться с обеими частями интересующей нас задачи. Начнем с экспликации той информации, которая, как нам кажется, имплицитно присутствует в схеме классической синтаксической структуры.

Сравнивая предложение *Я был учителем* с предложением *Я учитель*, можно отметить, что *был* в смысловом плане играет роль характеристики *учителем*, а не характеристики *я*: *был* как бы выступает в роли свободной морфемы при слове *учитель* (ср. с предложением 3). А это значит, что смысловой связи между *я* и *был* нет: имеются смысловые связи между *я* и *учителем* и между *был* и *учителем*. Однако это смысловые связи разного типа: одна морфологическая, другая — синтаксическая:



В формальном плане, как об этом мы говорили выше, можно зафиксировать три связи:



Все эти связи одного порядка — между словоформами.

Сопоставляя формальный и смысловой планы, можно отметить, что классическая схема позволяет эксплицировать всего пять связей: между *был* и *учителем* — формальную и смысловую («морфемную»), между *я* и *учителем* — формальную и смысловую (синтаксическую), между *я* и *был* — формальную. Таким образом, получается, что в синтаксическом плане *я* связано с *учителем* формально и по смыслу, *был* с *учителем* и *я* с *был* — только формально:

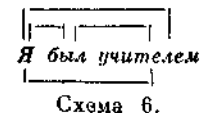


Схема 6.

Отсюда понятно, что поставленную выше задачу можно считать решенной только в том случае, если удастся найти способ представления синтаксической структуры данного предложения, при котором будут зафиксированы все четыре синтагматические связи. В качестве такого способа представления может выступать схема 6. В дальнейшем мы будем называть ее объединенной схемой формально-смысловой структуры предложения 2, полученной при экспликации информации классической схемы 1 методом перехода от анализа по словосочетаниям к анализу по словам. Возможен и другой способ представления, фиксирующий интересующие нас связи, — мы останавливались на нем выше: расцепив формально-смысловую структуру на две составляющие, можно изобразить те же четыре синтагматические связи с помощью двух схем — схемы формальной структуры предложения 2 и схемы его смысловой структуры (схема 4). Этот тип представления синтагматической структуры предложения 2 ниже мы будем называть дуалистическим.

Следовательно, мы показали, что нивелирование уровней не связано с типом передаваемой информации. Поэтому остается признать, что оно вызвано тем, что в классическом анализе отсутствует принцип выбора элементов синтаксических структур (узлов и связей). А это, как нам кажется, обусловлено, в свою очередь, отсутствием изоморфизма между элементами формальных и смысловых структур.

5. Выше мы остановились на четырех принципиально различных способах представления синтаксической структуры предложения: а) «классическом» (формально-смысловой анализ на уровне слов-и-словосочетаний) — схема 1; б) «формалистском» (формальный анализ на уровне слов) — схема 5; в) «дуалистическом» (формально-смысловой анализ на уровне слов) — схема 4; г) «объединенном» (формально-смысловой анализ на уровне слов) — схема 6. Схемы, построенные по методам а, в, г, служат для передачи одной и той же («классической») информации; схема, построенная по методу б, — для передачи информации, отличной от «классической». Естественно, что возможны и другие методы представления синтаксической структуры. Например, это может быть: д) формально-смысловой анализ на уровне словосочетаний типа г, т. е. анализ, принципы которого можно выявить в процессе решения задачи, обратной по отношению к той, которой занимались мы при переходе от уровня словосочетаний к уровню слов; е) смысловой анализ на уровне словосочетаний (или других смысловых блоков) и т. д.

Таких методов может быть много. Их выбор должен обуславливаться практической задачей, для которой строится данный анализ. Например, выше мы показали, что для решения задачи об инварианте и спецификаторе перевода метод *a* оказывается непригодным; с помощью метода *в* та же задача решается довольно просто. По-видимому, для каждого метода можно указать область его применения. При этом следует учитывать такие соображения.

Говоря о различных методах, прежде всего необходимо выяснить, о каких методах идет речь: о методах описания объектов или о методах познания. Например, методы *a, в, г* (см. схемы 1, 4 и 6) различаются между собой только в сфере изменения типов изображения одного и того же объекта (результаты познания одни и те же, хотя фиксируются они по-разному); в отличие от этого, между методами *a, в, г*, с одной стороны, и методом *б*, с другой, отношения сложнее: при переходе от метода *б* к методам *a, в, г* изменяется не только тип изображения, но и тип фиксируемой информации. И значит, в этом случае и результаты познания разные, и фиксируются они по-разному.

Установление типа метода важно, в первую очередь, для решения вопроса о критериях оценки. По-видимому, методы познания оцениваются в зависимости от того, какое количество информации они позволяют извлечь. С этой точки зрения методы познания, с помощью которых извлечена информация, изображенная при помощи схем 1, 4 и 6, лучше метода познания, с помощью которого извлечена информация, представленная схемой 5.

Для методов описания обычно применяются другие критерии. Здесь вступают в действие принципы удобства, простоты, экономии, адекватности, непротиворечивости, однозначности и полноты. С позиции этих критериев, метод *a* кажется наиболее экономным. Однако он, очевидно, не отвечает принципу однозначности: классический анализ допускает не одно толкование, а несколько; поэтому при оперировании со схемами классического анализа все время приходится оговариваться: «по-видимому», «нам кажется». Метод *в* — более сложный, но его применение дает возможность не нарушать принцип однозначности. Кроме того, с его помощью можно непротиворечиво представить те явления, которые такого представления в рамках метода *a* не имеют.

Составляя указанные методы и вырабатывая критерии их оценки, важно отметить также следующее. Каждый из методов связан с выбором системы обозначений. При этом результаты решения одной и той же задачи с помощью одного какого-то метода могут быть представлены по-разному — в терминах различных систем обозначений. В этом случае наиболее трудным оказывается установление корректных соотношений между терминами обеих систем.

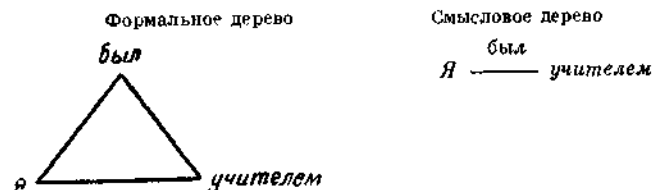
Так, в лингвистике обычно пользуются как линейным, так и древовидным изображением синтаксической структуры. Оба типа

изображения взаимопереводимы. Эта взаимопереводимость сохраняется независимо от того, по какому методу (*a, б, в* или *г*) построена сама синтаксическая структура. Покажем это.

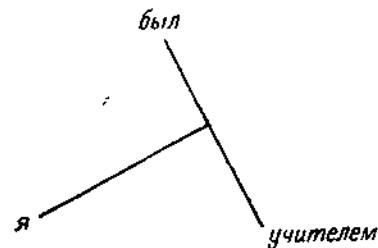
Начнем со схемы предложения 2 схемы 5. В зависимости от договоренности о выборе вершины, дерево, соответствующее этой схеме, будет иметь один из двух видов:



Для простоты будем далее говорить только о дереве с вершиной *был*. Теперь обратимся к предложению 2 на схеме 4. Соответствующие деревья (с вершиной *был*) будут иметь вид:



Перейдем к предложению 2 на схеме 1. Представить эту схему в виде дерева нельзя — в дереве не должно быть связей, упирающихся в середину других связей. Поэтому применительно к схеме на стр. 27⁹ речь может идти не о дереве, но только об изображении следующего вида:



Очевидно, что эту иерархию можно трансформировать в дерево любым из двух способов: сдвинув стрелку вниз, мы получим одну схему, сдвинув вверх — другую:



И значит, говоря, что «классическая» иерархия может быть заменена двумя деревьями, мы фактически снова подчеркиваем неоднозначность описания, построенного на основании классического анализа. Не менее интересны приведенные деревья и в плане соотношения классических и формалистских методов анализа, о котором мы говорили выше: классическое «дерево» эквивалентно двум деревьям, и только одно из них соответствует формалистскому дереву.

Таким образом, видно, что различие в типах передачи синтаксической информации, зафиксированное при линейном изображении, при древовидном остается тем же. Нужно только корректно подобрать системы обозначений.

И. И. Волосевич

МАШИННАЯ СЕГМЕНТАЦИЯ АНГЛИЙСКОГО ПРЕДЛОЖЕНИЯ

§ 1. АНАЛИЗ ПРЕДЛОЖЕНИЯ НА ОСНОВАНИИ СТРУКТУРНЫХ СИГНАЛОВ

1.1. Постановка задачи

Обычно, когда человек читает длинное предложение на иностранном языке, которым он владеет, он читает его не все целиком, а по частям. При этом создается впечатление, что сегментация помогает человеку воспринять и осознать смысл предложения, т. е. выступает как средство, лежащее на поверхности и помогающее ориентироваться в глубинном уровне.

Сегментация происходит даже в тех случаях, когда перевод целого предложения или отдельных его частей остается неясным. Поэтому возникает предположение, что в указанной ситуации сегментация осуществляется на основании каких-то формальных сигналов, имеющих в поверхностной структуре предложения и опознаваемых человеком без обращения к плану содержания. Читая предложение *The iggle squigs trazed wombly in the harlish goor*¹ читатель фиксирует только строевые элементы языка (например, служебные слова *the, in* и морфемы *-s, -ed, -ish*) и на основании законов их функционирования в поверхностной структуре предложения делает выводы о взаимоотношении элементов глубинной структуры. Другими словами, эти элементы выступают в качестве формальных сигналов смысловой сегментации текста. И значит, если с ними «ознакомить» ЭВМ, то она будет в состоянии сегментировать текст, подобно человеку; тем самым будет сделан еще один шаг на пути эвристического программирования, которое ставит своей целью моделирование не только результатов «человеческого» анализа текстов, но и воспроизведение механизма этого анализа на ЭВМ.

¹ См.: Г. Глпсон. Введение в дескриптивную лингвистику. Русск. пер. М., 1959, стр. 193.

Практические преимущества такого подхода к построению алгоритмов автоматического перевода можно кратко охарактеризовать так.

Выделение групп, полученных в результате предварительной обработки текста, позволяет значительно сократить число потенциальных выходных эквивалентов, выбираемых для многозначной входной словоформы из автоматического словаря. Это несколько облегчает решение вопроса о снятии многозначности входного текста. Кроме того, такое выделение помогает решению задач, возникающих при автоматическом реферировании текстов в тех областях науки, описание которых возможно на базе данного подязыка. Ведь большинство выделяемых именных групп является сложными терминами, а некоторые из них могут выступать в качестве дескрипторов. Уже только этих примеров достаточно, чтобы оценить ту роль, которую может сыграть в практике автоматического анализа текста предварительная его сегментация по типу той, которую проводит человек.

При условии, что мы знаем, как сегментирует текст человек, задача научить машину аналогичной сегментации может быть сформулирована так. Предложив человеку разбить текст на контекстные группы, лингвист должен исследовать результаты его деятельности и построить такое описание языка, в котором для каждой группы были бы зафиксированы диагностирующие ее формы. Обладая этим описанием, можно затем составить алгоритм и программу, на основании которой ЭВМ по диагностическим формам будет выделять те же контекстные группы, что и человек.

При такой формулировке задачи четко выделяются два этапа исследования:

1) процесс исследования языка человеком; при этом подготавливается материал для составления алгоритма и программы для ЭВМ; это путь от значения к форме: сначала выделяются контекстные группы, затем — диагностирующие их формы.

2) процесс реализации результатов первого этапа на ЭВМ, который состоит в том, что машина осуществляет сегментацию текстов на основании диагностирующих форм; здесь имеет место переход от формы к значению: по диагностическим формам выделяются организованные семантико-синтаксической связью сегменты,² из которых можно строить дерево предложения.

Отсюда следует, что форма и значение интересующих нас языковых элементов на обоих этапах должна фиксироваться отдельно, т. е. что само существование указанных этапов как бы предполагает тезис о дуалистическом описании языковых элементов, участвующих в сегментации. Действительно, приняв в качестве единицы синтаксического анализа отрезок текста, кото-

рый мы условно называем синтаксемой,³ и выделив в нем форму и значение, мы сможем организовать различные процедуры, в процессе которых можно будет следовать любому из интересующих нас путей: от формы к значению или от значения к форме. Мы сможем, если захотим, работать только с формами или только со значениями. Во избежание недоразумений подчеркнем, что, говоря о двух типах процедур (от значения к форме или от формы к значению), мы имеем в виду оценку указанных процедур человеком. Для ЭВМ противопоставление формы значению irrelevantно: она работает только с формами, и качественное различие в типе данных, обуславливающее противопоставление формы и значения для человека, для ЭВМ не существует. В коммуникативной системе «человек — машина — человек» возникает особая ситуация, сущность которой состоит в том, что ЭВМ не может самостоятельно оперировать единицами плана содержания. Машина воспримет их только тогда, когда элементы плана содержания взаимно-однозначно соотнесены с элементами плана выражения. Отсюда следует, что смысловая интуиция, на которую опирается человек при выполнении синтаксического анализа, или должна быть устранена из машинного синтаксического анализа, или она должна быть воплощена в элементы плана выражения.

При переводе синтаксеме как дуалу входного языка человек сопоставляет синтаксему как дуал выходного языка: идентичность означаемых позволяет переходить от одного означающего к другому (см. статью Л. З. Сова, здесь, стр. 278—292). Если перевод осуществляет ЭВМ, то она имеет дело только с означающими обеих синтаксем, а их означаемые остаются как бы за кулисами процесса автоматического перевода: означаемые синтаксем «не выходят из рук» исследователя, искавшего те означающие, которые он передал ЭВМ. Поэтому при анализе текста машиной, когда синтаксема «расстается» с человеком и поступает в распоряжение ЭВМ, ее означаемое «испаряется», а на выходе, когда синтаксема вновь «встречается» с человеком, означаемое снова «возвращается» к ней. С точки зрения человека, фиксировавшего в синтаксеме означаемое и означающее, в машине нет никакого семиозиса: ЭВМ имеет дело только с фигурами, а не со знаками. Но можно заставить ЭВМ «приобрести с в о ю точку зрения» (например, имея в виду возможность «общения» ЭВМ) и сконструировать единицы совершенно особого типа, но тоже имеющие знаковый характер. Например, это могут быть знаки такого типа: означающее равно означающему синтаксемы входного языка, а означаемое — означающему синтаксемы выходного языка. Воз-

³ Ср. с понятием синтаксемы в работе: А. М. Мухина. Синтаксема как функциональная синтаксическая единица. Научн. доклады высш. школы, 1961, 3, стр. 59.

² Ср.: Л. З. Сова. Аналитическая лингвистика. М., 1970, стр. 102.

можно и другие варианты, например: означающее равно означающему синтаксемы входного (или выходного) языка а означаемое — символу языка-посредника и т. д.

А теперь возвратимся к задаче сегментирования текста.

К сожалению, лингвистам не ясно, как сегментирует текст человек, — доказательством тому является множественность типов сегментации, которую производят различные исследователи при анализе текста по членам предложения или непосредственным составляющим.⁴ Поэтому обучение ЭВМ сегментации текста становится в то же время и исследованием того, как сегментирует текст человек.

Действительно, в зависимости от того, какие мы зададим диакритические формы, ЭВМ выделит сегменты. Сравнивая эти сегменты с теми, которые выделяет носитель языка, мы сможем уяснить, как работает интуиция человека при тех или иных видах анализа, а значит сможем понять, каким интуитивным процессам соответствуют, например, различные типы анализов по членам предложения или непосредственным составляющим.

Ставя такие эксперименты, мы можем начать с самого простого. Например, для начала мы можем взять в качестве граничных сигналов, на основании которых выделяются контекстные группы в предложении, формальные элементы, полученные на основании статистической обработки текста. Так, это могут быть самые частотные словоформы в данном подязыке. Посмотрим, какие контекстные группы можно выделить с их помощью.

Выберем в качестве граничных сигналов 10 первых словоформ из частотного словаря по электронике: the, of, and, a, is, in, to, he, for, with. Введем эти словоформы списком в память машины,⁵ сообщив ей при этом информацию следующего характера:

а) прочитав слово и отождествив его со словом в списке, накапливать группу;⁶

б) прочитав слово и отождествив его со словом в списке, печатать накопленную группу.

Теперь предложим ЭВМ разобрать текст:⁷

⁴ Ср.: P. Postal. Constituent Structure: A Study of Contemporary Models of Syntactic Description. The Hague, 1964.

⁵ О применении списков для сообщения информации ЭВМ в практике автоматического синтаксического анализа см.: И. А. Мельчук, Р. Д. Равич. Синтаксический анализ. В кн.: Автоматический перевод, М., 1967, стр. 264—328.

⁶ В данном эксперименте под группой понимается последовательность словоформ, ограниченная слева и справа одной из десяти указанных словоформ.

⁷ См.: Converter puts FM in your car. In: Radio-Electronics, August, 1959, № 8, p. 49.

With selector switch SI in the AM position, the incoming signal is fed through capacitor C22 to the AM receiver. In this position power supplies to the translator is turned off.

Placing switch SI in the FM position supplies power to the converter and connects the antenna input to the cathode circuit of the first rf amplifier, half of a 12EC8.

Прочитав первую словоформу with и отождествив ее со словоформой списка, машина запоминает словоформу with и читает следующую словоформу selector. Не обнаружив тождества прочитанной словоформы со словоформами, указанными в списке, ЭВМ также запоминает и ее. Аналогичным образом машина поступает и со словоформами switch SI. Прочитав словоформу in, машина отпечатает накопленную группу with selector switch SI, так как словоформа in в данном эксперименте обуславливает печать накопленной группы. Машина запоминает словоформу in и читает следующее слово. Отождествив словоформу the со словоформой, указанной в списке, ЭВМ отпечатает словоформу in, приняв ее за накопленную группу. Подобный анализ будет осуществляться до конца текста.

Итак, текст после машинного анализа приобретает следующий вид:

With selector switch SI	(I)
in	(II)
the AM position	(III)
the incoming signal	(IV)
is fed through capacitor C22	(V)
to	(VI)
the AM receiver	(VII)
In this position power supplies	(VIII)
to	(IX)
the translator	(X)
is turned off	(XI)
In	(XII)
the FM position supplies power	(XIII)
и т. д.	

Таким образом, располагая минимальным количеством граничных элементов, число которых определено совершенно произвольно, и минимальной информацией, предписывающей обращение с этими элементами, мы в результате машинного анализа в некоторых случаях получили неожиданно правильные группы. Например, группы I, IV, XI. Вместе с тем в результате анализа в тексте возникли лакуны из-за того, что, прочитав слово, машина не смогла отождествить его ни с одним из формальных структурных сигналов, данных в списке. Так, например, машина опустела при анализе группу placing switch SI. Не располагая информацией о природе комбинации «предлог + артикль», ЭВМ выделила предлоги отдельными группами.

Подводя итог этого маленького эксперимента, можно заключить, что машина может провести самостоятельную переработку

синтаксиса печатного текста только в том случае, когда она снабжена всеми данными о структурных сигналах, позволяющих проводить синтаксический анализ, и информацией о характере взаимодействия этих сигналов в зависимости от их позиции в структуре предложения.

Обсуждая результаты этого эксперимента, мы замечаем, что полученные сегменты отличаются от тех, которые хотелось бы получить, если основываться не на структурных сигналах, зафиксированных выше, а на данных интуиции информанта. В связи с этим возникает два вопроса: 1) какие сегменты мы хотим получить?; 2) какие структурные сигналы диагностируют эти сегменты?

Ответ на эти вопросы как раз и составляет задачу данной работы. Поиск решения, как это самоочевидно, велся в направлении от вопроса 1 к вопросу 2. Но при фиксации результатов решения и характеристике алгоритма сегментации (а именно он и явился целью наших поисков) оказалось удобнее избрать другую последовательность — от вопроса 2 к вопросу 1. Поэтому прежде всего мы зафиксируем формальные сигналы, которые учитываются при сегментации текстов, затем покажем, как из этих формальных сигналов составить алгоритм, и только после этого перейдем к описанию алгоритма сегментации и результатов его работы, т. е. к характеристике сегментов, которые с его помощью можно получить и из исследования которых мы исходили при составлении алгоритма.

1.2. Граничные сигналы

Предварительному обследованию подверглось 4000 словоформ, взятых из английских текстов по радиоэлектронике. В результате исследования было выявлено, что английская именная группа представляет собой «рамочную конструкцию», левой границей которой являются детерминативы (определенный и неопределенный артикли и притяжательные местоимения), глагольные формы, если именная группа является дополнением к глаголу, отдельные морфемы и т. д.⁸ Правой границей, не входящей в состав группы, считались личные местоимения, глаголы, предлоги и т. д. В качестве ядра именной группы принималось крайнее правое существительное, т. е. существительное, соседствующее с правым граничным сигналом.

Граничные сигналы именной группы выделялись с опорой на дистрибутивный метод анализа текста. При этом различались

⁸ Ср.: Л. Г. Кравец, Ю. М. Эмдина. Автоматический анализ английских номинативных групп. В кн.: Труды III Всесоюзной конференции по информационно-поисковым системам и автоматизированной обработке научно-технической информации. Семипалатинские проблемы автоматизированной обработки информации, т. 2, М., 1967, стр. 441—442.

мелкая (контактная) дистрибуция, т. е. окружение на глубину одного элемента, и глубокая, при которой учитывается несколько рядом стоящих элементов.

Различие контактной и глубокой дистрибуции сохраняется и в алгоритме. Сохранение этого принципа позволяет ЭВМ определять функцию граничного сигнала, который не дает возможности однозначно установить, является ли он сигналом, указывающим на начало или конец группы, или этот сигнал сам принадлежит какой-либо группе.

Ниже приводится таблица левых и правых граничных сигналов именной группы. Сигналы (границы) упорядочены в порядке убывания частоты употребления в структуре предложения.

Левые границы	Частоты	Правые границы	Частоты
1) Детерминативы	46%	1) Глаголы	26%
2) Предлоги	16%	2) □	20%
3) Местоименные слова	14%	3) Предлоги	19%
4) Глаголы	8%	4) Отдельные морфемы	7%
5) Сочинительные союзы	4%	5) Личные местоимения	4%
6) Морфема -ed; Part. II	4%	6) Детерминативы	4%
7) Подчинительные союзы	3%	7) Морфема -ed, Part. II	4%
8) Отдельные словоформы	3%	8) Отдельные словоформы	4%
9) Морфема -ing	2%	9) Подчинительные союзы	4%
		10) Местоименные слова	3%
		11) Сочинительные союзы	3%
		12) Морфема -ing	2%

Граничные сигналы были сгруппированы в классы, которые в виде списков вводятся в память машины.

Границы в списках упорядочены по количеству графем в словоформе или морфеме. Подобное распределение граничных сигналов сделано с учетом конструктивных особенностей ЭВМ.

Вместе с выделением граничных сигналов (границ) выделялась и информация о характере их взаимодействия в структуре предложения в зависимости от занимаемой сигналом позиции в предложении. Эта информация вводится в память машины в виде списка правил (предписаний), обуславливающих «поведение» ЭВМ в каждом отдельном конкретном случае.⁹

Опора на списки граничных сигналов, информация об их взаимодействии в структуре предложения и специальные предписания, предусматривающие логическую последовательность каждого шага машины, позволяют ЭВМ проводить переработку синтаксиса английских печатных текстов без обращения к семантике языка.

⁹ Подробное описание машинного анализа структуры предложения приводится на стр. 306.

1.3. Безусловные формальные граничные сигналы

Выделенные и сгруппированные в классы формальные граничные сигналы можно подразделить на безусловные и промежуточные. Критерием для такого деления является «поведение» ЭВМ при анализе структуры предложения входного языка.

Безусловной формальной границей считается такой структурный сигнал, использование которого позволяет машине однозначно определить начало или конец какой-либо группы.

Так, например, правой безусловной границей именной группы являются детерминативы (определенный и неопределенный артикли, притяжательные местоимения), предлоги, личные формы глагола и т. п.

Безусловные формальные границы именной группы в свою очередь подразделяются на инклюзивные, т. е. входящие в группу, и эксклюзивные, не входящие в группу. Разделение по этому принципу можно рассмотреть на примере следующего предложения: *If there is a voltage, the oscillator is working, although not necessarily at the right frequency* (1).

Здесь неопределенный артикль, вводящий именную группу *a voltage*, рассматривается как компонент именной группы, т. е. инклюзивно, а во второй группе *the oscillator* — безусловный граничный сигнал рассматривается двояко. С одной стороны, он — эксклюзивный, т. е. не входящий в предыдущую именную группу *a voltage*, так как он указывает на ее конец, а с другой стороны, он является инклюзивным сигналом вводимой им группы *the oscillator*.

Таким образом, критерием для определения инклюзивности или эксклюзивности безусловного формального сигнала именной группы является местонахождение этого сигнала относительно данной группы.

Естественно, что безусловные инклюзивные формальные границы именной группы являются эксклюзивными границами глагольной группы. Так, в нашем примере неопределенный артикль — безусловная эксклюзивная граница глагольной группы *there is*. Другой эксклюзивной границей глагольной группы *is working, although not necessarily* является предлог *at*, рассматриваемый одновременно и как инклюзивный граничный сигнал именной группы *at the right frequency*.

Безусловными инклюзивными сигналами глагольной группы — безусловными эксклюзивными сигналами именной группы — являются личные местоимения в общем падеже, как в приводимом ниже примере: *Last month we saw how some common accepted thoughts on transistors are not true* (2).

Аналогично рассматриваются указательные, неопределенные и другие местоимения,¹⁰ вводящие глагольную группу. Например,

¹⁰ См. сд. 2 в приложении к алгоритму.

this в предложении *This is not an unusual bias voltage in transistor circuits* (3).

В нашем алгоритме эти местоимения являются членами глагольной группы и рассматриваются как некоторые глагольные морфемы.

Подобное рассмотрение этих местоимений в составе глагольной группы обуславливается наличием особой грамматико-семантической связи, требующей дополнения вводимого глагола личным местоимением в общем падеже, указательными и неопределенными местоимениями. Эта связь определяется грамматическими свойствами как глагола, так и местоимения, диктующими их постановку в определенной форме.¹¹ Грамматическими показателями этой связи являются глагольная морфема *-s* для 3-го л. ед. ч. в *Present Indefinite*, дифференциальное использование вспомогательного глагола при образовании *Present Perfect*, времен группы *Continuous*, времен группы *Future* и существование в языке особых грамматических сокращений типа: *I'll, I'd, we've, that'll, he's* и др.

Рассмотрение личных местоимений в общем падеже, указательных и других местоимений, вводящих глагол в качестве особых глагольных морфем, позволяет машине:

- а) передавать полную грамматическую информацию, фиксируемую глаголами при машинном переводе глагольных групп;
- б) однозначно определять начало глагольной группы и снимать лексико-грамматическую омонимию словоформ внутри предложения.

1.4. Промежуточные граничные сигналы

Промежуточными формальными границами являются такие структурные сигналы, опознание которых не позволяет ЭВМ окончательно и однозначно как определить границу группы, так и установить принадлежность самого сигнала к какой-либо группе.

Действительно, в приводимых примерах:

Technicians working with transistors will frequently encounter circuits which have no counterpart in vacuum-tube work (4);

The second interesting feature is the method used to supply DC operating currents (5);

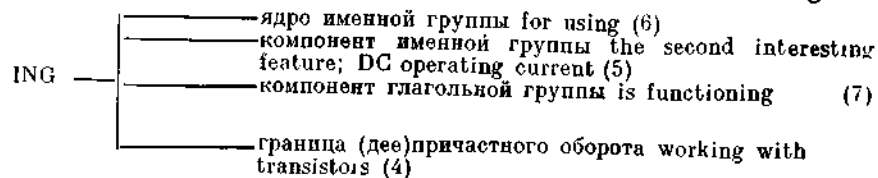
There are several reasons for using such an arrangement (6);

Base bias voltage remains essentially unchanged whether or not the oscillator is functioning (7),

опознав граничный сигнал *-ing* и опираясь только на него, маши-

¹¹ Ср.: В. Г. А д м о я н. Завершенность конструкции как явление синтаксической формы. ВЯ, 1958, №1, стр. 112.

на сможет установить, чем в каждом конкретном случае является словоформа с морфемой -ing, — ср. схему многозначности -ing:



Выявление промежуточного граничного сигнала является указанием на то, что необходим дальнейший анализ, во время которого машина, привлекая на помощь левую и правую контактную дистрибуцию, определит природу опознанного граничного сигнала. В этом случае ЭВМ устанавливает:

а) является ли опознанный граничный сигнал безусловным граничным сигналом, указывающим на конец или начало другой группы;

б) указывает ли опознанный граничный сигнал на его принадлежность к анализируемой группе.

Так, анализируя в предложении (4) правую и левую дистрибуцию, машина устанавливает, что словоформа с морфемой -ing — безусловная формальная граница именной группы technicians, с одной стороны, а с другой — что эта словоформа указывает на начало (дее)причастного оборота.

Опознав граничный сигнал -ing в предложениях (5), (6) и (7) и проверив правую дистрибуцию, ЭВМ установит, что этот сигнал указывает на принадлежность анализируемой словоформы в предложениях (5) и (6) к именной группе, а в предложении (7) — к глагольной группе.

Таким образом, определяя природу промежуточного граничного сигнала, сущность которого состоит в его полифункциональности, ЭВМ проводит формальный анализ контактных (поверхностных) связей между словоупотреблениями, не удаляясь далеко от анализируемой словоформы.

1.5. Задачи ЭВМ и требования к алгоритму и программе

Машина, располагая информацией о формальных структурных сигналах и порядке их расположения, может:

а) установить синтаксическую функцию слова как внутри рассматриваемой группы, путем проверки контактных внутригрупповых связей между словоупотреблениями, так и функцию граничного сигнала внутри предложения, путем проверки межгрупповых связей;

б) однозначно установить по формальным признакам начальную и конечную границы именных и глагольных групп и провести внутригрупповую индексацию компонентов группы;

в) однозначно определить начало придаточных предложений, причастного и герундиального оборотов.

Иначе говоря, на ЭВМ возлагается максимально возможный объем работ по формально-дистрибутивному анализу структуры предложения входного языка.¹² При этом различается понятие микроконтекста, или контекста группы (ядро группы и исходящие от него лексико-грамматические внутригрупповые связи) и контекста предложения — межгрупповых связей внутри предложения. Предложение на данном этапе работы является максимальной единицей анализа.

Описанный метод автоматического анализа накладывает определенные ограничения на описание структуры предложения:

1) границы должны быть формальными не столько с точки зрения лингвиста, сколько с точки зрения машины, т. е. они должны однозначно опознаваться ЭВМ;

2) списки граничных сигналов не должны быть слишком длинными (использование громоздких списков может привести к перегрузке памяти машины);

3) весь алгоритм синтаксического анализа должен быть компактным, чтобы также не перегружать память машины.

Таким образом, составление программы связано с решением конфликтной ситуации, сущность которой состоит в том, что она должна предусмотреть всю логическую последовательность шагов алгоритма и вместе с тем должна быть краткой.

Хотя мы и не можем предусмотреть абсолютно всех случаев выделения именных групп, однако путем статистических экспериментов мы обязаны выделить наиболее типичные случаи и обеспечить высокий процент правильных решений при машинном анализе текста.

1.6. Понятие именной и глагольной группы и задачи алгоритма

В качестве именной группы в алгоритме будут рассматриваться:

- 1) имя существительное нарицательное или собственное;
- 2) сочетание существительного с препозитивными определениями;
- 3) сочетание существительного с определяющими его предложными определительными конструкциями, выполняющими атрибутивную функцию;

¹² Предложение рассматривается как последовательность словоупотреблений, причем левой границей предложения является либо начало текста, либо комбинация «пробел — точка — пробел», а правой — комбинация «пробел — точка — пробел». Это определение предложения принято в качестве рабочего варианта.

4) сочетание существительного с постпозитивными определениями.¹³

В состав глагольной группы входят:

- 1) личные формы глагола (простые);
- 2) сложные (аналитические) формы глагола, а также глагольные словосочетания, в состав которых входят личные, указательные и другие местоимения (эти местоимения, таким образом, рассматриваются как своеобразные препозитивные аналитические форманты глагольных словоформ);
- 3) простые и сложные формы причастия;
- 4) герундий;
- 5) инфинитив;
- 6) прилагательные в функции предикативного члена составного именного сказуемого.

Алгоритм самостоятельного синтаксического анализа предложения рассчитан на однозначное опознавание ЭВМ:

- 1) именной группы (с последующей автоматической индексацией компонентов группы);
- 2) глагольной группы (с ее последующей индексацией);
- 3) начала придаточного предложения;
- 4) начала причастного и герундиального оборотов;
- 5) инфинитива (с его последующей индексацией);
- 6) устойчивых словосочетаний типа *to and fro, to and beneath, as well as, as far as, as soon as, as nearly as possible* и т. п., а также вводных слов и групп типа *first, in general, thus* и др.

Неизменным условием работы алгоритма является полное отсутствие предварительной ручной обработки текста, поступающего в ЭВМ. Алгоритмы автоматического синтаксического анализа можно подразделить на два типа: 1) алгоритмы, предусматривающие анализ целого предложения,¹⁴ 2) алгоритмы, анализирующие предложение, предварительно разделенное на отдельные сегменты.¹⁵

¹³ Ср. определение именной группы в работе: А. Л. Васильевский, Л. Г. Кравец, В. А. Москович, М. В. Самойлович, С. К. Поволоцкая, Г. А. Тарасова, Ю. М. Эмдина. *Англо-русский перевод*. М., 1967, стр. 146.

¹⁴ См.: Т. Н. Молошная. Алгоритм перевода с английского языка на русский. ПК, 1960, № 3; О. С. Кулагина. О машинном переводе с французского на русский. Там же; В. М. Лейкина, Г. С. Цейтин. Система англо-русского машинного перевода, разработанная в Вычислительном центре ЛГУ. В сб.: Тезисы докладов к XXV научно-технической конференции. . . Л., 1970; Ю. Г. Приймаков, В. А. Соркина. Алгоритм автоматического выделения именной группы в техническом тексте на английском языке. СТ II; М. В. Данейко, В. М. Петровская. Алгоритм автоматической сегментации английского предложения. Там же; S. Kuno, A. Oettinger. Multiple-path Syntactic Analyser. In: Proceedings of International Federation for Information Processing Congress 62. Amsterdam, 1963.

¹⁵ См.: И. А. Мельчук. Автоматический синтаксический анализ, т. I. Новосибирск, 1964; Л. Н. Иорданская. Автоматический син-

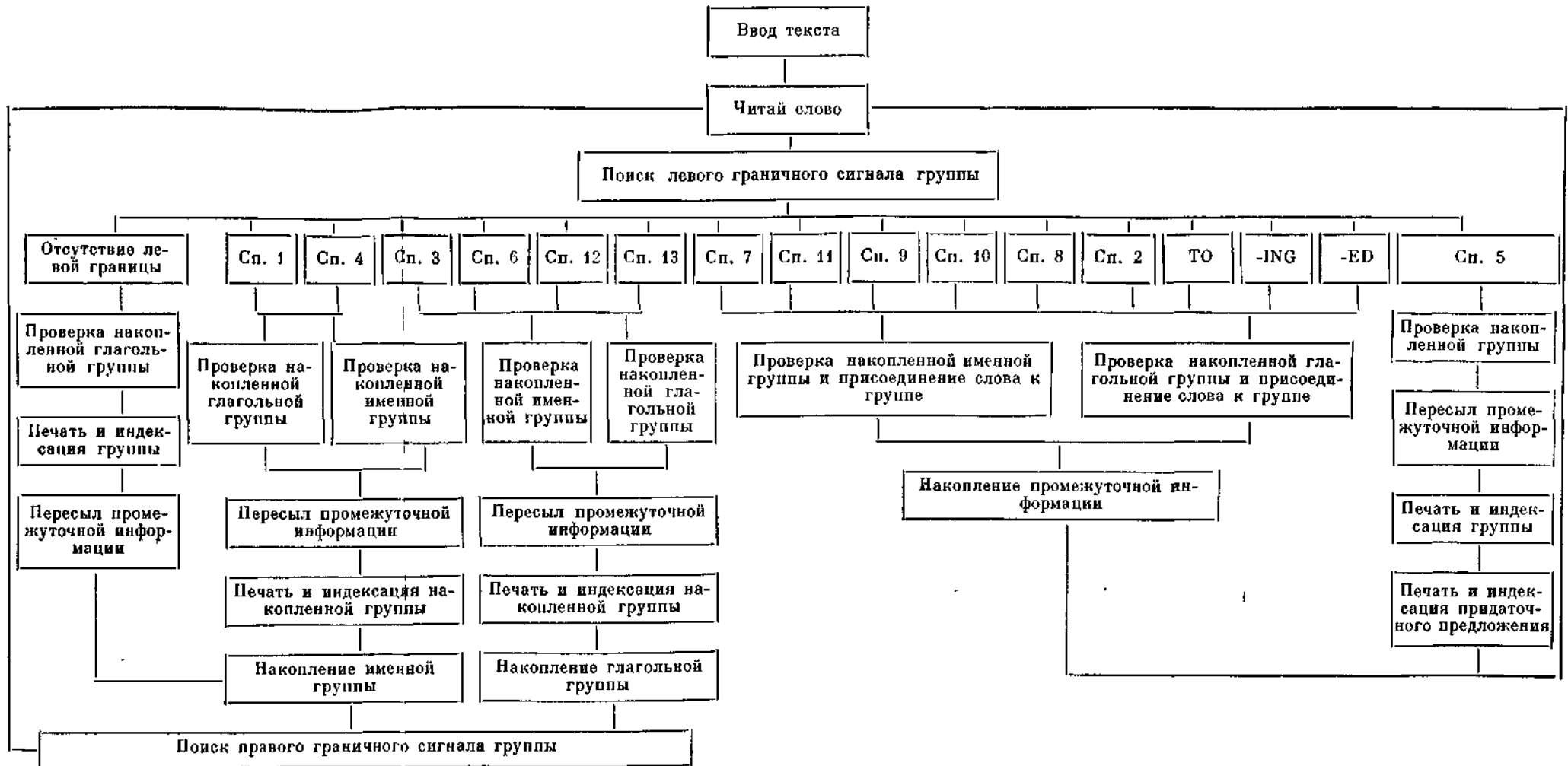


Рис. 2. Общая схема алгоритма.

Предлагаемый алгоритм анализа структуры предложения входного языка ориентирован на обработку целого предложения.

§ 2. АЛГОРИТМ СЕГМЕНТАЦИИ АНГЛИЙСКОГО ТЕКСТА

2.1. Алгоритм и принципы его работы

Принципиальная схема алгоритма анализа структуры предложения входного языка состоит из семи блоков — Б1, Б2, Б3, Б4, Б5, Б6, Б7 (рис. 1).

Б1	Ввод исходной синтаксической информации и текста
Б2	Поиск и индексация именной группы
Б3	Поиск и индексация глагольной группы
Б4	Поиск и индексация (дее)причастного оборота
Б5	Поиск и индексация начала придаточного предложения
Б6	Поиск и индексация устойчивых словосочетаний
Б7	Вывод индексированных групп

Рис 1.

Рассмотрим, как осуществляется анализ структуры предложения английского языка. Исходная синтаксическая информация и текст, кодированные на перфораторе, вводятся в оперативную память (МОЗУ) ЭВМ.

В МОЗУ предусмотрено наличие трех буферов информации: буфера именной группы (буфер 1), буфера глагольной группы (буфер 2) и буфера промежуточной информации (буфер 3), где накапливается информация во время анализа какой-либо группы.

Под буфером информации понимается определенное количество ячеек в оперативной памяти ЭВМ, откуда обработанная информация поступает на алфавитно-цифровое печатающее устройство.

таксический анализ, т. 2. Новосибирск, 1967; F. Alt, I. Rhodes. The Recognition of Clauses and Phrases in Mechanical Translation of Languages. Teddington, 1961; I. Rhodes. A New Approach to the Mechanical Syntactic Analysis of Russian. In: Mechanical Translation, 1961, № 6.

2.2. Общая схема алгоритма (рис. 2)

2.3. Инструкция по перфорации списков

1) На расстоянии 1.5—2.0 м пробить начало ввода в виде последовательности комбинаций: «русский — латынь» (граница ввода).

2) Пробить 10—15 см «русский».

3) Пробить комбинацию «латынь».

4) Очередное слово перфорируется вплотную в коде М-2.

5) Количество перфорируемых пробелов должно дополнять слово до длины, кратной 6. Например:

IN
THROUGH

6) Пп. 4 и 5 повторяются до исчерпания списка.

7) Признак конца ввода перфорируется в виде последовательности комбинаций: «латынь — русский» и 1.5—2.0 м пустой ленты («русский»).

2.4. Перфорация окончаний (сп. 8)

1) В целом, инструкция сохраняется за исключением пп. 4 и 5.

2) П. 4 — перед тем как перфорировать очередное окончание, необходимо дополнить пробелами длину окончания до длины, кратной 6.

3) П. 5 — окончание перфорируется, когда длина окончания вместе с пробелами составляет число, кратное 6. Например:

 LESS
 IC

2.5. Пример работы алгоритма

Рассмотрим работу алгоритма на анализе предложения The schematic for the audio amplifier used in the Westinghouse H-602P7 receiver is shown in Fig. 1 (8).

Читая текст слева направо, машина ищет левую границу первой группы, последовательно сравнивая прочитанную словоформу со списками формальных границ. Опознав граничный сигнал (в нашем примере the¹⁶), машина будет выполнять следующий цикл операций:

¹⁶ См. сп. 1 в приложении к алгоритму. Словоформы, включенные в этот список, являются безусловными инклюзивными или эксклюзивными границами именной группы, предусматривающими автоматическое занесение непосредственно следующей словоформы в буфер именной группы.

а) проверит наличие информации в буфере 1, т. е. в буфере именной группы;

б) проверит наличие информации в буфере 2 (глагольной группы);

в) проверит наличие информации в буфере 3 (промежуточной информации).

«Убедившись», что информация отсутствует в каждом из трех буферов, машина занесет the в буфер именной группы. Занесение the в буфер 1 обуславливает занесение в этот же буфер и словоформы schematic.

Выполнив этот цикл операций, машина читает следующее слово, начиная тем самым поиск правого граничного сигнала группы, часть которой уже находится в буфере 1. В анализируемом предложении следующим словом будет предлог for.¹⁷ Прочитав слово и начав последовательную проверку наличия информации в буферах информации, машина, естественно, обнаружит ее только в буфере именной группы.

В данном случае машина оказывается перед альтернативой, сущность которой состоит в том, что словоформа for является безусловным правым эксклюзивным граничным сигналом накопленной группы the schematic, с одной стороны, а с другой стороны — for может оказаться компонентом именной группы, если за for следует словоформа из сп. 2 или из сп. 7.¹⁸

Для того чтобы принять однозначное решение, машина временно засылает словоформу for в буфер промежуточной информации и читает следующее слово, которое необходимо для определения функции словоформы for, и последовательно сравнивает прочитанное слово the со сп. 2 и 7.

Не обнаружив тождества the со словоформами в указанных списках, машина «делает вывод», что словоформа for — безусловный эксклюзивный граничный сигнал именной группы the schematic и печатает накопленную именную группу.

Прежде чем осуществить печать именной группы, машина проводит ее анализ, т. е. устанавливает ядро группы и определяющие его компоненты.

Суть этого анализа состоит в следующем. Хотя ЭВМ и информирована, что последняя словоформа накопленной именной группы является ядром этой группы, на это условие накладываются следующие ограничения. Последняя словоформа группы при сравнении не должна отождествляться:

а) с \square ;

б) со словоформой из сп. 10.

Осуществив указанную проверку, ЭВМ определяет, что ядро группы является слово schematic, которому она сообщает

¹⁷ См. сп. 4. Словоформы, включенные в этот список, являются безусловными инклюзивными или эксклюзивными границами именной группы.

¹⁸ См. указанные списки в приложении к алгоритму.

индекс «п».¹⁹ Установив, что компонентом группы является в данном случае определенный артикль, который не индексируется, машина печатает всю группу, сообщив ей общегрупповой индекс «именная группа»:²⁰

THE SCHEMATIC_n

ИМЕННАЯ ГРУППА

Как указывалось выше, машина путем проверки контактной дистрибуции установила, что словоформа *for*, находящаяся в буфере промежуточной информации, является безусловным граничным сигналом предшествующей именной группы, и напечатала ее.

Теперь машина выполняет следующий цикл операций:

- а) пересылает словоформу *for* из буфера 3 в буфер 1;
- б) засылает словоформу *the*, прочитанную ранее для определения природы граничного сигнала *for*, в буфер 1;
- в) автоматически засылает сюда же следующую за *the* словоформу, т. е. *audio*.

Следующее слово *amplifier* машина также заносит в буфер именной группы. Подобное рассмотрение данного слова обусловлено тем, что, прочитав и последовательно сравнив *amplifier* со всеми списками граничных сигналов, ЭВМ не нашла ни одного структурного сигнала, предписывающего дальнейший анализ этой словоформы. Учитывая это обстоятельство, машина рассматривает *amplifier* на этом этапе анализа как компонент именной группы.

Анализируя словоформу *used*, ЭВМ находит, что данная словоформа оканчивается на *-ed*. Словоформа с такой морфемой может быть граничным сигналом, природа которого, однако, не позволяет однозначно установить:

- а) является ли словоформа с морфемой *-ed* компонентом именной группы, как в предложении *A figure X shows a forced draft blower limiting governor* (9), или
- б) анализируемая словоформа *used* — безусловная эксклюзивная граница именной группы *for the audio amplifier*.

Естественно, машина обращается к контексту. Запомнив слово *used*, вернее, поместив его в буфер 3, машина читает следующее слово *in*, которое помогает ЭВМ однозначно установить, что

¹⁹ Проведение внутригруппового анализа по определению ядра группы и определяющих его компонентов с последующей их индексацией диктуется необходимостью святия лексико-грамматической омонимии словоформ из машинного словаря при последующем машинном переводе выделенных групп. Данный алгоритм не получил практической реализации по техническим причинам.

²⁰ Присвоение общегруппового индекса обуславливается тем, как выделенная группа будет впоследствии переводиться на русский язык. По вопросу автоматической индексации см.: Н. Т. Carvell, J. Svartvic. *Computational Experiments in Grammatical Classification*. The Hague—Paris, 1969.

used — безусловная эксклюзивная граница накопленной именной группы.

Прежде чем напечатать группу и сообщить ей индекс, машина проводит внутригрупповой анализ и индексацию компонентов группы. Иными словами, машина устанавливает ядро группы *amplifier_n* и определяющий его компонент (ему приписывается индекс «а») — *audio₁*, а также фиксирует, что группа вводится предлогом *for_{pr}*. Как уже указывалось, артикль не индексируется. Вся группа приобретает следующий вид:

FOR_{pr} THE AUDIO₁ AMPLIFIER_n ИМЕННАЯ ГРУППА

Отпечатав эту группу, ЭВМ возвращается к дальнейшему анализу словоформы *used*, который необходим для установления природы безусловной границы. Теперь машина устанавливает:

- а) вводит ли словоформа *used* глагольную группу, как в предложении *The original plan suggested a rather conservative indirect cycle reactor system*, или,
- б) как в предложении (4), анализированном выше, эта словоформа вводит причастный оборот.

«Имея в виду» ранее прочитанную словоформу *in*, относящуюся к сп. 4, машина делает заключение, что словоформа с морфемой *-ed* вводит причастный оборот. Словоформа *used* пересылается из буфера 3 в буфер 2 (буфер глагольной группы) и печатается с общим индексом «начало (дее)причастного оборота»:

USED НАЧАЛО (ДЕЕ)ПРИЧАСТНОГО ОБОРОТА

Накопление и анализ именной группы *in the Westinghouse H-602P7 receiver* осуществляется аналогично накоплению и анализу двух вышеупомянутых именных групп.

Печать именной группы *in the Westinghouse H-602P7 receiver* осуществляется по словоформе *is*, отождествляемой со словоформой сп. 3.²¹ Выполняя печать именной группы, машина заносит словоформу *is* в буфер глагольной группы (буфер 2) и читает словоформу *shown*. Отождествив ее со словоформой сп. 11, ЭВМ присоединяет ее к словоформе *is* в буфере 2.

Внутригрупповой анализ с присоединением индекса «v»,²² печать и общегрупповая индексация осуществляются по следующей словоформе *in*:

IS SHOWN_v ГЛАГОЛЬНАЯ ГРУППА

Накопление именной группы *in Fig. 1* происходит по тому же принципу, что и накопление групп *the schematic, for the audio*

²¹ См. сп. 3 в приложении к алгоритму. Словоформы этого списка — безусловные эксклюзивные границы именной группы.

²² Индекс «v» получают простая глагольная форма, инфинитив или последний член группы, входящий в сложную аналитическую глагольную форму.

amplifier, in the Westinghouse H-602P7 receiver. Внутригрупповой анализ, общегрупповой индекс и печать осуществляется по □, которая воспринимается ЭВМ как отдельная словоформа, которая индексации не подлжит:

IN_{pr} FIG.1_n

ИМЕННАЯ ГРУППА²³

Выполнив печать группы, машина также проверяет наличие следующего за □ слова, в противном случае ЭВМ сообщает об окончании текста и о конце работы по его анализу.

Таким образом, предложение The schematic for the audio amplifier used the Westinghouse H-602P7 receiver is shown in Fig. 1 после автоматического анализа приобретает следующий вид:

THE SCHEMATIC_n

ИМЕННАЯ ГРУППА

FOR_{pr} THE AUDIO_n AMPLIFIER_n

ИМЕННАЯ ГРУППА

USED

НАЧАЛО (ДЕЕ)ПРИЧАСТНОГО ОБОРОТА

IN_{pr} THE WESTINGHOUSE_n H-602P7_n RECEIVER_n

ИМЕННАЯ ГРУППА

IS SHOWN_v

ГЛАГОЛЬНАЯ ГРУППА

IN_{pr} FIG. 1_n

ИМЕННАЯ ГРУППА

Такой автоматический синтаксический анализ вскрывает поверхностную структуру предложения входного языка, доступную непосредственному наблюдению, а общегрупповая индексация групп в терминах синтаксиса в х о д н о г о языка помогает снять лексико-грамматическую омонимию отдельных словоформ из машинного словаря.

Следует также отметить, что в созданном алгоритме предусмотрен вероятностный принцип анализа структуры предложения. Сущность этого принципа состоит в том, что при выделении именных групп без левого граничного сигнала алгоритм учитывает существующий в английском языке порядок слов.

Вероятностный принцип работы алгоритма можно проследить на предложении Last month we saw how some common accepted thoughts on transistors are not true. Прочитав словоформу last и последовательно проверив списки границ, машина не смогла отождествить ее ни с одной границей. «Учитывая» это обстоятельство и то, что первой группой в английском языке, как правило, является именная группа,²⁴ ЭВМ занесет last в

²³ На данном этапе внутригруппового анализа именной группы сочетания типа «Fig. 1» рассматриваются как единство. Это вызвано стремлением упростить автоматическую внутригрупповую индексацию.

²⁴ См.: А. И. Смирницкий. Синтаксис английского языка. М., 1957, стр. 61; Б. А. Ильиш. Строй современного английского языка. М.—Л., 1965, стр. 247 и сл.

буфер именной группы (буфер 1), туда же после анализа заносится и словоформа month. Накопленная именная группа после внутригруппового анализа выводится на печать по граничному сигналу we.

Сфера применения нашего алгоритма самостоятельного синтаксического анализа ограничена. Он рассчитан на автоматическую переработку синтаксиса английских научно-технических, деловых и публицистических печатных текстов и не предусматривает анализа текстов художественной литературы.

§ 3. МАШИННАЯ РЕАЛИЗАЦИЯ, ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ, АНАЛИЗ ОШИБОК И ПЕРСПЕКТИВЫ ДАЛЬНЕЙШИХ МАШИННЫХ ИССЛЕДОВАНИЙ

3.1. Машинная реализация

Алгоритм самостоятельного анализа структуры предложения входного языка состоит из стандартных блоков (см. рис. 2). Это дало возможность оформить эти блоки в виде стандартных подпрограмм (СП), которые включают в себе практически всю логику решения поставленной лингвистической задачи. Всего составлено семь СП совокупной длины примерно 1000 команд.

Работающие СП можно охарактеризовать следующим образом:

- СП-1 — «читай слово»;
- СП-2 — «проверь списки»;
- СП-3 — «засылка в буфер»;
- СП-4 — «печать буфера»;²⁵
- СП-5 — «проверь список окончаний»;
- СП-6 — «пересылка из буфера в буфер»;
- СП-7 — «опрос буфера».

Программа, состоящая исключительно из обращений к СП и команд условного перехода в зависимости от значения признаков, вырабатываемых СП, занимает 1400 команд. Программа рассчитана на реализацию на ЭВМ «Минск-22».

Программа, исходная синтаксическая информация и текст, кодированные на перфораторе, вводятся в оперативную память (ОЗУ) ЭВМ. Здесь же помещаются три буфера информации.

Реальное распределение объема оперативной памяти ЭВМ следующее:

- | | |
|----------------------------------|---------------|
| 1) рабочие ячейки с 0000 по 0130 | — 88 ячеек; |
| 2) перекодировка с 0130 по 0300 | — 104 ячейки; |
| 3) рабочее поле с 0300 по 0400 | — 64 ячейки; |
| 4) программа с 0400 по 3600 | — 1665 ячеек; |

²⁵ В СП-4 использована программа перекодировки, разработанная В. В. Мамонтовым.

исходит одновременно с установлением природы глагольной группы, т. е. в процессе анализа глагольной группы машина выясняет, что зависимой именной группой является словоформа *we*. Словоформе *we* присписывается индекс NP, словоформе *saw* — VP. Выполнив этот цикл операций, машина «смотрит», есть ли еще именная группа слева. Такой группой является именная группа *last_a month_n*. Проверив отношение этой группы к NP *we*, машина определит, что *last_a month_n* зависит не от *we*, а от *saw_v*.

«Убедившись», что за группой *last_a month_n* идет комбинация «пробел — точка — пробел», т. е. граничный элемент предложения, ЭВМ начинает обработку правой части предложения. Отметим, что справа идет словоформа *how_{НПН}*, машина начинает поиск главной глагольной группы, т. е. группы *are not true_v*. Определив ее природу, машина читает первую слева именную группу *on_{pr} transistors_n*. Поскольку группа вводится предлогом, ЭВМ читает следующую слева именную группу, которая должна подчинять группу *on_{pr} transistors_n*.

Такой группой в нашем примере является группа *some_a common_a accepted_a thoughts_n*.

В результате проведенного анализа машина «строит» дерево предложения (каждая группа внутри анализируется аналогично группе *we saw*) (рис. 3).

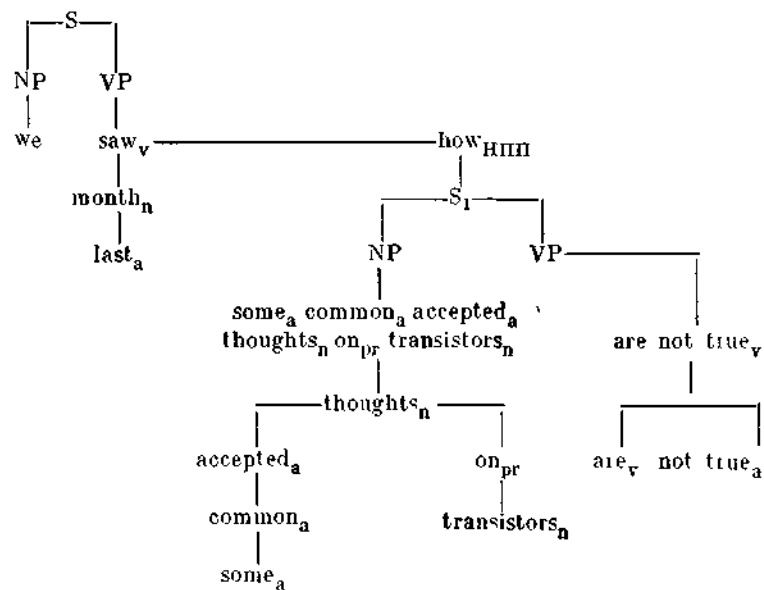


Рис. 3.

Автоматическое решение проблемы построения дерева предложения входного языка открывает возможность создания частот-

ного словаря нового типа, в котором будет отмечена частотность не отдельных словоупотреблений данного подъязыка, а частотность употребления отдельных типов деревьев входного языка.

Свернув деревья в символы, которым присписывается определенная синтаксическая функция, выполняемая ими в структуре предложения входного языка, и снабдив ЭВМ предписанием развешивания этих символов в структуре предложения выходного языка, можно будет значительно упростить и повысить эффективность бинарного перевода.²⁸

Другой путь использования дерева предложения в системе «человек — машина — человек» может быть следующим. Исходя из дуалистического (смыслового и формального) характера структуры предложения и возможности отделения плана формы выражения от плана формы содержания,²⁹ в процессе перевода можно выполнить следующие операции:

- 1) построить формальное синтаксическое дерево поверхностной структуры предложения входного языка;
- 2) соотнести только что полученное дерево с деревом глубинной структуры входного языка (или языка-посредника);
- 3) принимая глубинную структуру в качестве инварианта перевода, эксплицировать ее посредством поверхностной структуры предложения выходного языка.³⁰

3.5. Анализ ошибок

Обработка полученных результатов установила два типа ошибок:

- а) технические ошибки;
- б) лингвистические ошибки.

К первому типу следует отнести ошибки, связанные с кодированием текста на перфораторе. Искажение словоформы и знака препинания во время перфорации текста приводит к тому, что ряд словоформ, которые рассматриваются как граничные сигналы, вследствие их искажения, не опознаются ЭВМ как границы. Это в свою очередь, приводит как к неправильному анализу структуры предложения, так и к неправильной внутригрупповой индексации.

К текстам, имеющим такого вида ошибки, следует отнести тексты по судостроению, которые отперфорированы без знаков пунктуации.

Машинный анализ текстов по описанному алгоритму содержит и более серьезные ошибки — лингвистические. Эти ошибки в основном можно также подразделить на:

²⁸ Ср.: А. В. Гладкий, И. А. Мельчук. Элементы математической лингвистики. М., 1969, стр. 120—150.

²⁹ Ср.: Л. З. Сова. Аналитическая лингвистика, стр. 231.

³⁰ Ср.: О. З. Авербух. Формально-смысловые структуры английского языка. АКД. Калинин, 1970, стр. 1—4.

а) связанные с решением проблемы многозначности отдельных словоформ;

б) связанные с неполнотой некоторых списков граничных сигналов.

С ошибками первого типа, т. е. возникающими из-за многозначности отдельных словоформ, на данном этапе работы алгоритма пока приходится мириться. Во-первых, это обусловлено тем, что количество формализованной семантической информации, используемое в нынешнем алгоритме, недостаточно для успешного устранения ошибок этого типа. Во-вторых, в рамках данной работы не представляется возможным формализовать всю необходимую семантическую информацию, которая требуется для успешного устранения ошибок, связанных с проблемой многозначности отдельных словоформ.

Решение этой проблемы в рамках нынешнего алгоритма приводит к тому, что снятие многозначности отдельной словоформы и отнесение ее к конкретной группе в одном случае порождает неправильный анализ этой же словоформы в другом случае.

Сравним для примера два предложения:

All you do is disconnect the auto antenna from the AM radio and plug it into the receptacle.

The lead from the converter, terminated in an antenna plug, slips into the AM radio's antenna jack.

В обоих случаях мы будем рассматривать словоформу plug. Идеальным случаем анализа была бы следующая сегментация:

PLUG IT _v	ГЛАГОЛЬНАЯ ГРУППА
IN _{pr} AN ANTENNA _n PLUG _n	ИМЕННАЯ ГРУППА

В первом предложении словоформа plug — ядро глагольной группы plug it, во втором — ядро именной группы in an antenna plug.

Машина, опираясь на алгоритм в его нынешнем состоянии, рассматривает словоформу plug как компонент именной группы в обоих случаях.

Посмотрим, каковы пути улучшения анализа этой словоформы.

На первый взгляд кажется, что словоформу plug следует анализировать с применением этапа промежуточного анализа. При таком подходе словоформу plug можно анализировать двояко:

1) помещать словоформу plug полностью в буфер промежуточной информации (буфер 3) и исследовать ее контактную дистрибуцию как слева, так и справа;

2) рассматривать «искусственную» морфему -ug как промежуточный граничный сигнал.

Реализация обоих методов вызывает ряд возражений. Во-первых, если анализировать словоформу plug через буфер проме-

жуточной информации, аналогичной проверке необходимо подвергать каждую словоформу. Вместе с тем на данном этапе работы алгоритма трудно сказать, как проверка каждой словоформы через буфер промежуточной информации отразится на общей работе алгоритма. Вероятно, однозначность решения данного вопроса зависит от статистической обработки достаточно больших массивов текста.

Второй путь кажется еще менее приемлемым, так как, с одной стороны, он требует введения в список граничных элементов множества аналогичных «морфем» типа -ug, а с другой стороны, повышение эффективности анализа за счет увеличения списков граничных промежуточных элементов приводит в итоге к перегрузке памяти ЭВМ, что ставит вопрос о рентабельности всего алгоритма в целом.

Таким образом, для того чтобы устранить этот тип ошибки, необходимо или проводить статистическую обработку больших массивов текста, или проводить формализацию семантической информации, необходимой для устранения многозначности отдельных словоформ.

Ошибки второго типа, т. е. ошибки, связанные с неполнотой некоторых списков граничных элементов, в основном глагольных, устраняются путем обращения к двуязычному машинному словарю.

На настоящем этапе работы алгоритма ЭВМ мы пока еще не можем однозначно определить конец придаточного предложения, причастного и герундиального оборотов. Выделение формальных признаков, указывающих на завершенность этих групп, требует специального исследования, которое предполагается выполнить в дальнейшем.

Таким образом, обработка полученных машинных результатов показала, что чисто формальный анализ текста, осуществляемый ЭВМ, не может дать полного и вполне корректного синтаксического анализа. В каждом тексте после формального анализа обнаруживается остаток синтаксических явлений, который может быть обработан только путем обращения к семантической интуиции (форме плана содержания). В рамках МП это обращение может быть осуществлено либо путем обращения к автоматическому словарю или к программам снятия многозначности, либо путем предварительной индексации текста, поскольку во всех этих случаях мы имеем дело с формализацией этой семантической интуиции.

* * *

Оценивая алгоритм в целом, можно отметить, что он является компактным, т. е. он не ведет к перегрузке памяти ЭВМ, и рентабельным, позволяющим получать достаточно высокий результат анализа структуры предложения входного языка на ЭВМ, имеющейся в распоряжении лингвиста. Высокий положительный вы-

ход, т. е. правильный синтаксический анализ английского печатного текста, позволяет в дальнейшем использовать этот алгоритм в общей системе машинного перевода, создаваемой в группе «Статистика речи».

4.1 Приложение к алгоритму

- Сп. 1 — инклюзивные / эксклюзивные границы именной группы; эксклюзивные границы глагольной группы;
- » 2 — промежуточные границы именной и глагольной групп;
- » 3 — эксклюзивные границы именной группы;
- » 4 — инклюзивные / эксклюзивные границы именной группы; эксклюзивные границы глагольной группы;
- » 5 — инклюзивные границы именной и глагольной групп;
- » 6 — инклюзивные / эксклюзивные границы глагольной группы; эксклюзивные границы именной группы;
- » 7 — промежуточные инклюзивные границы групп;
- » 8 — » » » » ;
- » 9 — » » » » ;
- » 10 — » » » » ;
- » 11 — промежуточные границы именной и глагольной групп;
- » 12 — эксклюзивные границы именной группы;
- » 13 — инклюзивные / эксклюзивные границы глагольной группы; эксклюзивные границы именной группы.

Сп. 1

- | | |
|--------|----------|
| 1) a | 5) our |
| 2) an | 6) the |
| 3) my | 7) your |
| 4) its | 8) their |

— Указанные в сп. 1 словоформы входят в состав именной группы, если они являются левым граничным сигналом этой группы.

Сп. 2

- | | | |
|----------|------------------------|----------------|
| 1) all | 11) none | 21) anyone |
| 2) any | 12) same | 22) nobody |
| 3) few | 13) some | 23) anybody |
| 4) her | 14) such | 24) nothing |
| 5) his | 15) that ³¹ | 25) someone |
| 6) both | 16) this | 26) another |
| 7) each | 17) every | 27) anything |
| 8) many | 18) other | 28) somebody |
| 9) most | 19) these | 29) something |
| 10) much | 20) those | 30) everything |

³¹ Словоформа that после соответствующего анализа выводится на печать с пометой «начало придаточного предложения». Указанные словоформы выводятся в составе глагольной группы, если за ними следует глагол.

- | | | |
|----------|-------------|--------------|
| 1) be | 58) admit | 115) ionize |
| 2) do | 59) adopt | 116) obtain |
| 3) is | 60) agree | 117) occupy |
| 4) go | 61) allow | 118) passes |
| 5) add | 62) apply | 119) permit |
| 6) are | 63) arise | 12) remain |
| 7) buy | 64) avoid | 121) remove |
| 8) did | 65) began | 122) render |
| 9) fix | 66) begin | 123) reduce |
| 10) get | 67) bring | 124) rotate |
| 11) had | 68) carry | 125) modify |
| 12) has | 69) cause | 126) starts |
| 13) let | 70) cease | 127) suffer |
| 14) lie | 71) cover | 128) vanish |
| 15) mix | 72) drive | 129) varries |
| 16) put | 73) enter | 130) achieve |
| 17) say | 74) erase | 131) acquire |
| 18) see | 75) exist | 132) analyze |
| 19) adds | 76) holds | 133) applies |
| 20) buys | 77) imply | 134) believe |
| 21) came | 78) let's | 135) carries |
| 22) come | 79) leads | 136) collect |
| 23) deal | 80) leave | 137) compute |
| 24) does | 81) occur | 138) compare |
| 25) fall | 82) reach | 139) conduct |
| 26) feel | 83) refer | 140) consist |
| 27) fell | 84) relax | 141) diffuse |
| 28) felt | 85) serve | 142) discuss |
| 29) find | 86) spend | 143) examine |
| 30) gave | 87) think | 144) exhibit |
| 31) give | 88) write | 145) explain |
| 32) goes | 89) absorb | 146) express |
| 33) have | 90) accept | 147) furnish |
| 34) know | 91) adjust | 148) imagine |
| 35) knew | 92) affect | 149) implies |
| 36) keep | 93) anneal | 150) improve |
| 37) lies | 94) appear | 151) include |
| 38) live | 95) assume | 152) involve |
| 39) make | 96) attain | 153) observe |
| 40) meet | 97) became | 154) operate |
| 41) move | 98) become | 155) overlap |
| 42) obey | 99) behave | 156) perform |
| 43) puts | 100) cancel | 157) portray |
| 44) says | 101) denote | 158) prepare |
| 45) said | 102) depend | 159) prevail |
| 46) seem | 103) detect | 160) prevent |
| 47) show | 104) define | 161) proceed |
| 48) take | 105) differ | 162) produce |
| 49) tend | 106) employ | 163) protect |
| 50) took | 107) enable | 164) provide |
| 51) want | 108) ensure | 165) receive |
| 52) went | 109) equals | 166) release |
| 53) were | 110) exceed | 167) replace |
| 54) vary | 111) expect | 168) require |
| 55) uses | 120) extend | 169) suggest |
| 56) gets | 113) follow | 170) suppose |
| 57) sees | 114) happen | 171) sustain |

172) undergo	190) occupies	208) withstand
173) utilize	191) overcome	209) accelerate
174) compress	192) preclude	210) accomplish
175) concerns	193) reappear	211) contribute
176) consider	194) remember	212) coordinate
177) continue	195) restrict	213) correspond
178) decrease	196) solidify	214) degenerate
179) describe	197) alternate	215) distribute
180) discover	198) calculate	216) illustrate
181) endanger	199) determine	217) neutralize
182) estimate	200) eliminate	218) solidifies
183) evaluate	201) encounter	219) strengthen
184) generate	202) establish	220) understand
185) increase	203) intersect	221) understood
186) indicate	204) introduce	222) communicate
187) maintain	205) oscillate	223) concentrate
188) minimize	206) represent	224) familiarize
189) modifies	207) transform	225) incorporate

Сп. 4

1) at	15) over	29) during
2) by	16) upon	30) except
3) in	17) with	31) inside
4) of ³²	18) about	32) toward
5) on	19) above	33) versus
6) up	20) after	34) within
7) for	21) along	35) against
8) off	22) among	36) between
9) out	23) below	37) outside
10) via	24) under	38) through
11) down	25) around	39) towards
12) from	26) before	40) without
13) into	27) behind	41) according
14) onto	28) beyond	42) throughout

Сп. 5

1) as	6) when	11) while
2) if	7) whom	12) whose
3) how	8) until	13) unless
4) who	9) which	14) because
5) why	10) where	

Союзы и союзные местоимения выводятся на печать с пометой «начало придаточного предложения» и являются правой границей как именной, так и глагольной группы.

Сп. 6

1) I	4) we	7) they
2) it	5) she	
3) he	6) you	

Личные местоимения входят в состав глагольной группы и являются правым граничным сигналом глагольной и именной группы.

³² Предлог выводится в составе именной группы.

Сп. 7		
1) me	3) him	4) them
2) us		

Словоформы списка выводятся на печать в составе той группы, компонентом которой они являются.

Сп. 8

1) -al	7) -se	13) -ete	19) -ous
2) -ar	8) -age	14) -ful	20) -ult
3) -el	9) -ant	15) -ine	21) -ute
4) -ic	10) -ary	16) -ire	22) -able
5) -le	11) -ate	17) -ite	23) -ible
6) -ly	12) -ent	18) -ive	24) -less

Словоформы с выделенными морфемами после соответствующего анализа выводятся на печать в составе или именной, или глагольной группы.

Сп. 9

1) no	27) soon	53) direct
2) so	28) then	54) enough
3) due	29) this	55) indeed
4) not	30) thus	56) itself
5) now	31) true	57) present
6) too	32) very	58) rather
7) yet	33) well	59) seldom
8) able	34) what	60) slight
9) also	35) again	61) though
10) away	36) ahead	62) unique
11) back	37) apart	63) within
12) best	38) hence	64) already
13) dead	39) never	65) complex
14) easy	40) often	66) however
15) even	41) rapid	67) perhaps
16) fast	42) ready	68) thereby
17) full	43) right	69) somewhat
18) here	44) quite	70) straight
19) less	45) still	71) together
20) long	46) there	72) whenever
21) just	47) twice	73) according
22) main	48) abrupt	74) otherwise
23) once	49) almost	75) sometimes
24) only	50) always	76) therefore
25) poor	51) across	77) nevertheless
26) pure	52) common	

Словоформы выводятся в составе той группы, компонентом которой они являются.

Сп. 10

1) or	4) since	7) whereas
2) and	5) either	8) although
3) but	6) whether	

Словоформы выводятся в составе той группы, компонентом которой они являются.

Сп. 11

1) got	11) left	21) drawn
2) fed	12) lost	22) built
3) led	13) held	23) bound
4) met	14) kept	24) grown
5) won	15) shown	25) chosen
6) been	16) given	26) driven
7) made	17) found	27) written
8) seen	18) heard	28) brought
9) sent	19) taken	29) forbidden
10) done	20) known	

В алгоритме указанные словоформы могут явиться левым граничным сигналом именной группы. В этом случае словоформа выводится на печать с пометой «начало причастного оборота».

Сп. 12

1) can	4) must	7) shall
2) may	5) would	8) should
3) will	6) could	9) cannot

Глаголы являются левым или правым граничным сигналом именной группы; в составе сложной глагольной группы перечисленные словоформы являются компонентом этой группы.

Сп. 13

1) I'm	7) it'll	13) you're
2) I'll	8) she's	14) you've
3) I've	9) we're	15) they'll
4) he's	10) we've	16) they're
5) it's	11) she'll	17) they've
6) he'll	12) you'll	

Являясь граничным сигналом как именной, так и глагольной группы, указанные словоформы выделены в отдельный список для восстановления полной формы вспомогательного глагола.

О. А. Афаунова

АВТОМАТИЧЕСКОЕ УСТРАНЕНИЕ ГРАММАТИЧЕСКОЙ МНОГОЗНАЧНОСТИ АНГЛИЙСКИХ ING-ОВЫХ ФОРМ

§ 1. ЗАДАЧИ И МЕТОДИКА ИССЛЕДОВАНИЯ

Программа автоматического синтаксического анализа английского текста, описанная в статье И. И. Волосевича (см. здесь, стр. 293—322), дает наибольшее число неправильных решений при анализе ing-овых словоформ. В связи с этим возникла необходимость построить детальный алгоритм и машинную программу распознавания этих форм. Решению этой задачи и посвящена настоящая работа.

Широко представленные в современных английских текстах ing-овые формы¹ обладают большой многозначностью в рамках английской одноязычной ситуации. Эта неоднозначность намного увеличивается, когда мы обращаемся к двуязычной ситуации, в русле которой осуществляется англо-русский МП (ср. таблицу).

Чтобы добиться автоматического устранения двуязычной неоднозначности, мы должны задать ЭВМ исчерпывающий перечень формальных детерминант в контекстном окружении ing-овой формы, детерминант, с помощью которых эта форма может быть соотнесена с конкретным русским грамматическим эквивалентом.

К сожалению, предназначенные для русских грамматики английского языка и пособия по переводу (см. прим. к таблице) не всегда дают исчерпывающую информацию, опираясь на которую, можно было бы строить машинный алгоритм. Обычно для каждой из ing-овых форм предлагается по два и больше равноправных вариантов перевода, при этом не дается никаких указаний на то, что же определяет выбор того или иного эквивалента.

Чтобы получить информацию о типах дистрибуции ing-овых форм, приходится обратиться к исследованию параллельных текстов. Всего было просмотрено 4.5 тыс. употреблений ing-овых форм

¹ По нашим подсчетам на 1 000 словоупотреблений английского научного текста встречается обычно от 20 до 30 ing-овых форм.

Одноязычная ситуация		Двуязычная ситуация
морфологическая природа ing-овых форм	синтаксические функции ing-овых форм	способы передачи ing-овых форм в русском переводе
<i>pI</i> (active, passive)	а) опр., стоящее слева от определяемого	действительное прич. с суффиксами <i>-ущ, -ющ, -ащ, -ящ</i>
	б) опр., стоящее справа от определяемого	а) страдательное прич. с суффиксами <i>-им, -ем</i> слева от определяемого б) причастный оборот справа, иногда от глагола <i>подвергаться</i> + сущ. в) л. ф. глагола в составе определительного придаточного предложения
<i>pI</i> (active, passive, perfect)	обст. (времени, условия, причины, сопутствующих обстоятельств)	а) деепр. б) предложная группа сущ. в) деепр. сов. вида г) придаточное предложение с союзами <i>если, когда, поскольку, так как</i>
The Objective Predicative Construction <i>pI</i>		а) преимущественно придаточное предложение с союзами <i>что, как, чтобы</i> б) (<i>pI</i> от глагола <i>to be</i>) не переводится в) сущ. г) инф.
Participial Phrase	а) часть определительного причастного оборота б) обст.	а) прич. б) определительное придаточное предложение
Accusative with the Participle	а) часть сложного дополнения б) обст.	деепр. дополнительное придаточное предложение
<i>pI</i>	составная часть сказуемого	а) деепр. б) придаточное предложение
Nominative Absolute Participle Construction	опр. обст.	глагол в л. ф.
<i>pI</i> +to+Infinitive+Construction Gerund ger.	а) опр. б) подлежащее доп. (прямое или косвенное)	а) обстоятельное придаточное предложение б) самостоятельное предложение с союзами <i>причем, а, и</i> в) определительное придаточное предложение а) отглагольное сущ. б) инф. а) сущ. (предложная группа сущ.) б) отглагольное сущ. в) инф. г) придаточное предложение с союзами <i>то, что, что</i>

Одноязычная ситуация		Двуязычная ситуация
морфологическая природа ing-овых форм	синтаксические функции ing-овых форм	способы передачи ing-овых форм в русском переводе
ger. passive perfect, ger. + possessive pronoun	доп.	а) глагол в л. ф. в составе придаточного предложения с союзом <i>то, что</i> б) сущ.
ger. от глагола <i>to have</i>	доп.	а) придаточное предложение с союзом <i>то, чтобы</i> б) отглагольное сущ.
ger.	обст.	а) предложная группа сущ. со словами <i>путем, посредством, после</i> и т. п. б) сущ. в косвенном пад. без предлога
ger. (passive perfect)		а) придаточное предложение с союзом <i>то, что</i> б) сущ.
The Objective Predicative Construction with ger.	доп.	а) преимущественно придаточное предложение как для простых, так и для сложных форм герундия б) сущ. без изменения порядка слов в) инф.
Три конечные буквы ing-овой формы не являются формантом, а относятся к основе словоформы		а) л. ф. глагола б) сущ.

Примечание. Описание морфологической природы и синтаксических функций ing-овых форм в английском языке дано на основании информации, почерпнутой из следующих работ: С. С. Fries. *The Structure of English*. London, 1957; G. O. Curme. *Grammar of the English Language*. London-New York, 1931. Н. С. Н о u s e, I. E. H a r m a n. *Descriptive English Grammar*. New York, 1965.

Русские эквиваленты определены исходя из следующих работ: Ж. М. З е н и н а. Техника перевода английской специальной литературы. Казань, 1971; Е. Н. З в е р е в а. Грамматические соответствия английского и русского языков. Л., 1962; Грамматические трудности при переводе английской научной литературы. М., 1969.

в контекстах ² (это соответствует тексту длиной в 150 тыс. словоупотреблений) и их русских эквивалентов.

Опираясь на гипотезу глубины В. Ингве,³ мы исследовали левые и правые дистрибуции ing-овых форм длиной не более восьми

² Привлекались тексты разнообразной тематики: геофизика, электроника, приборостроение, органическая химия, прикладная лингвистика, логика.

³ V. Y n g v e. A. Model and a Hypothesis for Language Structure. In: *Proceedings of the American Philosophical Society, Philadelphia, 1960, № 5*; В. И. И н г в е. Вычислительные машины и перевод. Русск. пер. АП.

словоупотреблений каждая. При этом в контекстах *ing*-овых форм всегда имеются словоформы, которые никогда не выступают в качестве *ing*-овых детерминативов (например, наречия, отрицательные частицы, прилагательные). Словоформы такого типа из окружения *ing*-ов исключаются. Они устраняются и при автоматическом анализе в приводимой блок-схеме (см. блок-фильтр на рисунке).

Анализ и группировка дистрибуций *ing*-овых форм позволили выделить типовые «схемы» окружения *ing*-ов. Например, предложения типа: *Atomic oxygen was produced by subjecting molecular nitrogen to a microwave discharge* 'Атомарный кислород образовывался путем воздействия на молекулярный азот высокочастотным разрядом'; *The cratering may be implemented by fracturing of the rock* 'Кратерообразование может происходить вследствие разломов горных пород' объединялись моделью:

to be+ed (pp)+предлог (артиклъ)+ing-предлог (артиклъ)

Элементы схем задавались списками. Всего выделено 40 таких списков. Однако и анализ дистрибуций не всегда однозначно определяет выбор того или иного эквивалента *ing*-ов. Выясняется, что одна и та же *ing*-овая форма в одном и том же тексте может передаваться различными русскими грамматическими эквивалентами. Например, предложению *Water drops may freeze forming nucleation centers* соответствовал перевод: 'Капли могут замерзнуть и образовать зародыши градин'. Вместе с тем грамматики (см. таблицу) допускают и такой перевод: 'Капли могут замерзать, образуя зародыши градин'. При выборе эквивалента переводчик часто руководствуется соображениями стиля, благозвучности, опираясь при этом на свое лингвистическое чутье. Так, в указанном примере переводчик передал словоупотребление *forming* с помощью инфинитива потому, что в предыдущем предложении уже употреблялся деепричастный оборот.

Естественно, что реализовать все эти интуитивные возможности человека-переводчика на ЭВМ невозможно. Поэтому приходится закреплять за каждой *ing*-овой формой один эквивалент, руководствуясь при этом функцией *ing*-овой формы в английском предложении и частотой употребления каждой из ее возможных интерпретаций.

Вариативность в выборе переводчиком грамматических эквивалентов должна быть отнесена за счет многих переменных. Эту неясность мы попытались преодолеть с помощью тестового эксперимента.⁴

⁴ Ср.: A. Kaplan. An Experimental Study of Ambiguity and Context. MT, 1955, 2, № 2.

Из 4.5 тыс. проанализированных случаев употребления *ing*-овой формы было отобрано 80 наиболее типичных ситуаций, в которых одна и та же *ing*-овая форма была переведена разными способами. Эти микроконтексты были предложены для перевода пятидесяти студентам Ленинградского ГПИ им. А. И. Герцена. Тот вариант перевода, который выбрало наибольшее количество информантов — носителей русского языка, считался наиболее типичным и принимался за основу при построении алгоритма. Так, например, для сегмента . . .*may freeze forming*. . . были предложены следующие переводы:

...могут замерзать, образуя...	(24 чел.);
...могут замерзать и образовывать...	(16 чел.);
...могут замерзать с образованием...	(5 чел.);
...при замерзании образуется...	(3 чел.);
...замерзая, образуют...	(2 чел.);

Естественно, что наиболее подходящим для МП будет первый, набравший большее число «голосов», вариант.

Итак, при составлении алгоритма распознавания *ing*-овых форм используются свидетельства грамматик и пособий по переводу, данные по дистрибутивному анализу контекстов и, наконец, результаты тестового эксперимента.

§ 2. ОПИСАНИЕ АЛГОРИТМА

Алгоритм состоит из семи частей. Первую часть составляет предварительная обработка, или блок-фильтр. В нем окружение *ing*-овой формы подвергается следующим операциям.

1) Исключаются *ing*-овые формы, переводной эквивалент для которых можно установить однозначно без анализа окружения. Это словоформы с окончанием *-ings*; словоформы, в которых *-ing* не является суффиксом (сп. 1, 2 и 3). Такие случаи сразу выдаются на печать с пометой «существительное», «местоимение» или «глагол». Слова типа *foregoing*, *existing* (см. 9) сразу индексируются как причастия. Формы типа *neighbouring*, *startling* (сп. 21) индексируются как прилагательные и т. д.

2) Окружение *ing*-овой формы «очищается» от слов, не существенных для анализа. Они заданы списком словоформ (9) и списком суффиксов (8).

Центральной частью алгоритма является блок-диспетчер. Наличие такого блока позволило сократить алгоритм в 4—5 раз и значительно упростить анализ. Диспетчер делит все *ing*-овые формы по наличию (признак +) или отсутствию (признак 0) артикля и(ли) предлога в левом и правом окружении *ing*-овой формы. В результате работы блока-диспетчера каждая *ing*-овая форма подвергается испытанию на одну из четырех возможных ситуаций (по признаку: + *ing* +, + *ing* 0, 0 *ing* +, 0 *ing* 0). Такое

WHAT NICKNAME IS BE'NS USED ГЛАГОЛ.МВ.НС.В.
 AMONG OTHER THINGS KNOWS SOME PARTICULAR THING СУЩЕСТВ.ТЕЛНОЕ МНОЖ.
 AS PRIME IDENTIFYING CRITERIA СУЩЕСТВ.ЕД.У.
 WERE IDEALLY EFFICIENT AT KEEPING TRACK OF HIS СУЩЕСТВ.ЕД.У.
 CONDITION FOR TRANSFERRING KNOWLEDGE FROM ONE STATE ИМЕН.ДЕСЯТЫЙ.
 IS A TERM DENOTING A PERSON СУЩЕСТВ.ЕД.У.
 SEQUENCE WAS THE MEANING THE SET OF ЯДРО ПРМЧ.ОБОР.
 OBTAIN THE FOLLOWING SEQUENCE СУЩЕСТВ.ЕД.У.
 CONCEPTS ACCORDING TO THE THEORY ДЕЙСТВИТ.ПРМЧ.
 AFTER HAVING EXPRESSED THE VALUE ПРОЧМЕ ОБОРОТМ
 АЛФВР.ОБОР.С.В.

Рис. 1. Машинный результат.

ALGEBRAIC VALUE CORRESPONDING TO IT ПРОЧМЕ ОБОРОТМ
 OF PHILOSOPHY ARE MEANINGS OF WORDS СУЩЕСТВ.ТЕЛНОЕ МНОЖ. Ч
 SUCH STUDIES IS MISSING TO A CRESCEN ГЛАГОЛ.МВ.НС.В.
 LANGUAGES HE IS LOOKING IN THE GAME ГЛАГОЛ.МВ.НС.В.
 THE RELATIVITY OF MEANINGS OF WORDS ГЛАГОЛ.МВ.НС.В.
 ALL THE IMPROVING ARRAY OF EVIDENCE СУЩЕСТВ.ТЕЛНОЕ МНОЖ. Ч
 THE SCHOLLIUM FOLLOWING STATEMENT ДЕЙСТВИТ.ПРМЧ.
 THEY ARE REPRESENTING ДЕЙСТВИТ.ПРМЧ.
 BE A CROSSING POINT OF MANY ACTIVITIES ГЛАГОЛ.МВ.НС.В.
 СУЩЕСТВ.ЕД.У.

подразделение позволяет ограничить количество основных интерпретаций ing-овой формы до одной (существительное) в ситуации 1, двух (существительное и причастие) в ситуации 2, трех (существительное, причастный оборот, деепричастный оборот) в ситуации 4. Ситуация 5 не исключает ни одной из возможных интерпретаций, но снятие многозначности в пределах этой группы значительно упрощается в результате работы предыдущих блоков алгоритма. Между ситуациями 2 и 4 вклинивается не зависящая от признака ситуация 3, где рассматриваются аналитические формы глагола, причастия и деепричастия. Такое положение ситуации 3 в алгоритме объясняется тем, что ситуации 1 и 2 исключают появление в исследуемой группе аналитической формы, а ее появление в ситуациях 4 и 5 исключается потому, что все они реализуются в ситуации 3.

Дальнейший анализ внутри каждой из ситуаций проводится путем сравнения окружения анализируемой формы со списками признаков, характерных для одного из переводных эквивалентов. При этом анализ в этих блоках построен по методу исключения сначала менее вероятного варианта. Например, в ситуации 2 ing-овую форму можно перевести либо с помощью существительного, либо с помощью причастия. Причем наиболее вероятно появление в этой ситуации причастия. Тогда проверяются сначала все возможности по переводу ing-овой формы существительным, и в случае отрицательных ответов машина определяет ing-овую словоформу как причастие. Такой способ значительно сокращает поиск, уменьшает загрузку памяти ЭВМ.

На выходе алгоритм дает в терминах выходного языка следующую информацию об ing-овых формах:

- 1) существительное (с указанием числа);
- 2) прилагательное;
- 3) причастие действительного залога;
- 4) глагол (с указанием числа, времени, вида, залога);
- 5) инфинитив;
- 6) деепричастие (с указанием вида),
- 7) ядро причастного оборота;
- 8) ядро деепричастного оборота;
- 9) *не* + деепричастие;
- 10) нулевой перевод;
- 11) препозитивные, союзные и прочие обороты.

Реализация алгоритма осуществлена на ЭВМ БЭСМ-4 (см. рис. 1).

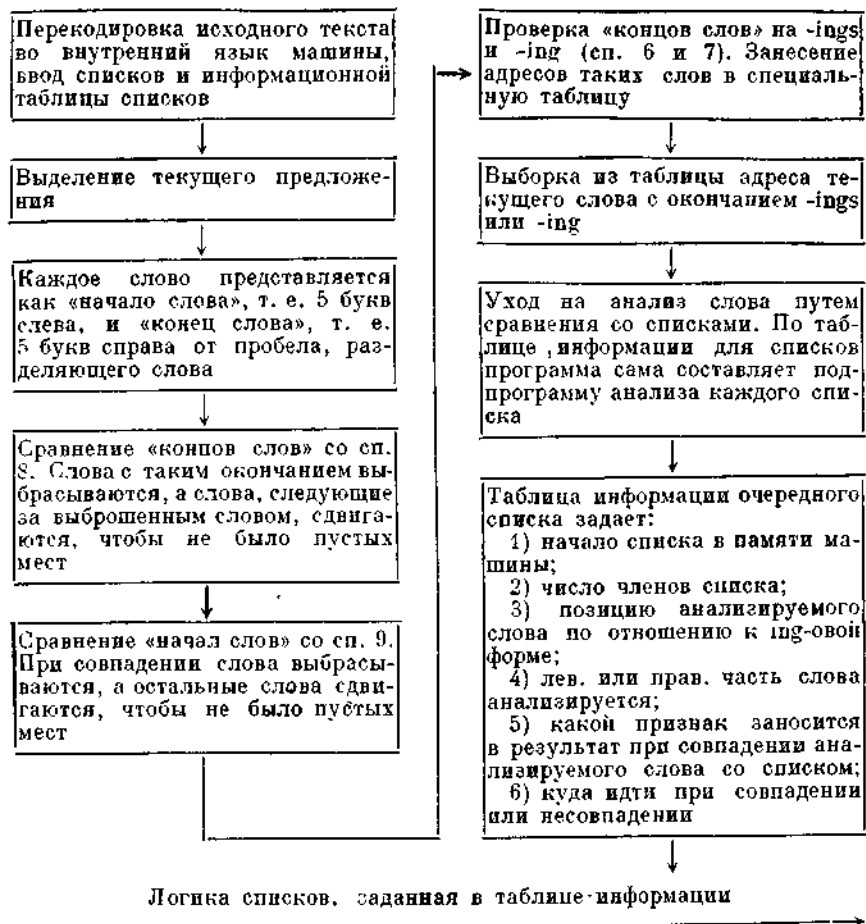
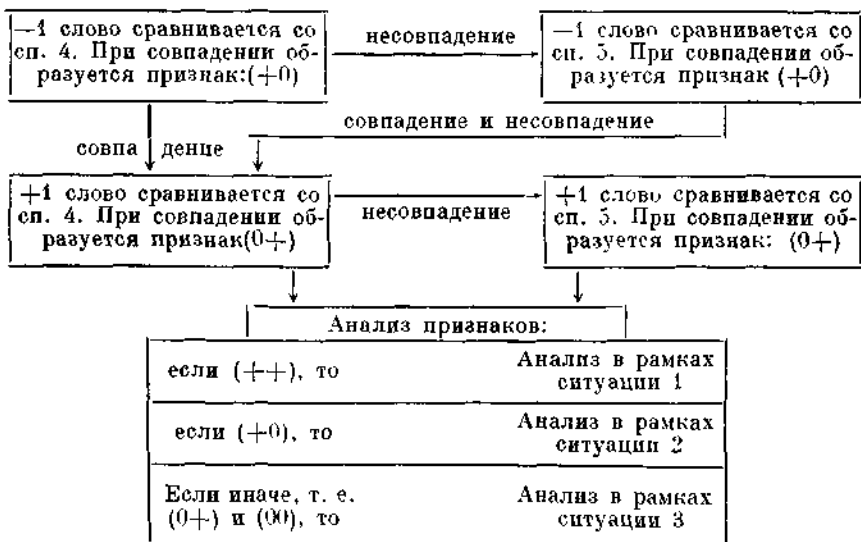


Рис. 2. Блок-схема алгоритма.

Блок-фильтр		
Позиция анализируемого слова	№ списка для сравнения	Что печатается при совпадении
0	1	глагол наст. вр., несов. вида
0	2	сущ. ед. ч.
0	3	местоимение
0	6	сущ. мн. ч.
0	15	анализ в ситуации 3
0	14	анализ в ситуации 3
0	37	препозитивные, союзные и проч. обороты

Блок-диспетчер



Ситуация 1		
Позиция слова	№ списка	Что печатается при совпадении
0	10	действительное прич.
-2	29	не+деепр. несов. вида, ядро оборота
-2	30	инф.
-2	28	деепр. несов. вида, ядро оборота
При несовпадении с перечисленными списками печатать: сущ. ед. ч.		

Ситуация 2

Позиция слова	№ списка	Что печатается при совпадении
0	10	действительное прич.
+1	32	сущ. ед. ч.
-1	36	» » »
-2	30	инф.
+1	13	сущ. ед. ч.
+1	12	» » »
+1	11	» » »
+1	18	» » »
-2	12	» » »
-2	13	» » »
+2	11	» » »
-1	29	деепр. несов. вида, ядро оборота
+2	26	сущ. ед. ч.
+2	12	» » »
+2	13	» » »
0	21	прил.
-1	28	не+деепр. несов. вида
+1	26	сущ. ед. ч.
-2	18	» » »
-1	38	глагол наст. вр., несов. вида
При несовпадении с перечисленными списками печатать: действительное прич.		

↓
Ситуация 3

№ операции	позиция анализируемого слова	№ списка для сравнения	что печатается при совпадении
1	0	21	прил.
2	0	10	действительное прич.
3	0	15	при несовпадении в № 18
4	+1	25	инф.
5	-1	16	глагол наст. вр., несов. вида
6	-1	17	» прош. » » »
7	+1	18	» наст. » » »
8	+1	19	» » » » »
9	-1	11	деепр. несов. вида, ядро оборота
10	+1	4	» » » » »
11	-2	12	» » » » »
12	-2	13	» » » » »
13	-2	33	» » » » »
14	+1	5	глагол наст. вр., несов. вида
15	-1	5	» » » » »
16	-1	11	» » » » »
17	При несовпадении с перечисленными списками: нулевой перевод		
18	0	14	при несовпадении в № 28
19	-1	18	деепр. сов. вида, ядро оборота
20	+1	25	инф.
21	+1	18	деепр. сов. вида, ядро оборота
22	+1	20	глагол прош. вр., несов. вида
23	+1	29	деепр. сов. вида, ядро оборота
24	-1	36	» » » » »
25	-2	12	» » » » »
26	-2	13	» » » » »
27	При несовпадении с перечисленными списками печать: ядро причастного оборота		
28	-1	16	глагол наст. вр., несов. вида
29	-1	17	глагол прош. вр., несов. вида
30	-1	20	при несовпадении в № 33
31	-2	34	неопределенная форма глагола
32	-2	35	глагол прош. вр., несов. вида
33	-1	22	при несовпадении в № 36
34	-2	23	глагол буд. вр., несов. вида
35	-2	24	глагол наст. вр., несов. вида
36	-1	31	неопределенная форма глагола
37	-33	39	неопределенная форма глагола

↓
Анализ ситуации

если (0+), то анализ в рамках ситуации 4
если (00), то анализ в рамках ситуации 5

Ситуация 4

позиция анализируемого слова	№ списка для сравнения	что печатается при совпадении
+1	25	сущ. ед. ч.
-1	26	действительное прич.
-1	28	деепр. сов. вида, ядро оборота
-1	26	сущ. ед. ч.
-1	12	ядро причастного оборота
-1	13	» » »
-1	33	» » »
0	27	деепр. несов. вида, ядро оборота
-2	26	сущ. ед. ч.
-1	38	глагол наст. вр., несов. вида
0	21	прил.
-1	11	деепр. несов. вида, ядро оборота
При несовпадении с перечисленными списками печать: ядро причастного оборота		

ситуация 5

позиция анализируемого слова	№ списка для сравнения	что печатается при совпадении
0	10	действительное прич.
0	21	прил.
+1	13	сущ. ед. ч.
+1	11	» » »
-1	30	глагол буд. вр., несов. вида
+1	18	сущ. ед. ч.
+1	32	» » »
+1	12	» » »
+1	33	действительное прич.
-1	12	ядро причастного оборота
-1	13	ядро причастного оборота
-1	33	ядро причастного оборота
0	27	деепр. несов. вида, ядро оборота
-1	29	» » » » »
-1	28	» » » » »
+1	26	сущ. ед. ч.
При несовпадении печать: ядро причастного оборота		

СПИСК И ЭЛЕМЕНТОВ ОКРУЖЕНИЯ
ИНГ-ОБЫХ ФОРМ

Ср. 1

bring, cling, fling, sing, spring, sting, swing, wring.

Ср. 2

evening, happening, hearing, morning, railing, shilling, stocking, string, wing, forecasting, launching, thing, tubing, ceiling, earnings. . .

Ср. 3

something, anything, nothing, everything.

Ср. 4

at, by, in, on, of, up, off, for, out, via, down, from, over, into, onto, upon, with, about, above, after, along, among, under, before, around, behind, to, beyond, during, inside, except, toward, versus, within, against, between, through, across, without, upward, downward, instead.

Ср. 5

an, a, the, this, that, these, one, it, any, both, several, such, my, your, our, its, them, those, since, because, if, as, until, till, what, which, when, where, whom, who, while, than, thus, his, her, each, every, same, some, though, although, other, another, further, together, preliminary, maximum, minimum, complete, constant, like, later, latter, first, many, few, all, former, no, but, there, so, certain, us.

Ср. 6

-ings.

Ср. 7

-ing.

Ср. 8

-ly, -less, -ous, -ive, -ible, -ic.

Ср. 9

not, nor, or, however, neither, either, already, ever, perhaps, never, itself, themselves, himself, yet, also, again, even, seldom, often, still, always, very, less, just, more, most.

Ср. 10

foregoing, existing, seeming, preceding, following, surrounding, prevailing, ensuing, dominating, exceeding, oncoming, incoming, consisting, ascending, descending, simulating.

Ср. 11

[] , []

Ср. 12

purpose, method, period, process, energy, velocity, stage, range, time, effect, rate, region, theory, tendency, system, condition, case, point, act, coefficient, contact, position, center, error, hour, technique, band, curve, ground, concept, feature, way, mechanism, mission, temperature, criteria, deformation.

Ср. 13

difficulty, possibility, means, object, alternative, sort, idea, cost, type, study, efficiency, advantage, complication, result, convenience, role, experiment, requirement, degree, mode, character, instrument, equipment, property, suitable, due, occurrence, kind, useful, effective, convenient, efficient.

Ср. 14

having.

Ср. 15

being.

Ср. 16

am, is, are.

Ср. 17

was, were.

Ср. 18

-ed.

Ср. 19

got, fed, let, met, won, been, made, seen, taken, sent, done, hold, known, kept, shown, given, left, found, heard, bound, shosen, driven, written, brought, forbidden, left, lost, drawn, built, come, become, shaken.

Ср. 20

been.

Ср. 21

surprising, startling, encouraging, astounding, conflicting, living, outstanding, convincing, concluding, nonliving, stimulating, remaining, counting, magnifying, distinguishing, neighbouring, promising, irritating, terrifying.

Ср. 22

be.

shall, will. Cr. 23
to. Cr. 24
of. Cr. 25
cause, undergo, involve, perform, permit, keep, ignore, acquire, produce, effect, mark, justify, prevent, insure, avoid, intensify, complicate, lead, expect, endure, begin, enjoy, arise, imply, formulate, constitute, find, possess, occur, depend. Cr. 26

Cr. 27
inflicting, choosing, confirming, noting, imposing, causing, producing, separating, including, excluding, assuming, integrating, liberating, dividing, meaning, differentiating, introducing, seeking, adopting, supporting, building, implying, believing, reflecting, equating, substituting, putting, recalling, analysing, exhibiting, awaiting, allowing, arguing, postulating, claiming, transforming, investigating, forming, commenting, denoting, postponing, omitting, neglecting, looking, enabling, inserting, inverting, using, ordering, indicating, saving, standardising, beginning.

Cr. 28
thus, when, while, though, although, but.

Cr. 29
without.

Cr. 30
capable, incapable, able, unable, view, hope, worth, reason, seem, basis, capacity, probability, succeed, insist, keep, help, start, begin, try, gift, risk, purpose, futility.

Cr. 31
dare.

Cr. 32
is, are, was, were, will, has, have, had, can, may, would, should, must, whether, could, cannot, do, don't, can't, shall, did.

Cr. 33
claim, image, definition, region, object, function, structure, edge, formation, motion, science, boundary, current, part, theory, expression, ratio, occasion, transformation, part, body, term, antenna, action, device, projectile, force, concept, group, discussion, proposition, reference, particle, value, law, material, property, difference, surface, liquid, agent, assumption, field, appearance, branch, barrier, factor, power, activity, feature, radar, cloud, spectrograph.

Cr. 34
has, have.

Cr. 35
had.
Cr. 36
on, upon, in, by, after, when, before.
Cr. 37
according, owing, regarding, disregarding, corresponding, respecting, depending, during, concerning.
Cr. 38
each, all, both.
Cr. 39
may, must, could.
Cr. 40
doing, making, initiating, keeping, dealing, questioning, saying, precipitating, synchronising, diluting, maintaining, giving, pursuing, clinging, satisfying, sizing, occurring, bringing, taking.

**СТАТИСТИКА РЕЧИ
И АВТОМАТИЧЕСКИЙ АНАЛИЗ ТЕКСТА
1972**

*Утверждено к печати
Институтом языкознания АН СССР*

Редактор издательства А. А. Эмрин
Художник М. И. Ралуевич
Технический редактор Г. А. Смирнова
Корректоры Н. П. Журавлева и Г. Н. Суворова

Сдано в набор 6/II 1973 г. Подписано к печати 13/Х
1973 г. Формат бумаги 60×90^{1/4}. Печ. л. 21^{1/4}+1 вкл.
(³/₄ печ. л.)=21,63 усл. печ. л. Уч.-изд. л. 24,03
Изд. № 4930. Тип. зак. № 71. М-13708. Тираж 3300.
Бумага № 3. Цена 1 р. 14 к.

Ленинградское отделение издательства «Наука»
199164, Ленинград, Менделеевская лин., д. 1

1-я тип. издательства «Наука»
199034, Ленинград, 9 линия, д. 13

ИСПРАВЛЕНИЯ И ОПЕЧАТКИ

Страница	Строка	Напечатано	Должно быть
43	9 снизу	reighbours	neighbours
46	5 сверху	reference s,	reference n,
72	15 »	erhãli	erhält
74	2 снизу	weiden	werden
77	5 сверху	werden	weiden
195	18 снизу	'инспирировать'	'инспирировать'
276	4-й стлб., 7 снизу	-ну-	-чу-

Статистика речи