

8. ОЗНАКОМИТЕЛЬНЫЙ ФРАГМЕНТ

И-75

МИНИСТЕРСТВО ПРОСВЕЩЕНИЯ РСФСР

Ленинградский государственный ордена Трудового
Красного Знамени педагогический институт имени А.И. Герцена

М.П. ЗОННИС

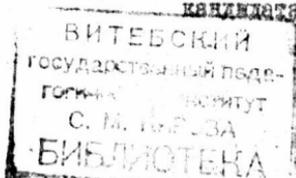
СТАТИСТИКА И ВЕРОЯТНОСТНЫЙ МАШИННЫЙ ПЕРЕВОД
ФРАНЦУЗСКОГО ГЛАГОЛА НА РУССКИЙ И МОЛДАВСКИЙ ЯЗЫКИ
(на материале французских текстов по публицистике).

(Диссертация выполнена на русском языке)

Ю664 Романские языки

Автореферат

диссертации на соискание ученой степени
кандидата филологических наук



Ленинград, 1972.

81.07.031
1175

ОЗНАКОМИТЕЛЬНЫЙ ФРАГМЕНТ

МИНИСТЕРСТВО ПРОСВЕЩЕНИЯ РСФСР

Ленинградский государственный ордена Трудового
Красного Знамени педагогический институт имени А.И. Герцена

М.П. ЮНИЦЭ

СТАТИСТИКА И ВЕРОЯТНОСТНЫЙ МАШИННЫЙ ПЕРЕВОД
ФРАНЦУЗСКОГО ГЛАГОЛА НА РУССКИЙ И МОЛДАВСКИЙ ЯЗЫКИ
(на материале французских текстов по публицистике).

(Диссертация выполнена на русском языке)

Ю664 Романские языки

Автореферат

диссертации на соискание ученой степени
кандидата филологических наук

Ленинград, 1972.



* 20504165 *

ОЗНАКОМИТЕЛЬНЫЙ ФРАГМЕНТ

Диссертация выполнена на кафедре
романской филологии Ленинградского
государственного ордена Трудового
Красного Знамени педагогического
института им. А.И.Герцена.

Научный руководитель: доктор филологических наук,
профессор Р.Г. ПИОТРОВСКИЙ.

Официальные оппоненты:

Доктор филологических наук, профессор Е.А.РЕФЕРОВСКАЯ
(ст.научный сотрудник Ленинградского Отделения Института
языкознания АН СССР, спец. романские языки).

Кандидат технических наук Е.П.ЛЮБИЧЕВ (начальник кафедры
Военной Краснознаменной ордена Ленина Академии Связи,
спец. техническая кибернетика).

Ведущее научное учреждение: кафедра иностранных языков
Московского экономико-статистического института.

Автореферат разослан *4 апреля 1972*

Защита состоится *6 июне 1972* на заседании

Ученого Совета факультета иностранных языков Ленинградского
государственного ордена Трудового Красного Знамени
педагогического института им. А.И.Герцена по адресу:

Ленинград, Д-186, ул.Мойка, 48, кор.14.

С диссертацией можно ознакомиться
в фундаментальной библиотеке института.

Ученый секретарь Совета
(доц. Е.С.НОВГОРОВОВА).

ОЗНАКОМИТЕЛЬНЫЙ ФРАГМЕНТ

Растущий поток научно-технической информации может быть "укрошен" только путем использования разных видов автоматической переработки текста. Машинный поиск, индексирование и каталогизирование, реферирование и перевод на ЭВМ текстов и документов позволяют в несколько раз сократить время доступа специалиста к научно-технической и деловой информации.

Всякая автоматическая переработка речи осуществляется в системе человек-ЭВМ-человек.

Несмотря на быстроту и устойчивость обработки информации, устойчивость ее хранения, современные ЭВМ обладают очень ограниченными объемами памяти, несоизмеримыми с богатством естественного языка. Поэтому, при решении на ЭВМ лингвистических задач необходимо искать какие-то лингвистические решения, чтобы уложиться в "лимит" памяти машины. В поисках этих решений можно опереться на два свойства языка.

Во-первых, каждый текст, написанный на естественном языке, обладает колоссальной избыточностью: в среднем эта избыточность равна 70-80%.

Во-вторых, язык распадается на разновидности (функциональные стили, "подязыки"), каждая из которых обладает ограниченным по сравнению с языком в целом числом лингвистических единиц и связей.

Указанные соображения приводят нас к выводу о том, что промышленный МП целесообразно строить в виде бинарного отраслевого перевода, опирающегося на некоторое сокращенное описание естественного языка - так называемый базовый язык (БЯ)¹⁾

1) Р.Г.Пиотровский. "Отраслевой вероятностный машинный перевод". Автоматическая переработка текста. Минск, 1970, стр.8.

ОЗНАКОМИТЕЛЬНЫЙ ФРАГМЕНТ

В настоящей работе таким БЯ является подъязык уголовной хроники и спорта публицистической речи французского языка.

При разработке проблем бинарного МП исходным приемом является определение общих информационных характеристик языка, а также выявление наиболее частых и информационно насыщенных элементов текста, которые дают необходимые и наиболее важные сведения при построении лингвистических алгоритмов и их реализации на ЭВМ. Иными словами, составлению алгоритмов автоматического анализа и МП предшествует статистическое описание текстов данного подъязыка. Для того чтобы увязать результаты статистического описания текста с построением алгоритмов его автоматической переработки необходимы лингвистические и машинные эксперименты. Именно на этом пути - пути проб и ошибок, можно получить более или менее надежную методику для результативного решения проблем автоматической переработки текста, в частности вероятностного МП. По такому руслу и выполнялась настоящая работа.

Г л а в а п е р в а я

Частотный словарь глагольных цепочек

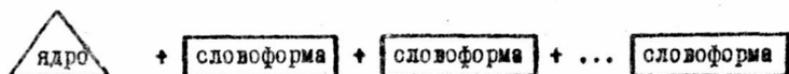
с опорным частотным глаголом

Так как наша работа имеет прикладной характер, то при определении единицы измерения нашего подъязыка большое значение имеет выбор грамматической концепции. Грамматика актуального членения, трансформационная грамматика, дискриптивная грамматика и другие грамматики не подходят для машинной реали-

ОЗНАКОМИТЕЛЬНЫЙ ФРАГМЕНТ

зации, так как они недостаточно формализованы. Каждая из них сохраняет в себе элементы семантической интуиции носителя языка.

В качестве исследуемой дистрибутивно-статистической единицы лингвистической совокупности, выделяемой формальным путем, устанавливается многоэлементная глагольная цепочка (ГЦ) (триада, тетрада, пентада, гексада) по схеме:



Такие многоэлементные цепочки вскрывают контактные связи и обнаруживают некоторые лексические валентности и грамматические особенности 100 наиболее частотных опорных глагольных словоформ (ядер), покрывающих около 50% французских текстов подъязыка уголовной хроники и спорта публицистической речи.

Нам нужна предельная формализация. Этим требованиям при выборе единицы измерения лингвистического эксперимента отвечает дистрибутивная грамматика, использующая некоторые идеи грамматики непосредственных составляющих. Для построения такой грамматики необходим статистический эксперимент, осуществленный на достаточной выборке.

Для расчета объема необходимой выборки (N) используется формулой:

$$N = \frac{(Z_p)^2}{\delta^2 f} \quad (1)$$

где: Z_p - константа, f - относительная частота,
 δ - величина относительной ошибки.

ОЗНАКОМИТЕЛЬНЫЙ ФРАГМЕНТ

ВЫДЕЛЕНИЕ ГЦ ИЗ ПРЕДЛОЖЕНИЙ И ИХ ПЕРЕВОД НА РУССКИЙ
И МОЛДАВСКИЙ ЯЗЫКИ

CRUT QUE L' - PASSE(SIMPLE ACTIF
ПОДУМАЛ (А О) - КРЕЗУ

E2TAIT ATTEINT DE - IMPARFAIT PASSIF
БЫЛ (А О) ПОРАЖЕН (А О) - ЕРА АТМНС

446 UN VERBE2 DE CM MOIS A E2TEZ ENLEVEZ DANS SON
LANDERU , ALORS QUE SA MEARE EST ALLEZE FAIRE
DES EMPLÈTTES .

A E2TEZ ENLEVE2 DANS - PASSE(COMPOSE(PASSIF
БЫЛ (А О) ПОЖИМЕН (А О) - А ФОСТ РЭПНТ

EST ALLEZE FAIRE - PASSE(COMPOSE(ACTIF
ПОВЕЛ (ШЛА ШЛО) - А МЕРС

447 MMARINA , FIL'E DU SCE2NORISTE ALLEMAND MMANS
MMARE , QUI AVAIT DISPARU , A E2TEZ TUEZE AB
COUPS DE COUTEAU .

AVAIT DISPARU - PLUS-QUE-PARFAIT ACTIF
ИСЧЕЗ (АА АО) - АМСПРВСЕ

A E2TEZ TUEZE AB - PASSE(COMPOSE(PASSIF
БЫЛ (А О) УБИТ (А О) - А ФОСТ УМНСО

448 LE LOURD VE2HICULE A DEZRAPE2 SUR UNE PLAQUE
DE VERGLAS AU CARREFOUR D' EENTRAIQUE , PUIS
A PERCUTEZ UN PLATANE ET S' EST RENVERBE2 AU
BEAU MILIEU DE LA CHAUSSEZE .

A DEZRAPE2 SUR - PASSE(COMPOSE(ACTIF
ЗАНЕС (АА АО) - А ДЕРАПАТ

A PERCUTEZ UN - PASSE(COMPOSE(ACTIF
УДАРИЛ (А О) - А ЛОВИТ

S' EST RENVERBE2 AU - PASSE(COMPOSE(PRONOMINAL
ПЕРЕВЕРНУЛСЯ (АСЬ ОСЬ) - С' А РЭСУРНАТ

ОЗНАКОМИТЕЛЬНЫЙ ФРАГМЕНТ

Основные положения диссертации отражены в следующих работах:

1. Об усвоении имперфекта на основе его значений способа действия (*mode d'action*). БГПИ им.А.Руссо. Тезисы международной конференции по вопросам методики преподавания иностранных языков в вузах и школах. Бельцы, 1965; стр.1,5 (совместно с Н.В.Шеремета).
2. Модальные перифразы во французском и молдавском языках. БГПИ им.А.Руссо. Материалы научной конференции преподавательского состава. Бельцы, 1965; стр.2.
3. Методика составления частотных словарей. БГПИ им.А.Руссо. Материалы научной конференции преподавательского состава. Бельцы, 1966, стр.2.
4. Студия статистика а вокабуларулуй. Тезеле реферателор конференцией шtiinificе, consacratе проблемelor де филологие молдовеняскэ ни чей 50 де ань ай Путерий Советиче. БГПИ им.А.Руссо, Бельцы, 1967, стр.1.(совместно с Н.Г.Матканом).
5. Комбинэриле де кувиите - обьекте де студиу ал "микросинтаксей". БГПИ им.А.Руссо. Тезеле реферателор конференцией шtiinificе, consacratе проблемelor де филологие молдовеняскэ ни чей 50 де ань ай Путерий Советиче. Бельцы, 1967, стр.2.
6. Анализ статистической модели французских и молдавских публицистических текстов. ЛГПИ им. А.И.Герцена. "Филологические науки". Краткое содержание докладов научно-теоретической конференции, посвященной 50-летию Великой Октябрьской Социалистической Революции. Ленинград, февраль, 1968 ; стр.2 (совместно с Н.Г.Матканом).

ОЗНАКОМИТЕЛЬНЫЙ ФРАГМЕНТ

7. Статистика трехсловных сочетаний в языке французской презы. ЛГПИ им.А.И.Герцена. "Филологические науки". Краткое содержание докладов научно-теоретической конференции, посвященной 50-летию Великой Октябрьской Социалистической Революции. Ленинград, февраль 1968; стр.3 (совместно с Н.М.Буравцовой и А.А.Орловой)

8. Диахрония глагольных словосочетаний во французских газетных текстах. ИГПИИЯ. Проблемы диахронии в изучении романских языков. Тезисы докладов всесоюзной конференции по изучению проблем романских языков. Минск, 1967, стр.2

9. Диатеза - капитал де базэ ал методичий де предраре а лимбий францезе. В жур. "Ыльэцэторул советик", Кишинев, № 6, 1968, стр.6.

10. Программа машинного перевода французских глагольных форм на русский и молдавский языки. БГПИ им.А.Руссо. Ученые записки, Бельцы, 1970 г., стр.30 (совместно с В.М.Петровской).

11. Пути устранения ошибок при дальнейшей работе над алгоритмом МП французского глагола на русский и молдавский языки. КПИ им.С.Лазо. "Автоматическая переработка текста методами прикладной лингвистики". Тезисы докладов всесоюзной конференции. Кишинев, 1971, стр.1 (совместно с В.М.Петровской).

12. Автоматическое различение омонимии в молдавских глагольных формах. КПИ им.С.Лазо. "Автоматическая переработка текста методами прикладной лингвистики". Тезисы докладов всесоюзной конференции. Кишинев, 1971, стр.1 (совместно с Н.Г.Маткашом).

34.

Отпечатано на ротопринте завода им. В.И.Ленина г. Бельцы МССР

Ваяз № 221 Тираж 200 экз. III - 1972 г. АБ 03100