

СПЕЦИАЛЬНЫЕ ТЕКСТЫ И ИХ ТЕРМИНОЛОГИЯ В ЧЕЛОВЕКО-МАШИННОМ ОБЩЕНИИ

В статье описываются методы АПТ (автоматической обработки текста), принадлежащих различным приоритетным областям и обладающих различными структурами. Особое внимание уделяется признакам отбора терминологической лексики в автоматический словарь.

Заманчивая идея общения человека и компьютера, а именно — идея переводить с одного языка на другой с помощью машины — родилась еще в середине 50-х годов. Одним из старейших коллективов, реализовавших эту цель, является

международная группа «Статистика речи» (СтР)¹. В ее состав входит Учебно-научный центр компьютерных исследований языка и речи (УНЦЯР) Российского государственного педагогического университета им. А. И. Герцена. УНЦЯР разрабатывает компьютерную систему машинного перевода (МП) устных и письменных текстов, относящихся к различным предметным областям (ПО) и неоднородных по своей структурированности (жестко нормированных текстов, текстов полужесткого и свободного построения). В задачи этой системы входит поддержка переводческой деятельности, лексико-грамматический анализ иноязычного текста, в том числе текста, извлекаемого из сети Интернет, организация помощи в преподавании языков². Системы МП являются центральными модулями многоязычного полифункционального лингвистического автомата (ЛА).

За более чем сорокалетнюю историю МП за разработку таких систем брались сотни групп в СССР, Европе, США, Японии и Китае³. Создать же работающие системы удалось лишь немногим коллективам. Здесь одно из ведущих мест наряду с американскими системами СИСТРАН и ГЛОБАЛИНК занимает многоязычная система СИЛОД, созданная в УНЦЯР. Конкурентоспособность СИЛОДа обеспечивается применением к языку информационно-статистической и многоуровневой методологии⁴.

Технология построения системы МП

Вышеназванная методология продиктована открытой, нечетко-множественной и ассоциативно-динамической природой естественного языка (ЕЯ) и онтологическими барьерами, отгораживающими его от языка компьютера⁵.

Пошаговое сокращение неопределенности в ходе переработки текста. Од-

ной из ключевых идей развития указанной технологии является принцип пошагового сокращения неопределенности, в которой оказывается ЛА, — неопределенности, порождаемой, с одной стороны, многозначностью словарных ЛЕ, морфологических форм и синтаксических схем, содержащихся в тексте, а с другой — недостатком лингвистических и энциклопедических знаний.

Первый шаг по снятию этой неопределенности осуществляется в блоке нормализации, задача которого заключается в адаптации входного текста к самому алгоритму анализа текста, а также к профессиональным интересам потребителя, на которого ориентирован перевод. Блок нормализации строится из двух подблоков.

Задачей первого является выявление в тексте иноязычных вставок, которыми чаще всего являются названия фирм, технических устройств или их модулей, рекламных фрагментов и т. п. (обычно на английском или французском языках). Не выявленные и не отсеянные иноязычные вставки становятся помехами на пути корректного лексико-грамматического анализа научно-технического текста. Ср. текст, написанный на французском языке (английские вставки выделены жирным шрифтом):

(...) **Keymage** 3.0 parvient à stocker tout type de document, quel que soit son format. Il accepte les documents bureautiques (**Word**, **Excel**, etc.), les documents provenant de messageries, du **Web** et de logiciels **ERP (enterprise resource planning)** ou de **data mining**. (...)

Au-delà du classement et de la recherche documentaire, **Keymage** propose des fonctions de **work-flow**. La gestion des tâches est disponible à travers la ressource fonctionnelle intitulée «**work cycles**» et enrichie dans cette troisième version. (...)

Dans ce **workflow**, l'utilisateur est alerté des nouvelles tâches à effectuer par l'affichage — discret — de leur nombre. Il accepte ou refuse de faire une nouvelle tâ-

che. Dans l'affirmative, il dépose le dossier sur son bureau; par ce **check out**, il devient le seul à pouvoir le modifier. Une fois le travail effectué, le dossier est replacé dans l'armoire (**check in**) et son indexation est mise à jour comme pour tout dossier intégrant l'arborescence des ressources. La fonctionnalité offerte par les **work cycles** paraît pratique, souple, adapté à des actions plutôt élémentaires.

Le recours à une messagerie, également proposée par **Keymage**, complète les fonctionnalités dans le sens du **groupware**. En même temps, le logiciel autorise l'usage de **Lotus Notes** et autres **Microsoft Exchange**.

(Archimag: les technologies de l'information (Paris).
Mai 1999. № 124. P. 35–36).

Алгоритмическая методика опознания этих вставок описана в работе: Р. Г. Пиотровского, с. 141—142 — см. п. 5а примечаний.

При необходимости эти вставки могут переводиться с помощью соответствующих иноязычно-русских систем МП.

Задача второго подблока состоит в определении жанрово-тематической специфики документа. Дело в том, что информация об этой специфике дает возможность ЛА использовать в ходе лексико-грамматического анализа терминологический словарь и грамматический алгоритм, строго соответствующий узкой тематике и жанру документа. Так, например, относящиеся к одной и той же предметной области научно-техническая статья и рекламный текст заметно различаются по своей синтаксической организации и по выбору лексики. Ср. французские тексты подъязыка информатики и вычислительной техники:

Научно-технический текст

(...) Le développement considérable qu'a connu Internet ces dernières années, notamment à la suite de l'apparition du World

Wide Web, a conduit à une croissance exponentielle du nombre d'utilisateurs du réseau et du nombre de documents accessibles aux utilisateurs. Des quelques centaines de milliers de textes accessibles au tout début du Web, en 1993, le stock d'information sur Internet atteignait près de 800 millions de documents en février 1999 soit un volume de données d'environ 15 téraoctets (millions de caractères), dont 6 téraoctets de texte pur. En fait, ces chiffres eux-mêmes sous-estiment la réalité car de très nombreux documents sont générés dynamiquement par programme à partir de bases de données, ce qui fait qu'il n'est pas totalement absurde de considérer le Web comme virtuellement infini.

Mais la complexité du Web est encore bien supérieure à ce que la taille seule du corpus pourrait laisser supposer. En effet, les documents sur le réseau sont très hétérogènes (sites commerciaux, pages personnelles, livres, articles et journaux en ligne, etc.), ne possèdent aucune véritable structure, sont écrits dans plus de cent langues différentes, utilisent sans toujours l'expliquer plusieurs encodages incompatibles des alphabets non latins, et sont très souvent dupliqués à plusieurs exemplaires, à l'identique ou avec des variations très mineures. Il s'agit, en fait, de toute une partie du savoir humain réuni, en vrac, sur le même support.

(La Recherche: Mensuel (Paris).
Février 2000. № 328. P. 66)

Рекламный текст

L'édition 1999 de
TERMIUM
sur CD-ROM
et le nouveau
TERMIUM Plus®

Il suffit de cliquer pour
accéder rapidement
à ce dictionnaire
bilingue du
gouvernement fédéral

L'édition 1999 de **TERMIUM® sur CD-ROM**

L'édition 1999 de **TERMIUM® sur CD-ROM** constitue l'édition la plus récente du dictionnaire électronique anglais-français, français-anglais du Bureau de la traduction du gouvernement du Canada qui a déjà reçu un prix d'excellence. (...)

L'une des plus grandes bases de données linguistiques au monde, **TERMIUM® sur CD-ROM** réunit plus de trois millions de termes et d'appellations utilisés dans presque tous les domaines du savoir, en particulier l'informatique, les affaires et l'administration, les sciences et la technologie. (...)

Plus de temps perdu à feuilleter des ouvrages de référence. Quelques secondes vous suffiront pour trouver l'information dont vous avez besoin, qu'il s'agisse de l'orthographe d'un toponyme canadien ou de l'emploi du point-virgule.

TERMIUM® sur CD-ROM est facile à utiliser. (...)

TERMIUM® sur CD-ROM est non seulement facile à utiliser, il permet aussi de gagner du temps et de travailler de façon plus efficace. Il ne faut que quelques secondes pour effectuer une recherche (...)

TERMIUM Plus® — Pour la terminologie la plus à jour

Si vous travaillez dans un domaine de pointe, où la terminologie évolue rapidement, **TERMIUM Plus®** est un outil dont vous ne pouvez vous passer. (...)

Et comme il est possible de consulter **TERMIUM Plus®** sur Internet, aucune installation n'est nécessaire. Il suffit de communiquer avec le Bureau de la Traduction pour obtenir un mot de passe.

Pour joindre les rangs des milliers de communicateurs, rédacteurs, réviseurs et traducteurs qui utilisent déjà **TERMIUM®** quotidiennement pour améliorer la qualité de leurs communications, il suffit de téléphoner

aux Services à la clientèle du Bureau de la traduction.

Le grand génie terminologique! LE dictionnaire bilingue par excellence!

(Reclama Бюро переводов в Канаде.

Site Internet:

www.bureaudelatraduction.gc.ca)

Рекламный текст, составленный канадской фирмой-производителем электронного англо-французского, французско-английского словаря, отличается от научно-технического текста вербальным кодом, а также наличием характерного для этого жанра невербального кода (к невербальному коду относится в данном случае цветовая гамма: сочетание черного, голубого и красного цветов). Оба кода приведенного рекламного текста служат одной цели — воздействовать на потенциального потребителя.

Для вербального кода свойственно употребление таких эмоциональных синтаксических моделей, как восклицательные предложения, эллиптические конструкции (Plus de temps perdu a feuilleter des ouvrages de référence. **TERMIUM Plus®** — Pour la terminologie la plus à jour.), повторы (il suffit de..., ...est facile à utiliser), употребление синонимичных конструкций (quelques secondes vous suffiront..., il ne faut que quelques secondes...). Лексика приведенного рекламного текста отличается употреблением в каждом абзаце неологизма **TERMIUM**, служащего названием выпущенного продукта. С целью привлечения внимания потребителя используется эмоционально-насыщенная лексика: d'excellence, par excellence, le grand génie. Все перечисленные приемы реализуют такие важные функции рекламного сообщения, как эмотивная, или экспрессивная и фатическая, или контактоустанавливающая.

Завершив нормализацию текста, МП-модуль ЛА приступает непосредственно к переводной переработке документа. Ее

основная идея состоит в пошаговом сокращении неопределенности, которое идет снизу вверх от лексических фактов к семантико-синтаксической и прагматической цели.

МП-модуль строится из следующих подблоков:

1) *словарный*, где производится перевод отдельных словоупотреблений и словосочетаний входного текста,

2) *микросегментный*, где выполняется перевод отдельных именных словосочетаний, а также вербальных групп,

3) *макросегментный*, где достигается опознание и обработка функциональных сегментов, т. е. групп подлежащего, сказуемого и т. д.,

4) *синтаксический*, на котором выявляется поверхностная синтаксическая структура входного предложения и подбирается соответствующая ей выходная схема,

5) *текстовый*, где, исходя из конкретной тематики, жанра и построения текста, производится итоговая корректировка перевода.

Каждый блок должен отвечать за опознание входных единиц соответствующего уровня, их описание и подбор выходных эквивалентов.

В словарном блоке словоупотребления, словосочетания и фрагменты из переводческой памяти (ПП) получают свои лексико-грамматические характеристики прямо из автоматического словаря (АС) — комплекса двуязычных словарей или одноязычных словников. Здесь одной из главных задач системы МП является устранение многозначности, свойственной не только общеупотребительной лексике, но и целому пласту терминологических слов и словосочетаний. Причем эта «терминологическая болезнь» характерна и для терминов, относящихся к различным ПО, и для терминов, обслуживающих одну ПО. Так, в подязыке информатики и вычислительной техники французский термин *base* имеет четыре значения:

1) основание (системы счисления), 2) база, 3) базовый [основной] адрес, 4) базис (в теории автоматов); термин *intégrateur* — пять значений: 1) интегрирующее устройство, интегратор, 2) блок интегрирования, 3) интегрирующий элемент, 4) разработчик комплексов или системы, 5) предприятие-изготовитель изделия из комплектующих (Французско-русский словарь по информатике и вычислительной технике. М.: ВИКРА, 1994).

В блоках более высоких уровней семантико-синтаксические характеристики передаются снизу, либо вырабатываются в самих блоках. При пакетной переработке больших потоков ненормализованных, а иногда и дефектных текстов ЛА постоянно сталкивается со «сбойными» ситуациями. Рассмотрим наиболее типичные случаи таких сбоев и приемы их преодоления:

1. Если на n -м уровне процесса МП происходит отказ от формирования структуры выходного текста, потребителю выдаются те более примитивные результаты перевода, которые были получены ЛА на предыдущем уровне.

2. Если на n -м уровне ЛА выработал несколько выходных схем, то на $(n+1)$ -й уровень передаются все эти варианты с тем, чтобы многозначность была разрешена на последующих уровнях пользователем или самим ЛА⁶. Поскольку децизивные и синергетические возможности ЛА ограничены, автомат в целях сокращения неопределенности выбирает обычно лишь тот выходной вариант, который, исходя из структурных и вероятностных критериев, ближе всего к семантико-синтаксической схеме входного текста и семантике обрабатываемой предметной области.

Значительно расширить возможности описываемой системы МП должна ПП. Она формируется на основе параллельного статистического анализа корпуса входных текстов и их переводов⁷. ПП включает не только фреймы словосочетаний и

предложений, в слоты которых ЛА должен вставлять словесную информацию, извлеченную из входного текста, но также часто повторяющиеся целые предложения или достаточно протяженные сегменты вместе с наиболее типичными их переводами. Кроме того, в ПП предполагается включать такие текстовые фрагменты на выходном языке, которые являются обязательными для документов определенного типа. Примером может служить русская стилистически нейтральная формула-фрейм для завершения письма:

<p><i>С уважением + слот</i> (для фамилии отправителя)</p>
--

Этот фрейм может использоваться в качестве стандартного эквивалента для самых разнообразных завершающих письмо французских или английских выражений типа:

Франц.

Recevez, cher Monsieur, l'expression de mes sentiments les plus respectueux;

Veillez croire, Madame, à l'assurance de mes sentiments distingués;

Veillez agréer, Madame, avec mes hommages, l'expression de mes sentiments les plus respectueux;

Je vous prie d'accepter, Monsieur, l'assurance de ma parfaite considération;

Soyez assuré, Monsieur, de mon respectueux dévouement;

Daignez agréer, Monsieur, l'expression de ma plus haute considération и др.

Англ.

Yours faithfully; Sincerely (yours); (My) best regards; Very truly yours.

Аналогичным образом нет необходимости требовать от ЛА точного лексико-грамматического перевода преамбулы делового соглашения типа:

Франц.

Contrat

La Société à responsabilité limitée "X", (...) dénommée ci-après le Client, d'une part, et la Société par actions de type ouvert "Y", (...) dénommée ci-après le Fournisseur, d'autre part, ont conclu le présent contrat.

(Матвишин В. Г., Ховхун В. П. Бизнес-курс французского языка. Киев: ЛОГОС, 1999. С. 217–218)

Англ.

Contract

This contract is made between Vserossijskoje Objedinenije "Rossexport", Moscow, hereinafter called "Sellers" and ... hereinafter called "Buyers", whereby it is agreed as follows: (...)

(Израилевич Е. Е. Коммерческая корреспонденция и документация на английском языке.

СПб.: Лениздат, 1992. С. 199)

Здесь достаточно воспользоваться стандартным фреймовым шаблоном типа:

Предприятие — **слот** (для наименования предприятия), — *именуемое в дальнейшем ПОКУПАТЕЛЬ, с одной стороны, и предприятие* — **слот** (для названия предприятия)), — *именуемое в дальнейшем ПРОДАВЕЦ, с другой стороны, заключили настоящий Контракт о нижеследующем:* — **слот** (для формулировки существа контракта и обязанностей ПОКУПАТЕЛЯ и ПРОДАВЦА)...

Такой шаблон можно использовать при переводе самых различных по форме вступительных фрагментов договоров и соглашений, например:

Англ.

Contract

Vneshnetorgovoje Objedinenije "Ros-simport", Moscow, hereinafter referred to as the "Seller" on the one part, and Messrs. ... hereinafter referred to as the "Buyer" on the other part, have concluded the present contract whereby it is agreed as follows: (...)

(Израилевич Е. Е. То же. С. 220)

Contract

Exportno-Importnoje Objedinenije "Rossexport", Moscow, hereinafter referred to as the "Sellers", and Messrs. ..., hereinafter referred to as the "Buyers", have concluded this contract to the effect that the Sellers have sold and the Buyers have bought on the terms and conditions set forth and subject to General Conditions of Sale endorsed hereon, the following goods: (...)

(Израилевич Е. Е. То же. С. 225)

Учитывая возможности неверного подбора переводных эквивалентов, смысловое соответствие последних входному оригиналу может проверяться через сопоставление с результатами их лексико-грамматического перевода, также выполненного ЛА.

Автоматический словарь. Лингвистическая база (ЛиБ) данных в ЛА должна строиться прежде всего как комплекс двуязычных словарей или одноязычных словариков. Кроме того, в эту базу входят функциональные грамматики, реализованные в виде расширенных сетей переходов. Центральной частью каждой системы МП является двуязычный АС. Его базовыми элементами являются словарные статьи (СлСт), содержащие входные (английские, французские и т. п.) лексические единицы (ЛЕ) (словоформы, основы слов, словосочетания) и их выходные (русские) эквиваленты, а также семантические, морфологические и синтаксические описания всех входных и выходных ЛЕ. В качестве ЛЕ могут выступать неспециальные и терминологические слова и словосочетания, а также топонимы, антропонимы и аббревиатуры. Отбор лексики в АС происходит на основе данных общепотребительных и отраслевых частотных словарей, а также с учетом исследования распределений этих ЛЕ в конкретной ПО. В этой связи следует особо отметить необходимость постоянного пополнения АС терминологической лексикой, так как в результате информацион-

ного взрыва именно она составляет свыше 90% новых слов, появляющихся в современных языках⁸.

Элементарной единицей АС могут быть:

– отдельно взятая словоформа (с/ф), например, франц. *maintenant*, англ. *now*, финск. *nyt* или русск. *теперь*,

– машинная основа (м/о), например, франц. *cheva-* (ср. пары *cheval* — *chevaux* 'лошадь — лошади') или англ. *potato-*, финск. *perun-*, русск. *картофел-*,

– словосочетание (с/с), называемое обычно машинным оборотом (м/об), например, франц. *tout autant*, англ. *so much*, финск. *saman verran* или русск. *столько же*. Каждая из указанных ЛЕ должна иметь СлСт.

Для аналитических языков с их слабо развитой морфологией в АС широко вводятся готовые с/ф. Например, китайский словарь целиком состоит из с/ф (иероглифов) и м/об (комбинаций иероглифов). Напротив, АС синтетических языков (например, финского и русского), строятся в основном из м/о и м/об. Если говорить об агглютинативных языках, то выделение м/о осуществляется обычно традиционно по легко обнаруживаемой границе между исходной формой слова и первым аффиксом. Сложнее обстоит дело с флективными (русский) и флективно-агглютинативными (финский) языками, для которых характерны ассимилятивные процессы на стыке основы и окончания, а также использование внутренней флексии, ср. финск. — *mdnty* — *mdnny* — *mdnnyt*, русск. *сосна* — *сосен*, *сухой* — *сушить* и т. п. Здесь выделение основы производится путем применения квазиагглютинативной технологии, согласно которой м/о считается последовательность букв от начала слова, общая для всех словоформ данной лексемы. Так, в приведенных примерах машинными основами будут: финск. *mdp-*, русск. *сос-* и *су-*.

Отбор ЛЕ в АС ведется лингвистами-разработчиками на основе:

– статистического критерия, требующего введения в АС таких ЛЕ, которые

обеспечивают 80%-е распознавание текстовых с/у;

– критерия релевантности, требующего помещения в АС всех терминологических единиц из обрабатываемой ПО независимо от частоты их появления в обучающей выборке текстов.

Выделение четко отграниченных от м/о и других грамматических формантов машинных флексий позволяет конструировать для каждого языка парадигмы аффиксов, которые представляются в ЛИБ в виде ориентированных графов. Связь м/о с парадигмой образуемых от этой основы с/ф осуществляется путем отсылок через грамматический код типовой парадигмы, указываемый в СлСт⁹. Этот подход оказался достаточно эффективным для русского, немецкого и романских языков, поэтому его целесообразно использовать и при построении систем МП для других языков.

Базовой единицей АС является СлСт. В ее состав включаются:

- сама входная ЛЕ;
- информация о переводе, включающая перечисление машинных основ переводных эквивалентов; каждая м/о снабжается признаком соответствующего формоизменения (номером типовой парадигмы);
- лексико-грамматический код (ЛГК), структура и наполнение которого определяются типологическими особенностями входного и выходного языков и наличием

реализованных в системе алгоритмов морфологического и семантико-синтаксического анализа.

Формирование ЛГК, которое может быть осуществлено и в интерактивном режиме, производится лингвистом с помощью таблиц кодирования, разработанных для каждой пары языков и включенных в ЛА.

Наряду с чисто морфологическими сведениями ЛГК может содержать синтаксическую и семантическую информацию. Синтаксическое описание охватывает функциональные характеристики ЛЕ, определяющие их потенциальные возможности выполнять определенную роль в синтаксической структуре предложения. Семантическое описание ориентировано на глубинный уровень содержания текста, связанный с экстралингвистической реальностью. Поэтому его формирование основывается на исследовании структуры текстов тех предметных областей, на анализ которых ориентирована данная система МП.

Описанная технология МП позволяет снизить барьер отторжения, отделяющий естественный язык от искусственного языка компьютера, и тем самым обеспечивает человеко-машинное общение на естественном языке. В дальнейшем основные усилия коллектива СтР должны быть сосредоточены на расширении автономных децизивных возможностей ЛА для сокращения неопределенности.

ПРИМЕЧАНИЯ

¹ *a.* Czyżakowski W.; Piotrowski R. Über den gegenwärtigen Stand der automatischen Textverarbeitung in der Forschungsgruppe 'Sprachstatistik' (Zum Problem des linguistischen Automaten) // *Glottometrika* 14. Trier: Wissenschaftlicher Verlag Trier, 1993. S. 161–189; *б.* Piotrowski R., Tambovtsev Y. Development of a Linguistic Automaton on the Basis of Statistics of Speech // *Literary and Linguistic Computing*. Vol. 9. 1994. No. 4. P. 291–294.

² Beliaeva L., Zaitseva N., Makarska S., Nymm V., Piotrowska K., Romanov Y. Oral SILOD — an Experimental System of Oral Machine Translation // *Speech and Computer. SPECOM'2000. International Workshop. Proceedings*. 25–28 September. St. Petersburg, 2000. P. 23–24.

³ *a.* Piotrowski R. Text — Computer — Mensch. Bochum: Brockmeyer, 1984. P. 10–53; *б.* Cole R., Mariani J., Uszkoreit H., Zaenen A., Zue V. (eds). *Survey of the Art in Human Language Technology*. Cambridge, MA: Cambridge University Press etc., 1997; *в.* Manning Ch. D., Schütze H. *Foundation of Statistical Language Processing*. Cambridge, MA: The MA Press, 1999; *г.* *Early Years in Machine Translation. Memoirs and biographies of pioneers*. Edited by John Hutchins // *Studies in the History of the Language Sciences*, 97, 2000.

⁴ Пиотровский Р. Г., Зайцева Н. Ю., Макарская С. П., Пиотровская К. Р., Романов Ю. В. Компьютер — переводчик, компьютер — учитель // ПРАВОинформ. Информационное обозрение / Информационно-юридический журнал. СПб., июнь 2000. С. 36.

⁵ а. Пиотровский Р. Г. Лингвистический автомат и его речемыслительное обоснование. Минск: МГЛУ, 1999. С. 6–22; б. Пиотровский Р. Г. Лингвистический автомат (в исследовании и непрерывном обучении). СПб.: РГПУ, 1999. С. 5–9, 27–34; в. Piotrowski R., Romanov Yu. Machine translation in the former Soviet Union and in the Newly Independent States (NIS) // Histoire — Epistémologie — Langage. Tome XXI, fascicule 1. Paris, 1999. P. 105–116.

⁶ а. Beliaeva L., Kondratjeva J., Piotrowski R., Sokolova S. Abstract from the Leningrad MT Project // Society for Conceptual and Content Analysis by Computer (SCCAC). The Newsletter of the European Network in Language and Speech '5, 1989/90. Bowling Green O.: State University, 1990. P. 26–35; б. Беляева Л. Н., Косарев Ю. А., Пиотровский Р. Г. Обработка речи и текста: достижения, проблемы, пути // Вестник СПб. отд. Российской Академии естественных наук (РАЕН). 1999. Т. 3. '1. С. 74–80.

⁷ а. Statistica Linguistica (con l'aggiunta di due appendici) / Alekseev P., Kalinin V., Piotrowski R. (eds). Bologna: Riccardo Patron, 1971; б. Beliaeva L. Machine Translation Method and Text Structure as a Source for Translator Competence Study // Across Language and Culture. Vol. 1, No. 1, 2000. P. 85–96.

⁸ Гринев С. В. Введение в терминоведение. М.: Московский лицей, 1993. С. 8.

⁹ Беляева Л. Н. Теоретические и практические аспекты разработки систем машинного перевода и лингвистических систем. Статистика речи и автоматическая переработка текста. Л.: ЛГПИ им. А. И. Герцена, 1988. С. 8–28.

N. Zaitseva

SPECIAL TEXTS AND ITS TERMINOLOGY IN THE HUMAN-MACHINE INTERCOURSE

In this article the Natural Language Processing system for the texts of different domains and structure is described. A special attention is given to the principles of terminological vocabulary selection in the automatic dictionary.