

возможны только на основе типовых технологических решений. Такие решения должны отражать опыт ведущих международных профессиональных консорциумов в области электронного обучения и базироваться на принципах открытых систем, обеспечивающих переносимость, интероперабельность и мобильность пользователей в интегрированной научно-образовательной среде.

#### БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Путилов Г.П. Концепция построения информационно-образовательной среды технического вуза. – М.: МГИЭМ, 1999.
2. Башмаков А.И., Башмаков И.А. Разработка компьютерных учебников и обучающих систем. – М: Синтег, 2002.
3. Норенков И.П., Зимин А.М. Информационные технологии в образовании. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2004.
4. Поляков А.А., Кузнецов Ю.М., Маслов С.И., Арбузов Ю.В. Концептуальные основы индустрии информационных ресурсов распределенного электронного обучения. – М.: МГИУ, 2002.
5. Декер С., Мельник С., ван Хермелен Ф., Фенсел Д. Semantic Web: роли XML и RDF – Открытые системы. 2001. № 9.
6. Янг М. XML. Шаг за шагом. – М.: ЭКОМ, 2002.

#### Ха Ти Чунг

### УЧЁТ ОСОБЕННОСТЕЙ ВЬЕТНАМСКОГО ЯЗЫКА В СИСТЕМЕ РУССКО-ВЬЕТНАМСКОГО И ВЬЕТНАМСКО-РУССКОГО МАШИННОГО ПЕРЕВОДА

**Введение.** *Машинный перевод* – действие по преобразованию текста на одном естественном языке в эквивалентный по содержанию текст на другом языке, а также результат такого действия [1]. Системы машинного перевода представляют собой сложные программные комплексы с разными видами обеспечений: лингвистическое, математическое и программное обеспечение [2]. К лингвистическому обеспечению системы машинного перевода относятся: словари слов и словосочетаний; базы синтаксических и грамматических правил; морфологические словари и т.д. Формально можно перечислить основные операции, обеспечивающие анализ и синтез в системе машинного перевода при преобразовании текста на одном естественном языке (входной текст) в текст на другом языке (выходной текст):

1. *Морфологический анализ*: приведение текстовых форм слов входного текста к словарным (к нормальной форме слов), приписывание словам ряда морфологических характеристик (часть речи, род, вид, число, падеж и т. п.).
2. *Синтаксический анализ*: выделение грамматических единиц, определение для них синтаксических связей. Цель синтаксического анализа – построение синтаксического дерева фразы, т.е. нахождение взаимозависимостей между разно-уровневыми элементами предложения.
3. *Семантический анализ*: выявление семантических соотношений между элементами текста, построение семантической сети, снятие семантической неоднозначности, что позволяет адекватно «понимать» содержание текста.
4. *Семантический синтез, синтаксический синтез и морфологический синтез*: Синтез выходных словоформ и предложений по синтаксическим, морфологическим и семантическим правилам на выходном языке.

Исследования по разработке систем машинного перевода, связанных с вьетнамским языком начались с конца 60-х годов XX века [3]. В 80-х годах начались попытки создания электронных словарей [4]. В результате исследований были созданы электронные словари, такие как "Just click and see", "Lac Viet", "Lingvo Ru-Vi-En dictionary". В настоящее время создаются первые системы англо-вьетнамского машинного перевода, но на данный момент, их эффективность и точность очень ограничены, и далеко не достигают до того уровня широко известных систем машинного перевода, которые пользуются популярностью в России и западных странах, такие как: СОКРАТ; ПРОМТ(<http://www.translate.ru/translator.asp>); мультязыковой онлайнный переводчик WorldLingo ([http://www.worldlingo.com/en/websites/url\\_translator.html](http://www.worldlingo.com/en/websites/url_translator.html)); мультязыковой онлайнный переводчик Babelfish (<http://babelfish.altavista.com/tr>) и т.п.

**Актуальность создания систем русско-вьетнамского и вьетнамско-русского перевода.** До последнего времени все еще не разработаны системы русско-вьетнамского и вьетнамско-русского перевода. Главные причины заключаются в следующем:

- ◆ неполнота исследований среди вьетнамских лингвистов в области машинного перевода;
- ◆ разногласия среди лингвистов по грамматическим правилам;
- ◆ неэффективность существующих методов автоматической обработки текстов на вьетнамском языке.

С подъемом экономического уровня страны и развитием экономических и политических отношений между Вьетнамом и Россией, создание систем русско-вьетнамского и вьетнамско-русского машинного перевода стало актуальной задачей. Об этом свидетельствуют ряд фактов:

- ◆ Возрастание количества студентов, специалистов, учащихся в России.
- ◆ Увеличение количества желающих овладеть русским языком среди вьетнамцев.
- ◆ Наличие огромного количества материалов на русском языке, используемых в вузах и учреждениях Вьетнама.
- ◆ Нужда специалистов в использовании материалов на русском языке.

**Постановка задачи.** При решении задачи лингвистического обеспечения системы русско-вьетнамского и вьетнамско-русского машинного перевода необходимо учитывать особенности русского и вьетнамского языка на разных уровнях репрезентации: графема, морфология, лексика, синтаксиса, грамматика. В данной работе приведен сравнительный анализ некоторых особенностей русского и вьетнамского языка на примерах, с целью выработки предпосылки для дальнейшей разработки метода анализа и синтеза в системе русско-вьетнамского и вьетнамско-русского машинного перевода.

**Сравнительный анализ вьетнамского и русского языков.** Русский язык является флективным по типологическому отношению [4], в русском языке для обозначения грамматических связей слова изменяют свои формы с помощью собственных им грамматических категорий: падежа, числа, времени и т.п. Таким образом, грамматические отношения предложения в русском языке выражаются «в пределах словоформы».

Вьетнамский язык, в отличие от русского языка, является тоновым моносyllабическим изолирующим [5]. Внутри предложения синтаксические отношения между словами не выражены, грамматические отношения выражаются преимущественно порядком слов и служебными словами, в то же время слова не изменяют свои формы.

Следующий пример отражает роль порядка слов в предложении, рассмотрим простое на вьетнамском языке предложение, состоящее из 5 слов, обозначим их соответственно А, В, С, D, E: A-*Nó*: он; B-*bảo*: сказать,...; C-*sao*,...; D-*không*: нет, не...; E-*đến*: приходить, приезжать...:

*Nó bảo sao không đến?* (ABCDE): Он спросил, почему вы не пришли?

Если порядок этих слов меняется, получаются и другие выражения, которые синтаксически и грамматически правильны, но носят совсем другие значения:

*Sao bảo nó không đến* (CBADE): Сказали, что он не придет.

*Không đến bảo nó sao* (DEBAC): Почему ты не пришел к нему и не рассказал ему об этом!

*Đến nó bảo không sao* (EABDC): Приди к нему и объясни всё нормально.

*Nó đến bảo không sao* (AEBDC): Он пришел и сказал всё было нормально.

*Bảo nó đến không sao* (BAEDC): Скажи ему, пусть он придет и т.п.

*Многозначность слов и проблема синтаксической омонимии.* Во вьетнамском языке, одно и то же слово в разных местах может служить глаголом, именем существительным, прилагательным и т.п. Соответственно, одно слово может носить разные, иногда и противоречивые значения.

*Проблема выделения слов.* Для русского языка границей слова является пробел или знаки препинания, не требуется реализации специальных процедур деления текста на слова. Для вьетнамского языка, где самой минимальной единицей предложений является не слово, а *морфема*. Морфемы могут разделяться пробелами. При этом *морфема* может являться слогом слова, а может и самостоятельным словом, либо соединяется с другими морфемами в составе длинных слов. Словообразование осуществляется примыканием [3], повторами (полными или частичными) морфем, максимальная длина слов (в морфемах-словах) может быть равна 4, минимальная – 1, а самая распространенная длина – 2. Требуется разработка эффективного алгоритма правильного выделения слов. Следующий пример показывает неоднозначность при разных вариантах выделения слов. Подобные случаи часто встречаются в вьетнамском языке: фраза, состоящая из 5 слов-слов *“ông già đi rất nhanh”*, обозначим эти слова, соответственно А, В, С, D, E. Рассмотрим их значение в словаре “Lingvo Ru-Vi-En Dictionary”:

*Ông*: твой, сударь, ты, я, дядя, мужчина, гражданин, господина, дед...

*Già*: старый, свыше, матёрый, старейший, старик, старческий...

*Đi*: сходить, уезжать, съездить, ходить, уходить, поехать, проходить...


*rất*: очень, крайне, сильно, отчаянно, чрезвычайно...

*Nhanh*: короткий, спешить, скоростной, быстрый, быстро, бегло, беглый.

Слова, преобразующие примыканием А и В, В и С: *Ông già*(AB): дед, старик, старейший; *Già đi*(BC): стареть, сдавать, стариться, старить.

Это приводит к тому, что данная фраза носит семантически несвязанные значения в зависимости от способа выделения слов, как понимать данную фразу зависит от контекста:

i).  Старик бегло шагает.

ii).  Этот мужчина быстро стареет.

Процесс синтаксического анализа и синтеза вьетнамского языка становится особо сложным (по сравнению с английским или русским) из-за синтактики и грамматики языка, как показано в следующем примере. На рис.1 изображено семантическое дерево выражения: “Эта система способна переводить некоторые простые выражения с русского языка на вьетнамский”, построенное программой RossDev- «Рабочее место лингвиста» (<http://www.aot.ru/download/shorttml.zip>), раз-

работанной Алексеем Сокирко факультета лингвистики РГГУ в рамках проекта «Диалинг».



Рис.1. Семантическое представление фразы на русском языке

Переводим это семантическое дерево в соответствующее дерево на английском языке. Здесь возникла небольшая проблема, как отражена на рис.2. Если учитываются морфологические и грамматические правила английского языка, то получим правильное выражение: “This system can translate some simple sentences from Russian to Vietnamese”.

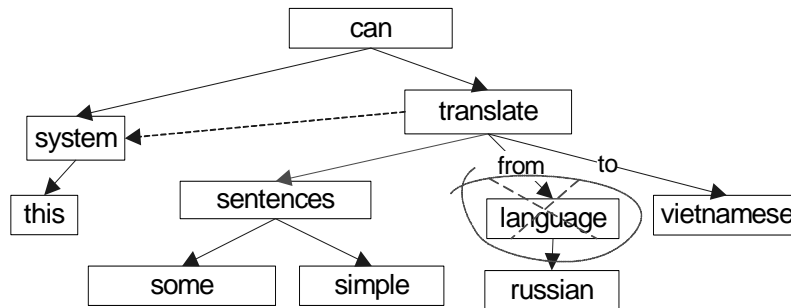


Рис.2. Семантический синтез на английском языке

Подобное действие для вьетнамского языка (как показано на рис.3) далеко не всегда дает приемлемый результат, получается совсем не понятная последовательность морфем: “Này hệ thống có thể dịch một vài đơn giản câu từ thuộc Nga ngôn ngữ sang thuộc Việt”.

Рис.4 показывает, как переставить порядок слов, порядок их связей, удалить ненужные морфемы, и добавить пропущенное слово. Результат такого действия отражен на рис.5.

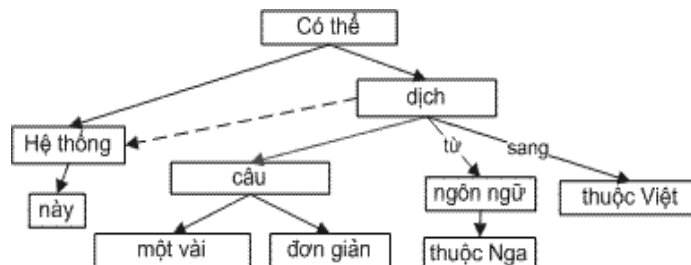


Рис.3. Семантический синтез на Вьетнамском языке

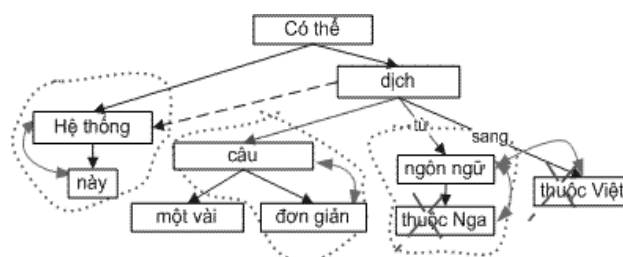


Рис.4. Семантический и синтаксический синтез на вьетнамском языке

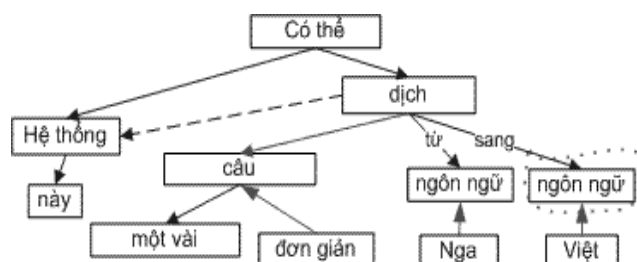


Рис.5. Правильное синтаксическое дерево совершенствование систем русско-вьетнамских словарей, вьетнамско-русских словарей;

Из этих примеров стало ясно, что, самые сложные этапы при реализации лингвистического обеспечения системы связаны с семантическим анализом и синтезом. Для решения этих проблем, на наш взгляд, необходимо решить следующие задачи:

- ◆ разработка эффективных алгоритмов правильного выделения слов;
- ◆ разработка баз грамматических правил и др.;
- ◆ разработка алгоритмов построения семантических сетей для текста на вьетнамском языке.

**Заключение.** Исследования показывают что, качество перевода зависит и от мощности словарных средств, от тематики и стиля исходного текста. Существуют предпосылки о возможности создания русско-вьетнамского и вьетнамско-русского машинного перевода. Возможность получения вполне приемлемого перевода для технических документов, путем создания специализированных машинных словарей по определённым тематикам текстов.

#### БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Машинный перевод. Энциклопедия Кругосвет® <http://www.krugosvet.ru/articles/82/1008256/1008256a1.htm>.
2. Башмаков А.И., Башмаков И.А. Интеллектуальные информационные технологии. – М.: Изд-во МГТУ, 2006. – С. 90-96.
3. Дао Хонг Тху. Проблема создания систем машинного перевода с вьетнамского / на вьетнамский язык // Прикладная лингвистика в науке и образовании. Третья международная научная конференция, 16-17 марта 2006, С-Петербург. – С. 49-54.
4. Дао Хонг Тху. Машинный перевод с вьетнамского языка и на вьетнамский язык (вьетнамско-английский и англо-вьетнамский машинный перевод) // Труды международной конференции «Корпусная лингвистика. 2006», 10-14 октября 2006 г., СПб: Изд-во СПбГУ, 2006. – С. 99-105.
5. Лингвистический энциклопедический словарь / Главн. ред. В.Н. Ярцева. – М., 1990.