

Говорящий «ЭТАП». Опыт использования синтаксического анализатора системы ЭТАП в русском речевом синтезе¹

The talking ETAP. Using the ETAP parser in Russian speech synthesis

Иомдин Л. Л. (iomdin@iitp.ru)

Институт проблем передачи информации РАН
им. А. А. Харкевича, Москва

Лобанов Б. М. (lobanov@newman.bas-net.by)

Гецевич Ю. С. (mix1122@gmail.com)

Объединенный институт проблем информатики НАН
Беларуси, Минск

Излагаются результаты работы по созданию экспериментальной гибридной системы синтеза русской речи, использующей в качестве промежуточного этапа поверхностно-синтаксический анализ читаемого текста. Синтаксическая структура предложения в виде размеченного дерева зависимостей, формируемая в ходе синтаксического анализа, обеспечивает лучшие качественные характеристики звучащей речи по сравнению с классической системой речевого синтеза, не учитывающей в явной форме информации о связях слов в предложении.

1. Вводные замечания

Данное исследование продолжает работу, начатую два года назад коллективами Лаборатории компьютерной лингвистики ИППИ РАН им. А. А. Харкевича и Лаборатории распознавания и синтеза речи Объединенного института проблем информатики НАН Беларуси и направленную на создание интегрированной системы русского речевого синтеза, в которой просодические и интонационные

¹ Авторы благодарят Российский фонд фундаментальных исследований и Белорусский республиканский фонд фундаментальных исследований, поддержавшие данную работу грантами РФФИ (№ 10-07-90001-Бел) и БРФФИ (№ Ф10Р-006) в рамках программы совместных исследований России и Беларуси. Авторы выражают также искреннюю благодарность активным участникам проекта В. Г. Сизову, осуществившему программную интеграцию синтаксического анализатора с речевым синтезатором, и О. Ю. Подлесской, в задачу которой входит акцентуирование большого морфологического словаря интегрированной системы: эта работа в настоящее время подходит к концу.

и акцентные характеристики генерируемого текста формируются с учетом информации о синтаксической структуре читаемого предложения.

В работе [1] были изложены результаты первых экспериментов в этом направлении. Эти эксперименты сводились к тому, что подлежащий синтезу текст пропускался через синтаксический анализатор многоцелевого лингвистического процессора ЭТАП-3 [2, 3], который формировал информацию об эмфатически выделенных элементах предложения.

В настоящем проекте речь идет о значительно более масштабном участии системы ЭТАП-3 в системе русского речевого синтеза, разработанного Б. М. Лобановым и его коллегами [4]. Синтезируемый текст в нормальной орфографической записи подвергается полному синтаксическому анализу, осуществляемому парсером ЭТАП-3, который (1) членит текст на отдельные предложения, (2) для каждого предложения строит его древесную синтаксическую структуру (СинтС), (3) с помощью специальных правил, применяемых к готовой синтаксической структуре, устанавливает границы речевых синтагм предложения и его эмфатически выделенные элементы. Система Мультифон обрабатывает эту информацию и определяет длительность пауз между синтагмами в зависимости от их синтаксического типа (на принципах, изложенных, в частности, в [5]). Попутно в формирующейся гибридной системе речевого синтеза успешно решается критическая для такой системы задача снятия омографии словоформ, которые различаются ударением и/или противопоставлением букв *е* и *ё*.

2. Синтаксический анализатор системы ЭТАП-3: современное состояние

Синтаксический анализатор (парсер) системы ЭТАП-3 используется в различных приложениях, разрабатываемых в Лаборатории компьютерной лингвистики ИППИ РАН, в том числе в системе машинного перевода с русского языка на английский, в системе синонимического перифразирования, для построения синтаксически размеченного корпуса русского языка СинТагРус [6, 7], а также, в последнее время, в целях создания онтологии для автоматической обработки текстов [8].

Этот парсер в значительной мере основан на лингвистической теории «Смысл \leftrightarrow Текст» И. А. Мельчука. Для каждого предложения письменного текста он строит его синтаксическую структуру (СинтС) (в терминах теории «Смысл \leftrightarrow Текст» — поверхностно-синтаксическую структуру) в виде дерева зависимостей. В дереве СинтС любого предложения имеется единственная вершина, которой непосредственно или опосредованно подчиняются все остальные узлы. Каждый узел такого дерева соответствует одному слову предложения (или некоторому словосочетанию, по тем или иным причинам трактуемому как слово, такому как *несмотря, по меньшей мере, во что бы то ни стало* и т. п.), а его дуги помечены именами синтаксических отношений (СинтО). Имена СинтО эксплицируют различные типы синтаксических связей между словами; в современной версии парсера используется 65–70 различных СинтО. Например, связь между

глагольным сказуемым в качестве вершины и именным подлежащим при нем в качестве зависимого члена (*старик ← получил*) представляется **предикативным Синто**; связь между предикатным словом и первым дополнением при нём (*получил → письмо, получение → письма*) представляется **1-ым комплетивным Синто**; связь между существительным и определяющим его прилагательным (*заказное ← письмо*) оформляется **определятельным Синто**, связь между глаголом и наречным обстоятельством (*неожиданно ← получил*) задаётся **обстоятельственным Синто**, а аналитические формы слов, рассматриваемые как синтаксические конструкции, оформляются с помощью **аналитического Синто** (*получил → бы, более ← интересный, будет → работать*).

Дерево СинтС предложения, генерируемое парсером ЭТАП-3, является упорядоченным — оно сохраняет информацию о порядке следования слов в предложении, который имел место в его исходной форме.

Алгоритм синтаксического анализа обращается к лингвистическим ресурсам двух основных типов: набору бинарных синтаксических правил, или синтагм², и так называемому комбинаторному словарю, содержащему богатую и разнообразную информацию о каждом входящем в него слове. Парсер работает пофрагментам и может функционировать в нескольких режимах, в частности, 1) в полностью автоматическом режиме, применяемом по умолчанию: в этом случае для каждого предложения строится ровно одна СинтС; 2) в режиме множественного анализа, когда пользователь может потребовать от системы построить для неоднозначного предложения несколько СинтС или даже все возможные СинтС; 3) в интерактивном режиме, когда в определенных точках алгоритма парсер, встретив неоднозначную лексическую единицу или омонимичную синтаксическую конструкцию, предлагает пользователю выбрать ту или иную морфологическую, лексическую и/или синтаксическую интерпретацию элементов предложения и тем самым направить работу по некоторому конкретному пути.

Система ЭТАП-3 в целом и ее синтаксический анализатор рассчитаны в первую очередь на тексты нейтрально-деловой прозы. Это, в частности, означает, что в составе некоторых приложений (в первую очередь, в машинном переводе) её нецелесообразно применять к стилистически окрашенному материалу, к авторской художественной прозе, поэзии или к разговорной речи. Однако, как показали наши эксперименты, в рамках рассматриваемой здесь задачи синтеза звучащей речи парсер ЭТАП-3 вполне применим для художественной прозы и публицистики: хотя СинтС предложений, образующих такого рода тексты, могут содержать ошибки, неприемлемые в задачах, требующих глубокой семантической переработки, эти ошибки не критичны для речевого синтеза, поскольку, как правило, информация о границах фонетических синтагм и эмфатически выделенных элементах предложения передаётся верно.

Современная версия русского парсера ЭТАП-3 характеризуется существенно лучшей производительностью и более высокой надёжностью по сравнению

² Тем самым термин «синтагма» используется здесь иначе, чем это принято в литературе, посвященной автоматической обработке устной речи (в том числе и в настоящей статье).

с предшествующими его вариантами благодаря включению в парсер достаточно развитого статистического компонента, основанного на материале синтаксически размеченного корпуса СинТагРус. Эти факторы оказываются весьма важными для разрабатываемой гибридной системы речевого синтеза.

3. Интерфейс «ЭТАП — МУЛЬТИФОН» и используемые правила

Разработанная в Лаборатории распознавания и синтеза речи ОИПИ НАН Беларуси система речевого синтеза «Мультифон» в целом хорошо справляется с членением и просодическим оформлением фонетических синтагм, идентифицируемых в тексте в первую очередь знаками препинания. Однако при чтении развёрнутых предложений с минимальным количеством знаков препинания (а встречаемость таких предложений в текстах весьма высока, см. [9]) система даёт ощутимые сбои, поскольку глубина их синтаксического анализа оказывается недостаточной. Эти сбои в значительной мере удаётся устранить, если прибегнуть к полному синтаксическому анализу предложения, осуществляемому парсером ЭТАП-3.

На рис. 1 представлена схема, реализующая интерфейс взаимодействия систем ЭТАП-3 и МУЛЬТИФОН. Интеграция двух систем осуществляется через стандартный способ взаимодействия SAPI 5.1 (The Speech Application Programming Interface), который предназначен для стыковки программ синтеза речи с другими программами, работающими в операционной среде Windows.

Прежде чем поступить на вход «Мультифона», входной текст в нормальной орфографической записи подвергается синтаксическому анализу, осуществляемому парсером ЭТАП-3. Специально сконструированный блок правил, применяемых к построенной парсером синтаксической структуре каждого предложения, формирует информацию о его просодически значащих элементах. Размеченный таким образом текст передаётся через SAPI в синтактико-просодический препроцессор (СПП), который на основе этой информации осуществляет членение предложений на фонетические синтагмы, определяет длительность пауз между ними и устанавливает интонационный тип полученных синтагм.

К настоящему времени как в постпроцессоре парсера ЭТАПа-3, так и в СПП задействован ограниченный массив фонетических правил, которые носят достаточно общий характер. Этот массив пока далеко не полон и будет совершенствоваться в дальнейшем. Тем не менее даже небольшое число синтаксических правил, применяемых при синтезе речи, даёт обнадеживающие результаты.

Так, для определения положения границ фонетических синтагм были использованы следующие типы синтаксических элементов предложения:

- 1) абсолютная вершина предложения;
- 2) вершины всех частей сложносочиненного предложения;
- 3) вершины всех придаточных предложений;
- 4) самые правые субстантивные элементы группы подлежащего, дополнения или обстоятельства при вершинах, перечисленных в пп. 1–3;

- 5) самый правый субстантивный элемент первой именной подгруппы в группах, перечисленных в п. 4;
- 6) отдельные классы лексических единиц и конкретные лексические единицы, стоящие в определенной позиции, такие как наречия-детерминанты в начале предложения типа *вовремя*, *наверняка*, *непрерменно*, числительные и количественные существительные типа *миллион*, *количество*, *часть* и пр).

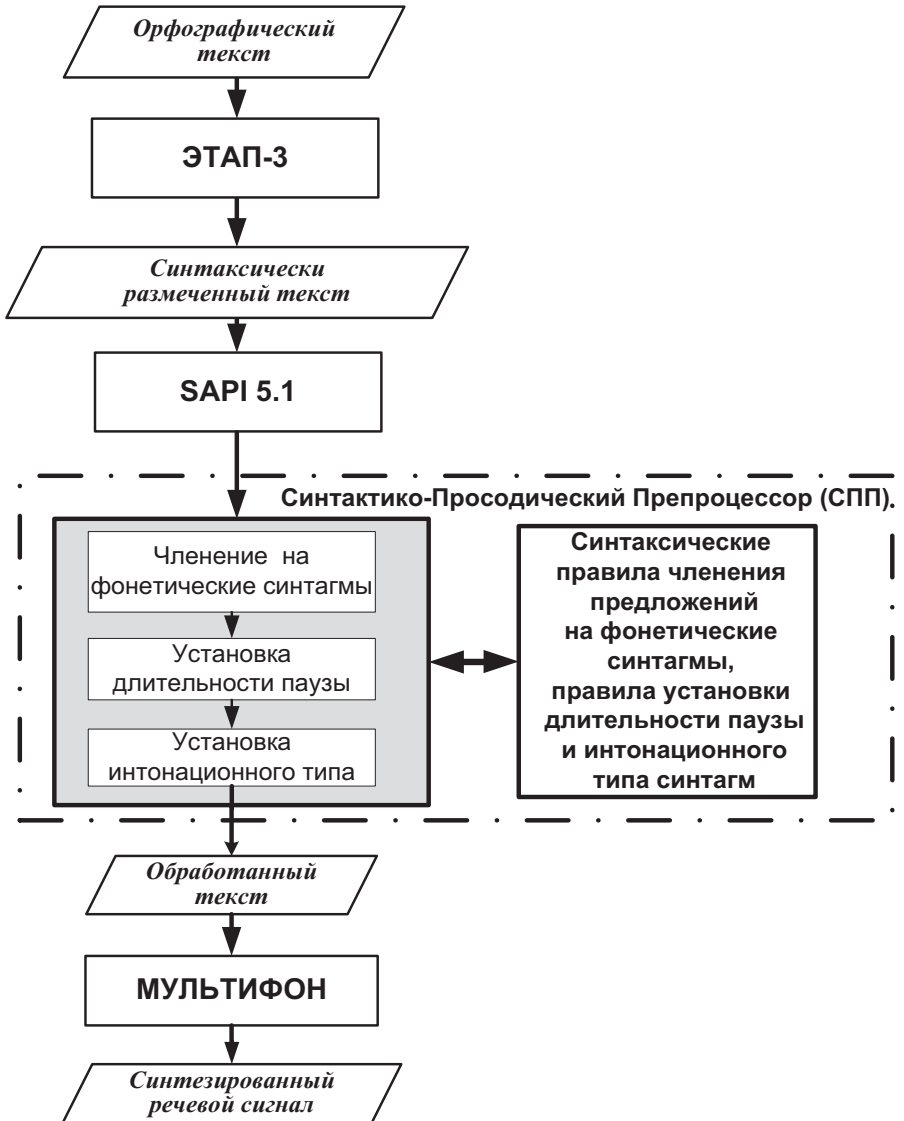


Рис. 1. Схема интегрированной системы «ЭТАП-3 — МУЛЬТИФОН»

Дополнительно опытным путём были определены также предпочтительные значения длительности межсинтагменной паузы и интонационного типа каждой из полученных синтагм в зависимости от используемых синтаксических типов элементов.

Эффективность разработанных правил хорошо видна из приводимых ниже примеров обработки небольшого отрывка текста (7 предложений) из книги Генри Форда «Моя жизнь, мои достижения». В приведенных примерах (1)–(7) после каждого из анализируемых входных предложений представлен обработанный интерфейсом «ЭТАП — МУЛЬТИФОН» выходной текст, размеченный на фонетические синтагмы и поступающий на вход системы «Мультифон». В конце каждой синтагмы присутствует один из интонационных знаков: С — интонация незавершённости, Р — интонация завершённости и Q — интонация вопроса. Каждый знак сопровождается цифровым индексом, по которому «Мультифон» выбирает один из подтипов интонационного типа (С, Р или Q) и соответствующую этому подтипу длительность межсинтагменной паузы. Кроме того, каждая синтагма разбивается на акцентные единицы (знак [/]), внутри которых выделяются фонетические слова с сильными, или главными акцентами (знак +), и со слабыми, или побочными, акцентами (знак =). Служебные и значимые слова в синтагмах объединяются в фонетические слова посредством твёрдого знака (Ъ).

(1) *Если бы имелось средство сэкономить время на 10% или повысить результаты на 10%, то неприменение этого средства означало бы десятипроцентный налог на все производство.*

1	е=слиббы име+лось/сре+дство/	С3
2	эконо+мить/вре+мя/	С3_1
3	набде=сять проце+нтов/	С2
4	и=ли повы+сить/результата+ты/	С3_2
5	набде=сять проце+нтов/	С7
6	то= непримене=ние э+того сре+дства/	С3
7	означа+ლობбы/десятипроце=нтный нало+г/набвсё= произво+дство/	Р4

(2) *Если, скажем, время одного человека стоит 50 центов в час, то десятипроцентная экономия составит лишний заработок в пять центов.*

1	е=сли ска+жем/вре+мя/одного= челове+ка/	С3_1
2	сто+ит/пядеся+т це+нтов/въча+с/	С7
3	то= десятипроце=нтная эконо+мия/	С3_2
4	соста+вит/ли=шний за+работок/	С01
5	въпя+ть/це+нтов/	Р4_1

(3) *Если бы владелец небоскреба мог увеличить свой доход на десять процентов, он отдал бы охотно половину этого добавочного дохода только для того, чтобы узнать это средство*

1	е=слиббы владе+лец/небоскрё+ба/	С3
2	мо+г/увели+чить/сво=й дохо+д/	С3_1
3	набде=сять проце+нтов/	С3_2
4	о=н отда+льббы/охо+тно/	С02

- | | | |
|---|--|----------|
| 5 | полови+ну/э=того доба=вочного дохо+да/ | C3 |
| 6 | то=лько для того+/
7 што=бы узна+ть/э=то сред+ство/ | C9
P9 |
- (4) *Почему он построил себе небоскреб?*
- | | | |
|---|---------------------------------|------|
| 1 | почему+/ | Q1_1 |
| 2 | о=н постро+ил/себе= небоскрё+б/ | Q2 |
- (5) *Потому что научно доказано, что известные строительные материалы, применённые известным образом, дают известную экономию пространства и увеличивают наемную плату.*
- | | | |
|---|---|------|
| 1 | потому= што= нау+чно/дока+зано/ | C8 |
| 2 | што= изве=стные строи=тельные материа+лы/ | C10 |
| 3 | применё+нные/изве=стным о+бразом/ | C3_1 |
| 4 | даю+т/изве=стную эконо+мию/простра+нства/ | C1 |
| 5 | и= увели+чивают/наё+мную/плату/ | P4_2 |
- (6) *Тридцатиэтажное здание не требует больше фундамента и земли, чем пятиэтажное.*
- | | | |
|---|------------------------------------|----------|
| 1 | тридцатиэта=жное зда+ние/ | C3_2 |
| 2 | не́тре+бует/бо+льше/фунда+мента/ | C1 |
| 3 | и= земли+/
4 че=м пятиэта+жное/ | C8
P8 |
- (7) *Следование старомодному способу постройки стоит владельцу пятиэтажного здания годового дохода с двадцати пяти этажей.*
- | | | |
|---|---|------|
| 1 | сле=дование старомо=дному спо+сому/постройки/ | C3 |
| 2 | сто+ит/владельцу/пятиэта=жного зда+ния/ | C3_1 |
| 3 | годово=го дохо+да/ | C3_2 |
| 4 | съдвадцати+/пяти+/этажей/ | P6 |

Нетрудно убедиться в том, что звучащий текст, синтезированный из приведенных выше синтагм, адекватно передаёт как межсинтагменные паузы, так и эмфатически выделенные слова.

Для сравнения приведём примеры разбиения на синтагмы предложения (7), осуществляемого синтезатором ЭТАП-Мультифон (7а), Мультифон (7б), а также двумя доступными в Интернете синтезаторами русской речи «Катерина» компании ScanSoft (7в) и «Алёна» группы компаний Asapela (7г).

- (7а) *Следование старомодному способу постройки // стоит владельцу пятиэтажного здания // годового дохода // с двадцати пяти этажей.*
- (7б) *Следование старомодному способу постройки // стоит владельцу // пятиэтажного здания годового дохода // с двадцати пяти этажей.*
- (7в) *Следование // старомодному // способу постройки // стоит владельцу // пятиэтажного здания // годового дохода // с двадцати пяти этажей.*
- (7г) *Следование // старомодному // способу постройки // стоит владельцу // пятиэтажного здания // годового дохода // с двадцати пяти этажей.*

Пример (7а), на наш взгляд, демонстрирует наилучший способ синтагматического членения. Кроме того, как видно из примеров (7в) и (7г), способы разбиения на синтагмы синтезаторами «Катерина» и «Алёна» оказались идентичными.

Идея применения синтаксического анализа текста к задаче синтеза звучащей речи высказывалась и раньше, хотя и нечасто. Так, в работе Ф. Кёна и соавт. [10] высказывалось предположение, что характер синтаксической структуры фразы и ее интонационный рисунок (включая межсинтагменные паузы) взаимосвязаны. Авторы провели небольшой эксперимент на материале английского языка, который это подтвердил. Дж. Тауберер [11] показал на материале достаточно объемного корпусного исследования звучащей английской речи, что между синтаксической структурой предложения и внутрисентенциальными паузами имеется безусловная корреляция. Наконец, в недавнем исследовании Ф. Кампилло Диаса и соавт. [12], выполненном на материале галисийского языка, обнаружена несомненная корреляция синтаксической структуры, интонационного контура и паузации.

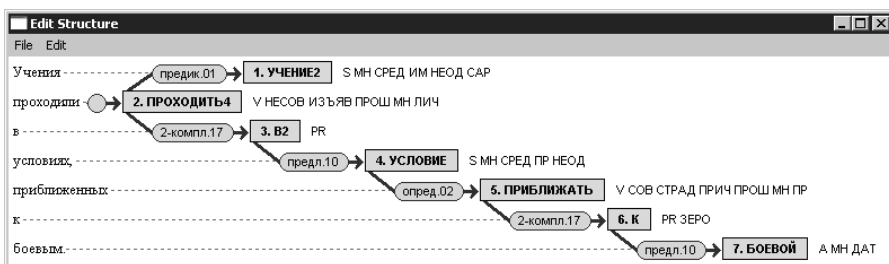
Насколько известно авторам, попытки непосредственной интеграции синтаксического анализатора в систему речевого синтеза до сих пор не принимались ни для какого языка.

4. Разрешение омографии

Использование синтаксического анализатора ЭТАП-3 в интегрированной системе речевого синтеза в значительной степени снимает проблему правильной передачи омографичных словоформ текста, которые различаются ударением и/или буквами е и ё. Дело в том, что в подавляющем большинстве случаев такие словоформы в результате синтаксического разбора получают однозначную лексико-морфологическую интерпретацию, исключающую необходимость применения каких-либо эвристических правил выбора вариантов произнесения. Так, построенная анализатором СинтС предложения

(8) *Учения проходили в условиях, приближенных к боевым.*

имеет вид



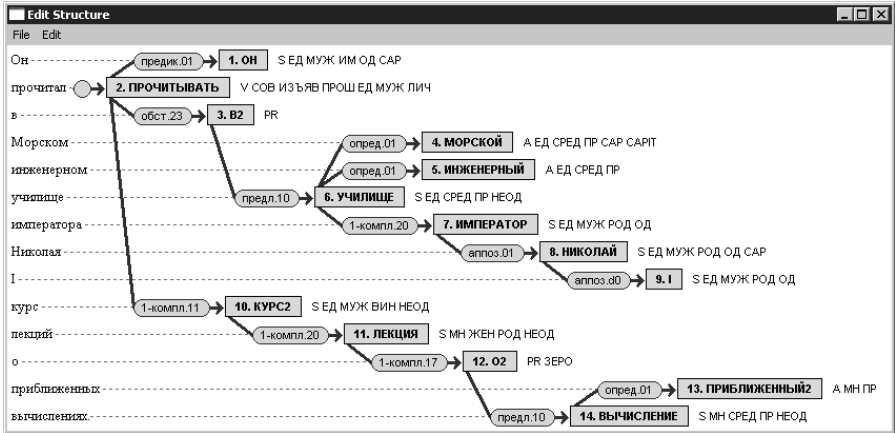
Как легко увидеть, словоформа *приближенных* интерпретируется ЭТАП-ом как страдательное причастие совершенного вида, прошедшего времени,

множественного числа и предложного падежа от глагола ПРИБЛИЖАТЬ, что соответствует произносительному варианту *приближенных*.

В то же время словоформа *приближенных* в предложении

- (9) *Он прочитал в Морском инженерном училище императора Николая I курс лекций о приближенных вычислениях.*

интерпретируется как прилагательное ПРИБЛИЖЁННЫЙ во множественном числе и предложном падеже:

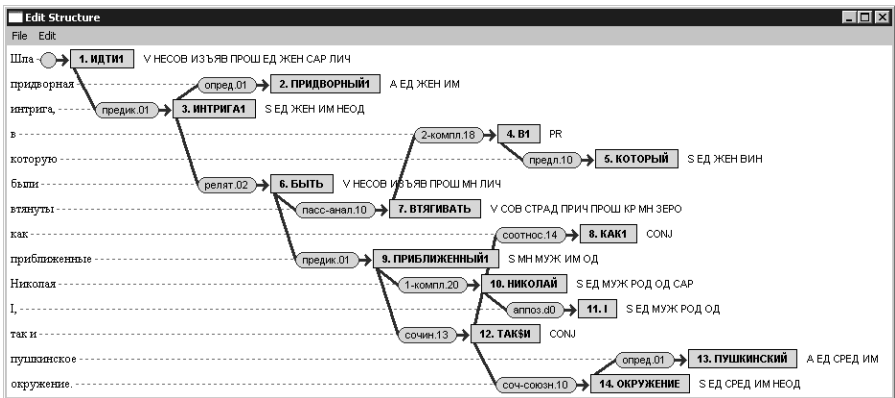


что соответствует произносительному варианту *приближённых*.

Наконец, во фразе

- (10) *Шла придворная интрига, в которую были втянуты как приближенные Николая I, так и пушкинское окружение,*

СинтС которого имеет вид



словоформа *приближенные* интерпретируется как существительное и, разумеется, произносится как *приближённые*.

Очевидно, что и синтаксический анализ не обеспечивает стопроцентного разрешения омографии: чтобы правильно выбрать интерпретацию словоформы типа *замок*, в общем случае необходимо обращение к глубокой семантике. Тем не менее и в таких ситуациях применение синтаксического анализатора нередко приводит к хорошим результатам, особенно с учетом того факта, что при разрешении лексической неоднозначности в современной версии ЭТАП-3 задействуется статистическая информация из размеченного корпуса, в том числе и информация о встречающихся там словосочетаниях (ср. *замки Луары и дверные замки*).

5. Использование модуля русского речевого синтеза в других задачах автоматической обработки текстов

Разрабатываемая интегрированная система речевого синтеза «ЭТАП-Мультифон» может использоваться не только для решения задачи выразительного чтения готового русского текста. В частности, она может быть применена и для озвучивания результата машинного перевода в системах, в которых русский язык является выходным, а также в любых других задачах, где необходимо или желательно произнесение сформированного компьютерной системой русского текста (перифразирование, общение с компьютером на естественном языке и т. д.). Первые эксперименты с фонетическим синтезом текста в составе системы англо-русского перевода ЭТАП-3 дали обнадеживающие результаты. В этих экспериментах правила идентификации значимых элементов предложения применяются не к готовой СинтС предложения русского текста, а к промежуточному результату работы перевода (непосредственно перед этапом морфологического синтеза).

Отметим в заключение, что ограниченный масштаб проведенных нами экспериментов не позволяет пока провести полноценную статистически обусловленную оценку полученных результатов. Эта задача предстоит авторам в ближайшем будущем.

Литература

1. Иомдин Л. Л., Лобанов Б. М. Синтаксические корреляты просодически маркированных элементов предложения // Труды Международной конференции «Компьютерная лингвистика и интеллектуальные технологии» Диалог 2009 (Бекасово, 27–31 мая 2009 г.). М.: РГГУ, 2009. Вып. 8 (15). ISBN 978-5-7281-1102-3. С. 136–142.
2. Апресян Ю. Д., Богуславский И. М., Иомдин Л. Л. и др. Лингвистический процессор для сложных информационных систем. М.: Наука, 1992. 256 с.

3. *Juri Apresjan, Igor Boguslavsky, Leonid Iomdin, Alexandre Lazourski, Vladimir Sannikov, Victor Sizov, Leonid Tsinman.* ETAP-3 Linguistic Processor: a Full-Fledged NLP Implementation of the MTT // MTT 2003, First International Conference on Meaning — Text Theory (June 16–18 2003). Paris: Ecole Normale Supérieure, 2003. P. 279–288.
4. *Лобанов Б. М., Цирульник Л. И.* Компьютерный синтез и клонирование речи. Минск: Белорусская Наука, 2008. — 342 с.
5. *Лобанов Б. М.* Алгоритм сегментации текста на синтаксические синтагмы для синтеза речи // Труды Международной конференции «Компьютерная лингвистика и интеллектуальные технологии» (Диалог 2008). — М.: Наука, 2008. С. 323–329.
6. *Богуславский И. М., Иомдин Л. Л., Валеев Д. Р., Сизов В. Г.* Синтаксический анализатор системы ЭТАП и его оценка с помощью глубоко размеченного корпуса русских текстов // Труды Международной конференции <Корпусная лингвистика — 2008>. СПб.: Санкт-Петербургский государственный университет, 2008. С. 56–74.
7. *Igor Boguslavsky, Leonid Iomdin, Svetlana Timoshenko, Tatiana Frolova* — Development of the Russian Tagged Corpus with Lexical and Functional Annotation. // Metalanguage and Encoding Scheme Design for Digital Lexicography. MONDILEX Third Open Workshop. Proceedings. Bratislava, Slovakia, 15–16 April, 2009. ISBN 978-80-7399-745-8. P. 83–90 (соавторы: S. Timoshenko, I. Boguslavsky, T. Frolova).
8. *Igor Boguslavsky, Leonid Iomdin, Victor Sizov, Leonid Tsinman, Svetlana Timoshenko.* Interfacing the Lexicon and the Ontology in a Semantic Analyzer. // In: COLING 2010. Proceedings of the 6th Workshop on Ontologies and Lexical Resources (Ontolex 2010). Beijing, August 2010. P. 67–76
9. *Лобанов Б. М.* Пунктуационная структура художественных произведений и её роль в синтезе выразительной речи по тексту // Труды Международной конференции «Компьютерная лингвистика и интеллектуальные технологии» (Диалог 2010), 26–30 мая 2010. — М.: Наука, 2009. С. 330–338.
10. *Koehn, P., Abney, S., Hirschberg, J., Collins, M.* Improving intonational phrasing with syntactic information. Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 2000, 3, p. 1289–1290.
11. *Tauberer, J.* Predicting Intrasentential Pauses: Is Syntactic Structure Useful? In Barbosa, P. A., Madureira, S., and Reis, C. (Eds.) Proceedings of the Speech Prosody 2008 Conference, May 6–9, 2008. Campinas, Brazil: Editora RG/CNPq, p. 405–408.
12. *Francisco Campillo Díaz, Jan van Santen, and Eduardo Rodríguez Banga.* Integrating phrasing and intonation modelling using syntactic and morphosyntactic information. Speech Communication. Volume 51, Issue 5, May 2009, p. 452–465.