

Анализируя лексико-семантические и морфологические трудности в обучении РКИ китайских учащихся, мы пришли к следующим выводам.

В НКРЯ и ККЯПУ входит ряд примеров, иллюстрирующих словарные значения единиц. Лингвистический корпус дает возможность китайским учащимся изучать прямые и переносные значения слов, правильную сочетаемость слов.

В связи с тем, что русский и китайский языки относятся к разным языковым семьям, в обучении китайских учащихся существует ряд грамматических трудностей. С помощью разнообразных видов разметки в НКРЯ можно решать такие проблемы. По запросу с конкретным словом в определенной форме мы получаем достаточно много контекстов с данной единицей, проявляющейся в живой речи. В китайском языке грамматика тесно связана с последовательностью слов, для изучения порядка слов мы можем использовать разные знаки с определенной целью.

В лингвистическом корпусе язык описывается в том виде, как он проявился в речи. В отличие от традиционной методики обучения иностранным языкам – на базе корпусных данных проводится работа со словоупотреблениями в том виде, в каком они встречались в контекстах. В обучении РКИ применяются высокорезультативные методы обработки эмпирического словесного материала. Активная работа на занятиях и во внеаудиторных ситуациях позволяет китайским студентам быстрее адаптироваться к обучению и жизни в русскоязычной среде.

С помощью НКРЯ преподаватель может создать разные типы упражнений:

- задания для изучения лексического значения слов;
- задания для изучения грамматики.

В разных подкорпусах НКРЯ возможно осуществлять разные виды поиска для разнообразных задач. Лингвистический корпус может оптимизировать процесс обучения русскому языку. Изучение русского, как и китайского языков может осуществляться учащимися самостоятельно. Лингвистические корпуса хорошо удовлетворяют индивидуальные запросы каждого учащегося.

Лингвистический корпус имеет широкие перспективы в обучении РКИ. По мере развития русского и китайского корпусов, он окажет большую помощь в самых разных областях преподавания при становлении у учащихся основных черт вторичной языковой личности.

©МГЛУ

АВТОМАТИЧЕСКАЯ ГЕНЕРАЦИЯ АНГЛОЯЗЫЧНЫХ КОНТЕКСТНО-ЗАВИСИМЫХ АННОТАЦИЙ

В.Н. ВЕРЕЩАКО, И.И. ЗУБОВА

The article deals with the problem of automatic generation of English Internet text context based abstracts. The formal model is created with this aim. It consists of the following parts: 1) the computer data base including English accessory words – prepositions, articles and conjunctions; 2) the algorithm. The Internet text sense compression is based on the statistic-distributional method. The computer experimental results don't contain any errors in case the user query includes one or two key words

Ключевые слова: генерация, Интернет-документ, компрессия, контекстно-зависимая аннотация, модель

Постоянный рост информационных ресурсов сети Интернет обусловил необходимость разработки дополнительных средств обработки информации, в первую очередь, средств смысловой компрессии текстовых документов и построения их аннотаций. В последнее время становятся популярными контекстно-зависимые вторичные документы, которые создаются поисковыми системами на основе текста запроса пользователя. Они содержат короткие выдержки из текста найденного в Сети документа и помогают пользователю принимать решение о необходимости обращения к тексту первоисточника.

Данная работа посвящена моделированию процесса формирования текста контекстно-зависимой аннотации англоязычного Интернет-документа. Материалом исследования послужили двадцать текстов англоязычных документов, взятых с разных сайтов сети Интернет. Предложенная в работе формальная модель состоит из двух частей: лингвистической базы данных и алгоритма формирования текстов данного типа. Лингвистическая база данных системы включает списки англоязычных предлогов, союзов и артиклей. Формальная модель базируется на статистико-дистрибутивном методе, позволяющем выделить из текста наиболее информативные предложения.

В процессе автоматического создания текста аннотации можно выделить три основные подзадачи: анализ запроса пользователя, поиск соответствующей запросу информации и формирование контекстно-зависимой аннотации. Рассмотрим подробнее процедуру решения третьей подзадачи. Пользователь вводит запрос на поиск определенной информации. После того как компьютер обработал запрос и нашел в Сети соответствующий документ, он создает по тексту этого документа частотно-

алфавитный словарь. Все словоформы словаря упорядочены по убыванию частоты их употребления в тексте. Далее с опорой на базу данных из словаря удаляется вся служебная лексика. Из оставшихся единиц словаря компьютер выбирает те, которые соответствуют словам запроса, а также те, которые обладают самой высокой частотой употребления (на основе сравнения со средним пороговым значением частоты). При формировании аннотации компьютер выполняет окончательный анализ найденных соответствий, анализируя контекст и местоположение слов запроса пользователя в тексте первичного документа. Возможны следующие действия системы. Если предложение начинается с поисковой строки, то после нее в аннотацию включается еще шесть слов. Если поисковая строка находится в середине предложения, то справа и слева к ней добавляются по три слова. Если поисковая строка находится в конце предложения, то к ней добавляется шесть слов, входящих в левосторонний контекст. Таким же образом формируется контекст самых частотных слов документа. Полученная последовательность лексических единиц и представляет собой текст аннотации. К сформированной аннотации добавляется имя первичного документа, и текст передается на страницу результатов поиска, которую воспроизводит браузер. Например, по запросу *snippet* пользователь получит следующую контекстно-зависимую аннотацию:

Snippet Conventions doc

... abilities described previously snippet management features can ... them. These snippet feature groups include ... elements within a snippet that are left ... text of the snippet is inserted during ... rest of the snippet text. Some systems ... part of the snippet and specify that ... In addition to the basic management abilities ... to the scope of interactivity between snippets ... In addition to the basic management the text editor or application that hosts.

Возможность продолжения исследования по данной тематике видится в совершенствовании формальной модели с целью корректного построения контекстно-зависимых аннотаций на основе длинных запросов пользователя (содержащих более двух слов).

©БрГУ им. А.С. Пушкина

УПРАВЛЕНИЕ ГЛАГОЛОВ ПСИХИЧЕСКОЙ ДЕЯТЕЛЬНОСТИ В РУССКОМ И ПОЛЬСКОМ ЯЗЫКАХ

Е.Е. ВЕРТЕЙКО, О.Б. ПЕРЕХОД

In the article there is an attempt to describe the divergence and convergence of the syntactic models of the verb government in Russian and Polish languages. Semantic area of verbs is limited to the verbs of mentality

Ключевые слова: управление глаголов, модель управления, коррелят

Компаративный семантико-синтаксический анализ глаголов, принадлежащих к лексико-семантической группе глаголов со значением психической деятельности, показывает, что каждая синтаксическая позиция как элемент структурной схемы обладает своим основным, наиболее типичным вариантом выражения. Для исследуемого поля коррелятивных глаголов характерно **15 общих и 39 дифференциальных моделей управления**. Большинство глаголов, принадлежащих к лексико-семантическому полю психической деятельности человека, образуют примерно одинаковое количество сходных и различных моделей управления. Выделяются глаголы восприятия (например, *приглядывать* – *patrzeć*, *слушать* – *łuchać*) и глаголы эмоционального состояния и эмоционального переживания (например, *грустить* – *żałować*, *мучиться* – *cierpieć*), у которых количество дифференциальных моделей значительно превосходит количество общих.

Общей обязательной позицией при глаголах со значением психической деятельности в исследуемых славянских языках является позиция субъекта/агенса, выраженная преимущественно именной частью речи в именительном падеже и позиция прямого объекта, выраженная винительным падежом без предлога (*чувствовать холод* – *czuć chłód*). Для называния адресата действия в русском языке при глаголах со значением психической деятельности используется предложно-падежная конструкция **k+dat** (*стремиться к кому-чему*), в то время как в польском языке используется генитивная конструкция **do+gen** (*dażyć do kogo-czego*). В русском языке глаголы с общим значением восприятия образуют две возможные (предложную и беспредложную) модели обязательного управления **za+instr** (*следить за кем-чем*) и **Sacc** (*воспринимать кого-что*), в то время как в польском языке глаголы этой же семантической подгруппы образуют четыре возможных варианта обязательного управления: беспредложные **Sacc** (*postrzegać kogo-co*) и **Sgen** (*patrzeć kogo-czego*), предложные – **nad+instr** (*czuć nad kim*) и **na+acc** (*czuchać na kogo-co*). Глаголы, обозначающие способность слухового восприятия, имеют в русском и польском языках различные синтаксические валентности. Объектные актанты при данных глаголах в русском языке имеют традиционную форму винительного падежа без предлога (*слушать кого-что*), в польском – родительного падежа без предлога (*łuchać kogo-czego*).