

## К ПРОБЛЕМЕ ГЕНЕРАЦИИ УЧЕБНЫХ МАТЕРИАЛОВ НА ОСНОВЕ ЛИНГВИСТИЧЕСКОЙ КЛАССИФИКАЦИИ

**Чжан Е.А., Личаргин Д.В.**

**научный руководитель канд. тех. наук Личаргин Д.В.  
Сибирский федеральный университет**

Актуальной на сегодняшний день является проблема автоматизации систем письменного и устного перевода для различных языков, экспертных, поисковых систем и систем реферирования. Для решения данных задач успешно реализуются многочисленные теории, концепции и программные системы. Особенно важной оказывается проблема автоматической генерации учебных материалов, как частного случая текстов на естественном языке.

Новизна данной работы состоит в нахождении некоторых принципов проецирования понятийного пространства слов языка на иерархическую структуру учебного курса, в частности по английскому языку и разработки соответствующих методов.

Основная идея работы состоит в проекции понятийного пространства единиц языка на структуру электронного учебного курса. Способы проецирования – многовариантны.

Цель работы – построить модель генерации текстов для учебных курсов по английскому и принципиально любому другому языку. Задачи работы состоят в применении данной модели на основе порождающих грамматик над ориентированным лесом строк, использование векторизованной классификации семантических понятий и слов естественного языка для генерации учебных заданий на основе деревьев строк.

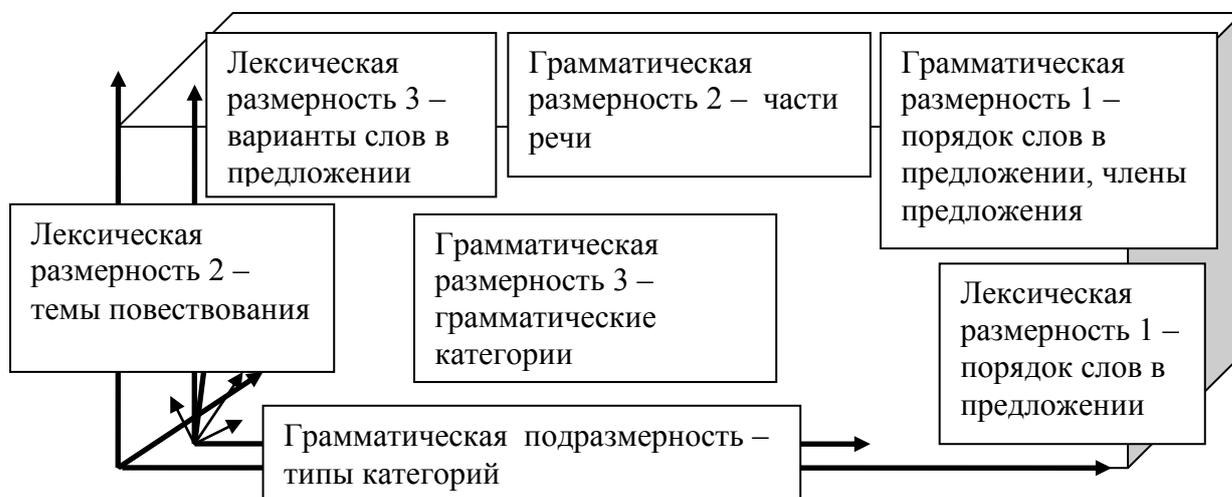


Рис. 1. Координаты многомерного лексико-грамматического подпространства леса данных естественного языка

Единицы естественного языка включаются в классы единиц естественного языка, пересечение и комбинаторика которых порождает его парадигматическую систему – фрагменты реляционной базы данных как срезов многомерной базы данных (рис. 1).

Парадигмы естественного языка являются подмножествами многомерных пространств, или (что эквивалентно) древесных иерархий естественного языка, представляемых на основе векторов признаков древесной классификации, т.е. векторов координат многомерного пространства для состояний единиц естественного языка.

Учебный материал есть подмножество естественного языка, задаваемое шаблонами особого вида: вопрос – ответ, вопрос – варианты ответов, текст – слова к тексту и т.д. Таким образом, для создания системы автоматической генерации учебных материалов, например, для уроков по английскому языку необходимо вначале зафиксировать, т.е. задать текстологический срез в системе естественного языка. Далее необходимо описать степени свободы текстологически незафиксированных единиц языка на основе задания подпространств состояний в виде: (1) подмножеств семантической классификации слов и понятий естественного языка; (2) множеств функций истинности над подмножествами семантической классификации слов и понятий естественного языка.

Слово в предложении является распределенной системой, так, например, слово wake ... up состоит из двух элементов распределенных в предложении. Рассмотрим следующую модель шаблонов для автоматической генерации учебных заданий. Например, ниже приводится дерево генерации учебных заданий со ссылками на источник слов в понятийном пространстве семантической базы данных.

1. Учебное пособие X1 [текст];
    - 1.1. Раздел Y1 [подраздел];
    - 1.2. Раздел Y2 [подраздел];
      - 1.2.1. Шаблон Z1 [подраздел] <текстологический текстовый шаблон>;
        - 1.2.1.1. «Ответьте на следующие вопросы:» [позиция в предложении];
        - 1.2.1.2. Вопрос [позиция в предложении];
          - 1.2.1.2.1. «Можно ли найти ...< в ...<<?» [вариант];
          - 1.2.1.2.2. «Находится ли ...< в ...<<?» [вариант];
          - 1.2.1.2.3. «Является ли ...< достопримечательностью ...<<?» [вариант];
          - 1.2.1.2.4. «Известен ли ...<< таким сооружением как ...<?» [вариант]
        - 1.2.1.3. Достопримечательности [позиция в предложении] <экспорт групп слов «здания», «памятники»>;
          - 1.2.1.3.1. Тауэр [вариант];
          - 1.2.1.3.2. Лондонский мост [вариант];
          - 1.2.1.3.3. Статуя Свободы [вариант];
          - 1.2.1.3.4. Биг Бэн [вариант];
          - 1.2.1.3.5. Царь Колокол [вариант];
        - 1.2.1.4. Страны [позиция в предложении] <экспорт группы слов «страны»>;
          - 1.2.1.4.1. Лондон [вариант];
            - 1.2.1.4.1.1. Столица Великобритании [синоним];
            - 1.2.1.4.1.2. Столица Туманного Альбиона [синоним];
          - 1.2.1.4.2. Вашингтон [вариант];
            - 1.2.1.4.2.1. Столица США [синоним];
          - 1.2.1.4.3. Москва [вариант];
- 1.3. Раздел Y3 [подраздел].

В результате работы порождающей грамматики над ориентированным лесом строк над шаблоном генерации осмысленных текстов получают тексты следующего вида: «Англоязычные страны имеют различные системы управления. Королева является главой Великобритании, в США глава страны – президент ...».

Такие деревья генерации текстов на естественном языке можно использовать для генерации строк символов текста на выходе системы с использованием порождающих грамматик над лесом строк. Приведем примеры необходимых для этого правил порождающей грамматики, где «0» означает нулевой символ.

Для генерации учебных материалов важным аспектом является генерация учебных текстов. Для генерации простейших учебных текстов будем использовать дерево генерации текста, представляющее собой дерево строк, например:

1.	Учебное пособие X2;	1.2.1.2.3.	Знак зодиака [позиция в предложении];
1.1.	Раздел Y1 [раздел];	1.2.1.2.4.	Характер [позиция в предложении];
1.2.	Раздел Y2 [раздел];	1.2.1.2.5.	Отношения [позиция в предложении];
1.2.1.	Шаблон Z1 [раздел];	1.2.1.2.6.	Занятия [позиция в предложении];
1.2.1.1.	Лицо [позиция в предложении];	1.2.1.2.6.1.	Книги
1.2.1.1.1.	Я [вариант];	1.2.1.2.6.1.1.	Позитивная модальность [варианты];
1.2.1.1.2.	Мой друг [вариант];	1.2.1.2.6.1.1.1.	Обожать;
1.2.1.1.3.	Моя подруга [вариант];	1.2.1.2.6.1.1.2.	Любить;
1.2.1.1.4.	Мой одноклассник [вариант];	1.2.1.2.6.1.1.3.	Нравится;
1.2.1.1.5.	Мой дядя [вариант];	1.2.1.2.6.1.1.4.	Часто;
1.2.1.2.	Аспект [позиция в предложении];	1.2.1.2.6.1.1.5.	Постоянно;
1.2.1.2.1.	Имя [позиция в предложении];	1.2.1.2.6.1.2.	Действия с книгами;
1.2.1.2.1.1.	Фамилия;	1.2.1.2.6.1.2.1.	Читать / Читает;
1.2.1.2.1.1.1.	Связка [варианты];	1.2.1.2.6.1.2.2.	Перечитывать / перечитывает;
1.2.1.2.1.1.1.1.	Имя ... -;	1.2.1.2.6.1.2.3.	Просматривать / просматривает;
1.2.1.2.1.1.1.2.	... зовут;	1.2.1.2.6.1.3.	Типы книг;
1.2.1.2.1.1.1.3.	... -;	1.2.1.2.6.1.3.1.	Книги;
1.2.1.2.1.1.1.4.	Фамилия ... -;	1.2.1.2.6.1.3.2.	Классику;
1.2.1.2.1.1.2.	Типы книг;	1.2.1.2.6.1.3.3.	Детектив;
1.2.1.2.1.1.2.1.	Иванов;	1.2.1.2.6.1.3.4.	Сказки;
1.2.1.2.1.1.2.2.	Петров;	1.2.1.2.6.2.	Музыка;
1.2.1.2.1.1.2.3.	Сидоров;	1.2.1.2.6.3.	Кино;
1.2.1.2.1.1.2.4.	Браун;		
1.2.1.2.2.	Год рождения [позиция в предложении];		

В результате генерации осмысленного подмножества естественного языка получаются тексты вида: «Меня зовут – Иван. Моя фамилия – Иванов. Я родился первого октября 1980 года. Мне 30 лет. Мой знак зодиака – Весы. Я люблю классическую музыку. ...». При этом строки на узлах деревьев разнородных данных должны быть объединены в лес данных, частично как подмножества понятийного пространства естественного языка, частично, как подмножества иерархии шаблонов электронного учебного курса.

На основе порождающих грамматик над лесом строк возможна дополнительная генерация дерева текстовых шаблонов с добавлением семантического шума и порождении текста на естественном языке, включающего информацию в иерархической системе текста, сгенерированную на основе знаний базы данных. Так, например, вместо фразы: «я люблю читать детективы» может быть употреблена разговорная фраза не приведенного вида: «я не прочь читкануть детективчик» с той же формально-базовой семантикой.

Классификация понятий естественного языка может служить источником лексических единиц, для составления шаблонов генерации осмысленных текстов, которые можно усложнять посредством добавления семантического шума на основе расширенных порождающих грамматик над лесом строк и деревьев разнородных данных.

**Построение электронных курсов на основе дерева времени** с учетом построения их как сложных гибких систем.

Рассматриваемая модель электронного курса – мульти-иерархическая система, представленная таким математическим объектом, как лес данных, т.е. множество деревьев данных с пересекающимися узлами. Данная мульти-иерархическая система представляет собой несколько вложений:

1. Иерархия времени: годы, месяцы и т.д.;
2. Иерархия процесса обучения: дисциплина, курс, модуль и т.д.;
3. Иерархия учебных материалов и их представления: база данных, подмножества, материалы урока, задания, состояние интерфейса обучающей программы, настройки.

Настройки предлагается представить как множества соответствий: параметр настройки – дерево учебного курса. Рассмотрим некоторые параметры настроек, задающих рамки для генерации учебного курса.

Траектория обучения определяется на основе опций настройки и управления параметрами генерации учебных заданий, модулей и курсов для различных этапов обучения иностранному языку.

**Выделение необходимых параметров для настройки мульти-иерархических электронных курсов.** К параметрам, задаваемым для генерации учебных курсов, относятся, в том числе, важнейшие понятия методики обучения иностранным языкам: подход к обучению, принципы обучения, цели и содержание обучения, методы и приемы обучения. Как показано в работе Личаргина Д.В., Таранчук Е.А. Иерархическая структура учебного электронного курса и ее вариабельность для обучения иностранному языку / Журнал Дистанционное и виртуальное обучение, 2011, № 4, стр. 56-76 электронный курс должен основываться на моделировании подходов, моделей и принципов обучения.

Подход к обучению (англ. approach – подход, подступ) выступает в качестве самой общей методологической основы обучения, характеризуя существующие точки зрения на предмет обучения (язык) и способы овладения им в процессе обучения, определяет деятельность по обучению ИЯ. Подход к обучению отражает стратегию обучения (что получим в результате), задает общую концепцию обучения. Существуют классификация подходов к обучению иностранному языку: бихевиористский подход, сознательный подход, когнитивный подход, интегрированный подход, языковой подход и речевой подход.

При разработке модели обеспечения гибкости построения и структуры обучающего электронного курса в той или иной степени учитываются все вышеназванные подходы к обучению иностранному языку. Выбрав в качестве доминирующего тот или иной подход, можно задать соответствующие параметры настроек.

Под принципами обучения понимаются исходные положения, которые определяют цели, содержание, технологию обучения и проявляются во взаимосвязи и взаимообусловленности. Принципы призваны определять стратегию и тактику обучения иностранному языку практически в каждой точке учебного процесса. В классической дидактике описаны общедидактические или безусловные принципы обучения, которых нельзя избежать. Помимо общедидактических принципов, в методике обучения ИЯ можно выделить несколько принципов, актуальных только для данной дисциплины: принцип коммуникативной направленности обучения, принцип дифференциации и интеграции, принцип учета родного языка.

Метод обучения – это направление в обучении, реализующее цели, задачи и содержание обучения языку и определяющее пути и способы их достижения. Слово «метод» означает путь к достижению поставленной цели.

Ключевым моментом любого обучения является практика. Практика предполагает тренировку учащихся в использовании изучаемого явления. Выполняются различные языковые упражнения: заполнение пропусков, дополнение предложений, воспроизведение с опорой на наглядность, заучивание коротких диалогов. Тренировка, предусматривающая многократную «встречу» с учебным материалом, обеспечивает запоминание материала, удержание его в памяти.

В заключение, необходимо отметить, что представление подмножеств естественного языка в виде деревьев строк позволяет осуществлять генерацию текстов, в частности, в качестве составляющих учебных заданий для уроков английского языка.