



OSTIS-2011

(Open Semantic Technologies for Intelligent Systems)

УДК 007.52

ПРИМЕНЕНИЕ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ОБУЧЕНИИ ЯЗЫКАМ

О.Е. Елисеева (*olae@open.by*)

Белорусский государственный университет, г.Минск, Республика Беларусь

В работе предлагается создание базы знаний естественных языков, предназначенной для самостоятельного изучения нескольких языков одновременно. Заложенные в основу данной работы принципы основаны на достижениях психологии, лингводидактики, лингвистики и искусственного интеллекта.

Ключевые слова: база знаний, когнитивная психология, компьютерная лингвистика, компьютерная лингводидактика, семантическая сеть.

Введение

В настоящее время, время межкультурных коммуникаций и международного обмена, время так называемого информационного взрыва, актуальным является знание различных языков. Трудно представить себе современного человека, не владеющего несколькими языками. Проблема изучения языков на сегодняшний день является важной практически для каждого человека, независимо от сферы его профессиональной деятельности. Несмотря на огромное количество различных методик обучения языку, изучение иностранного языка (и тем более, нескольких языков) зачастую становится крайне сложным и долговременным процессом.

В данной работе предпринята попытка обобщения результатов исследований в областях когнитивной психологии, компьютерной лингводидактики, педагогического дизайна, компьютерной лингвистики, современных информационных технологий и технологий искусственного интеллекта, объединение которых в единую компьютерную систему обучения может позволить достичь нового качества в процессе изучения нескольких языков одновременно.

1. Научно-методический базис

1.1. Психология

Исследованием вопросов усвоения знаний человеком, запоминания и хранения информации занимается **когнитивная психология** [Солсо, 2006]. Ещё в 60-е годы 20-го века в рамках данной науки рассмотрено и предложено достаточно много теорий восприятия, моделей репрезентации знаний и методов, способствующих лучшему запоминанию информации.

В рамках так называемого ассоцианистского подхода рассматривается одна из моделей репрезентации знаний – **модель иерархии понятий**, в которой понятия некоторой предметной области представлены в виде взаимосвязанных подструктур, отображающих явные иерархические зависимости этих понятий. Сегодня известны практические применения данной модели в процессе обучения различным учебным дисциплинам [Бершадский, 2010]. При этом применяется так называемый **метод карт понятий**, предложенный Д. Новаком и А. Канасом. Данный метод получил своё воплощение в виде специализированной компьютерной программы IHMC SmartTools.

В когнитивной психологии рассматриваются так называемые **когнитивные модели семантической памяти**, которые, как считают авторы, в наибольшей степени соответствуют способу хранения информации в памяти человека. Особое место в этих моделях занимают **пропозициональные сети**. При этом под **пропозициями** понимаются некие абстракции, формируемые в мозгу человека и основанные на наблюдениях, таких как чтение текста или слушание говорящего. С использованием пропозиций психологи и лингвисты пытаются объяснить механизмы запоминания человеком информации и усвоения знаний, представленных, в частности, в форме естественного языка и речи.

В процессе изучения любой новой информации огромное значение имеет скорость её запоминания и усвоения. Специалистами разрабатываются всевозможные методы, позволяющие расширить возможности нашей памяти и повысить тем самым эффективность запоминания. Одним из наиболее известных и распространённых таких методов является **мнемоника** – это метод или способ, например рифма или образ, основанный на использовании знакомых ассоциаций для улучшения сохранения информации в памяти и ее припоминания [Солсо, 2006].

Еще одним методом, способствующим лучшему усвоению и запоминанию информации, является **метод интеллект-карт** (mind mapping), который в последнее десятилетие получил очень широкое распространение в различных сферах деятельности – от образования до компьютерного проектирования [Бьюзен, 2003]. Некоторые авторы считают этот метод универсальным, позволяющим представить в виде сети ассоциаций практически любую информацию, тем самым переработав и обобщив её. Распространению и популярности данного метода способствует наличие соответствующего программного обеспечения, позволяющего создавать интеллект-карты и использовать их в интерактивном режиме.

1.2. Компьютерная лингвистика

Базируясь на исследованиях психологов и лингвистов, специалисты в области **компьютерной лингвистики** создают и совершенствуют всевозможные формальные языки описания, моделирования и обработки компьютером естественного языка. Важнейшей их заслугой является формализация и систематизация естественного языка, благодаря чему уже сейчас существуют реально действующие компьютерные системы, в рамках которых реализованы многие теоретические достижения.

Одним из важнейших достижений в рамках лингвистического направления когнитивной психологии является выделение Ноэмом Хомским и другими лингвистами **поверхностных и глубинных структур**, соответствующих различным уровням кодирования информации в естественном языке.

В настоящее время существуют вполне качественные системы автоматического перевода текстов. Использование таких систем в некоторой мере упрощает проблему перевода текстов с одного естественного языка на другой. Однако, для получения качественного перевода сложного текста по-прежнему нужен специалист, владеющий языком. Поэтому проблема изучения иностранного языка остаётся актуальной.

Ещё большую актуальность знание иностранного языка приобретает в случае необходимости живого общения (коммуникации) с носителями других языков. Как указывалось выше, в настоящее время это особенно актуально, так как современный рынок немислим без международных контактов.

1.3. Компьютерная лингводидактика и педагогический дизайн

Вопросами изучения и преподавания языков занимается **лингводидактика**, применение современных компьютерных технологий в котором привело к появлению **компьютерной лингводидактики**. И если в рамках лингводидактики разрабатываются методики эффективного изучения языка, то основной задачей компьютерной лингводидактики является дальнейшее повышение эффективности усвоения языка с использованием всех современных возможностей **компьютерных информационных технологий**. В частности, одним из очевидных плюсов применения компьютера в обучении языку являются широкие возможности **мультимедиа** и **сетевых технологий**. Убедиться в этом можно, например, посетив сетевое сообщество по изучению языков busuu.com [Busuu, 2010].

Особый интерес у разработчиков вызывает использование мультимедиа-технологий. Однако процесс подготовки и использования иллюстративного материала для компьютерных систем обучения языку требует не только технических навыков использования графических редакторов и пр. инструментов, но и детальной методической проработки [Елисеева, 2010]. Современный уровень развития вычислительной техники позволяет эффективно использовать в учебном процессе аудио- и видеоинформацию, что особенно востребовано в процессе изучения языков.

Компьютерная лингводидактика, таким образом, предполагает применение различных **методик обучения языку** в рамках **компьютерных средств и систем обучения**. При этом очевидно, что это влечёт за собой потребность в разработке новых методик обучения, а также применения современных достижений психологии и лингвистики (в том числе компьютерной лингвистики).

Особую актуальность в современных условиях приобретает также **педагогический дизайн**. Именно в рамках данного направления уделяется максимум внимания проработке структуры и содержания учебного материала, построения системы обучения в целом. При этом максимальное внимание уделяется **логике учебного процесса**, а не деталям и возможностям используемых технологий.

1.4. Искусственный интеллект

Нельзя не отметить актуальность применения технологий **искусственного интеллекта** в разработке компьютерных систем обучения. Очевидно, что процесс обучения насыщен задачами, которые принято называть **интеллектуальными**. В учебном процессе участвуют два субъекта: учитель и ученик. В связи с этим имеет смысл различать учебные задачи ученика и учителя. В данной работе будем рассматривать учебные задачи в контексте изучения языков.

К основным задачам ученика относятся:

- изучение и запоминание новых слов – этому способствует также решение задачи понимания взаимосвязей (семантических, иерархических) между отдельными словами и группами слов;
- усвоение особенностей употребления слов в словосочетаниях и предложениях;
- изучение правил построения словосочетаний и предложений;
- приобретение навыков произношения отдельных слов, словосочетаний, слитного текста (с переходом в естественную речь).

Основными задачами учителя языка являются:

- предъявление наборов изучаемых слов в методически обоснованной последовательности;
- объяснение особенностей произношения и употребления слов в словосочетаниях и предложениях;
- управление процессом обучения, контроль и оценка знаний, исправление ошибок учащегося.

Выделение субъектов учебного процесса и разделение их задач на две группы позволяет говорить о двух способах применения компьютерных информационных технологий (и технологий искусственного интеллекта, в частности) для поддержки процесса изучения языков. Можно утверждать, что все функции учителя должны моделироваться в рамках компьютерной системы обучения. Учитывая сложность многих этих функций, применение технологий искусственного интеллекта здесь вполне уместно. Это привело к появлению целого класса систем, названных интеллектуальными обучающими системами (ИОС) [Голенков и др., 2001].

Современный уровень развития технологий искусственного интеллекта позволяет многие из перечисленных учебных задач **моделировать** или **поддерживать**, способствуя тем самым повышению эффективности обучения. Так, в рамках интеллектуальных обучающих систем создаются следующие **модели**: учащегося, учителя, управления обучением, предметной области и др. [Елисеева, 2000]. Создание таких моделей в конечном итоге приводит к формированию соответствующих **баз знаний** (БЗ), обработка которых позволяет моделировать учебный процесс в целом. При этом в базе знаний представлена некая эталонная модель знаний учащегося, к которой требуется приблизить реальное состояние знаний обучаемого.

Создание адекватных моделей, сопутствующих реализации учебного процесса, является очень непростым и трудоёмким процессом. В некоторых случаях проблему можно решить путём привлечения технологий *машинного обучения* и *методов мониторинга и адаптации*. В качестве ещё одной из альтернатив решения указанной проблемы в данной работе предлагается обеспечение максимально комфортной *поддержки* учебного процесса с привлечением технологий искусственного интеллекта.

Анализ результатов исследований психологов даёт основания полагать, что представление знаний в базах знаний интеллектуальных систем в виде *семантической сети* является одним из наиболее соответствующих представлению знаний в памяти человека. Упомянутые выше пропозициональные сети, методы карт понятий и интеллект-карт визуализируются в учебном процессе в виде графовых структур, подобных тем, которые используются при создании семантических сетей. Следовательно, есть все основания полагать, что предлагаемые психологами модели в полной мере могут быть отображены в виде семантической сети. Создаваемые в рамках проекта OSTIS (Open Semantic Technology for Intelligent Systems) семантические языки представления знаний и методы их визуализации являются хорошей базой для представления знаний об изучаемых языках в наглядной форме с поддержкой интерактивности и мультимедийности.

2. Особенности предлагаемого подхода

В данной работе предлагается подход к созданию базы знаний, моделирующей лексический и грамматический состав нескольких естественных языков. База знаний описывается на семантическом языке представления знаний и обеспечивает следующие возможности:

- визуализация лексического состава любого естественного языка или их совокупности в виде соответствующих иерархических структур, в рамках которых явным образом обозначены общие для различных языков черты;
- визуализация грамматического состава естественных языков с использованием достаточного количества реальных практических примеров словоупотреблений в составе словосочетаний, предложений, диалогов, текстов;
- поддержка лексического и грамматического состава языков визуализацией графических иллюстраций – рисунков, фотографий и пр. – для наглядного интуитивного восприятия информации учащимся;
- возможность прослушивания аудиозаписей произношения различных языковых единиц (слов, словосочетаний, предложений, текстов);
- интерактивное взаимодействие с элементами сети и поиск отдельных фрагментов по различным признакам.

Предлагаемая концепция базируется на следующих перечисленных ниже основных положениях.

Учащийся становится центральным объектом учебного процесса, он его реализует и развивает самостоятельно, имея мотивацию, а также стремление к познанию и исследованию. Ему при этом предоставляется максимум свободы, которая поддерживается всеми перечисленными выше возможностями работы с семантической сетью. Портрет описанного учащегося наиболее полно соответствует достаточно взрослому, образованному человеку, владеющему базовыми навыками работы с вычислительной техникой. Заметим при этом, что такой образ учащегося вполне уместен именно в случае изучения нескольких языков одновременно, т.к. такую задачу ставят перед собой именно взрослые люди, что подтверждается также мониторингом аудиторий, изучающих различные иностранные языки с использованием сети Интернет.

Из учебного процесса извлекаются лидирующая роль учителя и учитель в классическом понимании. Учащийся становится полноценным и единственным участником учебного процесса. При этом на него возлагаются функции самоконтроля и самооценки. Здесь также учитывается немаловажный психологический фактор – по сведениям психологов большинство взрослых людей характеризуются тем, что боятся публичных ошибок, а это неизбежно происходит, если язык изучается в аудитории с преподавателем. Взрослый человек не любит также, когда ему «диктуют» некую иную линию поведения, которой он не готов либо не хочет

следовать. Именно эти два фактора являются основными причинами неудач в процессе изучения иностранных языков.

Интерактивные возможности семантической сети позволяют реализовать деятельностно-ориентированный исследовательский подход к обучению, что в значительной мере способствует более эффективному усвоению материала.

Наличие адекватной семантической модели языка и её визуализация в виде интерактивной семантической сети с поддержкой мультимедиа-технологий создают эффект погружения в виртуальную исследовательскую языковую среду. Это способствует формированию более устойчивого понимания того или иного языка на интуитивном уровне, без необходимости дополнительных глубоких объяснений. Именно таким образом, по всей видимости, происходит усвоение языка ребёнком, которому никто не сообщает правил и грамматических особенностей языка.

Таким образом, представление знаний о языке в виде семантической сети является адекватной *поддержкой* процесса его изучения. В данной работе, кроме того, предлагается рассматривать *интегрированную семантическую модель для нескольких естественных языков*, в рамках которой учащийся получит возможность на практике усвоить межязыковые сходства и различия. При этом информация о языке должна быть описана таким образом, чтобы обеспечить решение всех перечисленных выше учебных задач учащегося. Учебные же задачи учителя могут частично «решиться сами собой». Это, в частности, можно предположить относительно первых двух, перечисленных выше, т.к. при создании модели языка очевидным образом должна образоваться сеть, содержащая все присущие каждому конкретному языку семантические, иерархические и др. структурные связи, на основе которых и вырабатывается логика учебного процесса. Третью же группу задач преподавателя можно либо оставить без внимания, возложив их решение на учителя, либо создать соответствующие модели в составе БЗ по тем принципам, которые предлагаются в рамках ИОС, либо вообще устранив их из учебного процесса, предоставив при этом учащемуся полную свободу действий.

Итак, в данной работе будем говорить о создании *Системы поддержки самостоятельного изучения нескольких естественных языков* путём формирования базы знаний на семантическом языке представления знаний.

3. Структура знаний о естественных языках и изучение языка

Прежде чем рассматривать структуру базы знаний о естественных языках, отметим, что любой язык имеет две основные составляющие: лексика (словарный состав) и грамматика (набор правил построения конструкций данного языка). С другой стороны, классический подход к построению баз знаний предполагает наличие так называемых декларативных и процедурных знаний. Таким образом, можно соотнести лексику языка с декларативным уровнем знаний, а грамматику – с процедурным. При таком подходе появляется возможность говорить о создании среды поддержки формирования у учащихся как декларативного, так и процедурного уровней знаний о языках.

Изложенные ниже подходы к представлению знаний о нескольких естественных языках одновременно можно также рассматривать как некий ориентировочный сценарий изучения языков. В качестве иллюстраций ниже будут приводиться примеры для четырёх языков: русский, белорусский, английский и французский.

Следует также отметить, что в данной работе не ставится задача формализации естественного языка [Елисева, 2009]. Здесь предлагается не столько мета-описание языка, сколько систематизация его употребления в реальной жизни с целью предоставления учащимся соответствующих иллюстраций – как языковых, так и изобразительных, в виде рисунков, фотографий и пр. визуализаций объектов реальной действительности.

3.1. Декларативные знания о языках

Для получения знаний о лексическом составе любого языка достаточно иметь доступ к словарю этого языка. С учётом современных подходов к организации обучения языку с использованием информационных технологий становится очевидным, что изучаемый словарь обязательно должен быть, как минимум, иллюстрированным и, как максимум, поддерживать звуковое (речевое) воспроизведение слов.

На рис.1 представлен пример формирования отдельной «статьи» словаря в виде фрагмента семантической сети, включающей в качестве узлов слова из разных языков, которые связаны с соответствующей иллюстрацией (одной и той же). Кроме того, т.к. все перечисленные в данном фрагменте слова обозначают одно и то же, это означает, что они эквивалентны (синонимичны) друг другу. Указанный факт изображён в виде двойного ребра между узлами-словами.

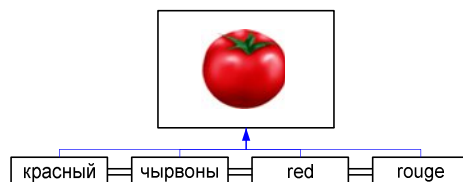


Рисунок 1 – Слово «красный» на 4-х языках и с иллюстрацией

Аналогичным образом в БЗ описываются все остальные слова «объектного» (предметного) уровня, т.е. те, которые используются в качестве наименований конкретных предметов, явлений, состояний, качеств и пр.

Известно, что наряду с «объектным» уровнем в развитых языках существуют различные уровни абстракций (языковые абстракции). Так, например, для слов «красный», «жёлтый», «зелёный» и т.д. имеется одно слово-абстракция «цвет». Слово «цвет» невозможно изобразить в виде иллюстрации. Поэтому в БЗ это может быть отображено и предъявлено учащемуся в виде фрагмента семантической сети, представленной на рис.2.

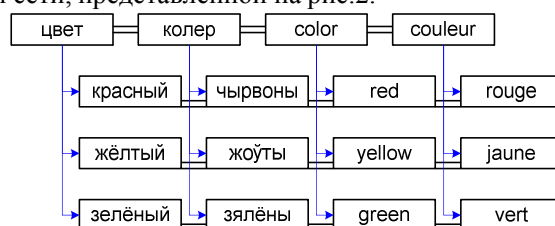


Рисунок 2 – Переход от «предметного» уровня к абстрактному

На рис.3 приведен пример формирования словосочетания на декларативном уровне.

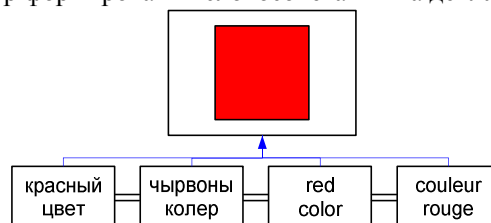


Рисунок 3 – Словосочетания на декларативном уровне

На рис. 4 продемонстрированы возможности естественного языка к описанию одного и того же объекта с использованием различных синтаксических конструкций, которые, очевидно, являются синонимичными.

Для обеспечения полноценной исследовательской работы учащегося с базой знаний функционально рассматриваемая Система поддержки самостоятельного изучения языков на декларативном уровне изучения языка помимо изобразительных (статичных) возможностей должна поддерживать следующие интерактивные функции:

- поиск отдельных слов на «предметном» уровне;
- переход от «предметного» уровня языка к абстрактному и наоборот;
- переход от отдельных слов к словосочетаниям и предложениям;
- переход от словосочетаний и предложений к иллюстрациям отдельных слов, входящих в их состав;
- переход от словосочетаний и предложений к грамматическим правилам их построения (процедурному уровню БЗ).

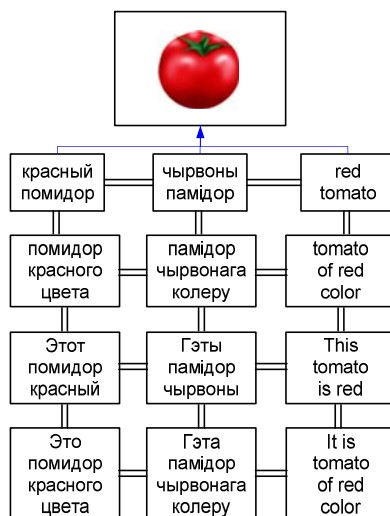


Рисунок 4 – Синонимичные синтаксические конструкции

3.2. Процедурные знания о языках

Процедурный уровень описания знаний о языках имеет целью обеспечить учащегося возможностью получить объяснения правил формирования конструкций различных естественных языков. Очевидно, что на синтаксическом уровне для различных языков эти правила могут быть разными. Процедурный уровень рассчитан на более продвинутый уровень изучения языка.

На рис.5 приведён фрагмент семантической сети, отображающей простейшее грамматическое правило построения однотипных словосочетаний. Данное правило имеет вид простой «формулы», т.к. такой способ записи должен быть понятен любому грамотному человеку.

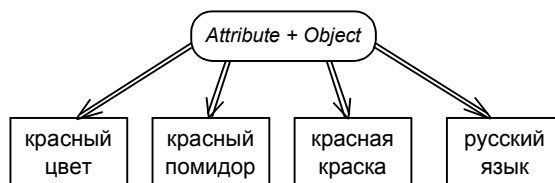


Рисунок 5 – Простейшие грамматические правила-формулы для словосочетаний

На рис. 6 – более сложное правило, описывающее структуру предложения с обозначенными синтаксическими отношениями.

В идеале, для полноценной работы с предлагаемыми конструкциями учащемуся необходимо предварительно объяснить используемые в них обозначения синтаксических отношений. С другой стороны, если использовать одни и те же обозначения в одинаковых ситуациях, учащийся может усвоить предлагаемые ему «формулы» на интуитивном уровне, подобно тому, как происходит усвоение лексики языка на наглядных примерах.

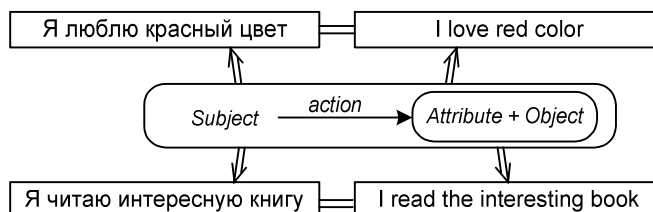


Рисунок 6 – Грамматические правила для предложений

Следует также отметить, что приведённые примеры синтаксических правил записаны в упрощённой форме, понятной именно учащемуся. На уровне баз знаний естественного языка эти правила могут быть раскрыты (представлены) по всем правилам организации предлагаемого в данной работе в качестве базиса семантического языка представления знаний.

4. Дополнительные замечания

Изначальная открытость и универсальность используемого в данной работе семантического языка представления знаний позволяет сделать следующие дополнительные замечания.

В случае наличия в БЗ достаточного объема синтаксических правил мы получаем возможность применять их для автоматического синтеза практически любых (в соответствии с номенклатурой правил) словосочетаний и предложений на разных естественных языках с использованием той лексики, которая описана на декларативном уровне. Это означает то, что формируя базу знаний, предназначенную для наглядного отображения и исследования учащимся, мы одновременно создаём основу для полноценной базы знаний о естественных языках, которая по существу является формализацией этих языков. А это, в свою очередь, даёт основания предполагать, что в перспективе эта же база знаний может быть использована для реализации полноценного естественно-языкового интерфейса интеллектуальной системы [Елисеева, 2009].

В рамках предложенного подхода в процессе формирования базы знаний языка на декларативном уровне (лексики, см. п.3.1) явным образом описываются семантические связи между словами естественного языка (например, при задании так называемых языковых абстракций). Это, в свою очередь, позволяет утверждать, что описывая лексику языка, мы тем самым описываем фрагменты различных предметных областей. И это естественно, т.к. любой естественный язык является идеальным языком представления знаний. Таким образом, предлагаемую базу знаний можно рассматривать как основу для формирования баз знаний различных предметных областей.

В заключение следует отметить, что предложенные в работе методы требуют практической реализации и апробации.

Библиографический список

- [Бершадский, 2010] Бершадский, М. Е. Психолого-педагогические основания метода карт понятий [Электронный ресурс]. – 2004. - Режим доступа: http://www.bershadskiy.ru/index/metod_kart_ponjatij/0-34 – Дата доступа: 15.10.2010
- [Бовтенко, 2006] Компьютерная лингводидактика: теория и практика: Курс лекций / М.А. Бовтенко, А.Д. Гарцов, С.И. Ельникова и др. / Под ред. А.Д.Гарцова. – М.: РУДН, 2006. – 211 с.
- [Быюзен, 2003] Быюзен, Т., Быюзен, Б. Супермышление – Минск: Попурри, 2003. – 320 с.
- [Голенков и др., 2001] Интеллектуальные обучающие системы и виртуальные учебные организации: Монография / В.В.Голенков, В.Б. Тарасов, О.Е. Елисеева и др.; Под ред. В.В. Голенкова, В.Б. Тарасова – Мн.: БГУИР, 2001. – 488 с.
- [Елисеева, 2000] Елисеева, О.Е. Графодинамические модели переработки сложно структурированной информации в интеллектуальных обучающих системах : дис. ... канд. техн. наук ; спец. 05.13.17 - теоретические основы информатики / Елисеева Ольга Евгеньевна ; науч. рук. В. В. Голенков ; БГУИР, ВЦ Российской Академии Наук, Рос НИИ ИТ и АП. – Минск, 2000. – 226 с.
- [Елисеева, 2009] Елисеева, О.Е. Естественно-языковой интерфейс интеллектуальных систем : учеб. пособие / О. Е. Елисеева ; под науч. ред. проф. В. В. Голенкова . – Минск : БГУИР, 2009. – 152 с.
- [Елисеева, 2010] Елисеева, О.Е. Эффективное использование возможностей компьютерной графики и анимации в обучении языкам // Мат-лы Международной научной конференции «Современная образовательная среда: приоритетные направления развития» (Минск, 22-23 октября 2009 г.) – Минск: НМУ «Национальный институт образования», 2010. – С. 35 – 40.
- [Солсо, 2006] Солсо, Р. Когнитивная психология / Р. Солсо. – 6-е изд. – СПб.: Питер, 2006. – 589 с.
- [Busuu, 2010] The Language Learning Community [Электронный ресурс]. – 2004. - Режим доступа: <http://www.busuu.com> – Дата доступа: 19.11.2010