

на правах рукописи

Дяченко Павел Владимирович

**РАЗРАБОТКА КОМПЬЮТЕРНЫХ МЕТОДОВ ОБУЧЕНИЯ ВЛАДЕНИЮ
ЯЗЫКОМ С ПОМОЩЬЮ АППАРАТА ЛЕКСИЧЕСКИХ ФУНКЦИЙ**

Специальность 05.13.17 - теоретические основы информатики

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата технических наук

Москва - 2008

Работа выполнена в Институте проблем передачи информации им. А. А.
Харкевича РАН

Научный руководитель:

доктор филологических наук,
профессор

И. М. Богуславский

Официальные оппоненты:

доктор технических наук,
ст. научный сотрудник

В. Л. Стефанюк

кандидат филологических наук,
научный сотрудник

Б. Л. Иомдин

Ведущая организация – Институт проблем информатики РАН.

Защита состоится «__» _____ 2008 года в ____ часов на заседании диссертационного совета Д002.077.01 при Институте проблем передачи информации им. А. А. Харкевича РАН по адресу: 127994, Москва, Большой Каретный пер., д. 19, стр. 1, конференц-зал.

С диссертацией можно ознакомиться в библиотеке Института проблем передачи информации им. А. А. Харкевича РАН по адресу: 127994, Москва, Большой Каретный пер., д. 19, стр. 1.

Автореферат разослан

«__» _____ 2008 года

Ученый секретарь
диссертационного совета
доктор физико-математических наук

И. И. Цитович

Общая характеристика работы

Актуальность диссертации. Степень владения естественным языком можно оценить по тому, насколько свободно человек может выражать на данном языке свои мысли и варьировать свои высказывания в зависимости от речевого контекста без потери или искажения смысла фразы. Для этого ему необходимо не только правильно понимать значения слов языка, но и уметь правильно (идиоматично) сочетать слова между собой.

При решении подобной задачи человек сталкивается с явлением несвободной сочетаемости слов. Это явление находит отражение, в частности, в том, что одна и та же характеристика, приписываемая денотату, может выражаться по-разному для различных лексических единиц. Это обстоятельство делает задачу построения правильного и идиоматичного высказывания особенно сложной, причем не только для программ порождения текста и не только для лиц, изучающих данный язык как иностранный, но нередко даже и для самих носителей языка.

Одним из наиболее совершенных инструментов описания несвободной сочетаемости слов, доказавшим свою практическую значимость при построении систем автоматической обработки текстов (а именно, вошедших в состав лингвистического процессора ЭТАП системы автоматического перевода и системы перифразирования, о которой будет сказано ниже), является аппарат лексических функций (ЛФ), который был предложен И. А. Мельчуком и А. К. Жолковским¹ в серии работ по модели «Смысл – Текст». Теория лексических функций была существенным образом уточнена и дополнена Ю. Д. Апресяном².

Аппарат лексических функций позволяет исчерпывающим образом описать несвободную сочетаемость любого естественного языка. Лексические функции делятся на стандартные, которые определены на достаточно большом количестве лексических единиц и при этом имеют разнообразные значения, и нестандартные, для которых одно или оба этих требования не выполняются. В существенном числе случаев несвободная лексическая сочетаемость может быть описана с помощью стандартных ЛФ. Кроме того, сопоставление наборов значений ЛФ для различных языков может способствовать выявлению наиболее характерных ЛФ для конкретного языка и тем самым идентификации глубоких особенностей его типологического строя. До разработки теории ЛФ и составления словарей их значений анализ подобных языковых особенностей был весьма затруднен.

Лексические функции являются мощным средством формализации и представления знаний о несвободной лексической сочетаемости в естественных языках. В силу этого они играют важную роль при построении когнитивных моделей естественных языков и при разработке компьютерных систем автоматической обработки текстов и человеко-машинного общения³. Однако потенциал ЛФ этим далеко не исчерпывается. Овладение несвободной лексической сочетаемостью является одной из важнейших задач, которые должны быть решены при развитии навыков

¹ И.А. Мельчук. Опыт теории лингвистических моделей "Смысл $\Leftrightarrow$ Текст". М., Наука, 1974.

А.К. Жолковский, И.А. Мельчук. О возможном методе и инструментах семантического синтеза. НТИ, 1965, № 6, с. 23-28.

И.А. Мельчук, А.К. Жолковский. Толково-комбинаторный словарь современного русского языка. Wiener Slawistischer Almanach, Sonderband 14, 1984.

² Ю. Д. Апресян. О лексических функциях семейства REAL – FACT. Nie bez znaczenia ... Prace ofiarowane Profesorowi Zygmuntowi Saloniemu z okazji jubileuszu 15000 dni pracy naukowej. Białystok, 2001, 23-40.

Ю. Д. Апресян. О семантической непустоте и мотивированности глагольных лексических функций. Вопросы языкознания, 2004, № 4, 3-18.

³ Ju. D. Apresjan, I. M. Boguslavsky, L. L. Iomdin, L. L. Tsinman. Lexical Functions in NLP: Possible Uses. Computational Linguistics for the New Millennium: Divergence or Synergy? Proceedings of the International Symposium held at the Ruprecht-Karls-Universität Heidelberg, 21-22 July 2000. Manfred Klenner / Henriëtte Visser (eds.) Frankfurt am Main, 2002, pp. 55-72.

использования естественного языка, как родного, так и иностранного. Эта задача осложнена тем, что материал, относящийся к области несвободной лексической сочетаемости, очень слабо структурирован и поэтому особенно труден для формализации и усвоения.

Овладение языком – один из важнейших информационных процессов, происходящих в каждом из нас, как на начальном этапе жизни, так и при изучении иностранного языка. В силу этого разработка методов обучения языку с помощью средств вычислительной техники позволяет внести определенный вклад в исследование информационных процессов, в которых участвуют индивидуальные пользователи.

Несмотря на то, что аппарат ЛФ давно нашел признание как в теоретической лингвистике, так и среди практиков преподавания языка, до сих пор существует крайне мало словарей, в которых соответствующие явления были бы представлены в достаточной степени системно. В какой-то мере этот пробел заполняют толково-комбинаторный словарь русского⁴ и французского⁵ языков, английские словари Oxford Collocations Dictionary⁶, BBI⁷ и LTP⁸, испанский REDES⁹. Среди этих словарей особое место занимает составленный Ю. Д. Апресяном, но пока не опубликованный, словарь ЛФ русского и английского языков, на котором базируется настоящая работа. Еще меньше идеи и материал лексических функций проникли в методики изучения языков. Опытные педагоги всегда отдавали себе отчет в важности изучения несвободной сочетаемости слов, однако на сегодняшний день систематических пособий, основанных на современной теоретической базе, насколько нам известно, ни для одного языка не существует. В настоящее время идет разработка компьютеризованных методических материалов по изучению несвободной сочетаемости (для французского языка в г. Левен¹⁰ (Бельгия) и в Монреале¹¹ (Канада), для испанского языка в Ла Корунье¹² (Испания)).

Основная цель и задачи исследования.

Наша рабочая гипотеза состоит в том, что структуризация лексической сочетаемости на основе аппарата ЛФ позволяет вывести процесс овладения языковыми навыками на качественно новый уровень. Этим определяется **основная цель** нашей диссертации: исследование методов работы с одним из важнейших компонентов языковых знаний - лексическими функциями, разработка метода обучения сложным аспектам лексики на основе ЛФ и проверка эффективности этого метода. Мы исходим из того, что наиболее эффективным средством обучения сложным и слабо

⁴ И.А. Мельчук, А.К. Жолковский. Толково-комбинаторный словарь современного русского языка. Wiener Slawistischer Almanach, Sonderband 14, Вена, 1984.

⁵ Mel'čuk Igor, Arbachevsky-Jumarie Nadia, Elnitsky Léo, Iordanskaja Lidija, Lessard Adèle. Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexic-sémantiques I-IV. Montréal, 1984-1999.

⁶ J. Crowther, S. Dignen, D. Lea (eds.). Oxford Collocations Dictionary for Students of English. Oxford: Oxford University Press, 2002.

⁷ M. Benson, E. Benson, R. Ilson. The BBI Combinatory Dictionary of English: a Guide to Word Combinations. Amsterdam/Philadelphia, 1986.

⁸ J. Hill, M. Lewis (eds.). LTP Dictionary of Selected Collocations. Hove, UK: Language Teaching Publications, 1997.

⁹ Ignacio Bosque. Redes : Diccionario Combinatorio del Español Contemporáneo, Madrid, Ediciones SM, 2004.

¹⁰ S. Verlinde, Th. Selva, J. Binon. ALFALEX : un environnement d'apprentissage du vocabulaire français en ligne, interactif et automatisé. Romanesque 1, 2003, 42-62.

¹¹ Alain Polguère. Towards a theoretically-motivated general public dictionary of semantic derivations and collocations for French. Proceedings of the Ninth EURALEX International Congress. Volume II. Stuttgart, 2000, 517-527, Igor Mel'čuk, Alain Polguère. Dérivations sémantiques et collocations dans le DiCo/LAF. Langue française. Special issue on collocations "Collocations, corpus, dictionnaires". P. Blumenthal & F. J. Hausmann (eds), 2006, N 150, 66-83.

¹² Margarita Alonso Ramos. Construction d'une base de données des collocations bilingue français-espagnol. Langages 143, 2001, 5-27, Margarita Alonso Ramos. Elaboración del Diccionario de colocaciones en español y sus aplicaciones. P. Bataner and J. de Cesaris (eds.). De Lexicographia. Actes del I Symposium internacional de Lexicografía. Barcelona. IULA-Edicions Petició, 2004, 149-162.

структурированным аспектам лексической системы является построение компьютерной системы, позволяющей организовать этот процесс в игровой форме.

Опыты построения подобных обучающих систем уже существуют. Первоначально идея использовать аппарат ЛФ в целях обучения естественным языкам была сформулирована в начале 90-х годов 20 века Ю. Д. Апресяном и Л. Л. Цинманом - появилась идея компьютерных лингвистических игр, была разработана первая версия лингвистического обеспечения и программных модулей, реализующих обучающие алгоритмы. Работа над лингвистическими основами и концепцией учебника была продолжена в середине 90-х годов в рамках двух проектов программы INTAS трех групп исследователей: группой лингвистов и математиков лаборатории компьютерной лингвистики ИППИ РАН под руководством Ю. Д. Апресяна (словари слов русского и английского языков, словари ЛФ русского и английского языков, идеология лингвистических игр), группой лингвистов Гейдельбергского университета под руководством П. Хельвига (словарь немецких слов объемом в 500 единиц) и группой лингвистов Клагенфуртского университета под руководством Т. Ройтера (словарь немецких слов объемом в 500 единиц и словарь немецких ЛФ).

Таким образом, для достижения основной цели исследования необходимо решить следующие задачи:

1. Разработать схему данных для представления знаний о моделируемых аспектах владения языком средствами реляционных баз данных. Решение должно быть независимо от естественного языка и допускать работу с данными любого языка с любой системой письменности.
2. Разработать пакет прикладных программ для системы обучения сложным аспектам лексики, учесть развитие всех основных идей CALLEX с момента появления этой системы до сегодняшнего дня и реализовать их на современном технологическом уровне. В нашей системе речь будет идти в первую очередь об обучении несвободной лексической сочетаемости, описываемой аппаратом лексических функций. Кроме того, создаваемая система должна помочь пользователю овладеть значениями слов данного языка и их переводными эквивалентами в другом языке, родном или иностранном. Таким образом, она будет представлять собой разновидность компьютерного учебника лексики.
3. Изучить возможность обучения с помощью разработанной системы разных категорий пользователей и проверить эффективность такого обучения.
4. Разработать универсальный инструмент для наполнения словаря и коррекции лингвистических данных, который был бы применим для всех добавляемых в словарь естественных языков.

Методы исследования. Анализ синтагматических и парадигматических связей лексических единиц, метод лексикографического портретирования. Методы теорий объектно-ориентированного программирования и реляционных баз данных.

Материал исследования. Работа проводилась на материале словарей русского и английского языков, составленных Ю. Д. Апресяном специально для построения описываемой системы. Словарь разработанной системы состоит из двух пересекающихся логических частей – Словаря ЛФ и Словаря слов. Таким образом, словарь системы состоит из четырех Словарей. Объем русского Словаря слов равен примерно 3000 лексических единиц, английского – 1500 лексем. Оба Словаря ЛФ содержат описания 116 ЛФ каждый. Дополнительно были использованы материалы испанского языка, разрабатываемые для данной системы совместно с Марией Ауксиладорой Барриос Родригес (Мадридский университет Complutense).

Научная новизна.

Исследована возможность применения аппарата ЛФ как средства представления знаний о несвободной лексической сочетаемости для задач обучения владению языком. При этом одни и те же словари ЛФ используются как для целей изучения языка (в

составе обучающей системы), так и для целей автоматической обработки текстов (в составе системы ЭТАП).

В процессе выполнения работы были получены следующие новые **научные результаты**:

- Разработана схема данных для представления знаний о ЛФ, соответствующая структуре словаря. Осуществлен перенос содержания лингвистических словарей в формат реляционной базы данных.
- На материале словарей русского, английского и испанского языков построена многоязычная система обучения лексике, реализующая набор из 5 лингвистических игр.
- Установлена эффективность разработанной обучающей системы.
- Разработана система для наполнения и редактирования словаря.

Теоретическая и практическая значимость диссертации.

Теоретическая значимость выполненной нами работы определяется тем, что в ходе использования системы была доказана эффективность изучения лексической системы языка и, в частности, его несвободной сочетаемости - исключительно важного аспекта языковой компетенции - с помощью аппарата лексических функций, на основе игровых методов. Было показано, что компьютерная система, реализующая комплекс лингвистических игр, способна заполнить существенный пробел в современных методиках обучения языку.

Разработанный в рамках исследования пакет прикладных программ предназначен для пользователей, совершенствующих навыки активного владения иностранным или родным языком. Работать с обучающей системой можно самостоятельно или под руководством преподавателя. С помощью этого продукта возможно обучение переводным эквивалентам слов, их толкованиям, а также лексическим функциям в нескольких различных режимах. Наши эксперименты показывают, что лексический материал может быть постепенно освоен обучаемым при работе с отдельными играми, при этом параллельно происходит овладение самими ЛФ. В то же время наиболее эффективной методикой является комбинация всех реализованных игр при работе с фиксированным списком словарных статей.

Инструментарий системы предоставляет возможность редактирования словаря, добавления в него словарных статей, в том числе для новых языков, и установления связей между словарными статьями различных языков, а также настройки материала для использования в тех или иных лингвистических играх.

Сохранение результатов работы пользователей с системой позволяет анализировать как процесс обучения конкретного игрока, так и содержание словаря на предмет его пополнения за счет накопленных ответов.

Разработана динамическая библиотека, которая содержит реализацию структур данных и оперирующих с ними методов, на основе которых возможно добавление в обучающую систему новых методик, а также дальнейшее развитие инструментария системы по вводу, коррекции и анализу лингвистических данных. Также в пакет прикладных программ вошла вспомогательная библиотека, с помощью которой на основе уже разработанных игр можно создавать приложения для специализированных образовательных задач.

Опубликованные работы. По теме диссертации опубликовано 3 работы.

Основные положения диссертации, выносимые на защиту:

1. Схема данных для представления лингвистических знаний.
2. Программная реализация набора лингвистических игр, предназначенных для обучения трудным аспектам лексической системы языка.
3. Универсальная система коррекции словаря системы, применяемая для подавляющего большинства естественных языков.

4. Динамическая библиотека, организующая работу прикладных программ со словарем системы.

Апробация результатов диссертации.

Основные положения и результаты исследования докладывались на постерной сессии конференции IV International Conference on Multimedia and Information and Communication Technologies in Education (m-ICTE) (Испания, 2006), а также на заседании научного семинара в ВИНТИ РАН в январе 2007 года.

Экспериментальное тестирование системы было проведено в октябре-декабре 2006 года в Институте лингвистики РГГУ и в феврале-марте 2007 года на Факультете славянских филологий Софийского университета им. Св. Климента Охридского.

В ходе тестирования была проверена пригодность реализованного набора игр для задач обучения несвободной сочетаемости слов, а также работоспособность системы в реальных учебных условиях. По результатам экспериментов была усовершенствована методика расчета оценок результатов работы. Обработка экспериментальных данных позволяет уверенно говорить о прогрессе в знаниях пользователей системы, а также сформулировать ряд идей по усовершенствованию программы.

Структура диссертационной работы. Диссертация состоит из введения, четырех глав, заключения, списка литературы и Приложения. Она содержит 138 страниц основного текста, 22 рисунка и 8 таблиц, расположенных в тексте диссертации. Список литературы включает 92 наименования.

Основное содержание работы

Во **введении** обосновывается актуальность темы диссертации, ее научная новизна, практическая ценность.

В **первой главе** мы привели обзор существующих на настоящий момент компьютерных систем обучения несвободной лексической сочетаемости, проанализировали их преимущества и недостатки. При этом ключевое место в обзоре заняли коллокации.

Термин *коллокация* относится к характерной, регулярно встречающейся полудиоматичной связи L1 + L2, в которой одна часть – *коллокат* - выбирается для выражения заданного значения и играет особую синтаксическую роль, определяемую второй частью – *основой* коллокации.

Важность изучения несвободной лексической сочетаемости в терминах коллокаций пользуется широким признанием. Подтверждением значимости коллокаций может служить, к примеру, тот факт, что им уделяется значительное внимание в материалах Кембриджского экзамена по английскому языку (CAE - Certificate in Advanced English). Таким образом, задача изучения коллокаций является актуальной проблемой. Мы обратили внимание, что на сегодняшний день учащиеся преимущественно работают с классическими словарями на бумажных носителях (в том числе с уже упомянутыми выше Oxford Collocations Dictionary, BBI, LTP и REDES) или с их электронными версиями, что, безусловно, затрудняет процесс обучения.

На сегодняшний день существует ряд статических тестов по коллокациям, которые представляют их в более удобной форме и используют Интернет – например, набор тестов Пирсона Брауна. Несмотря на то, что эти упражнения носят статический характер, благодаря их большому количеству возможно освоить охваченный ими словарь. Преимущество таких тестов и в том, что пользователь проверяет свои знания онлайн.

Более совершенная методика реализована в Католическом университете г. Лёвен, Бельгия, где разработан набор упражнений ALFALEX, направленных на изучение французского языка и снабженных справочной системой, который опирается на лексическую базу DAFLES. Один из разделов ALFALEX ориентирован на изучение коллокаций и построен на использовании так называемых глаголов поддержки (*verbes*

support). Это понятие близко к понятию ЛФ, но гораздо более грубое. Глаголы поддержки охватывают одновременно несколько типов ЛФ семейства OPER-FUNC, не делая между ними никакого различия

Упражнения ALFALEX доступны в сети Интернет в виде динамических Web-страниц и позволяют сразу узнать оценку результата работы учащегося. Все упражнения относятся к одному типу – в них пользователю необходимо вставить пропущенное слово в предложение или фразу. Вопросы подбираются по базе DAFLES динамически, это является существенным преимуществом ALFALEX перед статическими тестами Брауна. Кроме того, следует отметить еще два достижения ALFALEX: реализованную обратную связь пользователя с авторами системы и возможность настраивать упражнения, формируя персональный словарь.

Другая работа, также относящаяся к французскому языку, но ориентированная непосредственно на описание и изучение семантических дериватов и коллокаций, проводится в лаборатории OLST Монреальского университета в Канаде И. А. Мельчуком и А. Польгером. Теоретический и описательный подход, который лежит в основе этой работы – толково-комбинаторная лексикология, являющаяся лексикологической ветвью теории «Смысл - Текст». В рамках работы создана формальная лексическая база данных семантических дериватов и коллокаций Dico (сокращение от французского *dictionnaire combinatoire*, комбинаторный словарь) и словарь LAF (сокращение от французского *Lexique actif du français*, активная французская лексика), который целиком составлен по материалам статей Dico на основе использования разъяснительных описаний. Одной из дополнительных особенностей данной работы является применение при описании лексических единиц иерархии семантических меток для цели их семантической классификации.

Основной целью создания базы Dico является описание комбинаторного потенциала французского языка с помощью механизма ЛФ. Dico изначально ориентирован на компьютерную обработку, не зависит от конкретной платформы и может быть преобразован в словарь для автоматической обработки текста (АОТ).

Отличительной чертой LAF является то, что формальный толково-комбинаторный подход в нем расширен для того, чтобы описание было понятно нелингвистам – в основном, учителям иностранных языков и студентам. Авторы LAF приняли решение использовать для описания лексико-функциональных связей так называемые *formules de vulgarisation* «разъясняющие формулы». Основная идея, стоящая за использованием таких формул – рассматривать все стандартные лексико-функциональные связи как нестандартные, и работать с ними с помощью таких же определяющих выражений, которые используются в ТКС для нестандартных связей. Каких-либо упражнений по словарю LAF на сегодняшний день нет.

В Университете Ла Корунья (Испания) М. Алонсо Рамос разрабатывает базу лексических сочетаний и коллокаций DiCE (сокращение от *Diccionario de colocaciones del español* – словарь коллокаций испанского языка) для испанского языка, а также связанный с DiCE интерактивный обучающий модуль.

База DiCE содержит коллокации, описывающие человеческие эмоции. Коллокации формализованы на основе теории ЛФ, с помощью которых описывается связь между основой коллокации и коллокатом. М. Алонсо Рамос подчеркивает, что ЛФ удовлетворяют следующим трем требованиям, необходимым для описания коллокаций: 1) они представляют значение коллокации; 2) описывают синтаксическую и актантную структуру коллокации; и 3) обозначают функциональную зависимость коллоката по отношению к основе коллокации. При этом сами обозначения ЛФ в DiCE дополняются их металингвистическими описаниями на естественном языке, так называемыми *гlossами*. Цель введения глосс та же, что и цель появления разъясняющих формул в Dico и LAF – сделать словарь в полной мере доступным для пользователей, не знакомых с толково-комбинаторной лексикологией (ТКЛ).

В своих работах М. Алонсо Рамос приходит к выводу, что изучение языка с использованием компьютера (CALL - Computer Assisted Language Learning) по аналогии с применением компьютера в лексикографии и в области перевода также требует разработки специальных инструментов с интегрированным интерфейсом, который в этом случае будет обеспечивать онлайн-доступ к словарю, корпусу и обучающему модулю. Пользователь именно такой системы приобретает автономность, а ведь именно изучение словаря относится к одной из тех немногих областей знания языка, которая может эффективно осваиваться вне учебных аудиторий.

В отличие от лексических ресурсов, являющихся электронной версией бумажных словарей, DiCE изначально задумывался как электронная лексическая база данных. Это позволило М. Алонсо Рамос хранить в DiCE более сложные типы информации и обеспечить гибкий доступ к данным, а также стало возможным создать ассоциативный обучающий модуль, связанный со словарем.

На основе выполненного в первой главе обзора мы сформулировали требования к системе обучения коллокациям: подбор материала при составлении словаря должен производиться на основе современных лингвистических теорий, объем накопленного материала имеет критическое значение, обучающий процесс должен поддерживаться компьютерной системой, предоставляющей максимально эффективную форму освоения материала и удобный вспомогательный инструментарий.

Во **второй главе** мы привели подробное описание разработанной системы, которая обеспечивает контроль и развитие у пользователя навыков владения выбранным естественным языком на основе хранящейся в системе лингвистической информации. Во второй главе рассказано о лингвистическом обеспечении системы и о приложении, используемом для коррекции и пополнения словаря. Также мы описали механизм игр, а именно все основные моменты, с которыми сталкивается пользователь при обучении, и подробно рассмотрели сами игры.

Понятие лексической функции впервые было введено И. А. Мельчуком и А. К. Жолковским в рамках лингвистической теории «Смысл $\leftrightarrow$ Текст». Впоследствии теория лексических функций была существенно дополнена Ю. Д. Апресяном. Во второй главе исследования приводится ряд основополагающих утверждений и понятий, сформулированных ими в различных работах.

Под лексической функцией (ЛФ) в модели «Смысл $\leftrightarrow$ Текст» понимается достаточно общее значение или семантическое отношение, которое может выражаться многими различными словами, причем выбор конкретного слова зависит как от его собственного лексического значения, так и от того, при каком ключевом слове (аргументе функции) выражается данное общее значение или отношение.

ЛФ связывает тройку зависящих друг от друга объектов: самое функцию, ее аргумент и ее значение. ЛФ в общем случае являются многозначными, т. е. могут иметь более одного значения для данного аргумента. Ни одно из значений ЛФ не может быть стопроцентно предсказано по значению аргумента, т. е. не является стопроцентно семантически мотивированным.

С точки зрения формальной математики ЛФ являются бинарными отношениями, заданными на множестве слов и словосочетаний при помощи перечисления пар «аргумент-значение». Эти отношения не обладают свойствами рефлексивности, транзитивности, симметричности или антисимметричности.

ЛФ в теории «Смысл $\leftrightarrow$ Текст» предназначены для описания двух важнейших аспектов феномена владения языком - несвободной сочетаемости слов и системы перифразирования, то есть правил построения различных высказываний, имеющих один и тот же смысл. Данная работа посвящена обучению несвободной сочетаемости слов, и о ней будет рассказано ниже. Перифразирование в описываемом компьютерном учебнике лексики охвачено в незначительной степени и сводится к понятию синонимических замен, то есть к изучению функции SYN и развитию способности

указать синонимичные значения ЛФ для данного аргумента. Таким образом, усваивая ЛФ данного языка, человек сразу приобретает два важных умения: умение правильно комбинировать слова и умение перифразировать высказывания.

Приведем несколько примеров перифраз, построенных на компьютере¹³: Монополии контролируют цены – Контроль цен осуществляется монополиями, Монополии осуществляют контроль цен, Цены находятся под контролем монополий, Цены держат монополии под контролем, Цены контролируются монополиями и т. д.

Как уже было сказано выше, словарь разработанной системы состоит из Словаря ЛФ и Словаря слов. При этом к Словарю ЛФ относятся имя функции, определение функции, ее примеры, а также массивы нетривиальных значений. К Словарю слов относятся лексикографическое определение лексемы, примеры ее использования, пояснение для различных лексем одного слова, частеречная характеристика, переводы лексемы на другие языки, семантические дескрипторы лексемы, ее модель управления и массив значений ЛФ для данной лексемы.

В определениях ЛФ используются переменные: X – ключевая лексема, или аргумент функции; P1 – первый участник ситуации (действия, деятельности, процесса, состояния, свойства, отношения и т. п.), обозначаемой данной ключевой лексемой; P2 – второй участник ситуации; P0 – неучастник ситуации, обычно тот, кто ее каузирует или ликвидирует.

В настоящий момент словарь обучающей системы содержит упоминания 116 лексических функций. Их полный список, извлеченный из Словаря ЛФ, содержится в Приложении к диссертационной работе. При этом, согласно гипотезе И. А. Мельчука, примерно 55 элементарных функций составляют универсальный набор для всех человеческих языков, а остальные функции, по-прежнему универсальные, являются сложными, и образуются на основе элементарных. Элементарные функции и образованные на их основе сложные (неэлементарные) функции относятся к описанным выше стандартным ЛФ. Кроме того, существует значительное число нестандартных ЛФ, которые не обладают свойством универсальности. В описываемом учебнике лексики содержатся только стандартные функции.

Для целей лингвистических игр ЛФ были разделены в учебнике на три уровня в зависимости от сложности их усвоения.

Если ЛФ имеет простой смысл и простые синтаксические свойства, она относится к первому уровню сложности. Если ЛФ имеет простой смысл, но нетривиальные синтаксические свойства, она относится ко второму уровню сложности. Если ЛФ имеет и нетривиальный смысл, и нетривиальные синтаксические свойства, она относится к третьему уровню сложности.

В одном из разделов второй главы мы описали структуру словарных статей Словаря слов и Словаря ЛФ. Приведем несколько примеров лексических функций, взятых из рабочего словаря разработанной системы.

Функция MAGN: «Прилагательное, наречие, сочетание предлога с существительным или сравнительный оборот с союзом, обозначающие большую степень или интенсивность X-а и выполняющие при X-е функцию синтаксического определения или обстоятельства».

Элементарная адъективная ЛФ-коллокат. Уровень сложности: 1.

MAGN (БРЮНЕТ): жгучий

MAGN (ТЬМА): крошечная

MAGN (ВЛАСТЬ): прочная / твердая / неограниченная / устойчивая

¹³ Ю. Д. Апресян, Л. Л. Цинман. Формальная модель перифразирования предложений для систем переработки текстов на естественных языках. Русский язык в научном освещении, 2002, № 2 (4), сс. 102-146.

MAGN (ВЫБОР): тщательный / продуманный

Функция ANTI: «Антоним – лексема той же части речи, что X, со значением, противоположным значению X-а».

Элементарная ЛФ-субститут. Уровень сложности: 1.

ANTI (ВЛИТЬ): вылить

ANTI (ПЕРВЫЙ): последний

ANTI (ТЕМНЫЙ): светлый

ANTI (ПРЕДОК): потомок

ANTIMAGN – неэлементарная функция, являющаяся суперпозицией функций MAGN и ANTI: «Прилагательное, наречие или сочетание предлога с существительным, обозначающее небольшую степень X-а и выполняющее при X-е функцию синтаксического определения или обстоятельства».

Неэлементарная адъективная ЛФ-коллокат. Уровень сложности: 1.

ANTIMAGN (АПЛОДИСМЕНТЫ): жидкие

ANTIMAGN (ДОВОДЫ): слабые

ANTIMAGN (ДОЛГ (ДЕНЕЖНЫЙ)): пустяковый / мелкий / небольшой

ANTIMAGN (МИССИЯ): легкая / несложная / простая

Функция CAUSOPER1: «Делать так, что P1 делает, испытывает или имеет X (лексически обусловленный глагол, при котором P0 или P2 выполняет функцию подлежащего, P1 - функцию главного дополнения, а X – функцию второстепенного дополнения)».

Неэлементарная глагольная ЛФ-коллокат. Уровень сложности: 2.

CAUSOPER1(БОРЬБА): поднимать <поднять> (кого-либо) на (борьбу) / втягивать <втянуть> (кого-либо) в (борьбу) / вовлекать <вовлечь> (кого-либо) в (борьбу)

CAUSOPER1(ПОЛНОМОЧИЯ): облекать <облечь> (кого-либо) (полномочиями)

CAUSOPER1(СМУЩЕНИЕ): приводить <привести> (кого-либо) в (смущение) / повергать <повергнуть> (кого-либо) в (смущение)

Дальнейшее развитие Словаря лексических функций – перспективная работа, в которой возможно задействовать множество языков, поскольку как указано выше, лежащая в основе системы модель универсальна и не зависит от конкретного языка. Практическая возможность такого развития проверена при добавлении в словарь испанского материала, который требует обработки особых символов, а также изменения правил лексикографического упорядочивания.

В одном из разделов второй главы мы описали процедуру коррекции содержимого статей. В нашей системе составитель Словарей получает доступ к словарным статьям в текстовом виде. Данный подход является традиционным для лингвистов, работающих с базой знаний лингвистического процессора ЭТАП. Представление материала в виде простого текста позволяет свободно обозревать несколько статей одновременно и использовать для внесения исправлений стандартные для любого текстового редактора функции. Также этот подход позволяет осуществлять экспорт и импорт данных в текстовом виде. По окончании редактирования текста проводится его разбор на словарные зоны с одновременной проверкой формата записи. Процедуру разбора текста словарной статьи по аналогии с процедурами, принятыми в системе ЭТАП, мы назвали компиляцией словарной статьи.

В настоящий момент система содержит 5 лингвистических игр:

- Лексические функции
- Толкование
- Перевод
- Слово
- Слово с подсказкой

Игры разбиты на отдельные шаги, каждый из которых соответствует одному вопросу системы пользователю. Количество таких шагов в каждом конкретном случае определяется либо самим пользователем, либо объемом подготовленного лексического материала в словаре.

Все вопросы с участием ЛФ (в играх «ЛФ», «Слово», «Слово с подсказкой») делятся на три уровня сложности в зависимости от характера этих ЛФ. Вопросы игры «Перевод» отнесены к первому уровню сложности, игры «Толкование» - ко второму.

В играх реализована система подсчета баллов, начисляемых игроку за правильные ответы с учетом сложности выполненных им заданий и общего количества правильных ответов на каждый вопрос в словаре. В случае существования более одного варианта правильного ответа система предлагает обучаемому возможность дать альтернативный ответ, что подразумевает знание синонимических средств языка. С помощью альтернативных ответов пользователь игры может набирать дополнительные баллы – чем больше внешне различных способов выразить свою мысль он может предложить, тем выше его уровень знания конкретного языка. Несмотря на то, что в отдельных случаях количество альтернатив может быть достаточно велико - в настоящий момент максимальное число альтернатив равно 28 (Игра «слово», лексема «делать2», ЛФ НУРО: ваять / валять / вить / выковывать / выкроить / выпиливать / вырезать / высекать / вытесывать / выткать / вязать / гнать / ковать / кроить / лепить / лить / метать / набирать / отливать / пилить / плести / прясть / резать / рубить / сбивать / ткать / чеканить / шить), было принято решение, что пользователю достаточно дать не более 3 альтернативных ответов на один вопрос. На основании полученных ответов система формирует три оценки – текущую, нормальную и максимальную. Текущей называется оценка, равная сумме баллов, начисленных за все правильные ответы участника. Нормальной (базовой) называется оценка, равная сумме баллов за один правильный ответ на каждый из предложенных вопросов. Максимальной называется оценка, равная сумме баллов в случае получения трех правильных ответов на каждый вопрос, допускающий от трех и более правильных ответов. Все три оценки вместе со вспомогательной информацией сохраняются системой и в дальнейшем могут быть использованы как для анализа работы конкретного пользователя, так и для анализа эффективности игры.

Если игрок затрудняется с ответом, он может запросить список правильных вариантов с потерей баллов, которые он мог бы заработать за правильный ответ. Также игроку предлагается 3 попытки ответа на каждый вопрос, которые не снижают количества баллов за правильный ответ – на случай неточности или опечатки.

В ряде случаев система может облегчить участнику поиск правильного ответа: для этого разработан инструмент подсказок. С его помощью участнику показываются обязательные (но не критические) элементы ответа и примерная синтаксическая схема с обозначением мест вставки недостающих элементов. При этом в процессе подготовки подсказки обязательные элементы и аргумент функции с помощью внутренних правил переводятся в их словарный вид, чтобы падежная или иная форма слова явным образом не определяла недостающие предлоги.

Пример:
Функция: CAUSOPER1
Аргумент: ДЕЙСТВИЕ
Подсказка: «- (кто-либо) - (действие)»

CAUSOPER1(ДЕЙСТВИЕ): «побуждать <побудить> (кого-либо) к [действию]».

При этом от игрока ожидается один из двух вариантов ответа, которые соответствуют одному смыслу, «побудить кого-либо к действию» или «побуждать кого-либо к действию», и различаются глагольным видом – совершенным или несовершенным.

Для четырех игр - «Слово», «Слово с подсказкой», «Толкование» и «Перевод» - предусмотрена возможность подбора участником языкового материала, который будет в них использоваться – своего рода настройки игр. Материал выбирается пользователем по ряду признаков, которым должны соответствовать лексемы, участвующие в игре. Это может быть часть речи лексемы, приписанные ей дескрипторы или ЛФ. Для игры «ЛФ» участнику предлагается выбрать общее количество аргументов. Поиск материала производится компьютером по всем статьям словаря, которые подготовлены для данной игры и содержат соответствующую метку. Набор таких меток для каждой статьи индивидуален, поэтому составители словаря, не меняя содержимого словарных статей, могут варьировать материал той или иной игры.

Далее мы рассмотрели содержание каждой игры.

«Лексические функции»

Игра «лексические функции» полностью основана на материале упомянутого выше Словаря ЛФ. Игра состоит в следующем. Пользователь выбирает функцию, с которой он будет работать, после чего система сообщает ему определение этой функции и предлагает несколько примеров с разными аргументами. После этого ему предлагается набор аргументов, подобранных случайным образом. Предполагается, что пользователь сможет найти одно или несколько значений данной лексической функции для каждого аргумента.

Для выработки правильного ответа игроку потребуется:

- Понять смысл функции, то есть то значение, которое соответствует аргументу, приведенному в вопросе.
- Подобрать именно те слова, которые наиболее точно и идиоматично выражают это значение при ключевом слове.
- Поставить подобранные слова в нужные грамматические формы и составить из них корректные синтаксические конструкции (словосочетания), то есть правильным образом сочетать друг с другом выбранные слова. При этом помимо отыскания падежных форм может потребоваться подбор предлогов.

Развитие этих трех навыков и составляет цель данной игры.

«Толкование»

Игра «Толкование» предлагает пользователю лексикографическое толкование слова и просит указать (угадать) само слово. В толкованиях используются те же переменные, что и в определениях ЛФ (X для обозначения ключевой лексемы или ее референта, P1, P2, ... Pn для обозначения ее семантических актантов).

Пример:

Толкование: «Рассуждение P2 человека P1 или ссылка на факт P2, которая предназначена для того, чтобы убедить собеседника P3 в своей правоте».

Лексема: АРГУМЕНТ.

«Перевод»

Игра «Перевод» предлагает перевести каждую из выбранных лексем с одного из языков системы на другой. При этом пользователь сам может выбрать пару языков и направление перевода. Однако в нынешней версии системы только для пары языков русский-английский материал подготовлен в достаточной степени.

В ходе игры пользователю для каждой из выбранных им лексем приводится ее лексикографическое толкование на выбранном рабочем языке. Игрок должен предложить один или несколько переводов данной лексемы на иностранный язык.

Пример:

Лексема: ОГРАДА

Толкование: «Сооружение в виде стены, со всех сторон отделяющее объект P1 от остального пространства и предназначенное для того, чтобы не допустить чьего-либо проникновения к P1».

Переводы: FENCE / FENCING / RING-FENCE / RAILS / RAILING / ENCLOSURE

«Слово»

Игра «Слово» предполагает работу участника как с отдельными лексемами, так и с группами лексем, которые могут быть подобраны произвольно или на основе набора признаков.

Для каждой лексемы из выбранной группы участнику предлагается ее лексикографическое толкование, которое отображается на экране на протяжении всего времени работы с данной лексемой. Затем система последовательно выводит имена ЛФ, а также их определения и примеры значений. Участнику предлагается указать значение или значения каждой функции для ключевой лексемы.

Внутри игры «Слово» возможен дополнительный режим, в котором участник перечисляет еще и варианты перевода лексемы.

Игра «Слово» имеет достаточно много общих черт с игрой «лексические функции» и использует лингвистический материал аналогичного содержания. Однако, как уже упоминалось, в настоящий момент эти игры опираются на два самостоятельных и независимых словаря – Словарь слов и Словарь Лексических Функций.

«Слово с подсказкой»

Игра «Слово с подсказкой» ориентирована на работу в двуязычном режиме. При этом один язык – тот, изучением которого занимается пользователь, а другой, на котором формулируются подсказки системы, должен быть в достаточной степени ему знаком. Как правило, это родной язык пользователя.

В игре используются пары словарных статей, по одной для каждого языка, между которыми автором в процессе составления словаря установлена связь. Также возможен вариант, когда две статьи связываются системой автоматически на основании внутренних правил на этапе анализа материала.

Как следует из названия, игра «Слово с подсказкой» идеологически очень близка игре «Слово», а по содержанию она полностью аналогична секции «Лексические функции» игры «Слово». Однако при этом поиск ответа на очередной вопрос облегчается тем, что компьютер приводит правильные ответы на него на языке подсказки.

Для успешной эксплуатации и развития системы необходима организация обратной связи пользователей игры с авторами словаря. Очевидно, что составители словарей не могут учесть все возможные значения данной ЛФ от данного аргумента,

все возможные переводы данной лексемы и т. п. – естественный язык слишком богат для этого. В ответах пользователя могут содержаться вполне правильные ответы на предлагаемые ему вопросы, не предусмотренные системой. Для учета и анализа этого материала в системе предусмотрен буфер ответов, в который сохраняются все ответы игроков, не совпадающие со словарными вариантами. При этом игрокам предоставляется возможность отметить такие варианты ответов, которые, по их мнению, заслуживают отдельного внимания составителей.

В **третьей главе** описана внутренняя организация разработанной системы и рассмотрены вопросы ее реализации.

Пакет прикладных программ, включающий программную оболочку учебника лексики, систему редактирования словаря и ряд вспомогательных программ, разработан в среде Visual Studio .Net 2003 на языке C++ с использованием библиотеки классов MFC (Microsoft Foundation Classes). Взаимодействие программ с базой данных реализовано через интерфейс ODBC (Open Database Connectivity), который предоставляет доступ к данным также через MFC-классы.

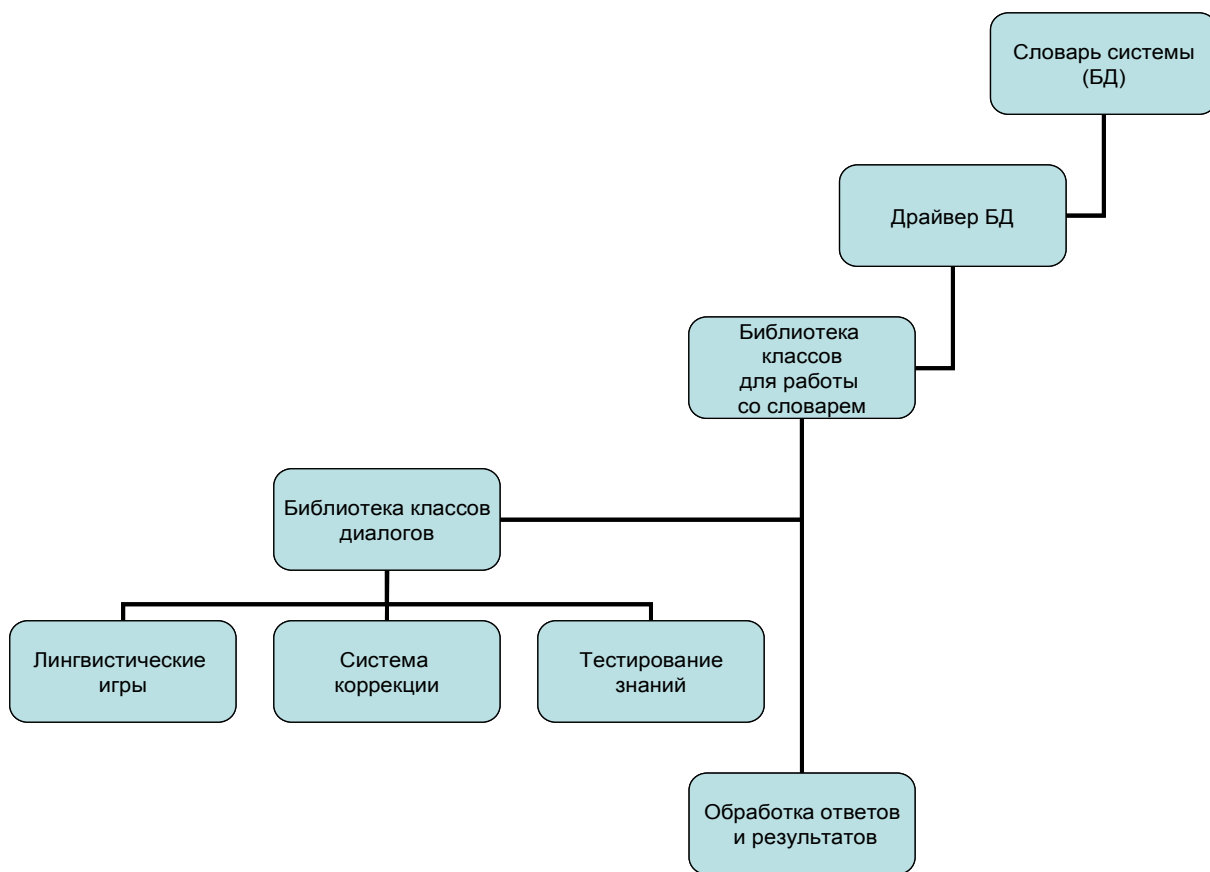


Рис. 1. Схема ППП

При проектировании пакета (см. Рис. 1) был выделен наиболее важный элемент - это словарь, содержащий лингвистическую информацию, которую используют обучающие игры. Поэтому вся система в целом проектировалась таким образом, чтобы наиболее полно реализовать все запланированные операции со словарем и при этом избежать их дублирования. Необходимо было обеспечить удобный доступ к лингвистическим данным нескольких прикладных программ, решающих ряд отдельных, самостоятельных задач, – самой обучающей системы, системы редактирования словаря и программ вспомогательного инструментария. При этом использование данных лексическими играми и редактирование этих данных

требовалось организовать так, чтобы гарантировать целостность и безопасное хранение информации.

В настоящий момент в системе применяется СУБД Microsoft Access, что обусловлено исторически и не является принципиальным выбором, так как использование интерфейса ODBC позволяет с минимальными трудозатратами заменить применяемую сейчас СУБД на любую другую, для которой существует ODBC-драйвер. Преимуществом MS Access является то, что она оснащена удобным графическим интерфейсом для работы с содержимым БД, и ODBC-драйвер для нее находится в режиме свободного распространения, его использование не требует дополнительных затрат и соглашений. Таким образом, разработанный ППП может использоваться без каких-либо ограничений, связанных с лицензированием, практически в любой современной Windows-системе. Обработываемые на сегодняшний день объемы данных не вызывают потребности в наращивании производительности или функциональности СУБД.

Структуры данных, соответствующие словарным статьям и их отдельным зонам, а также методы работы с ними, содержатся во вспомогательной динамической библиотеке `dictionary.dll`, которую для получения и обработки этих данных могут использовать прикладные программы. На практике это означает, что библиотека скрывает от приложений схему данных БД, и в случае изменения структуры словарей коррекции всех элементов системы не потребуются.

В ходе подготовки к экспериментальной эксплуатации системы, подробно описанной в главе 4, в состав ППП была добавлена еще одна библиотека - `interfaces.dll`, в которую были вынесены все классы диалогов, используемые в играх. В свою очередь диалоги обращаются к словарю, используя библиотеку `dictionary.dll`. Таким образом удалось избежать их дублирования помимо самих игр в приложениях для проверки знаний и тренировок.

Для проектирования системы важным моментом являлась возможность внесения плановых изменений в структуры словарных статей. Необходимость этого обусловлена развитием обучающих методик в ходе работы над системой и в процессе накопления и редактирования лингвистических данных. Исходя из такой достаточно нестандартной постановки задачи проектирования, для обеспечения всех описанных свойств конечной системы было принято решение, что оптимальным способом хранения словаря будет реляционная база данных. Такой вариант позволит создавать инструменты для поиска и анализа на основе использования стандартных SQL-запросов, а также позволит рассчитывать, что в будущем любое изменение структуры словаря, не затрагивающего его фундаментальных свойств, возможно будет реализовать без потери данных или функциональности системы в целом.

Для хранения словаря и вспомогательной информации с учетом требований, описанных выше, была разработана схема данных (см. Рис. 2), в которой каждому рабочему словарю – как Словарю слов, так и Словарю лексических функций сопоставлен набор схем отношений. При определении доменов было принято решение отказаться от строгих ограничений для того, чтобы в будущем возможные изменения форматов не вызвали необходимости вносить существенные изменения в схему данных. После этого стал возможен перенос лингвистических данных из текстового вида в формат базы данных.

В отдельном разделе третьей главы мы описали работу двух центральных приложений ППП – программы, содержащей лингвистические игры, и программы системы коррекции. Для этого мы перечислили разработанные для этих приложений структуры данных и параллельно в общих чертах рассказали об алгоритмах, в которых они используются. Как уже было сказано, при выполнении основных операций со словарем игры и система коррекции обращаются к динамической библиотеке `dictionary.dll`. Игры дополнительно используют классы диалогов, вынесенные в

библиотеку interfaces.dll. Обе эти библиотеки входят в состав ППП. Описанные и реализованные в dictionary.dll и interfaces.dll структуры данных и методы работы с ними и составили материал раздела.

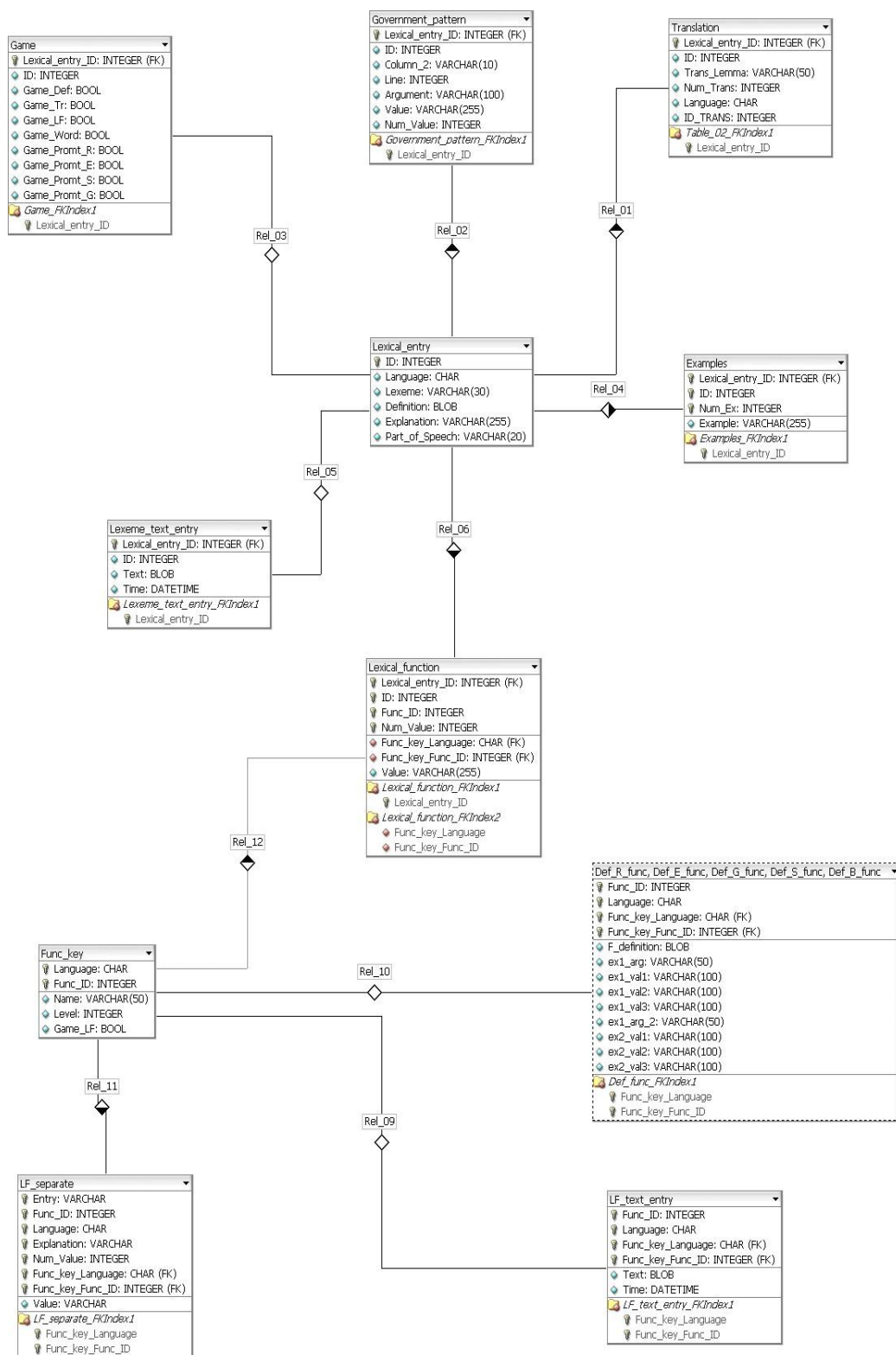


Рис. 2. Схема отношений БД

В **четвертой главе** исследования рассмотрен опыт применения разработанной нами системы. Ее свойства и возможности изучались в рамках экспериментов, представляющих собой работу с ППП нескольких групп пользователей по составленной заранее программе.

Эксперименты проводились на специально отобранном русском лексическом материале и были разделены на несколько серий, различающихся составом участников и характером заданий. По окончании каждой серии результаты экспериментов анализировались, что позволило делать промежуточные выводы и в ряде случаев корректировать методику экспериментов. Лексический материал при этом не менялся для того, чтобы результаты всех серий и всех участников имели идентичную структуру, и их можно было бы сравнивать между собой.

В процессе экспериментов участники работали с обучающей системой, используя ограниченный словарь русского языка и только одну игру – «Лексические функции». Целью экспериментов было выяснить, в какой степени участники владели несвободной лексической сочетаемостью до начала работы с системой, и определить, помогает ли она добиться прогресса в их знаниях. Дополнительной целью было узнать, насколько предлагаемый интерфейс ППП понятен и удобен для обучаемых, и сформулировать список задач по дальнейшему совершенствованию системы.

Всего было проведено три серии экспериментов. Во время первой серии с ППП работали студенты первого и четвертого курса Института лингвистики РГГУ и старшеклассники средней школы, участники лингвистического кружка при этом Институте. Во второй серии группа состояла из студентов четвертого курса Факультета славянских филологий Софийского университета им. Св. Климента Охридского. В третьей серии экспериментов участвовала группа, в которую вошли студенты московских ВУЗов разных возрастов различных гуманитарных и технических специальностей, а также люди среднего возраста с гуманитарным образованием.

Для экспериментов было подготовлено два отдельных приложения, которые представляют собой особые режимы работы с игрой «ЛФ» – режим тренировки и режим тестирования.

Режим тренировки представляет собой обычный вариант игры ЛФ, вынесенной в отдельное приложение и работающей только с материалом экспериментов.

В режиме тестирования программа сохраняет все ответы пользователей – как совпадающие со словарными вариантами, так и отличные от них, что дает возможность при анализе результатов восстановить полный ход работы каждого участника. В процессе тестирования участникам не сообщается о набранных ими баллах, правильные ответы также не показываются. При выполнении теста участники ограничены по времени, в течение которого для каждого аргумента они могут предложить вариант значения ЛФ. Задача теста – определить на некоторый момент времени степень владения участника предлагаемым лексическим материалом, что позволит рассчитать комплексную оценку его знаний, выявить пробелы в понимании и характерные ошибки. Данные начального тестирования дают представление о том, как человек до обучения справляется с несвободной сочетаемостью, описываемой с помощью аппарата ЛФ. Сравнение заключительного результата пользователя с начальным результатом дает возможность оценить изменения в его знаниях.

На этапе тренировки, когда участники экспериментов работают с ЛФ, им доступны правильные варианты ответов. В значительной степени освоение ими материала происходит за счет сравнения собственных вариантов значений ЛФ со словарными и запоминания верных. При этом, очевидно, участник старается понять, каким образом правильное значение ЛФ связано с ее аргументом, то есть он задумывается о смысле ЛФ. Мы выдвинули гипотезу, состоящую в том, что при работе в таком режиме параллельно запоминанию верных ответов, без жесткой привязки к тренировочному материалу, происходит более глубокое усвоение участниками самих

ЛФ. Конкретнее это означает, что можно ожидать, что после обучения повысится владение не только тем материалом, который составлял непосредственный объект тренировки, но и тем материалом, который в тренировках не был задействован. В процессе экспериментов мы проверили эту гипотезу. Для этого в общем материале экспериментов мы выделили контрольные данные - набор аргументов ЛФ и соответствующих им значений, которые встречаются участникам только на этапе тестирования. Предположение, что пользователи, успешно освоившие ту или иную ЛФ, смогут указать правильное значение и для этих аргументов по аналогии с материалом, проработанным на этапе тренировки, подтвердилось.

В результате экспериментов был получен ряд результатов, фиксирующих знания испытуемых на момент начала занятий. Эти данные выявили существенные сложности во владении механизмами несвободной лексической сочетаемости даже у пользователей с хорошей лингвистической подготовкой, являющихся носителями русского языка. Таким образом, логичным является вывод о необходимости реализации и внедрения методик, компенсирующих эти пробелы.

Анализируя результаты тренировок, мы можем с уверенностью констатировать прогресс в знаниях подавляющего большинства участников экспериментов, выполнивших запланированные задания в полном объеме - разработанная система помогает как изучить в игровой форме предлагаемый материал, так и освоить сам механизм ЛФ, что подтверждается данными по контрольным значениям. Также было отмечено, что с задачей освоения материала изучающие русский язык как иностранный справляются практически наравне с носителями русского языка, а освоение механизма ЛФ для первых представляет большую сложность чем для вторых.

В процессе экспериментов удалось выявить и ликвидировать несколько недостатков в интерфейсе системы, связанных с неоднозначным пониманием отдельных надписей. Также удалось усовершенствовать систему подсчета баллов, что позволяет отслеживать новый аспект в работе участника игры и проводить более детальный анализ результатов его работы.

Полученные во время экспериментов варианты значений ЛФ, отличные от словарных, были сохранены. Анализ этого массива позволяет сказать, что эти данные представляют интерес как минимум для двух задач – для выявления типичных ошибок, так как они повторяются у разных пользователей, и для задачи развития словаря, так как известны случаи новых удачных значений ЛФ, предложенных участниками.

В **Заключении** диссертации мы сформулировали основные результаты завершеного исследования и обозначили перспективы его продолжения:

1. Исследованы методы работы с лексическими функциями, метод обучения сложным аспектам лексики на основе ЛФ. Разработана схема данных для представления знаний о моделируемых аспектах владения языком. Осуществлен перенос рабочей версии лингвистических данных в новый формат их хранения.
2. Разработан пакет прикладных программ для системы обучения сложным аспектам лексики в игровой форме с заданными при постановке задачи характеристиками, которая может использоваться и как практический стенд для апробации обучающих методик. На материале словарей русского, английского и испанского языков построена многоязычная система обучения лексике, реализующая набор из 5 лингвистических игр.
3. В ходе выполнения работы был учтен опыт предыдущих реализаций компьютерных пособий, использующих формализм ЛФ для описания несвободной лексической сочетаемости. Также были проанализированы современные разработки в области обучения коллокациям, осуществляемые в различных научных центрах.
4. В состав пакета программ включен универсальный инструментальный для наполнения словаря и коррекции лингвистических данных, который применим для всех добавляемых в словарь естественных языков. Система коррекции протестирована

на материале словаря учебника и передана для эксплуатации автору словаря. Эксплуатация осуществляется в Лаборатории компьютерной лингвистики ИППИ РАН.

5. Проведено экспериментальное тестирование системы, которое подтвердило гипотезу о применимости аппарата ЛФ и разработанного на основе его использования ППП для целей обучения несвободной лексической сочетаемости и его эффективности. В рамках тестирования было создано несколько вспомогательных приложений, использующих те же механизмы, что и основные игры. Анализ результатов позволяет сформулировать требования по дальнейшим усовершенствованиям системы в зависимости от вариантов ее использования.

В частности, одним из возможных дополнений пакета мог бы стать Web-интерфейс, позволивший бы работать с обучающей системой удаленным пользователям, используя стандартный Internet-обозреватель. Эта идея требует разработки многопользовательского режима работы с системой, что потребует изменения механизмов записи в БД.

Также был выявлен ряд вопросов, по которым испытуемые не были согласны с хранящимися в словаре данными и настаивали на своих вариантах ответов. Для таких случаев планируется разработать механизм сохранения комментариев пользователей с целью их дальнейшего анализа составителями наряду с ответами, отличающихся от эталонных. Поскольку среди таких ответов вполне могут оказаться правильные, для режима просмотра этих данных целесообразно предусмотреть возможность включения их в словарь или коррекции соответствующей статьи.

6. Как указывалось выше, в настоящий момент методические вопросы обучения оставлены за кадром разработанного ППП, однако нами проделана подготовительная работа по организации механизма карточек пользователей, сохраняющих информацию о предыдущих сеансах их работы. Такие карточки необходимы для подключения к работе над учебником педагогов, потому что для внедрения обучающих методик, очевидно, требуется фиксировать результаты работы пользователя с различными единицами словаря в разные моменты времени и подбирать материал для игр на основе анализа этих данных. С использованием механизма карточек можно реализовать цикличность изучения материала, поэтапное освоение пользователем словаря и т. д. При этом механизм карточек может быть реализован без существенных изменений архитектуры системы, так как все ключевые для него сущности уже существуют.

Основные положения диссертации отражены в следующих **публикациях** автора:

1. Ю. Д. Апресян, П. В. Дяченко, А. В. Лазурский, Л. Л. Цинман, О компьютерном учебнике лексики русского языка, Русский язык в научном освещении, №2(14), 2007, с. 48-112.
2. Pavel Diachenko, Lexical functions in learning the lexicon. Current developments in Technology-Assisted Education (2006), Vol 1, p. 538-542.
3. Boguslavsky, I., Barrios Rodriguez M., Diachenko, P., CALLEX-ESP: A software system for learning Spanish lexicon and collocations. Current developments in Technology-Assisted Education (2006), Vol 1, p. 22-26.