Ю.Н. Филиппович, А.В. Сиренко

ПРОГРАММНЫЙ КОМПЛЕКС ИССЛЕДОВАНИЙ ПСИХОЛИНГВИСТИЧЕСКОЙ МОДЕЛИ ВЕРБАЛЬНОГО СОЗНАНИЯ НА ОСНОВЕ КОГНИТИВНОГО И АССОЦИАТИВНОГО ЭКСПЕРИМЕНТОВ

Введение

Одним из направлений научной школы «Русская языковая личность» [Ведущая научная школа Ю.Н. Караулова Русская языковая личность] является исследование и разработка компьютерных моделей вербального сознания — информационных технологий и лежащих в их основе формальных методов и алгоритмов, а также реализующих их программных систем и комплексов для моделирования вербального сознания.

Основные результаты работ данного направления носят по отношению к лингвистическим исследованиям школы «прикладной» (инструментально-вспомогательный) характер. Их назначение состоит в поддержке экспериментальной (теоретической и эмпирической) научной деятельности исследователей, а весь их комплекс представляет собой автоматизированную систему научных исследований психолингвистических экспериментов — АСНИ ПМ. Вместе с тем, выполняя инструментальную функцию «самопишущего интеллектуального пера и бумаги XX–XXI вв.», разрабатываемые информационные технологии претендуют и на относительно самостоятельную потенциальную роль действующего макета интеллектуальных компонент систем поиска информации, машинного перевода, автоматизированной обработки текста (text mining), извлечения и контроля знаний и др.

Объектом моделирования в рамках данного направления исследований и разработок является вербальное=языковое сознание (ЯС), которое само является *логико-лингвистической моделью* атрибутивной способности человека, его речемыслительной деятельности, и описывается следующими суждениями:

(а) Языковое сознание, являясь формой мышления человека с использованием языковых единиц (ЯЕ), возникает и развивается как процесс осознавания — перехода от неосознанного (неопределенного, «альтернативного») восприятия предмета реального мира к осознанному (определенному, «безальтернативному»). Предмет осознан — значит вербализован и ему в соответствие поставлена некоторая конкретная языковая единица [Филиппович, 2007: 123].

(б) «Языковое сознание складывается из вербально выраженных знаний о мире, т.е. содержанием языкового сознания является вербализованная часть картины мира. Языковая, или наивно-языковая, картина мира складывается не из слов, не из понятий, имеющих логико-лингвистическую природу, а из единиц когнитивной природы, обладающих различной системообразующей мощностью и вступающих одна с другой в различные иерархически-координативные отношения» [Караулов 2009: 168].

(в) Понятие элементарной единицы знаний — «фигуры знания», «когнемы» — введено и интерпретировано как минимальная когнитивная единица, представляющая собой пятикомпонентное отношение {<Знак> <Формула-смысла> <Способ> <Когнитивная область> <Функция>} [Караулов 2009: 27].

(г) «… языковое сознание представляет собой подвижное, динамическое образование, своего рода когнайзер, манипулирующий элементарными единицами знания (фигурами знания) и функционирующий в активном, смыслопорождающем (т.е. в направлении от знака — к смыслу) и пассивном, знакопорождающем (от смысла — к знаку) режимах» [Караулов 2004: 7].

В приведенных суждениях, представляющих и специфицирующих логико-лингвистическую модель ЯС, центральным понятием является — *когнайзер*. Сформируем «рамочное» (по отношению к исследовательским материалам, на основе которых написана данная статья) определение этого понятия в следующем виде:

Когнайзер — это лингвистическая база знаний, манипулирующая единицами знаний о мире (ЕЗМ) — единицами когнитивной природы (б), построенная на основе языковых единиц (а) и элементарных единиц знания (в) и функционирующая в режимах смысло… и знакопорождения (г).

В [Филиппович, 2006: 319–334; Филиппович, 2007: 123–143; Караулов 2009: 122–145] изложены основы построения *формальной модели когнайзера*. Модельными сущностями при этом являются: языковые картины мира (ассоциативный и когнитивный тезаурусы), полученные в результате проведения ассоциативного и когнитивного экспериментов и представленные в форме баз данных; два процесса вербализации знаний о реальном мире — активный (от слова к знанию) и пассивный (от знания к слову); когнайзер — база знаний, интегрирующая процессы вербализации знаний и базы данных ассоциативного и когнитивного экспериментов в форме процедуры осознавания — принятия решения о выборе вербальных альтернатив представления знаний о реальном мире.

Процедура осознавания (переход от неосознанного к осознанному = от невербализованного к вербализованному) в формальной модели когнайзера сведена к безальтернативным и альтернативным вариантам. Безальтернативные варианты осознавания — это переходы от априори известных ЕЗМ к ЯЕ и обратно. Альтернативные варианты характеризуются вариативностью, неопределенностью (вероятностью) и неточностью этих переходов. Иначе, имеется множество ЕЗМ и ЯЕ, а также процедура перехода между ними, которая состоит в выборе альтернативных единиц, традиционно разделяемая на две составные части — критериальная оценка альтернатив и принятие решения о предпочтении (выбор альтернатив).

Для описания функционирования когнайзера как инструментального средства компьютерного моделирования смысло… и знакопорождения можно предложить три типа формализмов: а) формально-логическое описание ассоциативных отношений вербальных единиц смыслопорождающей (сти-

мульно-реактивной) компоненты языкового сознания — *формально-логическая модель когнайзера*; б) описание знакопорождающей компоненты языкового сознания, как процедуры настройки (обучения) нейросетевой модели ассоциативно-вербальной сети — *нейро-сетевая модель когнайзера*; в) описание генетических алгоритмов смысло… и знакопорождения в ассоциативно-вербальной сети — *эволюционная модель когнайзера*.

В данной статье рассматривается программный комплекс, реализующий формально-логическую модель пассивного режима работы когнайзера. В первых реализациях этой модели внимание уделялось безальтернативному варианту осознавания, который сводился к моделированию перехода от Формулы смысла к Знаку посредством ассоциативно-вербальной сети. Программа «Когнайзер» [Сиренко, 2007: 264–277] реализовала только безальтернативный вариант работы когнайзера, выполняя поиск цепочек в ассоциативной сети между любыми двумя априори известными словами, представленными в таких компонентах когнем как Формула смысла и Знак. Последующие реализации формально-логической модели когнайзера более полно реализуют его логико-лингвистическое описание.

1. Программный комплекс «Лингвокультурный тезаурус русского языка»

Программный комплекс «Лингвокультурный тезаурус русского языка» [ЛКТ], который в настоящее время доступен для пользователей Интернет, предоставляет возможности построения цепочек во фрагменте АВС и просмотра фигур знания в базе когнитивного эксперимента (выбор в поле Навигация — «Ассоциативная сеть» и «Тезаурус» соответственно, рис. 2 и 1).

Просмотр фигур знаний (рис. 1) в базе данных когнитивного эксперимента реализован следующим образом. Для фильтрации по компоненте когнемы необходимо поставить флажок в поле перед соответствующим параметром. Если не указаны параметры, то его результатом будет все множество

фигур знания, имеющихся в базе данных. Значение параметра выбирается из списка (поля — Способ, Функция), отмечается подмножеством на множестве допустимых значений (поле — Область), либо вводится в текстовое поле (поля — Знак, Смысл). Перед выполнением поиска можно задать порядок сортировки результатов (поле — Сортировка). Поиск осуществляется по нажатию кнопки Пуск. Просматривать результаты поиска можно постранично (поле — Страница, кнопка — Перейти).

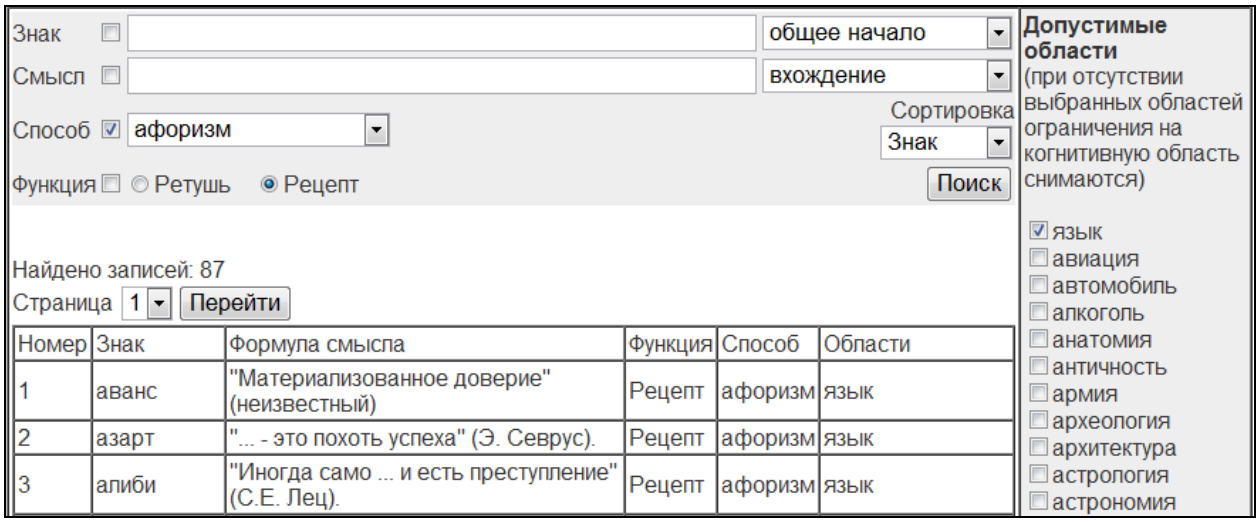

Рисунок 1. Запрос фигур знания

При поиске цепочек (рис. 2) поля Стимул и Реакция могут содержать по одной лексической единице. При моделировании перехода от Формулы смысла к Знаку исследователь должен самостоятельно выполнить запрос для каждой компоненты Формулы смысла, а затем объединить результаты. Возможным является задание длины искомых цепочек — от 1 до 5 или нахождение кратчайших путей в данном диапазоне длин.

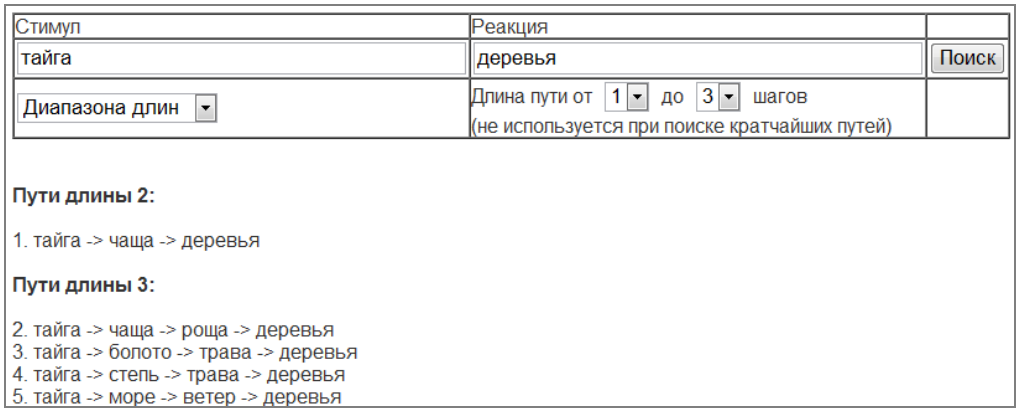

Рисунок 2. Запрос поиска ассоциативных цепочек

Существенными ограничениями при поиске цепочек являются: отсутствие лемматизатора исходных данных и элементов АВС, возможности осуществить синонимичные замены исходных данных и учета омонимии; при поиске цепочек не учитывается и не указывается общее число реакций стимула; возможно выполнение только единичных запросов. К числу ограничений следует также отнести и тот факт, что в программном комплексе используется только фрагмент РАС объемом 122059 стимульно-реактивных пар. Есть и другие ограничения, которые связаны с особенностями разработанных алгоритмов поиска, организационно-техническими возможностями вычислительной среды и сети обмена данными.

2. Программный комплекс «Терминал автоматизированной системы научных исследований психолингвистических экспериментов»

Для преодоления ограничений моделирования безальтернативного варианта работы когнайзера существующего программного комплекса «Лингвокультурный тезаурус русского языка», а также реализации варианта осознавания с альтернативой разработан новый программный комплекс «Терминал автоматизированной системы научных исследований психолингвистических экспериментов» (далее «Терминал», «Терминал АСНИ ПМ»). В нем используется следующее лингвистическое обеспечение: Ассоциативно-вербальная сеть на материале Русского ассоциативного словаря объемом 462530 стимульно-реактивных пар [РАС, 2002]; ~18 300 фигур знания трех этапов когнитивного эксперимента; Электронная версия словаря синонимов Абрамова (~4000 синонимичных рядов) [Онлайн словарь]; Орфографический электронный словарь iSpell (~1300 тыс. словоформ, ~129 тыс. лемм) [Russian ispell]; 1047 экспертных правил-инструкций обработки частичной омонимии АВС.

Новый программный комплекс по-прежнему моделирует пассивный, знакопорождающий режим работы когнайзера, т.е. переход от смысла к знаку, но в двух вариантах — безальтернативном и альтернативном.

В нем АВС представлена вероятностной контекстно-свободной грамматикой, а смена активных знаков при поиске и формировании цепочек принимает форму вывода. Статистические свойства ассоциативных пар изначально формируют вероятностные свойства правил грамматики. Затем грамматика дополняется контекстно-зависимыми правилами на основе фигур знания по принципу: Формула смысла соответствует левой части правила, Знак – правой части.

В программном комплексе возможна лемматизация и синонимичные замены вводимых данных, элементов АВС и фигур знаний; при поиске цепочек учитывается частичная омонимия элементов АВС и частотные характеристики стимульно-реактивных пар; возможной стала обработка множественных запросов. Вместе с тем программный комплекс реализован только как локальная информационная система [Сиренко, 2010: 118–119].

При запуске «Терминала АСНИ ПМ» отображается главная экранная форма (рис. 3). В ней представлены 7 наборов инструментальных средств: Словоформы, Ассоциативная сеть, Анализ данных в БД, Выбор когнем из БД, Создать когнему, Моделирование. Каждое инструментальное средство (технология) состоит из отдельных процедур и операций, доступных для исследователя. Главным инструментальным средством является Моделирование, использование которого и позволяет реализовать пассивный режим работы когнайзера в безальтернативном и альтернативном вариантах.

Лемматизация. Ассоциативно-вербальная сеть содержит многообразные грамматические формы. Узлы АВС могут быть стимульными (S), стимульно-реактивными (SR) и реактивными (R). Заметим, что узлы S и SR чаще представлены словами в основной форме, чем узлы R. Формула смысла фигуры знания также включает лексические единицы в разных грамматических формах. Знак, напротив, всегда представлен словом в основной форме.

Перед началом моделирования в безальтернативном режиме необходимо представить Формулу смысла в виде упорядоченного множества узлов S или SR типов, а Знак – узла R или SR типа. В результате, моделирование перехода от Формулы смысла к Знаку затрудняется отсутствием необходимых лексем среди стимулов и реакций. Лемматизация АВС позволяет более полно использовать лексемы Формулы смысла и Знака.

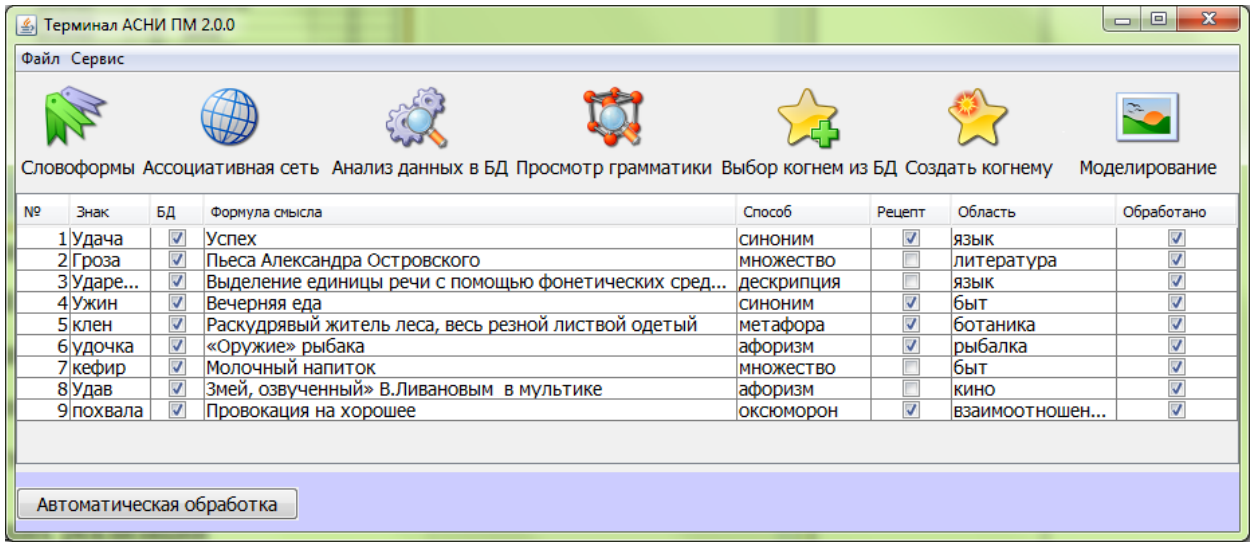

Рисунок 3. Главная форма Терминала АСНИ ПМ

Синонимия. Следующей доработкой программы моделирования является использование в ней словаря синонимичных отношений. При отсутствии Знака среди SR и R узлов, выполняется поиск узлов из ряда его синонимов. Аналогичные суждения применимы к Формуле смысла и стимульным узлам. Предположим, требуется моделирование для Формулы смысла, включающей слово «лицедей». Его среди S и SR узлов АВС на основе РАС нет, но есть синонимичный ему — «актер». Замена позволяет произвести моделирование.

Омонимия. В лингвистической базе данных Терминала узлы ассоциативной сети, имеют текстовое поле «Релятор», формально позволяющее различать полные омонимы. При лемматизации АВС было выделено 1047 частичных омонимов (банк – банка, белок – белка), что делает неопределенными 25880 связей. Неоднозначность частичных омонимов в АВС была устранена посредством диалога с экспертом. Результатом стали программные ин-

струкции для разделения ассоциативных связей по леммам частичных омонимов.

№ ассоциации	Лемма 1: цвет	Лемма 2: цветок
1	базис → цветов	аромат → цветов
2	выбор → цветов	вредитель → цветов
3	радуга → цветов	выставка → цветов
4	семь → цветов	ваза → цветов

Таблица 1. Фрагмент инструкций устранения частичной омонимии узла АВС "цветов".

Частичная омонимия также актуальна и для вводимых запросов пользователя на этапе поиска входов в АВС, но дальнейшее построение ассоциативной цепочки производится без смешения разделенных правил, что можем увидеть на примере лемматизированного фрагмента АВС, обработанного с помощью инструкций.

№ ассоциации	Лемма 1: цвет	Лемма 2: цветок
1	базис → цвет	аромат → цветок
2	выбор → цвет	вредитель → цветок
3	радуга → цвет	выставка → цветок
4	семь → цвет	ваза → цветок

Таблица 2. Пример лемматизированного фрагмента АВС, обработанного с помощью инструкций устранения частичной омонимии.

Полная омонимия не была устранена. Это касается как правил АВС, так и вводимых запросов пользователя. Внутренние структуры программы позволяют добавить к лексеме поясняющий релятор, но это требует разделения омонимичных узлов АВС, связанных с ними правил и соответствующих изменений в грамматике.

Поиск в АВС. В Терминале моделирование режимов работы когнайзера осуществляется с учетом частотных характеристик ассоциативных связей. Прежде всего, это используется для управления поиском и позволяет обрабатывать более устойчивые ассоциации в первую очередь. В результате повысилась эффективность вычислительных процедур.

Множественные запросы. Фигура знания (когнема) является ключевой структурой в программном комплексе, поскольку исходные данные модели-

рования представляются в форме когнем даже для вручную заданных пропозиций. Программный комплекс позволяет обрабатывать последовательно несколько когнем. На главной форме (рис. 3) отображаются когнемы, с которыми в настоящий момент работает пользователь.

3. Терминал АСНИ ПМ: безальтернативный вариант когнайзера

Диалоговое окно настроек моделирования вызывается кнопкой Моделирование главной формы.

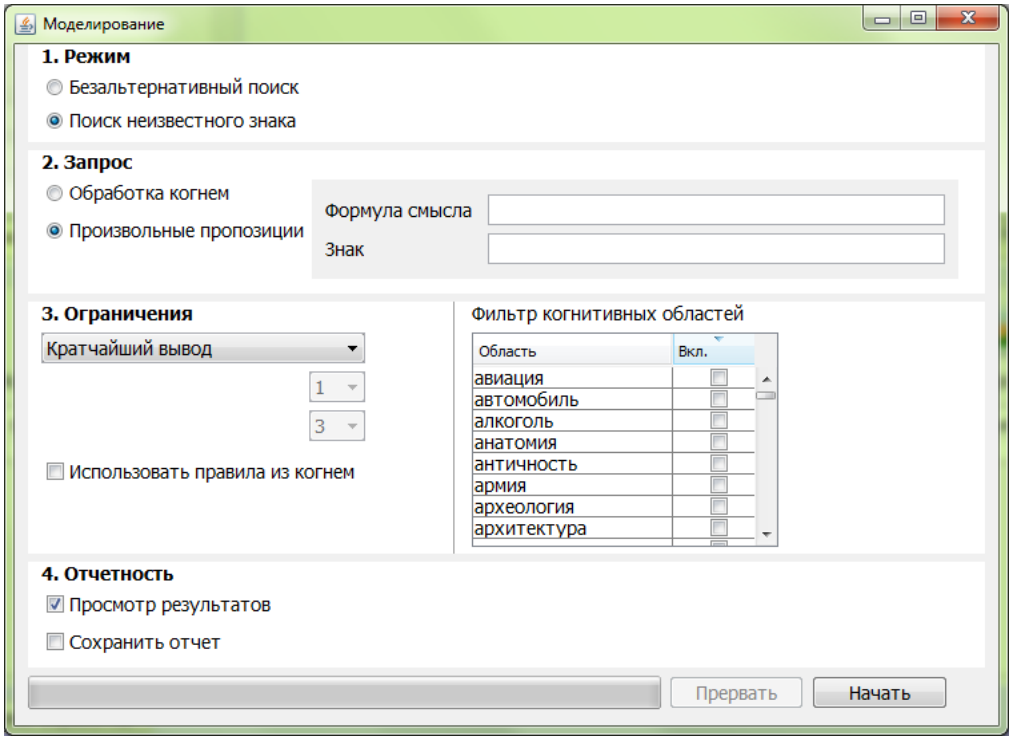

Рисунок 4. Установка параметров моделирования

Диалоговое окно позволяет установить:

- ограничения на длину вывода в грамматике: кратчайший вывод, фиксированное число шагов вывода, диапазон числа шагов вывода;
- источник данных для обработки: Обработка текущих когнем — используется список когнем, представленных в главной форме, Произвольные пропозиции — формула смысла и знак устанавливаются в соответствующих полях окна;
- режим поиска: Безальтернативный поиск или с альтернативой — Поиск неизвестного знака;

•когнитивные области, которыми может быть ограничен поиск в режиме с альтернативой — Область;

•сохранение результатов поиска во внешний файл: Сохранить отчет;

•отображение результатов поиска в программном окне терминала: Просмотр результатов.

В иллюстрации (рис. 5) представлено окно результатов моделирования в безальтернативном режиме для когнемы: {<Марка> <Торговый знак, клеймо на изделиях, товарах> <…> <…> <…>}.

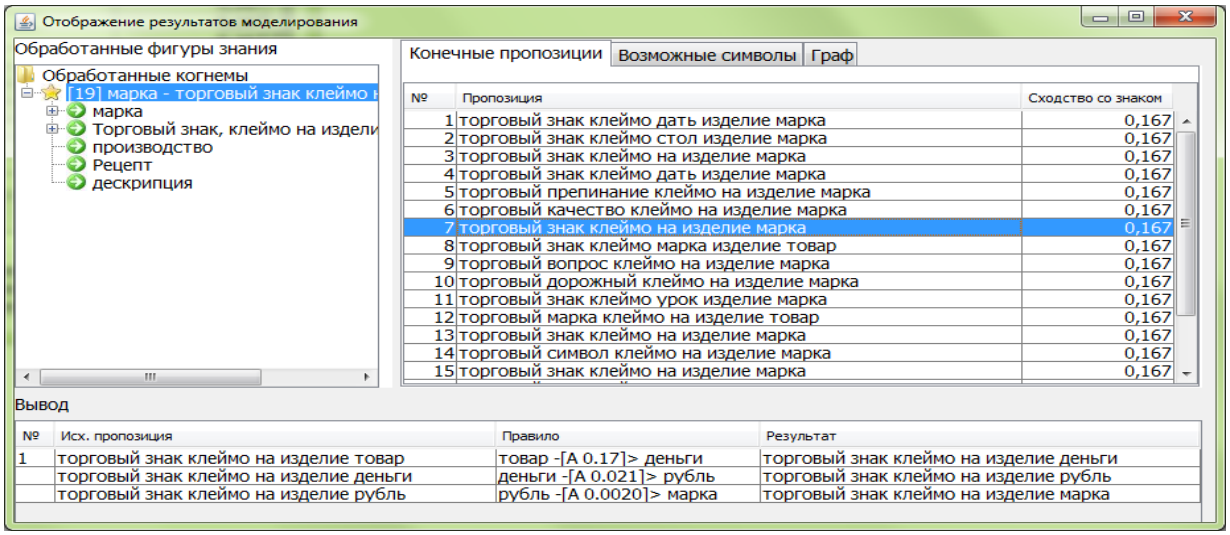

Рисунок 5. Результаты поиска в безальтернативном режиме.

Древовидная структура в левой части окна содержит параметры когнем. Выделение когнемы приводит к отображению результатов моделирования в таблицах справа. В таблице «Конечная пропозиция» перечислены сгенерированные грамматикой из Формулы смысла пропозиции, имеющие ненулевую степень сходства со Знаком, что соответствует построению цепочки от Формулы смысла к Знаку в АВС. Значение 0.167 в столбце «Сходство со знаком» соответствует наличию в пропозиции лексемы «марка». В качестве Знака может выступать естественно-языковая пропозиция, в этом случае «Сходство со знаком» для различных конечных пропозиций может отличаться. Таблица «Вывод» отображает вывод конечной пропозиции в грамматике.

На рис.5 таблица «Вывод» демонстрирует достижение Знака «марка» из лексемы Формулы смысла «товар» при помощи вывода, соответствующего ассоциативной цепочке:

«товар» → «деньги» → «рубль» → «марка».

4. Терминал АСНИ ПМ: альтернативный вариант когнайзера

В Терминале предлагается реализация моделирования осознавания с альтернативой. Конечный знак моделирования в этом режиме неизвестен, результатом будет ранжированный список знаков-кандидатов. Ранг знака должен быть связан с его релевантностью формуле смысла и, опционально, иным параметрам фигуры знания.

Поиск в альтернативном варианте предполагает генерацию пропозиций из формулы смысла, группировку достигнутых в процессе вывода знаков, их фильтрацию и ранжирование. На этапе оценки результатов проверяется наличие действительного знака среди достигнутых моделью и установка соответствующих значений в сводной таблице отчета (рис. 6).

Каждому знаку-кандидату на этапе ранжирования соответствуют две характеристики:

• *достижимость* — суммарная характеристика цепочек вывода, ведущих к знаку; Характеристика принимает значения в диапазоне $(0..+\infty)$. Большему значению достижимости соответствуют более короткие цепочки с меньшей стоимостью применения правил.

• *использование формулы смысла* — часть лексем формулы смысла, цепочки вывода от которых ведут к знаку. Характеристика принимает значения в диапазоне $(0..1]$. Нулевое значение невозможно, поскольку вывод всегда начинает работу с некоторого фрагмента формулы смысла.

Пусть цепочка вывода $Chain = A \to B \to C$. Тогда достижимость $Access(Chain) = invCost(A \to B) \times invCost(B \to C)$, где $invCost$ – инвертированная стоимость правила, определяемая эвристически для каждого типа правил и

нормированная в диапазоне [0..1]. Для ассоциативных правил $invCost$ принимает значение вероятности перехода в АВС от левой части правила к правой части (контекстно-свободные правила). Для правил на основе когнем $invCost$ определяется через редакционные операции, которые надо произвести над предложением в грамматике, чтобы применить рассматриваемое правило. Для всех подстрок предложения рассчитывается редакционное расстояние Дамерау-Левенштейна [Navarro, 2001: 35–37] между подстрокой и левой частью правила. Минимальное редакционное расстояние приведет к применению правила с равной редакционному расстоянию стоимостью. Инвертированной стоимостью будет обратно-пропорциональная редакционному расстоянию величина, нормированная к диапазону [0..1]. Полное совпадение подстроки и левой части правила ведет к нулевому редакционному расстоянию и применению правила с максимальным значением $invCost$ = 1.

Сводная таблица когнем, знак которых присутствует в результатах поиска [1 из 1]

Знак	Формула смысла	Всего в области	Найдено символов	Ранг по достижимости	Ранг по использованию ФС
клен	Раскудрявый житель леса, весь резной листвой одетый	149	50	8 из 47	1 из 7

Знак отсутствует в результатах поиска [0 из 1]

Знак	Формула смысла	Всего в области	Найдено символов	Ранг по достижимости	Ранг по использованию ФС

Детали:

Когнема: клен - Раскудрявый житель леса, весь резной листвой одетый

Свойства

Знак	ФС	Знак преобр.	ФС преобр.	Способ	Области
клен	Раскудрявый житель леса, весь резной листвой одетый	клен	раскудрявый житель лес весь резной листва одетый	метафора	ботаника

Настройки поиска

Режим	Диапазон	Мин. длина	Макс. длина	Фильтр областей	Лемматизация
Поиск неизвестного знака	Ближайший знак	1	6	AreaFilter: ботаника	Да

Найденные символы:

Код символа	Наименование	Стоп-слово	Достижимость	Использование ФС
43951	крона	false	3.528 (ранг 1)	7 / 7 (ранг 1)
26786	дуб	false	1.379 (ранг 2)	7 / 7 (ранг 1)
40083	кедр	false	1.11 (ранг 3)	7 / 7 (ранг 1)
115566	тополь	false	0.839 (ранг 4)	7 / 7 (ранг 1)
115810	трава	false	0.76 (ранг 5)	7 / 7 (ранг 1)
111447	ствол	false	0.725 (ранг 6)	7 / 7 (ранг 1)
145285	шелковица	false	0.697 (ранг 7)	7 / 7 (ранг 1)
40703	клен	false	0.509 (ранг 8)	7 / 7 (ранг 1)

Рисунок 6. Фрагмент отчета по поиску неизвестного знака

Приведем примеры результатов моделирования в режиме с альтернативой для фигуры знания. После генерации пропозиций в грамматике и фильтрации достигнутых Знаков по Когнитивным областям формируется список (в таблице и на рисунке выше приведены фрагменты списков Знаков с максимальными рангами):

Знак	Стоп-слово	Достижимость(cost)		Использование ФС (usage)	
		значение	ранг	значение	ранг
{<Клен> <Раскудрявый житель леса, весь резной листвой одетый> <Метафора> <Ботаника> <Рецепт>}					
крона	Нет	3,528	1	7 / 7	1
дуб	Нет	1,379	2	7 / 7	1
кедр	Нет	1,11	3	7 / 7	1
тополь	Нет	0,839	4	7 / 7	1
трава	Нет	0,76	5	7 / 7	1
ствол	Нет	0,725	6	7 / 7	1
шелковица	Нет	0,697	7	7 / 7	1
клен	Нет	0,509	8	7 / 7	1
лист	Нет	0,497	9	7 / 7	1
{<Спорт> <Физические упражнения, направленные на достижение высоких результатов в соревнованиях> <Дескрипция> <Спорт> <Рецепт>}					
кросс	Нет	11,468	1	9 / 9	1
стол	Нет	8,54	2	9 / 9	1
спорт	Нет	5,694	3	9 / 9	1
бег	Нет	2,603	4	9 / 9	1
рост	Нет	0,975	5	9 / 9	1
{<Тротуар> <Пешеходная дорожка вдоль трассы> <Дефиниция> <Город> <Рецепт>}					
метро	Нет	0,291	1	4 / 4	1
зебра	Нет	0,188	2	4 / 4	1
тротуар	Нет	0,099	3	4 / 4	1
смог	Нет	0,008	4	4 / 4	3
трава	Нет	0,005	5	3 / 4	2
{<Фокус> <Искусный трюк, основанный на обмане зрения> <Дефиниция> <Шоу, цирк> <Рецепт>}					
фокус	Нет	30,445	1	6 / 6	1
акробат	Нет	7,683	2	6 / 6	1
шапито	Нет	0,487	3	5 / 6	2
номер	Нет	0,144	4	4 / 6	3
место	Нет	0,013	5	3 / 6	4
ложка	Нет	0,006	6	6 / 6	1

ложа	Нет	0,005	7	5 / 6	2

Таблица 3. Моделирование с альтернативой фигур знания

Заключение

Разработанный программный комплекс АСНИ ПМ позволяет в исследованиях вербального сознания реализовать следующие цели: во-первых, верифицировать предложенную логико-лингвистическую модель работы когнайзера в пассивном режиме в двух вариантах — безальтернативном и альтернативном; во-вторых, манипулируя инструментальными средствами (технологиями и предусмотренными в них процедурами и операциями), изменять экспериментальные базы данных исследований — АВС и фигуры знаний; в-третьих, получить принципиально новый результат лингвистического конструирования — интегрированную базу лингвокультурных знаний типового носителя русского языка; в-четвертых, осуществить макетирование построенной на когнитивных принципах поисковой компоненты широкого класса информационных систем.

В настоящее время верификация когнайзера выполнена не полностью. Верифицированными являются только одна треть всех фигур знания (~6500 когнем первого этапа когнитивного эксперимента). Для них сначала были сформированы контекстно-зависимые правила, далее они были внесены в вероятностную контекстно-свободную грамматику АВС, в результате чего сформировалась контекстно-зависимая грамматика когнайзера. Тем самым в базе знаний когнайзера был «зафиксирован» безальтернативный вариант его работы, который можно интерпретировать как этап «обучения».

Верификация альтернативного варианта работы когнайзера состоит в получении «допустимого» результата моделирования, которым считается «удачное» сопоставление Знака и Формулы смысла. Оно состоит в выполнении двух условий: а) в попадании действительного Знака когнемы в число Знаков-кандидатов на «удачное» сопоставление — грубая оценка результата; б) в значении позиции Знака в списке Знаков-кандидатов по критериям использования Формулы смысла и достижимости — уточненная характеристика результата. Например (см. Рис.6 и Табл.3), полный набор знаков области

<Ботаника> в базе данных включает 149 единиц, из которых системой для моделирования Формулы смысла <Раскудрявый житель леса, весь резной листвой одетый> отобраны 50 потенциальных кандидатов, в числе которых есть Знак <Клен> (условие «а»). Он имеет ранг = 1 из 7 возможных по использованию Формулы смысла и ранг = 8 из 47 возможных по достижимости (условие «б»).

Когнитивная область знаков устанавливается на основе когнем и играет роль фильтра Знаков-кандидатов при моделировании режима с альтернативой. Допустим, в системе нет данных о принадлежности Знака <Клен> к области <Ботаника>. Пусть моделируется сопоставление Формулы смысла <Дерево с широкими, у большинства видов резными листьями> из области <Ботаника> с некоторым Знаком. Тогда Знак <Клен>, как не имеющий отношения к области <Ботаника>, может быть обнаружен системой только с отключенным фильтром когнитивной области. А это в свою очередь приводит к увеличению множества Знаков-кандидатов. Для исключения подобной ситуации необходимо: а) определение Когнитивной области и для тех элементов АВС, которые не представлены множеством когнем; б) разработка процедур «автоматического» определения когнитивных областей у фигур знания по Формуле смысла.

Изменение базы данных исследований достигается за счет представления базы знаний когнайзера в виде грамматики и внесения в неё новых правил. Ввод новых правил связан с тремя процессами: учетом омонимии — введены 1047 экспертных правил-инструкций обработки частичной омонимии АВС; учетом синонимии — фактическим добавлением в базу знаний когнайзера новых Знаков из синонимичных рядов; лемматизацией — переходом от Знаков-словоформ к Знакам-леммам с соответствующей перегруппировкой связей (правил вывода в грамматике).

В описании программной реализации представлено техническое решение только для устранения частичной омонимии АВС. Но этого недостаточ-

но, требуется различение частичных омонимов при вводе запросов пользователя, а также устранение неоднозначности полных омонимов. В настоящий момент, при вводе запроса с полными омонимами, например «Замок в Лондоне», моделью будут применены правила, относящиеся как к строению, так и к охранному устройству.

База знаний когнайзера — это интегрированная база лингвокультурных знаний «усредненного» носителя русского языка. Она получена в результате лингвистического конструирования и является принципиально новым лексикографическим объектом. Каждая ЯЕ в нем специфицирована не только множеством параметров и эксцерпций (компоненты когнем, грамматические формы, синонимичные ряды, омонимия), но и эксплицитно задает возможную динамику ее использования носителем языка в виде ассоциативных цепочек. Полная реализация базы знаний когнайзера возможна только в виде компьютерной программы (информационно-программного изделия) зафиксированной на электронном носителе данных. В традиционной печатной форме возможно представление только ее частных проекций.

Постановка задачи поиска с альтернативой в различных вариациях встречается в системах информационного поиска, вопросно-ответных систем, семантических информационных фильтрах, автореферировании и системах машинного перевода.

Наиболее близким к задаче моделирования осознавания с альтернативой видится класс вопросно-ответных информационных систем (далее ИС) ориентированных на сущностно-ориентированные вопросы (Entity Search). Рассмотрение представленного решения с этих позиций позволит: во-первых, использовать общепринятые в области информационного поиска способы оценки качества модели; во-вторых, реализации данного режима могут быть применены как интеллектуальные (когнитивные) компоненты информационных систем вышеуказанных классов. Внедрение компоненты потребует доработки представленного решения под задачи конкретной ИС в отношении:

•качества работы — релевантности Знаков-кандидатов Формуле смысла в конкретной Референтной области, при типичных в ИС способах представления смысла;

•доступности данных — наполнение баз данных модели для конкретной предметной области, доступность компонент когнем при моделировании, специальных данных ИС (классификаций, обратной связи с пользователем);

•ресурсных ограничений — объема доступной памяти и вычислительной мощности, интеграция с архитектурой ИС.

Программный комплекс АСНИ ПМ несмотря на существование Интернет-версии «Лингвокультурного тезауруса русского языка» разрабатывается и функционирует в течение 2008-2010 гг. в виде локальной информационной системы. Основными причинами этого является [Сиренко, 2010]: требования к производительности; гетерогенность используемых программных средств; интерактивность интерфейса пользователя; недостаток качественных каналов доступа к сети Интернет у пользователей; необходимость хранения персональных параметров работы пользователей.

Наращивание возможностей Интернет и распространение широкополосного доступа к сетевым ресурсам являются причиной ее использования в научных исследованиях. При этом необходимость установки программного обеспечения и его администрирование на стороне пользователя воспринимается нежелательной. Современной технологией совмещения этих противоречивых требований пользователей к средствам автоматизации исследований является технология «облачных вычислений», позволяющая организовать обработку удаленных заданий моделирования и доставку их результатов через сеть. Именно такой видится перспектива развития АСНИ ПМ.

Список литературы

1.	*Караулов Ю.Н.* Концептография языковой картины мира. Статья 1. Первый этап «восхождения» к образу мира: от элементарных фигур знания к предметно-референтным областям культуры // Проблемы прикладной лингвистики. Выпуск 2. Сборник статей. / под ред. Н.В. Васильевой. Москва: Азбуковник, 2004.- С. 7-17.
2.	*Караулов Ю.Н., Филиппович Ю.Н.* Лингвокультурное сознание русской языковой личности. Моделирование состояния и функционирования. М., 2009: Издательский центр «Азбуковник». – 336 с.
3.	Русский ассоциативный словарь В 2 т. / *Ю.Н.Караулов, Г.А.Черкасова, Н.В.Уфимцева, Ю.А.Сорокин, Е.Ф.Тарасов.* М.: Изд. «Астрель», 2002.
4.	*Сиренко А.В.* Лингвокультурный тезаурус русского языка // Интеллектуальные технологии и системы. Сборник учебно-методических работ и статей аспирантов и студентов. Выпуск 9 / Сост. и ред. Ю.Н.Филипповича. — М.: «CLAIM», 2007. — С.264–277.
5.	*Сиренко А.В.* Информационная технология анализа результатов ассоциативных экспериментов: переход от веб-ориентированной к локальной информационной системе // Язык и сознание: психолингвистические аспекты. Сборник статей / под ред. Н.В. Уфимцева, Т.Н. Ушакова. М.-Калуга: Эйдос, 2010. С. 118-119.
6.	*Ю.Н.Филиппович.* Эскиз информационной технологии лингвокультурного тезауруса // Язык. Сознание. Культура. Сборник статей / Под ред. Н.В.Уфимцевой, Т.Н.Ушаковой. — М.– Калуга: ИП Кошелев А.Б. (Издательство «Эйдос»), 2006.— С. 319–334.
7.	*Ю.Н.Филиппович* Моделирование работы лингвокультурного когнайзера русского языка // Вопросы психолингвистики, 2007, № 6, С. 123-143.
8.	Ведущая научная школа Ю.Н. Караулова Русская языковая личность. [Электронный ресурс]. Режим доступа: http://www.philippovich.ru/Projects/Karaulov/index.htm, свободный (дата обращения: 20.03.2011).
9.	ЛКТ – Лингвокультурный тезаурус русского языка. [Электронный ресурс]. Режим доступа: http://tesaurus.ru/, свободный (дата обращения: 22.03.2011).
10.	Онлайн словарь. Словарь русских синонимов и сходных по смыслу выражений Н. Абрамова. 1999. [Электронный ресурс]. Режим доступа: http://www.dict.t-mm.ru/abramov/, свободный. — Загл. с экрана (дата обращения: 14.03.2011).
11.	Russian ispell. [Электронный ресурс]. Режим доступа: http://www.opennet.ru/prog/info/59.shtml (дата обращения: 10.03.2011).
12.	*Navarro G.* A Guided Tour to Approximate String Matching. // ACM Computing Surveys, Vol. 33, No. 1, March 2001.— С. 31–88.

Авторы: Филиппович Юрий Николаевич, к.т.н., доцент МГТУ им. Баумана; Сиренко Александр Викторович, соискатель степени к.т.н. МГУП.

Аннотация статьи на русском языке (не более 800 символов): в работе представлено программное решение для исследования когнитивной модели языкового сознания научной школы "Русская языковая личность". Приведены краткие сведения о применяемом походе к моделированию вербального сознания. Описаны психолингвистические эксперименты, на результатах которых строится модель. Рассмотрены более ранние реализации, их режимы работы и ограничения, преодолению которых посвящена данная работа. Демонстрируется режим моделирования с альтернативой, отсутствующий у прежних программных решений. В заключение обозначены основные недостатки и ограничения программного комплекса, предложено направление дальнейшего развития.

Ключевые слова: программный комплекс, когнитивное моделирование, вербальное сознание, ассоциативный эксперимент, когнитивный эксперимент.

ФИО автора, название статьи, ключевые слова и аннотация на английском языке (70-80 слов).

Philippovich Yuriy Nikolaevich, Sirenko Alexander Victorovich, Software to research psycholinguistic model of verbal consciousness based on cognitive and associative experiments.

Research software, cognitive modeling, verbal consciousness, free associative experiment, cognitive experiment.

Paper presents software to research cognitive model of verbal consciousness, based on achievements of scientific school “Russian verbal personality”. At the beginning, there is brief information about modeling approach. Paper continues by description of psycholinguistic experiments, which are empirical base of modeling.

For previously presented solutions, modes and constraints are described. Different modes of new application are demonstrated, including modeling with alternatives, which weren’t implemented before. In conclusion shortcomings of current solution are noticed, followed by suggestion of future improvements.