

РАЗРЕШЕНИЕ ФУНКЦИОНАЛЬНОЙ ОМОНИМИИ В РУССКОМ ЯЗЫКЕ НА ОСНОВЕ КОНТЕКСТНЫХ ПРАВИЛ

Ю.В. Зинькина, Н.В. Пяткин, О.А. Невзорова,
Казанский государственный университет

Научно-исследовательский институт математики и механики им. Н.Г. Чеботарева
Казанский государственный педагогический университет

В работе рассматриваются прикладные проблемы разрешения функциональной омонимии в русском языке. Изложены результаты, полученные при использовании метода разрешения функциональной омонимии, основанного на контекстных правилах. Исследуются структурные характеристики минимальных контекстных правил для различных типов функциональной омонимии. Особое внимание уделено изучению управляющей структуры правила, обеспечивающей точность разрешения омонимии не ниже 95%. Построенные контекстные правила реализованы в системе анализа технических текстов.

Введение

В лингвистической литературе нет единства взглядов на явление омонимии. Дискуссии ведутся вокруг содержания понятия, принципах классификации, классификационных схемах. Наиболее общая классификация подразделяет омонимы на лексические, т.е. принадлежащие одной части речи и грамматические, т.е. принадлежащие различным частям речи. В лингвистической литературе известен термин “функциональная омонимия”. Этот термин, предложенный О.С. Ахмановой, был принят в работах В.В. Бабайцевой [8] и её последователей. Функциональные омонимы В.В. Бабайцева определяет как “слова, совпадающие по звучанию, этимологически родственные, относящиеся к разным частям речи” [8, с. 14] и распространяет явление функциональной омонимии не только на знаменательные, но и на служебные части речи. В настоящей работе исследуются методы автоматического разрешения функциональной омонимии различных типов.

Успешность прикладных исследований в компьютерной лингвистике во многом зависит от наличия соответствующих лингвистических ресурсов, прежде всего лексикографических. В последние годы изданы словари омонимов русского языка различных авторов [1, 2, 7]. В этих словарях явление омонимии представлено с различной полнотой, так, например, в словаре Колесникова Н.П. явление омонимии понимается расширенно, и в круг рассматриваемых явлений включаются помимо лексических омонимов омоформы, омофоны и омографы. Попытки описания функциональных омонимов предприняты в специальных словарях омонимов О.С. Ахмановой [1], О.М. Ким [7]. Появление интернет-ресурса Н.Г. Аношкиной [3], в котором была сделана попытка собрать воедино все грамматические омонимы, стимулировало дальнейшее развитие

теоретических и прикладных исследований, в том числе по классификации типов функциональных омонимов. В работе Кобзаревой Т.Ю. [6] приведена классификация 58 типов функциональных омонимов. Данная классификация послужила базовой основой для разработки правил контекстного разрешения функциональных омонимов, предложенных в настоящей статье. В процессе исследования в базовую классификацию типов функциональных омонимов были внесены изменения, связанные с выделением подтипов, добавлением новых типов, а также исключением некоторых типов.

Метод контекстного разрешения функциональной омонимии (проблемы построения контекстных правил)

Теоретические исследования по проблеме разрешения омонимии в текстах имеют многолетнюю историю. Еще в конце 50-х годов в работах К.Е. Harper [5], А. Caplan [4] основным способом снятия омонимии признавалось изучение и описание тех контекстных условий, в которых реализуется то или иное значение слова. При этом под контекстом понималось окружение слова в тексте или слова, с которыми данное слово употребляется. Актуальным для исследователей также являлся вопрос о минимальном разрешающем контексте. В этой связи заслуживают упоминания результаты, полученные А. Caplan [4] по исследованию минимального разрешающего контекста. В работе анализировались 140 многозначных употребительных английских слов (в основном, лексических омонимов), находившихся в различных контекстных условиях. Были выделены следующие виды контекстов:

Сочетание с предшествующим словом – P_1 (preceding);

Сочетание с последующим словом – F_1 (following);

Сочетание с предшествующим и последующим словом – B_1 (both);

Сочетание с двумя предшествующими словами – P₂;

Сочетание с двумя последующими словами – F₂;

Сочетание с двумя предшествующими и двумя последующими словами — B₂;

Все предложение – S (sentence).

Основной вывод заключался в том, что цепочка B1 по эффекту редуцирования многозначности (отношение количества значений слова в конкретном контексте к их количеству в нулевом контексте) более продуктивна, чем контекст, состоящий из двух предшествующих или двух последующих слов (P2 и F2) и приближается к эффекту, даваемому целым предложением (S). Другой вывод подчеркивал важное значение материального типа контекста, т.е. входят ли в непосредственное окружение знаменательные слова, или слова, называемые автором «particles» (предлоги, союзы, глаголы типа will или do, артикли, местоимения и наречия типа there и др.). Первый тип контекста дает значительно большую редукцию многозначности, чем контекст, содержащий слова без конкретного лексического наполнения. Общие выводы А. Сапран сводятся к тому, что наиболее практичным является контекст, состоящий из одного слова слева одного слова справа от анализируемой многозначной лексемы. Если же одно из слов окружения — «particle», то следует «усилить» контекст до двух слов с обеих сторон.

Несмотря на многочисленные упоминания в 60-х годах в западной литературе цитированных выше результатов, их применение для русского языка в реальных контекстах вряд ли возможно. Реальная ситуация с разрешением функциональной и лексической омонимии в русском языке значительно сложнее и не может быть разрешена на основе упрощенных схем. В работе [6] для разрешения функциональной омонимии разработан специальный словарь диагностических ситуаций (СДС), задающих описания линейной структуры минимального контекста необходимого для идентификации части речи омонима. Ситуации в СДС состоят из двух частей: 1) цепочка компонент предложения (может быть разрывной) - ярлык ситуации; 2) условия, накладываемые на ярлыки и их окружение, определяющие функционально-грамматическое значение омонима. Анализируя примеры, приведенные в статье [6], можно заметить, что границы минимального разрешающего контекста становятся подвижными, в качестве границ выступают символы из определенных множеств (различаются для разных типов омонимов). Кроме того, авторы допускают разрывные контексты – что еще больше усложняет ситуацию. Помимо усложнения структуры контекста следует отметить следующее обстоятельство. В ситуации множества правил, применимых для разрешения некоторого

функционального типа, главной проблемой, по нашему мнению, становится построение управляющей структуры, контролирующей порядок применения правил.

Другими словами, предлагаемый авторами данной статьи **метод контекстного разрешения функциональной омонимии** включает:

Построение полной классификации типов функциональных омонимов.

Выделение минимального множества разрешающих контекстов для каждого типа. Минимальность множества означает, что для каждого типа функционального омонима следует оценить сложность распознавания каждой части речи, принадлежащей данному типу. Затем построить множество разрешающих контекстов (МПК), имеющих минимальную сложность распознавания.

В алгоритмической записи данное требование будет выглядеть следующим образом:

Если для функционального омонима X, имеющего тип T₁ или T₂, применено правило из МПК, то тип омонима X определяется примененным правилом, иначе приписывается альтернативный тип.

Построение управляющей структуры для множества разрешающих контекстов, обеспечивающей максимальную точность распознавания.

Классификация типов функциональных омонимов

В основе базовой классификации типов функциональных омонимом лежит классификация, предложенная в статье [6]. Данная классификация построена на основе словаря грамматических омонимов Н.Г Аношкиной. Классификация выделяет 58 типов функциональных омонимов. При этом наиболее частотными являются первые десять типов (общее число омонимов - 2965), для 26 типов выделено менее пяти представителей каждого типа. Несмотря на безусловную важность данной классификации, она требует дальнейшего развития. Это развитие происходит как за счет добавления новых функциональных типов, так и за счет дополнительной детализации типов (выделения подтипов). Примером расширения первого рода является добавление нового типа "краткое прилагательное / категория состояния" – например, омоним готово, расширения второго рода, например, связаны с выделением трех подтипов в типе <Vf/N*>, где Vf - глагол в личной форме, а N* = {N - существительное, Npr - местоименное существительное}. Примеры омонимов данного типа - берег, вызову, души. Подтипы данного типа выделены на основе синтаксического критерия, т.е. распознавание представителей данных подтипов требует разработки существенно различных правил и управляющих структур. Примеры омонимов,

относящихся к подтипу $\langle Vf/N^* \rangle_1$: *заводы, задел, пойму*. Примеры омонимов, относящихся к подтипу $\langle Vf/N^* \rangle_2$: *льну, бегу, замер, запал*. Примеры омонимов, относящихся к подтипу $\langle Vf/N^* \rangle_{3.1}$: *стали, стану*. Примеры омонимов, относящихся к подтипу $\langle Vf/N^* \rangle_{3.2}$: *заросли, поросли*.

Правила разрешения функциональной омонимии некоторых типов (реализация)

В данном разделе будет рассмотрен пример разрешения функциональной омонимии типа N^*/A^* , где $N^* = \{N - \text{существительное, Npr} - \text{местоименное существительное}\}$, $A^* = \{A - \text{полные прилагательные, Av} - \text{полные причастия, Arg} - \text{местоименные прилагательные}\}$.

Введем следующую систему обозначений классов объектов:

X – функциональный омоним; P – предлоги;

Con_{XY}^{*YX} – запись обобщенной модели управления, в которой либо X управляет словоформой Y , либо X управляется Y . Для каждого омонима X вводятся конкретные модели управления, фактически модель управления определяет некоторое множество правил, задающих устойчивые синтаксические связи омонима X .

Выражение вида $X \bigcap_{p \in n} N^*$ означает, что омоним

X согласуется по указанным грамматическим характеристикам (p - падеж, g - род, n -число) с N^* .

В записи правила может присутствовать выражение вида (Z) , которое означает возможность наличия вставочных конструкций некоторых специальных типов. В русском языке допускается практически в любом месте предложения разрыв линейной последовательности обособленной вставочной конструкцией. Нами используется типология вставок, позволяющая идентифицировать вставочные конструкции. Так например, в текстах обособляются вводные слова или словосочетания, выражающие чувства и эмоции (типа *к счастью, к несчастью*); синтаксические конструкции, выражающие оценки степени достоверности (типа *конечно, несомненно*) и др. Распознавание вставочных конструкций различных типов является важной задачей, поскольку в контекстных правилах указывается максимальная граница контекста, задаваемая либо числом, либо ограничителями, к числу которых принадлежат знаки пунктуации. Следовательно, актуальной является задача распознавания функции знака пунктуации, т.е. требуется установить, является ли знак пунктуации ограничителем вставочной конструкции или несет другую синтаксическую нагрузку. В первом случае запятая не рассматривается как ограничитель контекста.

При разработке правил разрешения функциональной омонимии различных типов нами был предложен метод множественного разрешения

функциональных омонимов. Данный метод позволяет разрешать функциональную омонимию однородной группы омонимов. Правила множественного разрешения особенно характерны для некоторых функциональных типов, например для типа $D/Abr/Vsp$ (наречие/краткое прилагательное/категория состояния), типа N^*/A^* . Ниже приведен пример контекстных правил, разработанных для типа N^*/A^* , в структуру которых входят два правила (с номерами 1 и 4), использующих множественное разрешение. Правило 1) позволяет разрешать омонимию группы типа 'старые лысые горы', правило 4) применяется для разрешения группы вида 'мой дорогой больной'.

Тестирование полученных правил производилось на корпусе текстов из библиотеки Мошкова (<http://www.aot.ru/search1.html>). Точность распознавания функциональной омонимии каждого типа не ниже 95%.

Приведем пример группы контекстных правил для распознавания омонимов функционального типа N^*/A^* . Минимальное множество разрешающих контекстов было выделено для распознавания функционального омонима как A^* . Управляющая структура обобщенного правила определяет порядок применения контекстных правил (правила заданы своими номерами) и результаты распознавания (приведены в прямоугольниках на схеме).

- 1) if $[X_1 \bigcap_{p \in n} N^*, X_2 \bigcap_{p \in n} N^*, \dots, X_m \bigcap_{p \in n} N^*, N^*]$ then $X_1 = X_2 = \dots = X_m = A^*$
 - 2) if $[X \frac{(Z) \bar{P} N^* \bigcap_X N^* (Z) \bar{P}}{\leq 4/, \leq 2/}]$ then $X = A^*$
 - 3) if $[Con_{XY}^{*YX}]$ then $X = A^*$
 - 4) if $[X_1 \bigcap_{p \in n} X_n, X_2 \bigcap_{p \in n} X_n, \dots, X_n]$ then $X_n = N^*, X_1 = X_2 = \dots = A^*$
- else $X = N^*$

Рис. 1 Правила разрешения омонимов функционального типа N^*/A^*

Комментарий к правилам. Правило 1 выделяет в локальном контексте группу омонимов, согласованных по грамматическим характеристикам с существительным, непосредственно примыкающим к правой границе группы. В этом случае все омонимы группы разрешаются как A^* . Пример контекста правила 1: 'старые лысые горы'.

Правило 2, состоящее из двух альтернатив, в которых анализируется соответственно правый и левый контекст омонима и в допустимых границах контекста проверяется отсутствие предлога и наличие существительного, согласованного по грамматическим характеристикам с омонимом X . При выполнении всех условий омоним X разрешается как A^* . Пример контекста правила 2: 'дорогой человек'. Правило 3 проверяет наличие в контексте модели управления омонима X как A^* .

Правило 4 применяется к группе омонимов, согласованных по своим грамматическим характеристикам с последним омонимом. В данном контексте последний омоним группы X_n разрешается как N^* , а другие члены группы как A^* . Пример контекста правила 4: 'дорогой больной'.

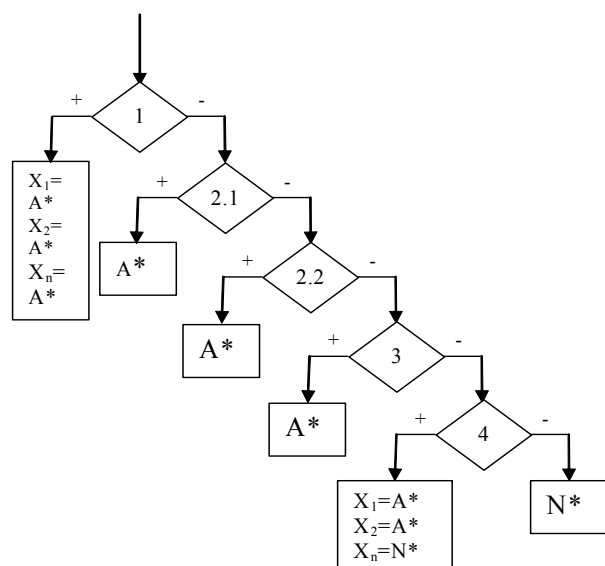


Рис. 2. Управляющая структура обобщенного правила N^*/A^*

Примеры разрешения функциональной омонимии типа N^*/A^* для омонима 'больной'

У нее был **больной** вид. (1)

На основе правила 2.1 омоним **больной** разрешается как A^* .

По закону **больной** проказой подлежит немедленному переселению в лепрозории. (2)

Омоним **больной** разрешается как N^* по ветви *else* последнего правила.

Заключение

В настоящее время заканчивается программная реализация модуля предсинтаксической обработки системы анализа технических текстов, который содержит процедуру разрешения функциональной омонимии. Реальные технические тексты не содержат некоторые типы функциональной омонимии, например типы, содержащие в своем составе междометия, а также типы омонимов с разговорно-сниженной стилистической окраской. Другие ограничения связаны с лексическим составом ограниченной предметной области. Коллекция технических текстов (размером свыше 100 Мб) не содержит, например, ни одного

вхождения омонима *при* как Vf (личная форма глагола). Тем не менее, контекстные правила позволяют разрешать функциональную омонимию типа Vf/P с единственным представителем – омонимом *при*. Настройки модуля разрешения омонимии позволяют менять количество типов омонимов, подлежащих разрешению. Тестирование программного модуля разрешения функциональной омонимии для реализованных типов дает хорошие результаты, позволяя для некоторых типов получать точность распознавания, равную 100 % при тестировании не менее 100 примеров, в наихудших случаях – точность не менее 95%. Причинами возникновения ошибочных ситуаций являются либо случайное согласование в пределах контекстного правила, либо недостаточность контекста, либо недостаточность ресурсов (например, отсутствие словаря моделей управления слов различных частей речи). Некоторые ошибки в разрешении омонимии могут быть сняты на последующих этапах анализа.

Благодарности

Данная работа выполнена при поддержке РФФИ, грант № 05-07-90257.

Список литературы

- 1) Ахманова А.С. Словарь омонимов русского языка. М., 1984
- 2) Колесников Н.П. Словарь омонимов русского языка. Тбилиси, 1978.
- 3) Аношкина Ж.Г. Словарь омонимичных словоформ русского языка. М: Машинный фонд русского языка Института русского языка РАН, 2001.
(<http://irlras-cfpl.rema.ru:8100/homoforms/index.htm>).
- 4) Caplan A. An Experimental Study of Ambiguity and Context // Mech. Translation, vol. 2, No 2, Nov. 1955.
- 5) Harper K.E. Contextual Analysis // Mech. Translation, vol. 4, No 3, Dec. 1956.
- 6) Кобзарева Т.Ю., Афанасьев Р.Н. Универсальный модуль предсинтаксического анализа омонимии частей речи в РЯ на основе словаря диагностических ситуаций // Труды междунар. конференции Диалог'2002. М., 2002. С. 258-268.
- 7) Ким О.М., Островкина И.Е. Словарь грамматических омонимов русского языка. М, 2004.
- 8) Бабайцева В.В. Переходные конструкции в синтаксисе. Воронеж, 1967.