

МЕТОД КОНТЕКСТНОГО РАЗРЕШЕНИЯ ФУНКЦИОНАЛЬНОЙ ОМОНИМИИ: АНАЛИЗ ПРИМЕНИМОСТИ¹

THE METHOD OF FUNCTIONAL HOMONYMY DISAMBIGUATION ON THE BASIS OF CONTEXTUAL RULES: THE STUDY OF METHOD APPLICABILITY

*О.А. Невзорова (Olga.Nevzorova@ksu.ru)
Казанский государственный университет*

*Ю.В. Зинькина (zjuliv@mail.ru)
Научно-исследовательский институт математики и механики им. Н.Г. Чеботарева*

*Н.В. Пяткин (nikolaip@mail.ru)
Татарский государственный гуманитарно-педагогический университет*

Статья посвящена анализу применимости метода разрешения функциональной омонимии на основе контекстных правил в русском языке. Рассматривается состояние лексикографических ресурсов, описывающих явление омонимии, и сложные случаи метода.

Введение

Настоящая статья является продолжением статьи [1], в которой были изложены первые результаты исследований по автоматическому разрешению функциональной омонимии в русском языке на основе метода контекстных правил. Дальнейшее развитие исследований способствовало более четкому выделению основных проблем, связанных, прежде всего, с описанием явления функциональной омонимии в существующих лексикографических источниках; выделением синтаксически сложных случаев разрешения омонимии. Некоторые результаты этих исследований будут изложены в настоящей статье. Предваряя основные выводы статьи, можно утверждать, что абсолютно точное разрешение омонимии на основе метода контекстных правил не представляется возможным в настоящее время, по крайней мере, по двум наиболее существенным причинам:

1) данное явление слишком противоречиво представлено в существующих лексикографических ресурсах русского языка, и эта ситуация, вероятно, не может быстро существенно измениться – существуют дискуссионные проблемы описания и принципиально сложные случаи;

2) синтаксически сложные случаи функциональной омонимии могут быть разрешены на основе процедур, опирающихся на использование сложных синтаксических структур, т.е., другими словами, требуются развитые методы синтаксического анализа, в то время как этап разрешения функциональной омонимии является все-таки предсинтаксическим анализом.

Несмотря на всю сложность рассматриваемого явления, авторы считают, что наиболее точные результаты могут быть получены в рамках предлагаемого контекстного метода, или на основе метода исчисления семантических контекстов (что представляется еще более сложным), однако затраченные усилия оправданы в тех лингвистических приложениях, где существенна точность автоматического анализа.

Метод контекстного разрешения функциональной омонимии: словарные ресурсы

В работе [1] метод контекстного разрешения функциональной омонимии включал несколько этапов:

1) построение полной классификации типов функциональных омонимов;

2) 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 необходимо построить множество разрешающих контекстов (МПК), имеющих минимальную сложность распознавания. В алгоритмической записи данное требование выражается следующим правилом: 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;

3) построение управляющей структуры обобщенного правила, обеспечивающего максимальную точность распознавания.

¹ Работа выполнена при поддержке Российского Фонда Фундаментальных Исследований, грант № 05-07-90257.

Дальнейшие исследования привели к усложнению структуры метода разрешения омонимии, к перечисленным выше основным этапам добавился ряд новых, позволяющих распознавать и учитывать при разрешении омонимии более сложные синтаксические конструкции.

Рассмотрим некоторые основные проблемы, затрудняющие применение контекстного метода. Безусловно, важным этапом является этап 1, который требует построения максимально полной классификации омонимов и уточнения грамматических характеристик омонимов различных типов.

Явление омонимии в русском языке описано в различных словарях [2-7]. Словарь Ахмановой предоставляет довольно большую коллекцию лексических омонимов при практически полном отсутствии грамматических (некоторые даны в приложении). Словарь Колесникова дает лексические и грамматические омонимы, однако не проводит разграничения между ними; принцип составления словника при этом не уточняется.

Словарь Аношкиной практически целиком посвящен грамматическим омонимам. В словаре указаны не только полные омонимы, т.е. те, у которых совпадает вся парадигма (например, лексические омонимы *ключ*, *коса*), но и частичные, т.е. имеющие лишь отдельные общие словоформы (*стих*, *стекло* – существительное/глагол), и неравнообъемные – с полным совпадением для одного и частичным для другого (*летом* – наречие/существительное). Однако в этом словаре отсутствуют примеры, что не позволяет с уверенностью сказать, в каких контекстах употребляется каждый конкретный омоним. Это особенно затрудняет дело, если учесть, что многие омонимы, приведенные в словаре, являются потенциальными, т.е. употребляются только в одной части речи всегда или почти всегда. Как и в словаре Ахмановой, лексические омонимы не отделяются от грамматических.

Словарь Ким – Островкиной рассматривает исключительно грамматические омонимы. На каждый из них дается по одному контексту употребления на каждый морфологический класс омонима. Однако и этот словарь не указывает принцип отбора слов, хотя явно отдает предпочтение некоторым определенным типам омонимов перед другими. Так, здесь представлена довольно значительная коллекция омонимов на '-о', а также типа "существительное/прилагательное", в то время как тип "существительное/глагол" практически отсутствует.

Электронный ресурс [6] созданный на основе словаря Зализняка, дает морфологические характеристики омонима и его английский перевод, что служит хорошим материалом при работе с редкими словами неизвестной семантики. Однако в словаре практически полностью отсутствуют наречия, что создает неудобства при определении характеристик омонимов на '-о'. Кроме того, отсутствуют примеры контекстов. Новым электронным ресурсом является Национальный корпус русского языка [7] с размеченной омонимией. Так, омонимам на '-о' здесь даются довольно точные характеристики, приводятся и примеры их использования в газетных текстах. Однако, пока омонимия разрешена не во всем корпусе, а специфика газетного корпуса иногда затрудняет определение всего спектра возможных контекстов.

Весьма непростая ситуация возникает при сопоставлении различных словарных источников. Так, в проведенном нами эксперименте по сопоставлению грамматических характеристик 560 омонимов, оканчивающихся на букву 'о' по ресурсам [4-7] были получены неожиданные результаты. Только в трех случаях из 560 все четыре ресурса приписали омонимам одинаковые характеристики. Это были омонимы *прямо* и *исключительно* с характеристиками "краткое прилагательное/наречие/частица", а также омоним *относительно* с характеристиками "краткое прилагательное/наречие/предлог". Были также случаи, когда в одном из перечисленных ресурсов слово отсутствовало, а оставшиеся три соглашались между собой. Ниже приводятся все эти омонимы с приписанными им характеристиками:

- *ненастно* – "краткое прилагательное / предикатив" (отсутствует в НКРЯ);
- *рано* – "наречие/ предикатив" (отсутствует в словаре Аношкиной);
- *рановато* – "наречие/ предикатив" (отсутствует в словаре Аношкиной);
- *ровно* – "краткое прилагательное/ наречие/ частица/ союз" (отсутствует в словаре Аношкиной).

Во всех прочих случаях (553 из 560) омонимам приписывались два, три и даже четыре различных набора характеристик.

Усложненная структура правил метода контекстного разрешения функциональной омонимии

Для каждого типа функциональной омонимии разрабатывается обобщенное правило разрешения омонимии данного типа. Обобщенное правило представляет собой упорядоченную совокупность правил, записанных на специальном формальном языке. Каждое правило внутри совокупности фиксирует некоторый разрешающий контекст. Структура задает порядок применения правил, который базируется на оценке частотности контекстов.

Пример обобщенного правила для регулярного подтипа типа N^*/Vf (существительное/личная форма глагола, например, *бегу*, *вызову*, *гнет*) приведен на рисунке 1. Данное правило применимо к подгруппе регулярных омонимов указанного типа, т.е. омонимов, которые условно равночастотно являются существительными или личными формами глагола, при этом правило фактически направлено на распознавание типа Vf , т.е. если найдены соответствующие разрешающие контексты, то омоним распознается как Vf , иначе – как N . Однако, распознавание омонима как N имеет два очень частотных контекста, поэтому соответствующие контексты включены в структуру обобщенного правила (правила 1 и 2). Для записи контекстных правил используется формальный язык со следующей системой обозначений: X – функциональный омоним; P –

предлог; Con - модели управления (или синтаксические конструкции), в которых омоним X однозначно разрешается. Выражение вида $X \bigcap_{pgn} N^*$ означает, что омоним X согласуется по указанным грамматическим

характеристикам (р - падеж, g - род, n-число, f - лицо) с N*. В записи правила может присутствовать выражение вида (Z), которое означает возможность наличия вставочных конструкций некоторых специальных типов. В качестве примера дадим комментарий к правилу 1) в составе обобщенного правила 1. Правило 1) позволяет распознать омоним X как существительное (N*), если в левом контексте длиной не более 3 слова обнаружено согласованное по роду, числу и падежу с X прилагательное (причастие или местоименное прилагательное) и нет другого согласованного по тем же характеристикам существительного.

Приведенная структура обобщенного правила, в целом, является примером простой структуры. Фактически, каждое правило в структуре данного обобщенного правила устанавливает наличие или отсутствие в определенном численном интервале слова определенных частей речи. Дальнейшее развитие метода связано с учетом контекстов сложной синтаксической природы, в частности, с анализом однородных групп. Выделение однородной группы позволяет искать разрешающий элемент за границами однородной группы; тем самым, реально увеличивается численный интервал разрешающего контекста. Такого рода правила анализа омонимов в составе однородной группы были включены в состав обобщенных правил различных типов, в частности, для типов D/Abr (наречие/краткое прилагательное), D/Abr/Vsp (наречие/краткое прилагательное/предикатив), N*/A* (существительное - местоименное существительное/прилагательное - причастие - местоименное прилагательное). Например, в предложении "Однако, разбить программу на отдельные подсистемы одним из определенных и корректных способов вполне возможно, важно и даже выгодно" присутствует однородная группа омонимов "вполне возможно, важно и даже выгодно", содержащая омонимы *возможно, важно, выгодно* типа "наречие/краткое прилагательное/предикатив", которые с помощью специального правила анализа однородной группы, заданного для данного типа, разрешаются как предикативы.

- 1) if $\left[\frac{A^* \cap_{pgn} x(Z)N \cap A^*}{\leq 3} X \right]$ then $X = N^*$
- 2) if $\left[\frac{P \cap X (Z)N \cap p}{\leq 3} X \right]$ then $X = N^*$
- 3) if [как X,] / [словно X,] / [будто X,] then $X = N^*$
- 4) Con
- 5) if $\left[\frac{D - Dmd - Dp(Z)V^* Av}{\leq 3/\otimes} X \right] / \left[X \frac{(Z) V^* Av D - Dmd - Dp}{\leq 3/\otimes} \right]$ then $X = Vf$
- 6) if $\left[\frac{V^*(Z)}{\leq 5/\otimes} X \right]$ then if (л.к. ≥ 3 then $X = Vf$ else goto 7) else $X = N^*$
- 7) if $\left[X \frac{V^*(Z)}{\leq 4/\otimes} \right]$ then if (н.к. ≥ 3 then $X = Vf$ else goto 8) else $X = N^*$
- 8) if $\left[\frac{N^{**} \cap_{p1gnf} x = Vf(Z)Vf \cap_{gnf} x}{\leq 5/\otimes} X \right] / \left[X \frac{N^{**} \cap_{p1gnf} x = Vf}{1} \right]$ then $X = Vf$
- 9) if $[X = Vf < Sv\& > Vf \cap_{gnf} X] / [Vf \cap_{gnf} X < Sv\& > X = Vf]$ then $X = Vf$
else $X = N^*$

Рис. 1. Обобщенное правило для регулярного подтипа типа N*/Vf

С другой стороны, синтаксическая однородность тяготеет к семантической однородности, т.е. метод приписывает всем членам однородной группы одни и те же характеристики по частям речи, рассматривая членов однородной группы как элементы одного множества. Так в предложении "Среди них были богатые и бедные, красивые и страшные, образованные и невежды" благодаря наличию однородной группы с парными членами "красивые и страшные", "богатые и бедные", "образованные и невежды" омонимы типа N*/A* *богатые*

и *бедные* разрешаются как существительные, т.к. одним из членов пары является однозначное существительное *невжды*. Такое разрешение осуществляется на основе специального правила анализа однородной группы с парными членами, которое включено в состав обобщенной группы указанного типа. При этом заметим, что в составе обобщенного правила типа N*/A* содержатся также и другие правила, в том числе и для анализа однородных групп другого строения. Еще одним препятствием метода контекстного разрешения омонимии является явление эллипсиса, которое имеет важнейшее значение для описываемого способа разрешения омонимии. Как уже упоминалось ранее, приписывание омониму той или иной характеристики части речи осуществляется на основе анализа наличия либо отсутствия в контексте определенной длины слов тех или иных частей речи. Опущение этих слов в контексте служит серьезным препятствием для осуществления синтаксического анализа и может существенно исказить результаты.

Тесно связано с эллипсисом явление субстантивации. Например, в предложении "*В магазине не было черного хлеба, а лишь дорогой белый*" применение правил разрешения омонимии может приписать словоформе *белый* характеристики существительного. Чтобы разрешить его как прилагательное, правило должно найти согласованное существительное *хлеб*, которое в данном контексте эллиптически опущено. Избежать такой ошибки можно, если не рассматривать *белый* как потенциальный субстантив. Возможно, некоторым выходом из этой ситуации – рассматривать как субстантивы либо слова, устойчиво употребляющиеся в качестве существительных (*больной, мороженое*), либо прилагательные среднего рода, употребляющиеся без определяемого слова, часто как подлежащее или прямое дополнение (*'носить белое', 'верить в сверхъестественное'*). Однако и такое решение не исключает ошибки.

Сравним два предложения:

- *Она не любит блеклость, даже носит только красное* – субстантив.
- *Я не люблю белое вино, я пью только красное* – эллипсис.

Таким образом, проблема эллипсиса является весьма значимой для синтаксического разрешения омонимии и требует поиска новых путей решения.

Заключение

Метод разрешения функциональной омонимии на основе контекстных правил по сути своей базируется на синтаксических моделях. Это обстоятельство определяет достоинства и недостатки метода. К числу преимуществ, прежде всего, следует отнести более высокую точность, по сравнению с вероятностными методами. Недостатки метода также следуют из его синтаксической природы – синтаксические структуры русского языка весьма гибки и многообразны. В практических приложениях, ориентированных на конкретные подязыки, целесообразно применение инженерного подхода с его прагматическими решениями – выделение группы потенциальных омонимов, т.е. практически не омонимичных; учет семантики конкретной предметной области, фактически снимающей большую часть многозначности и т.п. Однако следует понимать, что подобные решения уменьшают не сложность явления, а лишь частично снижают трудоемкость выхода на некоторую точность распознавания.

В заключение следует отметить актуальную проблему разработки нового словарного ресурса по функциональной омонимии, основанного на корпусных исследованиях, в котором будут уточнены грамматические характеристики функциональных омонимов.

Литература

1. Зинькина Ю.В., Пяткин Н.В., Невзорова О.А. Разрешение функциональной омонимии в русском языке на основе контекстных правил // Труды межд. конф. Диалог'2005.– М.: Наука, 2005. С. 198-202.
2. Ахманова А.С. Словарь омонимов русского языка. - М., 1984.
3. Колесников Н.П. Словарь омонимов русского языка. - Тбилиси, 1978.
4. Аношкина Ж.Г. Словарь омонимичных словоформ русского языка. – М.: Машинный фонд русского языка Института русского языка РАН, 2001. (<http://irlras-cfml.rema.ru:8100/homoforms/index.htm>).
5. Ким О.М., Островкина И.Е. Словарь грамматических омонимов русского языка. – М., 2004.
6. Сайт "Вавилонская башня" <http://starling.rinet.ru>
7. Национальный корпус русского языка <http://www.ruscorpora.ru>