

ОБ ОДНОМ ПОДХОДЕ К КЛАССИФИКАЦИИ ПРИЛАГАТЕЛЬНЫХ

Е.А.Каневский, Е.Н.Клименко, В.А.Тузов

С.-Петербург

kanev@emi.spb.su

1. Цели работы

Проблема обработки естественно-языковых текстов, проблема "понимания" текста компьютером была и остается актуальной. Цель нашей работы – извлечь содержательную информацию из текста "деловой" прозы и пополнить ею базу знаний.

Одним из основных источников информации для трансляции текстов с русского языка на формальный язык является семантический словарь. Его словарная статья содержит информацию о моделях управления и описание семантики слова, выраженное в терминах лексических функций и базисных понятий. В простейшем случае такая статья может содержать только указание на принадлежность лексемы к тому или иному классу [1, 2]. Результатом анализа текста является его семантико-синтаксическая модель. Одно из ее основных назначений – унификация формального представления предложений русского языка, близких по смыслу.

В базе знаний информация представлена в виде концептуальных моделей данных. Модель описывает содержание соответствующего ей текста набором утверждений об упоминаемых в нем объектах, об отношениях между этими объектами и их атрибутами. Средства этого языка [3] позволяют строго определять прагматическую интерпретацию описываемых понятий, экземпляров (примеров) этих понятий и отношений между ними. Набор утверждений на этом языке представляет собой модель реальной действительности в пределах данной предметной области.

Связь между двумя этими моделями в каждом конкретном случае устанавливается правилами преобразования семантико-синтаксических отношений между сущностями в концептуальные. Состав правил определяется спецификой конкретной предметной области и той информацией, которая уже содержится в базе знаний. Правила преобразования фиксируются в базе знаний, представляя собой существенную часть прагматического описания данной предметной области.

2. Постановка задачи

Источником семантического описания слова является толковый словарь. Естественно, что формальное описание лексемы в семантическом словаре существенно отличается от своего прообраза. При этом часть лексем являются базисными – их толкование или невозможно, или нецелесообразно [1]. Остальные лексемы выражаются через базисные с помощью лексических функций. Все базисные лексемы классифицированы. Заметим, что предварительный классификатор базисных прилагательных для 100-тысячного семантического словаря содержит около 65 классов [2].

Рассмотрим подробнее вопросы, связанные с классификацией прилагательных. Очевидно, что согласно [3] наиболее удобная форма представления в концептуальной памяти прилагательного и сочетающегося с ним существительного есть тройка элементов вида

<объект>(<атрибут>.<значение>)

Отсюда следует, что, вообще говоря, для каждой пары "существительное – прилагательное" следует определить соответствующий атрибут. Им может служить наименование класса данного прилагательного. Отсюда и возникает задача – классифицировать все прилагательные и указать, к каким существительным относятся прилагательные данного класса.

Далее следует учесть требования унификации результатов анализа. Очевидно, что словосочетания типа белый стол и стол белого цвета являются равнозначными. Их унификация возможна, например, в том случае, если наименованием класса данного прилагательного является атрибут ЦВЕТ. Естественно, что подобная ситуация может иметь место только в тех случаях, когда в русском языке не только существует, но и напрямую употребляется название соответствующего атрибута: цвет, форма, размер и т. п. (в любом падеже и числе). Такие прилагательные будем называть прилагательными первого рода. В приведенных ниже классах этих прилагательных в скобках даны примеры равнозначного преобразования (отметим, что не все подобные преобразования выглядят привычно). Названия других атрибутов не употребляются в русском языке, их мы будем обозначать латинскими буквами (Abst, Aktiv, Distan и т. п.). Это прилагательные второго рода. Встречаются случаи, когда одни и те же прилагательные, употребляясь с разными типами существительных, образуют разные классы. Тогда основной класс отмечен "*", а класс-аналог мы будем относить к прилагательным третьего рода.

Для простоты изложения будем пользоваться упрощенной классификацией существительных. Все существительные делятся на объекты и понятия. Выделяются объекты одноразового пользования, съедобные объекты и объекты, обладающие запахом. Объекты относятся к живой и неживой природе. Среди объектов живой природы выделяются живые существа, среди них – человек. Для последнего характерно наличие частей тела. Из объектов неживой природы выделены объекты, имеющие объем, и предметы. Последние можно обособить, они обладают размерами, формой и весом. Среди понятий выделяются процессы.

Таким образом, наша задача заключалась в классификации прилагательных с обязательным указанием, с какими существительными этот класс взаимодействует. На начальном этапе работы в качестве исходного материала взят малый словарь, содержащий около 10 тысяч наиболее употребительных слов русского языка.

3. Известные подходы к классификации прилагательных

Известны различные подходы к классификации прилагательных. Так, при работе с синтаксическим словарем системы АРТ были выделены семантико-синтаксические классы прилагательных [4]. В частности, выделены классы с семантикой цвета, исключительности, успешности / неуспешности, степени пригодности, отношения и потенциальности.

При работе с семантическим словарем системы ПОЛИТЕКСТ была исследована семантика прилагательных, взаимодействующих с параметрическими существительными [5]. При этом все прилагательные были разделены на два класса. Для первого класса приведено 9 групп прилагательных, относящихся к определению параметра. В них отражены привязка к физическому явлению, способ получения значения параметра, причина возникновения явлений и др. Для второго класса приведено 6 групп прилагательных, характеризующих числовое значение параметра. В них отражены результаты аналитической обработки, характеристики значения параметра и др.

Интересный подход к классификации относительных прилагательных предлагает Ю.Д.Апресян [6]. Он выделяет 3 типа каузативных значений, 2 типа целевых значений, 5 типов параметрических значений и 7 других типов. Для всех типов приведены многочисленные примеры с указанием существительных, сочетающихся с соответствующими прилагательными. Приводятся примеры равнозначных преобразований.

Поскольку семантические связи, скрытые в сочетаниях прилагательных и существительных, не отличаются от семантических связей соответствующих выражений с наречиями, обратимся к классификации наречий. В [7] приводится 9 классов наречий, которые в свою очередь делятся на группы. Наибольший интерес для нас представляют следующие группы: субъект контролирует / не контролирует ситуацию, отсутствие в данной ситуации некоторого компонента, субъект не осознает свое участие в данном действии, наличие подобию.

Более детально связь между прилагательными и наречиями проанализирована в [8]. Отмечается, что качественное прилагательное напрямую связано с существительным. Напротив, относительные прилагательные связаны с существительным посредством отношения (сделанный из ..., имеющий форму ... и т. п.). Среди прилагательных встречаются такие, которые обладают и теми, и другими признаками (прилагательные цвета и времени). Показано, что наречиям с плавающей сферой действия соответствуют качественные прилагательные, а наречиям с фиксированной сферой действия – относительные.

Наконец, интересные соображения по определению классов прилагательных можно получить при анализе классов существительных, предложенных в [9]. В частности, определенный интерес вызывают такие классы, как лицо, свойства, состояние, сфера.

Рассматривая зону лексической сочетаемости словарной статьи в толково-комбинаторном словаре, И.А.Мельчук [10] указывает, что несочетаемость лексем в выражении типа восхищение черного цвета "в словаре особо оговариваться не должна: она вытекает из несочетаемости их смыслов". Учет этого глубокого замечания на практике позволил существенно расширить класс допустимых существительных или, по некоторой аналогии с наречиями, сферу действия прилагательных (СДП).

4. Примеры классов прилагательных первого рода

Свойство ВИД – внешний вид (СДП – все). Например, жалкий пес, живописный пейзаж, зловещая мысль, нагой юноша, обнаженная статуя, угрюмый океан (голый человек – человек голого вида).

Свойство ВКУС* (СДП – съедобные объекты). Например, кислый суп, сладкий компот, соленый огурец (горькая вода – вода горького вкуса).

Свойство ВОЗРАСТ (СДП – все). Например, взрослый человек, младший брат, молодой сад, новорожденное животное, старая мысль, юный пионер (зрелый человек – человек зрелого возраста).

Свойство ВРЕМЯ (СДП – все). Например, вековой вопрос, весенний город, древний человек, октябрьский переворот, прошлогодний снег, прошлый год (античный город – город античных времен).

Свойство МАТЕРИАЛ* – состоит из (СДП – объекты неживой природы). Например, алмазный бур, восковая фигура, глиняный горшок, гранитный камень, каменная гора (белковый торт – торт из белкового материала).

Свойство (ПРЕД)НАЗНАЧЕНИЕ – предназначенный для чего-либо (СДП – предметы и понятия) или предназначенный для кого-либо (СДП – объекты одноразового использования). Например, бытовой предмет, декоративная панель, праздничный стол, столовый прибор, солдатская каша (контрольная работа – работа, предназначенная для контроля, командирский обед – обед, предназначенный для командира).

Свойство ПРИНАДЛЕЖНОСТЬ – принадлежащий кому-либо (СДП – все, кроме объектов одноразового использования). Например, барский тон, героический поступок, хозяйский дом, девичий платок, волчий вой (авторский гонорар – гонорар, принадлежащий автору).

Свойство РАЗМЕР* (СДП – объекты). Например, гигантский гора, короткое пальто, крупный порт, крохотный парк, микроскопический отпечаток (большая квартира – квартира большого размера).

Свойство РОСТ (СДП – живые существа). Например, исполинское дерево, невысокий юноша, низенький пони (высокий человек – человек высокого роста).

Свойство ТЕМПЕРАМЕНТ* (СДП – живые существа). Например, агрессивный лев, воинственный сосед, живой пес, мягкая женщина, холодный муж (горячий мужчина – мужчина с горячим темпераментом).

Свойство ФОРМА (СДП – объекты). Например, вытянутое лицо, квадратный сад, круглый стол, плоская коробка (граненый стакан – стакан граненой формы).

Свойство ХАРАКТЕР* (СДП – живые существа). Например, безжалостный зверь, беспощадный воин, добродушный кот, капризная болонка, легкомысленная девочка (гордая девушка – девушка с гордым характером).

Свойство ЦВЕТ* (СДП – объекты, кроме человека). Например, белый пароход, зеленый свитер, коричневая скала, лиловый туман, синее небо (желтый цветок – цветок желтого цвета).

Свойство ЧУВСТВО* (СДП – человек). Например, страстный юноша, вдохновенный поэт, влюбленная девушка, возмущенный сосед, восторженный, гневный критик, грустный водитель (радостная женщина – женщина с радостными чувствами).

5. Примеры классов прилагательных второго рода

Свойство ABST – абстрактный (СДП – понятия). Например, абсолютный закон, абстрактные знания, академический подход, гуманное решение, законодательный акт.

Свойство AKTIV – активность (СДП – живые существа и процессы). Например, активный труд, бодрый щенок, бойкое дитя, возбужденные пчелы, спокойное течение.

Свойство CHAST – частота процессов и изменения объектов (СДП – объекты и процессы). Например, непрерывный шум, неразрывное течение, нередкий гром, постоянное расписание, регулярный процесс, редкий запуск.

Свойство DAY – день (СДП – все). Например, воскресный семинар, ежедневный дождь, завтрашний поход, ночной разговор, дневная сиделка.

Свойство DISTAN – расстояние (СДП – все). Например, ближнее озеро, близкий друг, далекий край, дальняя баня, местное отделение, отдаленное дерево.

Свойство EFEKTO – тип или оценка действия (СДП – понятия). Например, внезапный взрыв, дружная работа, машинальный взмах, рациональное понятие.

Свойство FIZIKA – физика (СДП – понятия и объекты неживой природы). Например, атомный котел, инфракрасное излучение, квантовая физика, магнитное поле.

Свойство GEOGRAPH – место происхождения (СДП – все) азиатский грипп, армянская кухня, белорусский холодильник, волжский арбуз, питерский житель.

Свойство НОМО* – внутреннее состояние (СДП – человек). Например, тихий подросток, трезвый сосед, пьяный водитель, лихой пилот, растерянный прохожий.

Свойство LIVHAR* – характер (СДП – живые существа). Например, веселый щенок, бешеный пес, добрый юноша, злой волшебник, коварный, ласковый кот, преданный друг.

Свойство LOC – место (СДП – объекты). Например, альпийская фиалка, береговой маяк, деревенская изба, океанский патруль, степной волк.

Свойство MOV – вид или тип движения (СДП – объекты и процессы). Например, автоматический процесс, быстрый рост, медленное течение, скорый поезд.

Свойство NOPOSLED – отсутствие последствий (СДП – все). Например, безвозвратная потеря, безвозмездный труд, бесплатный обед, бесплодная корова, напрасные усилия.

Свойство POLIT – политика (СДП – предметы и понятия). Например, беспартийный сотрудник, буржуазные предрассудки, классовый враг, комсомольский билет, феодальный строй. Сюда же примыкают прилагательные цвета в сочетании с определенными существительными: белая пуля, красный пояс (при голосовании), зеленые банды, коричневый отряд (штурмовиков).

Свойство REAL – реальность (СДП – все). Например, волшебный сон, выразительный взгляд, действительный академик, достоверный результат, подлинный рай.

Свойство RELO – родственные отношения (СДП – человек). Например, вдовый родственник, двоюродный племянник, женатый мужчина, кровный брат, родной человек, семейный. Особняком стоит группа притяжательных прилагательных типа RELO1, обозначающая принадлежность родственников типа бабушкин платок, отцовский сапог, девичий сарафан, семейный портрет, братский поцелуй (СДП – все).

Свойство RELSITUAC – относительное положение (СДП – все). Например, верхняя полка, внешний мир, внутреннее море, нижняя палата, поперечная стойка.

Свойство SOUND – звук (СДП – все). Например, глухой человек, громкий стук, беззвучный голос, бесшумный двигатель, тихий водопад, шумный класс.

Свойство SPHERA – сфера деятельности человека и название всех наук (СДП – все) аграрный вопрос, астрономический календарь, биологический штамм, военно-морской флаг, полицейское государство, почтовый работник.

Свойство STOJK – стойкость (СДП – все). Например, выносливый мул, огнеупорный кирпич, стойкий солдатик, устойчивая тенденция.

Свойство UM – ум (СДП – все). Например, сумасшедший дом, безумный взгляд, остроумный разговор, дурацкая квартира, мудрый старец, глупый попугай.

Свойство VES* – вес (СДП – предметы). Например, легкий чемодан, нелегкий баул, массивная статуя, равновесные гири, тяжелый прибор.

Свойство VOLUME* – объем (СДП – объекты, имеющие объем). Например, глубокая речка, мелкая тарелка, полный стакан, полые рога, пустая канистра.

6. Примеры классов прилагательных третьего рода

Свойство CHARAKTER – оценка (СДП – части тела, понятия и объекты неживой природы). Например, беспощадный океан, капризный рот, гордый взгляд, добродушная улыбка.

Свойство CHUVSTVO – эмоциональная оценка (СДП – понятия). Например, страстный поцелуй, возмущенный протест, гневный взгляд, грустное прощание.

Свойство COLOR – эмоциональная окраска (СДП – понятия). Например, белое безмолвие, черная зависть, зеленая тоска.

Свойство HOMO2 – характеристика понятий, связанных с внутренним состоянием человека (СДП – понятия). Например, трезвая мысль, пьяная компания, собранный ум, лихой умысел, почтительное отношение.

Свойство LIVHAR2 – характеристика окружающих нас явлений и событий (СДП – понятия и объекты неживой природы). Например, веселая вечеринка, бешеный бег, добрый совет, злой навет, коварная погода, ласковое море.

Свойство MATERIAL – метафорический перенос (СДП – понятия). Например, бумажная теория, золотая молодежь, каменный век.

Свойство RAZMER – оценка (СДП – понятия). Например, большой ум, гигантская идея, короткая память, микроскопическое время, необъятный замысел.

Свойство TEMPERAMENT – активность чего-либо (СДП – понятия). Например, воинственный дух, вялый подъем, живой процесс.

Свойство VES2 – психофизиологическая оценка (СДП – понятия). Например, легкая грусть, нелегкий труд, равновесное положение, тяжелый состояние.

Свойство VKUS – приятность чего-либо (СДП – понятия). Например, горькая доля, кислый вид, сладкая жизнь.

Свойство VOLUME2 – объем или полнота чего-либо отвлеченного (СДП – понятия). Например, глубокая мысль, мелкий ум, полный упадок, пустая теория.

7. Заключение

Таким образом, анализ рассмотренных выше классов прилагательных позволяет сделать некоторые выводы. Прежде всего, такой подход к классификации прилагательных обеспечивает для каждого из них наличие атрибута, что в значительной степени упрощает фиксацию в концептуальной памяти результата анализа пары прилагательное – существительное. Далее, для прилагательных первого рода упрощается процесс унификации результатов анализа: встретив словосочетания высокий человек или человек высокого роста, мы сразу обнаруживаем их полную идентичность.

Более того, выясняется, что подобная классификация прилагательных предъявляет некоторые требования к классификации существительных. Так, например, объекты, имеющие объем, следует выделить в отдельные классы. Не следует объединять в один класс предметы и другие объекты (или понятия), хотя бы они и относились к одному роду человеческой деятельности или были похожи по каким-либо другим признакам.

Литература

Тузов В.А. Компьютерная обработка текстов в гуманитарных исследованиях // Информационные технологии в гуманитарных и общественных науках. СПб.: СПбЭМИ РАН, 1997. Вып. 6. С.16–23.

Тузов В.А. Компьютерная лингвистика. Опыт построения компьютерных словарей. СПб.: СПбГУ (в печати, 43 печ. л.).

Лезин Г.В., Боярский К.К., Каневский Е.А., Попова А.И. Анализ текстов: представление и обработка концептуальной информации // Труды Международного семинара Диалог'97 по компьютерной лингвистике и ее приложениям. М.: РосНИИ ИИ, 1997. С.170–174.

Агранат Т.Б. Классы русских прилагательных // Труды Международного семинара Диалог'96 по компьютерной лингвистике и ее приложениям. М.: РосНИИ ИИ, 1996. С.14–15.

Семенова С.Ю. Прилагательные в контексте параметрического имени // Труды Международного семинара Диалог'97 по компьютерной лингвистике и ее приложениям. М.: РосНИИ ИИ, 1997. С.238–243.

Апресян Ю.Д. Избранные труды, т. I. Лексическая семантика. М.: Школа "Языки русской культуры", 1995. 472 с.

Филипенко М.В. Русские глаголы в контексте наречий (к вопросу о лингвистическом обеспечении человеко-машинных систем) //НТИ, сер.2. 1992, № 10. С.25–31.

Филипенко М.В. Прилагательные с точки зрения наречий (к вопросу о противопоставлении качественных и относительных прилагательных) //Труды Международного семинара Диалог'98 по компьютерной лингвистике и ее приложениям. Казань: ООО "Хэтер", 1998. С.125–138.

Лингвистический процессор для сложных информационных систем /Ю.Д.Апресян, И.М.Богуславский, Л.Л.Иомдин и др. М.: Наука, 1992. 256 с.

Мельчук И.А. Русский язык в модели "Смысл $\Leftrightarrow$ Текст". Москва – Вена: Школа "Языки русской культуры" – Венский славистический альманах, 1995. 682 с.