

*М. Ф. БОНДАРЕНКО*, д-р техн. наук, *С. И. МАТОРИН*,  
*Е. А. СОЛОВЬЕВА*, канд. техн. наук

## **О ФОРМАЛИЗАЦИИ СЕМАНТИКИ ДЛЯ ПРОЕКТИРОВАНИЯ ПОНЯТИЙНОЙ БАЗЫ ЗНАНИЙ**

---

Цель работы — исследование и формализация временной семантики русского языка для последующего моделирования и применения в понятийной естественной языковой базе знаний. Математические модели могут быть представлены на языке алгебры конечных предикатов [1] и затем реализованы на ЭВМ.

Рассмотрим результаты исследований неразмытого времени. Размытые временные квантификаторы описаны ранее [2]. Основное ограничение на объект моделирования заключается в рассмотрении смысла элементов деловой прозы. Это важное для практики ограничение обосновано [3—6] на базе идей А. П. Ершова о роли деловой прозы. При моделировании не учитываются устаревшие, разговорные, диалектные элементы языка, употребление языковых конструкций в переносном смысле, которые практически не встречаются в деловой прозе, предназначенной для однозначного, буквального понимания. Рассматриваются не только отдельные слова, но и другие элементы языка, выражающие понятия, например, словосочетания. Многие временные понятия обозначены существительными, которые изучены в первую очередь. По результатам нейрофизиологии существительные занимают центральное место в лексиконе человека. По мнению А. Р. Лурии, ребенок выделяет вещественные слова—существительные—раньше других. Известны и другие данные о роли существительных, например, согласно психолингвистике, из речи при ряде заболеваний выпадают предлоги и местоимения, в тяжелых случаях — глаголы, а выпадение существительных свидетельствует об окончательном распаде личности.

Учеными выяснено, что слова с вещественным, предметным значением, т. е. существительные, возникли на первоначальной ступени развития членораздельной речи.

На основе предложенного подхода к построению понятийной базы знаний разработаны и апробированы методы формализации семантики [3—6 и др.], которые отличаются от общепринятых подходов. При формализации смысла определяемого слова следует [3—6] четко ограничивать его от смысла других слов, смысла контекста. Многочисленные толкования значений слов в толковых словарях [7 и др.], грамматиках, лингвистических описаниях зачастую относятся к контексту, но приписываются отдельному слову. Гипотеза об избыточности числа смыслов в лингвистических описаниях проверена на разнообразном лингвистическом материале [3—6 и др.]. Под контекстом в зависимости от рассматриваемых элементов деловой прозы понимаются словосочетание, предложение, текст, слово или его части.

Рассмотрим для примера ряд значений существительного *время* [7, с. 85]. Значение<sub>7</sub> «подходящая, удобная, благоприятная пора» («не время сидеть сложа руки») и значение<sub>8</sub> — «до-суг» («нет времени для прогулок») не отличаются от значения<sub>9</sub> «промежуток, ... в который совершается что-нибудь» («время работает на нас», «время покажет»), если не приписывать слову *время* смысла контекста-смысла других слов.

Проанализируем отдельные результаты исследований и формализации временной семантики на базе Словаря русского языка [7] и некоторых других источников. Рассмотрим сначала группу временных понятий — единиц измерения времени: *секунда, минута, час, сутки, неделя, декада, месяц, квартал, полугодие, год, век, столетие* и некоторые др. Большая часть из них относится [7] к единицам исчисления времени («исчисление» — книжный термин [7], поэтому будем применять слово «измерение», которое также годится в данном случае), например, *секунда, час, декада* и т. д.; но некоторые — *год, сутки* — трактуются как «промежутки времени». Это не соответствует действительности, так как именно *год* и *сутки* приняты в качестве основных единиц измерения времени (см., например, Малую Советскую Энциклопедию), потому что устанавливаются на базе естественных периодических процессов, которые повторяются без изменения периода и определяются вращением Земли вокруг Солнца и вокруг своей оси соответственно. Странно, что именно основные единицы измерения названы в словаре [7] промежутками. В Международной системе единиц СИ основной единицей измерения времени принята *секунда*. Не останавливаясь детально на этих вопросах, что может понадобиться при моделировании конкретной предметной области, связанной с точным измерением времени, будем считать *год, сутки* и *секунду* — основными единицами измерения времени. Дополнительными единицами измерения времени будем считать временные промежутки, имеющие конечную

определенную длительность, соотносимые с другими единицами измерения, используемые как единицы измерения времени.

Рассмотрим единицы, которым приписывается единственное значение [7]: *неделя, сутки, полугодие*. Неделя трактуется [7, с. 324] как «единица исчисления времени, равная семи дням, от понедельника до воскресенья включительно», *сутки* — как «промежуток времени, равный 24 часам, продолжительности дня и ночи» [7, с. 639], *полугодие* — «промежуток времени в полгода» [7, с. 451]. Такое понимание в принципе соответствует деловой прозе с учетом некоторых оговорок, например, *сутки* — основная единица измерения времени.

Проанализируем единицы измерения времени, которые трактуются многозначно. Значение<sub>1</sub> слова *квартал* — временное: «четвертая часть отчетного года» [7, с. 221], значение<sub>2</sub> — нет, но оно близко значению<sub>1</sub> по смыслу, по крайней мере этимологически. Определения [7] остальных единиц измерения отличаются неформальностью и смешиванием смыслов определяемого слова и контекста.

Рассмотрим слово *час*, которое в Словаре русского языка [7, с. 716] имеет пять толкований. Охарактеризуем каждое из них с точки зрения формального описания для модели временной семантики.

1. «Промежуток времени в 60 минут». Такое определение смысла слова *час* соответствует специфике деловой прозы, например, *опоздать на час, на два часа* [7], хотя толкование другого примера в этом же пункте (расти не по дням, а по часам — очень быстро [7]) к деловой прозе не относится.

2. «Срок в 60 минут, исчисляемый от полудня или полуночи». Например, *час ночи, два часа ночи, час дня*. Во-первых, подобные выражения в деловой прозе имеют вид 1.00, 14.00, 3.15 и т. п. Во-вторых, значение<sub>2</sub> эквивалентно значению<sub>1</sub>, так как в обоих случаях говорится о промежутке, а промежуток характеризуется определенной длительностью и имеет две границы — нижнюю и верхнюю. Следовательно, значение<sub>2</sub> полностью подходит под значение<sub>1</sub> и является его частным случаем, при котором левая граница промежутка фиксируется на времени 0 часов или 12 часов. Промежуток, о котором идет речь в приведенных примерах, занимает не одно, а множество мест на временной оси, а его конкретное место, привязка к «абсолютному» времени, возможно, определится из контекста. Тот факт, что слово *час* обозначает класс промежутков, соответствует определению смысла [3—6]. Следует отметить, что смысл словосочетаний *час ночи, два часа дня* — множество точек временной оси (в конкретном контексте, может быть, единственная точка), отстоящих на определенный промежуток (указанное число часов) от точек отсчета — *полуночи* или *полуночи*. Рассматриваемые точки представляют собой верхние границы соответствующих промежутков.

3. «Промежуток времени, отводимый на урок, лекцию, занимающий с отдыхом 60 минут». Справедливости ради отметим,

что в нашем вузе такой час длится 45 мин, а вместе с отдыхом — 50 или 55, хотя суть дела не в этом. Очевидно, что в примере типа *оплата по часам*, если он взят из деловой прозы, должно быть ясно, о каком *часе* идет речь: во-первых, это может быть *час* в смысле<sub>1</sub>, с другой стороны — эллиптическая конструкция, например, от словосочетания *академический час*. Во втором случае слово *час* является частью единого понятия, которое должно быть видом того же рода, что и *час*.

4. «Пора, время (высок.)». Например, грозный час войны. Как справедливо отмечено в [7], в этой и аналогичных конструкциях слово *час* используется в «высоком» смысле, т. е. является литературным, художественным приемом, который в деловой прозе не встречается, следовательно, эту трактовку не следует учитывать в модели понимания деловых текстов.

5. «Время, предназначенное для чего-нибудь». Например, *часы занятий, приемные часы*. Подобные (обычно эллиптические) конструкции, подразумевающие «время, предназначенное для чего-нибудь», и длительность этого «чего-нибудь» в течение нескольких часов, а не минут или дней, т. е. «часы, предназначенные для...». Значит, первоначально в таких случаях под *часом*, видимо, понимался промежуток времени в 60 мин, т. е. *час* в смысле<sub>1</sub>, этот смысл не изменился, но появилось новое понятие, выраженное словосочетанием. Подобные конструкции могут быть учтены в модели как выражающие отдельное понятие определенной предметной области.

На основании проведенного разбора можно считать, что в деловой прозе *час* имеет единственный смысл<sub>1</sub>, за исключением использования этого слова в словосочетаниях, выражающих другие понятия, которые следует учитывать при моделировании конкретной предметной области.

Рассмотрим временной квантификатор *год*, определяемый в [7, с. 111] также многозначно.

1. «Промежуток времени, равный периоду обращения Земли вокруг Солнца — 12 календарным месяцам». Например, *три года*. Такое толкование соответствует пониманию *года* в деловой прозе.

2. «Промежуток времени, в который завершается цикл каких-нибудь работ, занятий». Например, учебный год, сельскохозяйственный год и пр. Приведенные и другие немногочисленные примеры употребления *года* в смысле<sub>1</sub> в составе словосочетаний, выражающих единое понятие, могут быть учтены в модели.

3. «Период времени в некоторое количество лет». Например, *героические годы гражданской войны, детские годы*. Это толкование использует множественное число от слова *год*, которое также можно рассматривать в смысле<sub>1</sub>. Нетребовательность к точности отсчета времени в подобных случаях характерна только для деловой прозы, поэтому нет необходимости выделять отдельное толкование *года*, отличное от смысла<sub>1</sub>.

4. «Мн. в сочетании с порядковым числительным — промежуток времени в пределах десятилетия». Например, *40-е годы*. В данном толковании наглядно видно распространение смысла контекста на смысл слова, так как длительность временного интервала в «десятилетие» получается после учета смыслов числительного и *года*. Фактически *40-е годы* заменяются интервалом 1940—1949 гг. Само слово *год* по-прежнему может рассматриваться в смысле<sub>1</sub>, тем более, что для деловой прозы характерна четкость выражения смысла.

5. «Возраст». Например, *он уже в годах*. Подобные выражения в деловой прозе не встречаются, хотя и они могут быть проанализированы с применением толкования<sub>1</sub>.

Обобщим аналогичные исследования смысла временных единиц *секунда, минута, декада, месяц, век*. Значение<sub>1</sub> этих слов [7] соответствует, как и значение<sub>1</sub> *часа* и *года*, деловой прозе. Остальные одно-два толкования [7] можно обобщенно охарактеризовать так.

1. Смыслы, которые не используются в деловых текстах. Например, трактовки: *минута* — «короткий промежуток времени, мгновение» (придет с минуты на минуту); *век* — «жизнь, существование» (много повидал на своем веку, век живи, век учись) и *месяц* — «диск луны или его часть».

2. Смыслы, относящиеся к узкой области знания и которые следует учесть при ее моделировании. Сюда мы отнесли употребление *секунды* и *минуты* в качестве специальных математических терминов — единиц измерения углов.

3. Смыслы, которые характеризуют другие понятия, выраженные словосочетаниями, например, *каменный век*. Приведенный пример, кроме того, можно отнести и к специальным терминам.

4. Смыслы, которые характеризуют контекст, но приписаны слову. Например, *декада* — «промежуток времени, посвященный чему-нибудь» (декада грузинского искусства). «Посвященный чему-нибудь» здесь обозначается не *декадой*, а остальными словами.

Результаты проведенных исследований применены при моделировании семантики единиц измерения времени.

Рассмотрим лексическую группу *день, ночь, утро, вечер*. В толковании этих слов в [7] имеется ряд противоречий. Так, временной смысл *утра* — «начало дня» [7, с. 689], смысл<sub>1</sub> *дня* [7, с. 130] — «часть суток от утра до вечера, т. е. исходя из первого определения, — *утро* — часть *дня*, а следуя второму — *утро* — часть суток, не входящая в *день*, а предшествующая ему».

Из определений *вечера* и *ночи* ясно, что это части суток, следующие друг за другом: *ночь* — «часть суток от вечера до утра» [7, с. 338], *вечер* — «часть суток перед наступлением ночи, следующая после окончания дня» [7, с. 64]. Смысл<sub>2</sub> *вечера* и *утра* и смысл<sub>4</sub> *дня* [7] относятся к словосочетаниям, например: *танцевальный вечер, литературное утро, день рождения* и являются примерами перенесения смысла контекста на смысл слова.

Смысл<sub>2</sub> дня — «сутки», применяется в деловой прозе, и является, на наш взгляд, основным смыслом, выражаемым этим словом. В неделевой прозе день употребляется также в смысле<sub>1</sub> и в смысле<sub>3</sub> [7] — «промежуток времени, занятый какой-нибудь деятельностью», например рабочий день, если не забывать, что смысл<sub>3</sub> выражается не одним словом, а словосочетанием. Следует учесть, что даже в смысле<sub>1</sub> («часть суток») день входит в сутки один раз, т. е. число дней всегда совпадает с числом суток.

Смысл<sub>5</sub> дня — «время, период» (дни юности) не употребляется в деловой прозе, но в принципе может быть сведен к смыслу<sub>2</sub>.

Понимание дня как светлой части суток между восходом и заходом Солнца и аналогичное понимание ночи учитывает полярный день (ночь), длительность которого может быть больше 24 ч, а на полюсах день длится полгода (см.: Большую Советскую Энциклопедию). Учесть этот момент при работе с соответствующей предметной областью можно, рассматривая понятия, выраженные словосочетаниями.

В результате проведенных исследований сделан вывод, что понятие утро, вечер и ночь не применяются в деловой прозе. Эти слова не вошли, например, в Малую Советскую Энциклопедию. При необходимости можно учесть их как временные интервалы определенной длительности (например, по 4 ч или 8 ч) или рассмотреть понятия, которые обозначены словосочетаниями, включающими одно из этих слов. День в деловой прозе применяется в смысле<sub>2</sub> «сутки» или как эллипсис от рабочий день и т. п.

Число таких дней будет соответствовать числу прошедших суток, следовательно, можно рассматривать день в смысле<sub>2</sub>, при необходимости учитывая словосочетания с этим словом как отдельные понятия. При необходимости, учитывая привычность понимания дня как части суток, можно рассматривать день в таком понимании.

Проанализируем такие однотипные временные понятия, как месяцы года (сентябрь — девятый месяц года [7, с. 581]), дни недели (вторник — второй день недели [7, с. 88]) и времена года. Определения времен года (лето — самое теплое время года, следующее за весной) [7, с. 261]) не будет соответствовать, например, пониманию лета в южном полушарии, но главное, что эти определения не соответствуют деловой прозе. Если времена года понадобится учесть в модели, то естественно приписать им общепринятое понимание, связанное с их длительностью в месяцах и конкретными границами.

В отличие от единиц измерения, конкретные месяцы года, дни недели, времена года можно привязать к шкале с постоянной точкой отсчета, которую принято называть абсолютной. На такой шкале каждый из рассмотренных промежутков времени, например сентябрь, вторник, лето, изобразится любым элементом множества периодически повторяющихся промежутков соответ-

ствующей длительности. О каком именно из этих элементов идет речь, в большинстве случаев становится ясно из контекста.

Аналогичным образом можно понимать *дату* и *годовщину*. Рассмотрим эти слова подробнее. *Дата* трактуется [7, с. 123] следующим образом.

1. «Календарное время какого-нибудь события. Исторические даты».

2. «Помета, указывающая время (год, месяц, число) написания чего-нибудь...».

Эти определения аналогичны, если не учитывать, что второе характеризует не саму *дату*, как время, а ее обозначения.

Каждая конкретная *дата* (в контексте) обозначает конкретные *сутки*, т. е. конкретный отрезок временной оси такой длительности, а *дата* вообще — любой отрезок абсолютной временной шкалы с границами, соответствующими границам суток и длительностью в сутки (любые сутки). Поэтому родом *даты* будем считать *сутки*. Родом *годовщины* указывается [7, с. 111] «календарная дата, отмечающая, что со времени какого-нибудь события прошел очередной год». *Годовщину* можно представить как *дату*, но не любую, а отстоящую от некоторой даты, как точки отсчета на некоторое число лет.

Рассмотрим слово *сезон*, которому приписывается два значения [7, с. 578].

1. «Одно из четырех времен года. Летний сезон... Моды сезона».

2. «Часть года, характеризующаяся какими-нибудь явлениями природы или подходящая для какой-нибудь деятельности, работы ...театральный сезон. Купальный сезон».

Первое определение не годится для деловой прозы, если *время года* (см. выше *лето*) понимать так, как в [7]. Здесь, видимо, в противоречии с определениями *времен года* [7], *время года* понимается четко, но в приведенных примерах это не является необходимостью. В примерах определения<sub>2</sub> также видно перенесение смысла контекста на смысл слова. *Сезон* можно определить как промежуток времени, длительность которого не превышает год. Конкретная длительность словосочетания со словом *сезон* обычно определяется из контекста.

Ряд определений из [7] соответствует родовидовым отношениям между понятиями. Так, *перерыв* определяется как промежуток, на время которого прекращается какая-нибудь деятельность [7, с. 412], а *перемена*, *пауза* и *антракт* определяются через *перерыв*. Отметим, что значение<sub>2</sub> *антракта*, как музыкального произведения, относится только к конкретной предметной области, а первые три смысла *перемены* необходимо рассмотреть детально, тем более, что *перемена*, как перерыв между уроками [7, с. 409], видимо, произошла от первых значений.

Известны временные существительные, которые обозначают не промежутки времени, а точки временной оси, например *полдень*. В деловой прозе используется *срок*, который можно трак-

товать и как промежуток времени и как временную точку. Значения же *срока* в [7, с. 621] определяются через *промежуток* и *момент*, который, в свою очередь, определяется через *промежуток*.

Проведенные исследования по формализации понятий временной семантики русского языка, примеры которых описаны выше, являются составной частью исследований, необходимых для моделирования системы понятий деловой прозы [5 и др.]. Существование, необходимость моделирования, структура и форма представления системы понятий обоснованы в системном бионическом подходе к моделированию понимания, разработанном [4—7 и др.] на основе системологии [8]. Разработанные методы [4—7 и др.] позволяют определить достаточность исследований смысла. Модель системы понятий является основой понятийной естественно-языковой базы знаний [5, 6 и др.].

При описании исследований использованы термины «значение», «смысл» и некоторые другие в одном и том же общепринятом смысле. Уточнение их определений как объектов моделирования дано в работах [3—6 и др.]. Формализованные описания смысла могут быть представлены в виде графовых структур, таблиц смысловых признаков и их значений (с большой избыточностью), совокупности различных классификаций. Нетрудно показать, что родовидовая классификация (даже без учета видовых различий) является наиболее общей, включающей любые понятия. В классификацию типа «часть—целое» временной лексики войдут только те слова, которые имеют количественные характеристики времени и могут сравниваться между собой по длительности обозначаемых временных промежутков. Ситуативные классификации, которые часто применяются в системах искусственного интеллекта, не являются порождением логики, словесного мышления, а опираются, как известно, на наглядный опыт человека.

Математические модели, полученные на основе проведенной формализации, реализованы на ЭВМ, исследованы и могут применяться на практике после сравнительно быстрой настройки на конкретную предметную область.

**Список литературы:** 1. Шабанов-Кушнаренок Ю. П. Теория интеллекта. Математические средства. Х., 1984. 144 с. 2. Бондаренко М. Ф., Маторин С. И., Соловьева Е. А. О моделировании времени в некоторых ситуациях с неполной информацией//Пробл. бионики. 1988. Вып. 42. С. 3—6. 3. Соловьева Е. А., Маторин С. И. О смысле слов как объекте моделирования//Уч. зап. Тартуского ун-та. Тр. по искусств. интеллекту. Тарту, 1987. Вып. 751. С. 150—160. 4. Соловьева Е. А., Маторин С. И. Об одном подходе к моделированию понятийной структуры базы знаний. К., 1987. 13 с. Деп. в УкрНИИНТИ 05.05.87. № 1374. 5. Соловьева Е. А., Маторин С. И. О построении базы знаний для модели понимания деловой прозы. К., 1987. 27 с. Деп. в УкрНИИНТИ 01.03.88. № 601. 6. Соловьева Е. А., Маторин С. И. О моделировании системы понятий деловой прозы//Базы данных и знаний в автоматизированных региональных системах. К., 1988. С. 25—28. 7. Ожегов С. И. Словарь русского языка. М., 1987. 749 с. 8. Мельников Г. П. Системология и языковые аспекты кибернетики. М., 1978. 368 с.

Поступила в редколлегия 28.03.88