

О ФОРМАЛИЗАЦИИ СЕМАНТИКИ СЛОВСОЧЕТАНИЙ С ИНСТРУМЕНТАЛЬНЫМ ЗНАЧЕНИЕМ

Машинная обработка информации на естественном языке требует анализа семантики сообщения. Для этого необходимо определить значение единиц сообщения. Вопросы о взаимодействии значений слов между собой, а также со знаниями о внешнем мире при определении значения сообщения, его элементов, в частности, словосочетаний минимальной длины исследованы недостаточно. В данной работе рассматриваются некоторые аспекты этой проблемы на примере двухсловных сочетаний с инструментальным значением.

Класс инструментальности включает в себя существительные, выражающие как инструмент, так и средство действия. Однако имеется четкое различие между этими двумя подгруппами существительных одного класса. Использование некоторого средства приводит к его расходованию, в то время как применение инструмента оставляет его в первоначальном состоянии [1]. Здесь рассмотрим только словосочетания со значением «действие + инструмент действия». Существительные-инструменты могут заполнять валентность инструментальности при глаголе разнообразными способами. Поэтому словосочетания «действие + инструмент действия» могут иметь такую структуру [2]: $\Gamma + C_{им.}$; $\Gamma + из + C_{род.}$; $\Gamma + с + C_{род.}$; $\Gamma + в + C_{вин.}$; $\Gamma + на + C_{вин.}$; $\Gamma + C_{тв.}$; $\Gamma + на + C_{пр.}$, где Γ — глагол конкретного действия или акциональный глагол; $C_i (i = \{им., род., вин., тв., пр.\})$ — существительное, выражающее инструмент действия, в соответствующем падеже.

Обычно значения слов классифицируют по некоторому набору априорных признаков. В результате такой классификации, например, среди глаголов выделяются группы глаголов движения, бытия, трудовой деятельности, сознания и др. [3]. Такая группировка может быть полезна в некоторых системах обработки информации на естественном языке. Однако она не позволяет отразить реальную сложность взаимосвязей в лексической системе языка. При этом нельзя рассчитывать на полноту и однозначность такого разбиения.

Вместе с тем из всей совокупности глаголов русского языка класс глаголов трудовой деятельности выделяется довольно четко. Существительные со значением «инструмент действия» чаще всего сочетаются с глаголами лексико-семантической группы трудовой деятельности. Остановимся на анализе таких словосочетаний.

Класс «инструмент действия» может иметь более частные варианты: отчуждаемый инструмент (рубить топором, резать ножницами), неотчуждаемый инструмент (хлопнуть ладонью, ударить

кулаком), окказиональный инструмент (копать ложкой, забивать табуреткой) [4]. Для рассмотрения у глаголов валентности инструментальности особый интерес представляет первый вариант, так как лишь существительные этой группы могут быть охарактеризованы как «орудия, предназначенные для воздействия на данный объект». Существительные этой лексико-семантической группы тесно связаны по значению с глаголом, а идея действия — с представлением о любом орудии, инструменте [5].

Предметы и их свойства неразрывно связаны между собой. Если предметы проявляют себя через свои свойства, то свойства, в свою очередь, наблюдаются у предметов, благодаря которым они становятся явными. В частности, действие определяется по орудью, с помощью которого оно производится, и, наоборот, орудие может быть определено по действию, для совершения которого оно предназначено [6]. Поэтому большинство глаголов трудовой деятельности имеют свои, более или менее жестко прикрепленные к ним, инструменты. Например, копать лопатой, пилить пилой, забивать молотком. Но это не означает, что указанные действия нельзя выполнять при помощи другого инструмента. Более того, существуют глаголы для выражения таких действий, выполнять которое можно различным инструментами. Например: резать ножом (дерево), резать ножницами (бумагу), резать проволокой (масло). Применение того или иного инструмента в этом случае зависит от объекта действия. Можно привести также примеры, когда использование определенного инструмента при выполнении действия не зависит ни от объекта действия, ни от некоторых других факторов: брить бритвой, брить ножом.

Для исследования семантики словосочетаний указанного типа будем представлять значения входящих в него слов с помощью некоторого формального аппарата. Наиболее наглядным способом представления значений языковых единиц являются семантические сети и связанный с ними математический аппарат. Семантическая сеть для каждого слова должна содержать признаки, существенные для некоторого класса слов.

Рассмотрим построение семантической сети словосочетания как средства формализации семантики двухсловного сочетания «колоть иглой». Отметим, что глагол «колоть» многозначен и может принимать следующие значения: «раздроблять, рассекать, делить на куски» и «касаться чем-нибудь острым, причиняя боль» [7]. В данном примере принимается во внимание лишь второе значение анализируемого глагола.

Семантические сети для каждой компоненты словосочетания представлены в виде графов G_1 и G_2 на рис. 1, 2 соответственно. Вершины графов имеют следующие значения (каждая цифра соответствует номеру вершины): 1 — колоть, 2 — предмет; 3 — небольшой для субъекта действия, 4 — твердый, 5 — острый конец, 6 — блестящий, 7 — игла.

Вес каждой дуги $z_i (i=1, 5)$ учитывает тип связи между вершинами: z_1 — иметь инструментом, z_2 — иметь размер, z_3 — иметь признак, z_4 — иметь частью, z_5 — быть элементом класса.

Таким образом, графы G_1 и G_2 содержат наборы признаков, присутствующих глаголу «колоть» и существительному-инструменту «игла» соответственно. Безусловно, граф G_2 мог бы иметь другой вид,

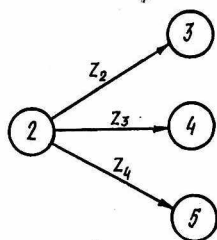
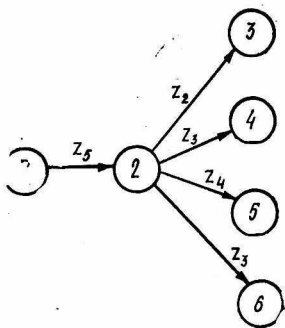
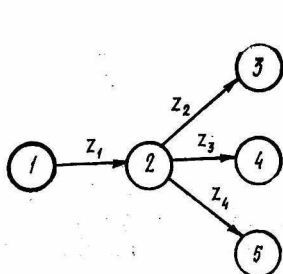


Рис. 1. Граф G_1

Рис. 2. Граф G_2

Рис. 3. Граф G_3

так как анализируемое существительное обладает еще рядом признаков. Однако эти признаки опущены, поскольку они не являются существенными для выполнения действия «колоть». При построении семантической сети глагола основная задача заключалась в выделении совокупности признаков, которыми должен обладать инструмент для совершения данного действия.

Семантическую сеть, передающую значение всего словосочетания «колоть иглой», можно интерпретировать как результат некоторых операций над графами G_1 и G_2 , соответствующих словоформам. На первом шаге строится граф G_3 пересечения графов G_1 и G_2 .

Граф G_1 задан множеством вершин $V(G_1)$ и множеством дуг $E(G_1)$: $V(G_1) = \{1, 2, 3, 4, 5\}$; $E(G_1) = \{(1z_1 2), (2z_2 3), (2z_3 4), (2z_4 5)\}$. Граф G_2 имеет: $V(G_2) = \{2, 3, 4, 5, 6, 7\}$; $E(G_2) = \{(7z_5 2), (2z_2 3), (2z_3 4), (2z_4 5), (2z_3 6)\}$.

Тогда граф G_3 будет иметь следующие множества вершин и дуг: $V(G_3) = V(G_1 \cap G_2) = V(G_1) \cap V(G_2) = \{2, 3, 4, 5\}$; $E(G_3) = E(G_1 \cap G_2) = E(G_1) \cap E(G_2) = \{(2z_2 3), (2z_3 4), (2z_4 5)\}$.

Этот граф изображен на рис. 3. Заметим, что построенный таким образом граф пересечения может быть несвязным, т. е. иметь изолированные вершины. Такие вершины отбрасываются и в выполнении последующих действий не участвуют.

На языке семантики G_3 трактуется как графическое представление множества признаков, общих для компонент исходного соче-

тания слов. Однако располагая лишь таким графом, нельзя проследить связь данной совокупности признаков непосредственно с каждым словом словосочетания. Для слияния значений двух слов и образования осмысленного словосочетания необходимо восстановить пути на графах, берущие начало в G_1 и G_2 и заходящие в G_3 . После присоединения полученных последовательностей вершин и дуг к графу G_3 получаем графы G_4 и G_5 соответственно. При этом.

$$V(G_4) = \{1, 2, 3, 4, 5\}; E(G_4) = \{(1z_1 2), (2z_2 3), (2z_3 4), (2z_4 5)\};$$

$$V(G_5) = \{2, 3, 4, 5, 7\}; E(G_5) = \{(7z_5 2), (2z_2 3), (2z_3 4), (2z_4 5)\}.$$

Объединенно-пересеченный граф G_6 , заданный множеством вершин $V(G_6)$ и множеством дуг $E(G_6)$, отражающий смысл исходного словосочетания «колоть иглой», строится путем объединения графов G_4 и G_5 :

$$V(G_6) = V(G_4 \cup G_5) = V(G_4) \cup V(G_5) = \{1, 2, 3, 4, 5, 7\}; E(G_6) = E(G_4 \cup G_5) = E(G_4) \cup E(G_5) = \{(1z_1 2), (2z_2 3), (2z_3 4), (2z_4 5), (7z_5 2)\}.$$

Результирующий граф изображен на рис. 4.

Применение семантических сетей для представления значений слов и словосочетаний требует создания специальных процедур для решения каждой отдельной задачи на семантических сетях,

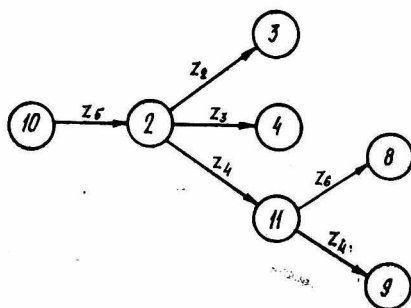
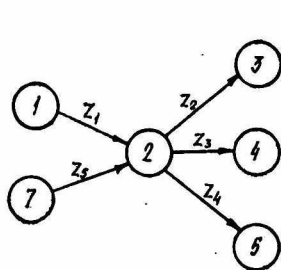


Рис. 4. Результирующий граф

Рис. 5. Граф G'_2

таких, как определение смежности или достижимости вершин в сети, нахождение всех путей, связывающих вершины, и т. п. Однако это не всегда удобно для использования в машинных системах обработки естественного языка универсального назначения. Желательно использовать такой формальный аппарат для представления семантики, который позволил бы решать все эти задачи еди-

ным способом. В качестве такого формального аппарата удобно использовать алгебру конечных предикатов [8].

Представим полученные результаты в терминах алгебры конечных предикатов. Пусть X — множество всех вершин графов G_1 и G_2 , из которых выходит хотя бы одна дуга, Y — множество всех вершин этих графов, в которые входит хотя бы одна дуга, и Z — множество весов дуг этих графов. Для данного примера эти три множества содержат следующие элементы: $x \in X = \{1, 2, 7\}$; $y \in Y = \{2, 3, 4, 5, 6\}$; $z \in Z = \{1, 2, 3, 4, 5\}$.

Каждому из графов G_i , полученных в результате выполнения описанных действий, поставим в соответствие предикат

$$P_i(x, y, z) \quad (i = \overline{1, 6}). \quad \text{Тогда } P_1(x, y, z) = x^1 y^2 z^1 \vee x^2 y^3 z^2 \vee x^2 y^4 z^3 \vee x^2 y^5 z^4; \\ P_2(x, y, z) = x^7 y^2 z^5 \vee x^2 y^3 z^2 \vee x^2 y^4 z^3 \vee x^2 y^5 z^4 \vee x^2 y^6 z^3.$$

Пересечение графов G_1 и G_2 в терминах алгебры конечных предикатов соответствует конъюнкции предикатов $P_1(x, y, z)$ и $P_2(x, y, z)$:

$$P_3(x, y, z) = P_1(x, y, z) \cdot P_2(x, y, z) = (x^1 y^2 z^1 \vee x^2 y^3 z^2 \vee x^2 y^4 z^3 \vee x^2 y^5 z^4) \cdot (x^7 y^2 z^5 \vee x^2 y^3 z^2 \vee x^2 y^4 z^3 \vee x^2 y^5 z^4 \vee x^2 y^6 z^3) = \\ = x^1 y^2 z^1 x^7 y^2 z^5 \vee x^2 y^3 z^2 x^7 y^2 z^5 \vee x^2 y^4 z^3 x^7 y^2 z^5 \vee x^2 y^5 z^4 x^7 y^2 z^5 \vee \\ \vee x^1 y^2 z^1 x^2 y^3 z^2 \vee x^2 y^3 z^2 x^2 y^3 z^2 \vee x^2 y^4 z^3 x^2 y^3 z^2 \vee x^2 y^5 z^4 x^2 y^3 z^2 \vee \\ \vee x^1 y^2 z^1 x^2 y^4 z^3 \vee x^2 y^3 z^2 x^2 y^4 z^3 \vee x^2 y^4 z^3 x^2 y^4 z^3 \vee x^2 y^5 z^4 x^2 y^4 z^3 \vee \\ \vee x^1 y^2 z^1 x^2 y^5 z^4 \vee x^2 y^3 z^2 x^2 y^5 z^4 \vee x^2 y^4 z^3 x^2 y^5 z^4 \vee x^2 y^5 z^4 x^2 y^5 z^4 \vee \\ \vee x^1 y^2 z^1 x^2 y^6 z^3 \vee x^2 y^3 z^2 x^2 y^6 z^3 \vee x^2 y^4 z^3 x^2 y^6 z^3 \vee x^2 y^5 z^4 x^2 y^6 z^3.$$

Используя тождества алгебры конечных предикатов, получаем

$$P_3(x, y, z) = x^2 y^3 z^2 \vee x^2 y^4 z^3 \vee x^2 y^5 z^4. \quad (1)$$

Построим предикаты $P_4(x, y, z)$ и $P_5(x, y, z)$ для описания графов G_4 и G_5 соответственно:

$$P_4(x, y, z) = x^1 y^2 z^1 \vee x^2 y^3 z^2 \vee x^2 y^4 z^3 \vee x^2 y^5 z^4; \quad (2)$$

$$P_5(x, y, z) = x^7 y^2 z^5 \vee x^2 y^3 z^2 \vee x^2 y^4 z^3 \vee x^2 y^5 z^4. \quad (3)$$

С использованием формулы (1) равенства (2) и (3) упрощаются:

$$P_4(x, y, z) = x^1 y^2 z^1 \vee P_3(x, y, z); \quad (4)$$

$$P_5(x, y, z) = x^7 y^2 z^5 \vee P_3(x, y, z). \quad (5)$$

Образовав дизъюнкцию предикатов $P_4(x, y, z)$ и $P_5(x, y, z)$, получим предикат $P_6(x, y, z)$, который описывает объединенно-пересеченный граф G_6 , отражающий смысл словосочетания «колоть иглой»:

$$P_6(x, y, z) = P_4(x, y, z) \vee P_5(x, y, z). \quad (6)$$

В результате подстановки в (6) значений $P_4(x, y, z)$ и $P_5(x, y, z)$ из (4) и (5) получаем $P_6(x, y, z) = x^1 y^2 z^1 \vee x^7 y^2 z^5 \vee P_3(x, y, z)$.

Если построение графа G_6 и соответствующего ему предиката $P_6(x, y, z)$ в принципе возможно, то это означает, что рассматриваемая комбинация слов образует осмысленное словосочетание. Однако при этом необходимо учитывать, что семантические сети слов были построены так, что все признаки глагола присутствовали в результирующей семантической сети исходного словосочетания. В то же время у существительного-инструмента оказался «лишним» один признак, наличие которого не обязательно для совершения действия «колоть» (острием). Вообще говоря, у существительного таких признаков может быть сколь угодно много. Но если хотя бы один из признаков глагола окажется «лишним», это может привести к допущению сочетаний слов, лишенных смысла в обычных ситуациях. Примером может служить словосочетание «колоть молотком».

Построим семантическую сеть для существительного-инструмента «молоток». Воспользуемся графом G_2 , введя при этом недостающие вершины и признаки: 8 — железный, 9 — плоская поверхность, 10 — молоток, 11 — наконечник, z_6 — состоять из материала.

Граф G'_2 , построенный с учетом принятых обозначений, изображен на рис. 5. Соответствующий этому графу предикат $P'_2(x, y, z) = x^{10} y^2 z^5 \vee x^2 y^3 z^2 \vee x^2 y^4 z^3 \vee x^2 y^{11} z^4 \vee x^{11} y^8 z^6 \vee x^{11} y^9 z^4$.

Согласно изложенной процедуре получения предиката результирующего графа словосочетания на первом шаге необходимо образовать конъюнкцию предикатов $P_1(x, y, z)$ и $P'_2(x, y, z)$:

$$P'_3(x, y, z) = P_1(x, y, z) \cdot P'_2(x, y, z) = (x^1 y^2 z^1 \vee x^2 y^3 z^2 \vee x^2 y^4 z^3 \vee x^2 y^5 z^4) \cdot (x^{10} y^2 z^5 \vee x^2 y^3 z^2 \vee x^2 y^4 z^3 \vee x^2 y^{11} z^4 \vee x^{11} y^8 z^6 \vee x^{11} y^9 z^4) = x^1 y^2 z^1 x^2 y^3 z^2 \vee x^1 y^2 z^1 x^{11} y^8 z^6 \vee x^1 y^2 z^1 x^{11} y^9 z^4 \vee x^1 y^2 z^1 x^{10} y^2 z^5 \vee x^1 y^2 z^1 x^2 y^3 z^2 \vee x^1 y^2 z^1 x^2 y^4 z^3 \vee x^2 y^3 z^2 x^{10} y^2 z^5 \vee x^2 y^3 z^2 x^2 y^3 z^2 \vee x^2 y^3 z^2 x^2 y^4 z^3 \vee x^2 y^3 z^2 x^2 y^{11} z^4 \vee x^2 y^3 z^2 x^{11} y^8 z^6 \vee x^2 y^3 z^2 x^{11} y^9 z^4 \vee x^2 y^4 z^3 x^{10} y^2 z^5 \vee x^2 y^4 z^3 x^2 y^3 z^2 \vee x^2 y^4 z^3 x^2 y^4 z^3 \vee x^2 y^4 z^3 x^2 y^{11} z^4 \vee x^2 y^4 z^3 x^{11} y^8 z^6 \vee x^2 y^4 z^3 x^{11} y^9 z^4 \vee x^2 y^5 z^4 x^{10} y^2 z^5 \vee x^2 y^5 z^4 x^2 y^3 z^2 \vee x^2 y^5 z^4 x^2 y^4 z^3 \vee x^2 y^5 z^4 x^2 y^{11} z^4 \vee x^2 y^5 z^4 x^{11} y^8 z^6 \vee x^2 y^5 z^4 x^{11} y^9 z^4 = x^2 y^3 z^2 \vee x^2 y^4 z^3.$$

Описанная процедура на втором этапе предполагает построение предикатов $P'_4(x, y, z)$ и $P'_5(x, y, z)$, соответствующих графам G'_4 и G'_5 , аналогичным G_4 и G_5 , из предыдущего примера:

$$P'_4(x, y, z) = x^1 y^2 z^1 \vee x^2 y^3 z^2 \vee x^2 y^4 z^3 = x^1 y^2 z^1 \vee P'_3(x, y, z);$$

$$P'_5(x, y, z) = x^{10} y^2 z^5 \vee x^2 y^3 z^2 \vee x^2 y^4 z^3 = x^{10} y^2 z^5 \vee P'_3(x, y, z).$$

Получим описание объединенно-пересеченного графа словосочетания «колоть молотком» в терминах алгебры конечных предикатов. Воспользуемся для этого формулой (6):

$$P'_6(x, y, z) = P'_4(x, y, z) \vee P'_5(x, y, z) = x^1 y^2 z^1 \vee \vee P'_3(x, y, z) \vee x^{10} y^2 z^5 \vee P'_3(x, y, z).$$

Применим тождества алгебры конечных предикатов. Тогда

$$P'_6(x, y, z) = x^1 y^2 z^1 \vee x^{10} y^2 z^5 \vee P'_3(x, y, z). \quad (7)$$

Формула (7) — формальное представление семантики словосочетания «колоть» молотком. Если не учитывать условие обязательного наличия всех признаков глагола в результирующем графе, то можно прийти к ошибочному выводу, что исходное словосочетание семантически верно. Учет указанного условия ставит все на свои места. Следовательно, последним шагом в процедуре выяснения осмысленности комбинации слов должна быть проверка наличия всех признаков глагола в объединенно-пересеченном графе. Другими словами, должно быть обязательно выполнено следующее условие: граф G_1 есть подграф G_6 , или в терминах алгебры конечных предикатов

$$P_1(x, y, z) \supset P_6(x, y, z) = 1. \quad (8)$$

Учет (8) не позволит пропустить как семантически правильные словосочетания, лишенные смысла.

Отметим, что это правило действует лишь в случае, когда глагол однозначен или заранее оговорено, в каком значении употреблен многозначный глагол. В частности, в рассмотренных примерах пришлось ограничить смысл многозначного глагола «колоть» лишь одним из его значений. Если же учесть значение, которое раскрывается в словосочетаниях «колоть орехи клещами», «колоть дрова топором», то семантическая сеть для этого глагола будет иметь более разветвленный вид. Это даст возможность воспринимать как семантически правильные еще целый ряд допустимых словосочетаний. Однако построение результирующего графа в этом случае потребует более сложной процедуры.

Объединенно-пересеченный граф словосочетания, а также соответствующее ему уравнение алгебры конечных предикатов позволяют отвечать не только на вопрос об осмысленности словосочетания, но и на некоторые вопросы относительно его значения, например: установить, обладает ли какой-либо объект некоторым свойством; определить, какими признаками должен обладать инструмент для совершения действия; восстановить существительное — инструмент или глагол по набору признаков и т. д.

Выбранный способ представления семантической сети также позволяет довольно просто формулировать вопросы относительно смежности вершин и отвечать на них. Для решения более слож-

ных задач целесообразно выбрать иное представление графа в виде уравнений алгебры конечных предикатов.

Список литературы: 1. *Апресян Ю. Д.* Лексическая семантика. М., 1974. 128 с. 2. *Золотова Г. А.* Синтаксический словарь. М., 1988. 383 с. 3. *Кузнецова Э. В.* Итоги и перспективы семантической классификации русских глаголов//Семантические классы русских глаголов. Свердловск, 1982. С. 3—10. 4. *Богданов В. В.* Семантико-синтаксическая организация предложения. Л., 1977. 67 с. 5. *Лазарева Э. А.* Морфосемантические связи существительных лексико-семантической группы «Орудия, предназначенные для воздействия на объект» с глаголами//Семантические классы русских глаголов. Свердловск, 1982. С. 90—96. 6. *Янценецкая М. Н.* Мотивационные отношения в лексике и лексическое гнездо//Семантическая структура слова. Кемерово, 1984. С. 3—17. 7. *Ожегов С. И.* Словарь русского языка. М., 1987. 232 с. 8. *Шабанов-Кушнарченко Ю. П.* Искусственный интеллект: Математические средства. Х., 1984. 144 с.

Поступила в редколлегию 12.04.89