

структура, то необходимо помнить, что чрезмерное усиление жесткости управления и порядка, как это ни странно, может привести к явлениям неустойчивости в управлении, кризису развития. Негативное воздействие на устойчивость управления оказывает увеличение скорости упорядочения ситуации. Эта неустойчивость может редуцироваться другими факторами, например, повышением информационной открытости национально-го образования.

Тезис 5. Каждый отдельный субъект национального образования (личность) должен иметь гарантированную возможность обособления (в том числе — информационного) от других индивидов, т. е. охраняемую частную жизнь, свободу совести и вероисповедания. При этом необходимо развивать способы и средства поддержки активного информационного взаимодействия субъектов.

Перечень тезисов можно продолжить и конкретизировать с учетом изложенных выше факторов. При этом необходимо ориентироваться на конкретную социально-политическую обстановку, в которой они применяются. Приведенные тезисы достаточно противоречивы — только творческий подход лица, принимающего решение, может определить приоритеты их применения. Для более эффективной реализации творческих подходов в управлении социально-политическими ситуациями целесообразно использовать соответствующие интеллектуальные информационные технологии [1].

Интересно видеть, что некоторые из представленных тезисов хорошо эвристически обусловлены, хотя методологический источник построения этих тезисов имеет совсем другую природу — природу создания интеллектуально-информационных систем.

* * *

В работе обосновывается ряд информационных факторов, существенно влияющих на устойчивость развития социально-политических ситуаций. Показывается, что эти факторы имеют в своей основе фундаментальные физические и математические закономерности.

При подготовке управленческих воздействий на процессы развития социально-политических ситуаций информационные факторы должны учитываться наряду с историческими, экономическими, культурными и прочими традиционно используемыми закономерностями.

СПИСОК ЛИТЕРАТУРЫ

1. Райков А. Н. Интеллектуальные информационные технологии в аналитических исследованиях социально-политических объектов // НТИ. Сер. 2.— 1994— № 11.— С. 1—8.
2. Raikov A. N., Ulyanov S. V. Fuzzy modelling of support decision as intelligent technology in control of complicated social-political object // Proceedings of The Int. Conf. on Application of Fuzzy Systems (ICAFS-94), Tabriz, Iran, 17—19 October.— Tabriz, 1994.— P. 281—284.
3. Захаров В. Н., Ульянов С. В. Нечеткие модели интеллектуальных промышленных регуляторов и систем управления. II. Эволюция и принципы построения // Техническая кибернетика.— 1993— № 4.— С. 189—205.
4. Кон Г. Алфавит национализма. Проблемы восточной Европы: Сб. статей. № 41—42 / Под ред. Ф и Л. Сялиничих.— Вашингтон, 1994.— С. 88—135.

Материал поступил в редакцию 20.01.95.

УДК 81'37

Э. Ф. Скороходько

СЕМАНТИЧЕСКАЯ СЛОЖНОСТЬ СЛОВА (ТЕРМИНА):

СЕТЕВЫЕ ПАРАМЕТРЫ И КОММУНИКАТИВНЫЕ ХАРАКТЕРИСТИКИ

Семантическая сложность лексической единицы, в частности, термина, определяется рядом факторов: количеством семантических элементов, из которых состоит ее значение; степенью их разнородности; количеством семантических отношений между ними; степенью структурированности, в частности, иерархичности значения и т. д. Предлагается восемь параметров, вырабатываемых автоматизированной системой сетевого анализа СЕАНС. Экспериментальное обследование общелитературной и терминологической лексики показало, что существует прямая зависимость между этими параметрами и такими характеристиками, как распространенность слова (термина), трудность его понимания, грамматическая сложность. Установленные закономерности могут найти практическое применение для определения оптимальной длины термина в зависимости от сложности выражаемого им понятия.

1. ПОСТАНОВКА ЗАДАЧИ

Слову, как элементу лексико-семантической системы, присущ ряд свойств, которые характеризуют его с точки зрения взаимосвязей с остальными элементами этой системы. Эти свойства нередко остаются вне поля зрения лингвиста, поскольку выявить их без специальных методов исследования невозможно: человек не в со-

стоянии удерживать в памяти и учитывать все разнообразные и протяженные семантические связи между словами. Одним из методов, предназначенных для выявления именно системных свойств слова, является сетевое моделирование [1—3].

Основное преимущество семантической сети состоит в том, что она позволяет исследовать лексику с той «разрешающей способностью», которая определяется

конкретным заданием: объект исследования может изменяться от отдельного слова до группы любого объема, включая лексику в целом. При этом, исследуя отдельное слово, мы сохраняем возможность учитывать все его связи с окружением, объем которого, в свою очередь, можно устанавливать по желанию. Окружение может быть представлено как ближайшими «соседями» слова, так и сколь угодно большим фрагментом лексики. Такая возможность обеспечивается тем, что сеть фиксирует, во-первых, все лексические значения и каждое значение в отдельности, во-вторых, все непосредственные и опосредованные связи между значениями. «Разрешающая способность» исследования, его глубина и детальность определяются тем, какие именно значения и отношения принимаются во внимание, на какое расстояние от слова прослеживаются его связи (это проявляется особенно наглядно при использовании графического способа представления семантической сети).

С другой стороны, исследуя лексико-семантическое объединение слов в целом, мы оставляем в поле зрения и каждое отдельное слово, поскольку можем учитывать роль и место любого слова и любой группы слов в данном лексико-семантическом объединении. В результате последнее предстает перед нами не как неупорядоченное множество слов, но как структура, в которой каждый элемент занимает определенное место (разумеется, в той мере, в какой это допускает язык, где существует большая или меньшая размытость). Таким образом, сетевое моделирование лексики оказывается особенно плодотворным при исследовании вопросов, связанных с соотношением между отдельными словами или словом и лексическим объединением. Одним из благодарных объектов для применения сетевого моделирования является терминология, где степень системной организации выше, а размытость семантических отношений ниже, чем в общелитературной лексике. Тем не менее и в последнем случае применение семантических сетей позволяет получать нетривиальные результаты. В данной статье для краткости используются обобщенные наименования *слово* и *лексика*, которые относятся как к общелитературному, так и к терминологическому уровням языка. Таким образом, под словом может подразумеваться и терминологическое словосочетание.

Сетевое моделирование дает возможность определять качественные и количественные характеристики, которые отражают свойства отдельных слов и лексики в целом с позиций системно-структурного подхода. Под **качественным сетевым параметром** слова *A* здесь подразумевается организованное множество слов, которые находятся в определенном отношении к слову *A* и характеризуют, таким образом, тот или иной аспект его системной семантики. Примерами качественных параметров являются: перечень слов, которые представляют собой декомпозицию значения данного слова, перечень слов, семантически производных от данного и т. д.

Количественный сетевой параметр — это число, представляющее собой меру определенного свойства слова, которая определяется его местом в лексико-семантическом объединении, составом и структурой последнего. Количественные сетевые параметры относятся к двум типам. Параметр первого типа — это число, выражающее объем соответствующего качественного параметра, например: количество слов, являющихся декомпозицией значения определенного слова; количество слов, семантически производных от данного. Количественный параметр второго типа не имеет аналогов среди качественных.

Большинство этих параметров лингвистически релеванты, но было бы ошибкой пытаться интерпретировать каждый параметр в традиционно привычных терминах.

Рассмотрим некоторые возможности сетевой модели на примере параметров, характеризующих семантическую

сложность слова (здесь речь идет не о полисемии, а об отдельном значении, в частности, значении термина, который, как правило, моносемичен).

Понятие семантической сложности само по себе имеет реляционный характер, оно предполагает соотносительность разных объектов. Сложность (одного объекта) предполагает простоту (другого объекта). Прилагательное *сложный* имеет по меньшей мере четыре значения: 1. Состоящий из нескольких частей. 2. Образующий систему взаимосвязанных частей. 3. Характеризующий совокупностью многих переплетенных между собой явлений, признаков. 4. Трудный для понимания [4]. Сюда можно добавить и такое значение: «выводимый из других единиц». Обратим внимание на то, что эти значения организованы по принципу матришки: каждое последующее включает предыдущее. Кроме того, они имеют характер шкалы, т. е. предполагают наличие некоторой градации [5].

В термине *семантическая сложность слова* актуализируются все эти значения. Утверждая, что некоторое слово семантически сложно, мы подразумеваем, что его значение состоит из каких-то иных, более простых, элементов, что эти элементы связаны друг с другом, что они образуют определенную структуру и т. д. Поэтому для адекватного описания значения слова необходимо иметь возможность оценивать разные аспекты этого явления, в частности:

- количество семантических элементов, из которых состоит значение слова;
- степень их разнородности;
- количество семантических отношений между ними;
- степень разнородности семантических отношений между элементами значения;
- степень структурированности, в частности, иерархичности значения слова;
- трудность понимания слова.

В определенной мере эти аспекты семантической сложности слова можно оценивать с помощью сетевых параметров. Для облегчения восприятия последующего материала предположим ему краткий обзор некоторых основополагающих понятий сетевой модели (термины, отмеченные звездочкой *, более подробно поясняются в [1—3]).

2. СЕТЕВАЯ МОДЕЛЬ ЛЕКСИКИ

Семантический компонент и семантический дериват. Если значение слова *A* непосредственно или опосредованно содержится в значении слова *B* (это проявляется, в частности, в том, что слово *B* определяется через слово *A*), первое слово называется **семантическим компонентом *** (СК) второго, а второе — **семантическим дериватом *** (СД) первого. В дальнейшем будем использовать сокращенные варианты этих терминов — *компонент* и *дериват*.

Если слово *A* явно присутствует в определении (толковании) слова *B*, то слово *A* считается непосредственным семантическим компонентом (НСК) слова *B*, а *B* — непосредственным семантическим дериватом (НСД) слова *A*. Если *A* присутствует в определении (толковании) третьего слова, через которое определяется *B*, то оно — опосредованный компонент *B*, а *B* — опосредованный дериват *A*. НСК может представлять собой как элементарную, так и производную семантическую единицу. Некоторые слова одновременно играют роль и непосредственных, и опосредованных компонентов (дериватов).

Компоненты, значения которых в избранном лексико-семантическом объединении (в частности, в избранной терминологии) приняты за элементарные, первичные, называются **конечными (КСК)**. Дериваты, значения которых не включаются в значение какого-либо из иных слов объединения, называются **конечными (КСД)**.

Семантическая сеть * лексики представляет собой граф, вершины которого сопоставлены значениям слов (терминов). Вершины соответствующие НСК и НСД, соединены дугой, направленной от первого слова ко второму (рис. 1). Для упрощения и большей наглядности на рисунке представлены лишь основные связи между несколькими словами, связанными с понятием «дерево». Не учитывается и содержание этих связей.

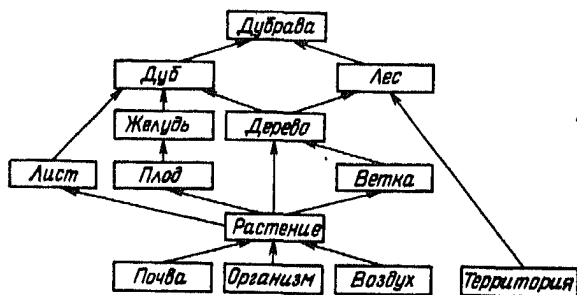


Рис. 1. Семантическая сеть

На семантической сети компонентам слова *A* соответствуют вершины, из которых достижима вершина, соответствующая слову *A*, т. е. вершины, пути из которых ведут к *A* (компонентами слова *дерево* являются *растение*, *ветка*, *организм*, *почва*, *воздух*). Поскольку вершины семантической сети сопоставлены словам, для простоты вместо «вершина, которая соответствует слову *A*», будем иногда писать просто «слово *A*»). Дериватам отвечают вершины, достижимые из *A*, т. е. вершины, к которым ведут пути из *A* (*дуб*, *лес*, *дубрава*). Непосредственным компонентам и дериватам отвечают соседние вершины, примыкающие к *A* (соответственно, *ветка*, *растение* и *дуб*), конечным — наиболее удаленные вершины (*почва*, *организм*, *воздух* и *дубрава*).

Семантическая деривация/декомпозиция. Между словом и его дериватом существует отношение семантической деривации, между словом и его компонентом — отношение семантической декомпозиции. Эти два отношения по существу представляют собой одно, рассматриваемое, однако, с противоположных направлений. Отношение семантической деривации/декомпозиции связывает, в частности, определяемое слово со словами, присутствующими в его определении (толковании).

Отношение семантической деривации/декомпозиции мы понимаем как чисто синхронное явление. Наличие этого отношения между парой слов свидетельствует лишь о том, что значение одного из них является составной частью значения другого. Например, исходя из того, что слово *лес* имеет значение '(большая) территория, покрытая лесом', мы констатируем наличие отношения семантической деривации в паре слов *дерево*→*лес*.

В отдельных случаях отношение семантической деривации совпадает с отношением словообразовательной деривации (*самолет*→*самолестостроение*), иногда оно отражает реальный процесс образования понятия при отсутствии словообразовательных связей между словами (*самолет*→*крыло самолета*→*элерон*). Однако иногда оно противоположно истинному направлению процесса образования значения [6]. Так, понятие «летательный аппарат тяжелее воздуха» образовалось в результате обобщения понятий «самолет», «вертолет», «автожир» и под. Следовательно, в диахронии становление значения шло от «самолет» к «летательный аппарат тяжелее воздуха». Однако современное определение термина *самолет* гласит: «летательный аппарат тяжелее воздуха с неподвижными относительно аппарата несущими поверхностями и с силовой установкой, обеспечивающей

поступательное перемещение аппарата в воздухе». Таким образом, сейчас мы воспринимаем значение термина *самолет* как производное от значения термина *летательный аппарат тяжелее воздуха*. Именно в этом смысле мы и говорим о семантической деривации: как о минимом процессе образования значения слова *самолет* от значения термина *летательный аппарат тяжелее воздуха*.

На семантической сети отношение семантической деривации/декомпозиции эксплицируется в форме дуги или цепочки дуг, направленных в одну сторону.

Отношение семантической деривации/декомпозиции между словом и его НСД (НСК) назовем элементарным, а между словом и его опосредованным дериватом (или компонентом) — сложным. Переход от слова к его НСД составляет один такт семантической деривации. Иногда в таких случаях будем говорить об элементарной семантической деривации.

Замену слова таким множеством слов — его компонентов (непосредственных или опосредованных), которые в сумме передают его интенциональное значение (или содержание значения), назовем семантической декомпозицией. Если семантический компонент слова *A* рассматривается как член такого множества, т. е. как результат процесса декомпозиции, он называется семантическим декомпозитом слова *A*.

Обратный процесс, т. е. замену множества слов их дериватом (непосредственным или опосредованным), значение которого сводится к значениям этих слов, назовем семантической композицией.

Деривационная цепочка. Деривационной цепочкой * называется такая последовательность слов, в которой между каждой парой смежных членов существует отношение непосредственной семантической деривации/декомпозиции. *почва*→*растение*→*дерево*→*лес*→*дубрава*. Первый член цепочки называется исходным словом *, последний — терминальным *, остальные слова — промежуточными.

На семантической сети деривационной цепочке соответствует простой путь, т. е. последовательность вершин, связанных однонаправленными дугами.

Семантическое дерево. Та часть (подграф) семантической сети, которая включает слово *A* вместе с его компонентами и связывающими их дугами, т. е. все деривационные цепочки, входящие в избранную вершину, называется семантическим деревом слова *A* (вся сеть, изображенная на рис. 1, представляет собой семантическое дерево слова *дубрава*, фрагмент сети, образованный словами *дерево*, *желудь*, *лист*, *плод*, *ветка*, *почва*, *организм*, *воздух* представляет собой семантическое дерево слова *дуб*).

Семантическое дерево слова отображает структуру его значения: состав семем, их иерархию, связи между ними (здесь имеется в виду не иерархия семем в полисемантическом слове, а иерархия элементарных и неэлементарных значений — семем, — образующих единое, более сложное, значение).

Каждая вершина в семантическом дереве отражает некоторое значение, входящее составной частью в значение целого. Некоторые слова входят не в одну, а в несколько деривационных цепочек (например, слово *дерево* входит, кроме названной выше, и в такую цепочку: *почва*→*растение*→*дуб*→*дубрава*). Это означает, что значение соответствующего слова не одно, а несколько раз входит в состав значения деривата (слова *A*), т. е. несколько раз принимает участие в процессе семантической деривации (значение слова *дерево* вошло в значение слова *дубрава* один раз опосредованно через *лес*, другой раз — через *дуб*, значение слова *растение* вошло в значение слова *дерево* один раз непосредственно и один раз опосредованно — через *ветка*).

3. ПАРАМЕТРЫ СЕМАНТИЧЕСКОЙ СЛОЖНОСТИ СЛОВА

Семантическую сложность слова определяют количественные параметры, которые вычисляются по семантическому дереву. Рассмотрим здесь те из них, которые не требуют учета различий в значении семантических связей между словом и его компонентами.

Количество НСК c_1 — количество слов, принадлежащих анализируемому лексико-семантическому объединению, значения которых непосредственно образуют значение данного слова (в частности, число слов, явно присутствующих в толковании или определении). Характеризует семантическую сложность слова с точки зрения количества непосредственных семантических составляющих его значения (для слова *дубрава* $c_1=2$). Следует учесть, что значения параметров, вычисленных по небольшой, к тому же более или менее произвольно выбранной группе слов, как в нашем примере, играют лишь иллюстративную роль и не всегда отражают реальные особенности языкового объекта).

Количество КСК c_2 — количество слов, принадлежащих анализируемому лексико-семантическому объединению, значения которых в конечном итоге сводятся к значению данного слова. Характеризует семантическую сложность слова с точки зрения количества элементарных (исходных) семантических составляющих, которые образуют значение данного слова (для *дубрава* $c_2=4$).

Количество СК c_3 — количество всех слов, принадлежащих анализируемому лексико-семантическому объединению, значения которых непосредственно или опосредованно содержатся в значении данного слова. Характеризует семантическую сложность слова с точки зрения объема его семемного состава (для *дубрава* $c_3=12$).

Д-порядок слова l_1 — максимальная глубина иерархии семантической структуры значения слова, т. е. количество слов, которые входят в деривационную цепочку максимальной длины, соединяющую данное слово с одним из его КСК. Характеризует количество тактов семантической декомпозиции, за которые значение слова полностью сводится к элементарным (в пределах данного лексико-семантического объединения) значениям (для слова *дубрава* $l_1=5$).

К-порядок слова l_2 — минимальная глубина иерархии семантической структуры значения слова, т. е. количество слов, которые входят в деривационную цепочку минимальной длины, соединяющую данное слово с одним из его КСК. Характеризует количество тактов семантической декомпозиции, за которые значение слова сводится к «ближайшему» элементарному (в пределах данного лексико-семантического объединения) значению (для слова *дубрава* $l_2=2$).

С-порядок слова l_3 — средняя глубина иерархии семантической структуры значения слова, т. е. среднее количество слов, которые входят в деривационные цепочки, соединяющие данное слово с его конечными компонентами. Характеризует среднее количество тактов семантической декомпозиции, за которые значение слова сводится к элементарным (в пределах данного лексико-семантического объединения) значениям (для слова *дубрава* $l_3=3,5$).

Кроме того, Д-, К- и С-порядки характеризуют максимальное (соответственно — минимальное, среднее) количество тактов семантической деривации на пути становления значения слова, начиная от семантически первичных (в пределах данного лексико-семантического объединения) слов.

Количество входных цепочек s_1 — количество деривационных цепочек, которые входят в данное слово, т. е. количество «путей», которыми шло становление данного значения. Характеризует значение слова главным образом с точки зрения сложности его семантической струк-

туры (для слова *дубрава* $s_1=19$; цепочки: *почва* → *растение* → *лист* → *дуб* → *дубрава*, *почва* → *растение* → *плод* → *желудь* → *дуб* → *дубрава* и т. д.).

Суммарный вес входных цепочек w_1 — сумма слов, входящих во все деривационные цепочки, для которых данное слово является конечным. Характеризует семантическую сложность слова с учетом как количества семем, образующих значение слова, так и структуры значения. Этот параметр более чувствителен, по сравнению с предыдущим, к увеличению числа семантических компонентов слова и сложности структуры семантического дерева (для слова *дубрава* $w_1=83$).

Различие между количеством СК и суммарным весом входных цепочек состоит в том, что в первом случае каждый семантический компонент учитывается один раз, а во втором — столько раз, сколько он входит в деривационные цепочки. Например, при вычислении параметра w_1 для слова *дубрава* компонент *растение* учитывается 18 раз, поскольку он входит в 18 деривационных цепочек, компонент *лес* — семь раз и т. д. Здесь можно провести аналогию между этими параметрами, с одной стороны, и количеством разных слов и числом словоупотреблений в тексте, — с другой.

Перечисленные параметры определяют разные аспекты семантической сложности слова. Сопоставим два семантических дерева (А и В на рис. 2). Слова, семан-

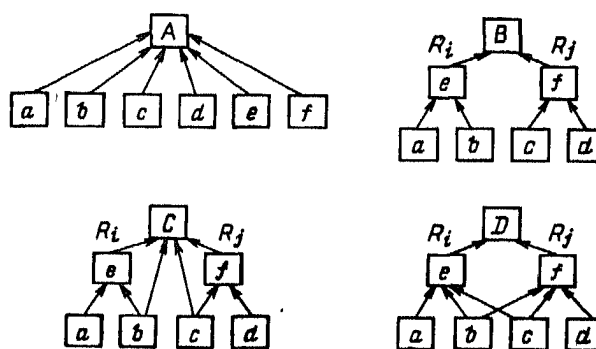


Рис. 2. Типы семантических структур значения

тические структуры которых здесь представлены, имеют равное число семантических компонентов — *a, b, c, d, e, f*. По количеству КСК, т. е. по числу элементарных семантических единиц, образующих значение, более сложным является первое слово, но по глубине иерархии семантической структуры, т. е. по степени семантической производности, — второе (его значение явилось результатом двух тактов семантической деривации. К тому же семантические отношения R_i, R_j во втором слове несут двойную нагрузку, поскольку связывают слово с БСК, каждый из которых, в свою очередь, включает значения двух других слов). Это различие отражают соответствующие параметры — количество КСК (шесть против двух) и Д-порядок (один против двух).

В другой паре слов (В и С на рис. 2) совпадают количество НСК, КСК, СК и Д-порядок. Однако семантическая структура второго из этих слов включает большее число деривационных цепочек. Это означает, что одна и та же элементарная семантическая единица (*b* и *c*) дважды участвует в становлении значения слова С: один раз как непосредственная составляющая (что манифестируется дугами $b \rightarrow C$ и $c \rightarrow C$) и второй раз как опосредованная (соответственно через *e* и *f*). Различие между значениями этих слов отражает соответствующий параметр — количество входных цепочек. В первом слове он равен четырем ($a \rightarrow e \rightarrow B, b \rightarrow e \rightarrow B,$

$c \rightarrow f \rightarrow B$, $d \rightarrow f \rightarrow B$), во втором — шести (дополнительно: $b \rightarrow C$, $c \rightarrow C$).

Сравним еще одну пару слов — *C* и *D* (рис 2). Структура значения второго слова отличается более сложным переплетением отношений семантической деривации НСК первого слова имеют более простые значения, ибо каждый НКС образован двумя семантическими единицами, тогда как НСК второго слова — тремя. Соответственно, ребра R_i и R_j второго слова несут тройную нагрузку. Это различие находит свое отражение в суммарном весе входных цепочек (количества входных цепочек здесь равны) — десять против двенадцати.

Среди перечисленных параметров особого внимания заслуживают три: количество СК, Д-порядок и сум-

Таким образом, мы располагаем набором количественных параметров, которые позволяют оценивать некоторые из перечисленных выше аспектов семантической сложности слова, а именно:

количество семантических элементов, образующих значение слова (определяется через количество НСК, КСК, СК);

степень семантической производности (выводимости) значения (определяется через Д-, К- и С-порядки);

степень иерархичности структуры значения (определяется через Д-, К- и С-порядки);

степень «переплетенности» семантических связей между элементами, образующими значение слова, (определяется через количество входных цепочек и суммарный вес входных цепочек).

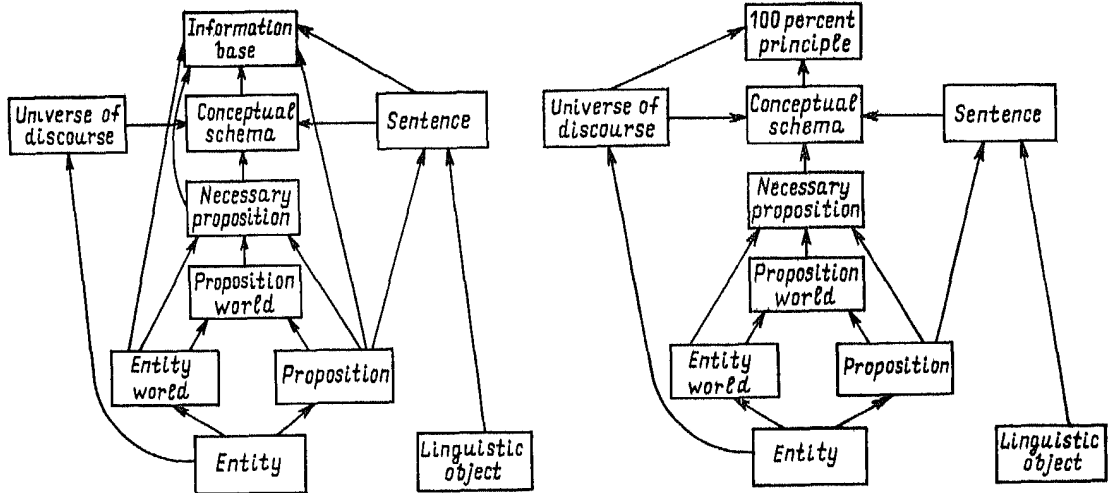


Рис. 3 Семантические деревья терминов *information base* и *100 percent principle*

марный вес входных цепочек. Во-первых, они отражают важнейшие аспекты семантической сложности слова. Во-вторых, от них в известной мере зависят и другие параметры. Как разные значения термина *сложный*, так и разные аспекты семантической сложности слова не независимы друг от друга. Между некоторыми из них ощущаются статистические зависимости (например, обычно с увеличением Д-порядка растет число СК и количество входных цепочек). Три параметра связаны аналитической зависимостью: $w_1 = l_3 \times s_1$. Это позволяет в некоторых случаях по одним параметрам оценивать (приблизительно или точно) другие параметры [1]. Тем не менее для точной характеристики семантической сложности слова необходимо вычислять все восемь параметров. На рис. 3 представлены семантические деревья терминов *information base* и *100 percent principle* (терминология концептуального проектирования баз данных). Значения обоих терминов образованы одними и теми же семантическими единицами, если не учитывать компоненты, лежащие за пределами данной терминологии. Эти компоненты не отражены в деревьях и не принимаются в расчет при вычислении количественных параметров. Различия между значениями заключаются главным образом в количестве и содержании семантических связей между терминами и их компонентами. Семантическая сложность данных терминов характеризуется такими показателями.

	c_1	c_2	c_3	l_1	l_2	l_3	s_1	w_1
<i>information base</i>	5	2	9	5	2	3,5	15	52
<i>100 percent principle</i>	2	2	9	5	2	3,5	8	28

Комплексную оценку семантической сложности слова можно производить с помощью формулы $c_3 \times \ln(w_1 + 1)$. В этой формуле учтены основные аспекты семантической сложности слова: количество разных семантических единиц, присутствующих в значении слова (c_3), и число вхождений этих единиц в значение (w_1). Косвенным образом, через параметр w_1 , учтен и третий важный аспект — степень иерархичности структуры значения.

4. КОЛИЧЕСТВЕННЫЕ СЕТЕВЫЕ ПАРАМЕТРЫ И СУБЪЕКТИВНОЕ ВОСПРИЯТИЕ ЗНАЧЕНИЯ СЛОВА

Проверка научной модели чаще всего заключается в сравнении результатов, полученных с использованием этой модели, с результатами, полученными путем прямого наблюдения или прямого измерения соответствующих параметров. Но в нашем случае объект исследования, т. е. семантическая сложность слова, недоступен прямому наблюдению и измерению. Человек не в состоянии оценивать такие качества значения, как его сложность в целом или количество семантических составляющих значения, тем более степень иерархичности его структуры.

Единственный аспект семантической сложности слова, о котором человек может судить хотя бы в общих чертах, — это трудность понимания слова. К сожалению, именно это свойство не находит прямого отражения среди перечисленных параметров. Однако

естественно предположить, что трудность понимания зависит, в частности, от остальных аспектов семантической сложности слова. Можно ожидать также, что и некоторые другие свойства слова, доступные наблюдению, находятся в определенной зависимости от тех аспектов семантической сложности, которые покрываются перечисленными параметрами. К ним относятся, прежде всего, такие коммуникативно-значимые характеристики, как легкость восприятия, усвоения и употребления слова, его частотные характеристики, грамматическая структура.

Несомненный интерес представляла бы проверка наличия корреляции между семантической сложностью слова и частотой его употребления в тексте. Однако объектом нашего исследования служила, главным образом, терминологическая лексика разных отраслей знания, для которой нет частотных словарей. Поэтому мы ограничились проверкой следующих трех гипотез:

существует зависимость между семантической сложностью слова и субъективным восприятием слова как знакомого или незнакомого;

существует зависимость между семантической сложностью слова и трудностью его понимания;

существует зависимость между семантической сложностью слова и его грамматической структурой.

Под семантической сложностью слова здесь имеются в виду те ее аспекты, которые находят свое отражение в перечисленных параметрах. Восприятие слова как знакомого в какой-то мере позволяет судить о его распространенности, а понятие распространенности может рассматриваться как нестрогий аналог понятия частотности. Понимание слова (термина) отождествлялось с умением сопоставить его значение с соответствующим понятием или объектом действительности.

Эксперименты проводились на материале украинской, русской и английской общелитературной и терминологической (общенаучной, общетехнической, специальной) лексики. Специальная лексика относилась к следующим областям: информационные системы, базы данных, обработка текста, валюта и финансы и др. Всего экспериментами было охвачено 400 слов общелитературного языка (четыре выборки) и 1600 терминов (восемь терминологий от 24 до 350 терминов в каждой). Для каждого из слов (терминов) с помощью автоматизированной системы СЕАНС [2] были вычислены сетевые параметры.

При проверке названных гипотез учитывались количество СК, Д-порядок и (частично) суммарный вес входных цепочек.

В первом эксперименте респондент получал список, содержащий 40 слов (терминов) без определений, половина из которых имела минимальное значение соответствующего параметра, а половина — максимальное. Ему предлагалось ответить относительно каждого слова (термина) на такой вопрос: «Знакомо ли Вам это слово?» Принимались ответы: «Да», «Нет», «Не уверен».

Во втором эксперименте список терминов, предъявляемый респонденту, содержал с их определения, а вопрос был сформулирован таким образом: «Понимаете ли Вы определение данного термина?» Ответы: «Да», «Нет», «Понимаю частично (или: не уверен в правильности понимания)».

Лексический материал, предъявляемый каждому респонденту, подбирался с учетом его специальности. В обоих экспериментах респонденту не предлагали наукоемкую терминологию далеких от него специальных отраслей. Во втором эксперименте была исключена, кроме того, общелитературная лексика и терминология отраслей, в которых работали респонденты, поскольку в обоих этих случаях понимание значения (учитывая наличие определений) было заведомо 100%. В обоих экспериментах центральное место занимала общенаучная, общетехническая лексика, терминология отраслей,

смежных с теми, в которых работают респонденты, а также терминология, представленная толковыми словарями научно-популярного типа, в которых отсутствуют узкоспециальные термины, а определения несколько упрощены.

Результаты экспериментов подтверждают и первую, и вторую гипотезы. Все слова (термины) с минимальной семантической сложностью были восприняты как знакомые, а их толкования (определения) — как полностью понятные. Среди лексических единиц с максимальной семантической сложностью в среднем 30% были признаны знакомыми, 55% — незнакомыми, а 15% — такими, в отношении которых не было уверенности. 45% определений были признаны полностью понятными, 23% — непонятными, 32% — частично понятными. Для иллюстрации приведем несколько примеров. Слова (термины) с минимальным количеством СК и минимальным Д-порядком: *адрес, алгоритм, анализ, банк, валюта, единица, естественный язык, индекс, капитал, клиент, окно, отношение, цена, ценные бумаги, building, deer, face, floor, narrow, rock, room, valley*. Слова (термины) с максимальным количеством СК и максимальным Д-порядком: *дескрипторный язык* — «искусственный информационный язык, лексическими единицами которого служат дескрипторы»; *инкассо* — «банковская операция, с помощью которой банк по поручению клиента на основании расчетных документов получает от предприятий и организаций денежные суммы, которые ему причитаются за товары и услуги, и зачисляет их на счет клиента»; *контокоррент* — «единый счет, на котором учитываются все операции банка с клиентом»; *маржа* — «дополнительная доля по депозиту, залоговому обеспечению или допустимым колебаниям курса валют»; *релевантность* — «соответствие документа информационному запросу абонента»; *язык индексирования* — «искусственный информационно-поисковый язык, предназначенный для индексирования документов и информационных запросов»; *gorge* («узкое ущелье»), *dingle* («глубокая лощина»), *dell* («лесистая долина»), *strath* («широкая горная долина с протекающей по ней рекой»), *combe* («ложбина»).

Особенно наглядно проявляется различие в семантической сложности, если сравнивать слова (термины) в пределах одного лексико-семантического объединения (одной терминосистемы). Так, в терминологии концептуального проектирования баз данных минимальную семантическую сложность имеют такие термины: *лингвистический объект (linguistic object)* — «грамматически допустимая конструкция языка»; *сущность (entity)* — «любой конкретный или абстрактный объект, в частности связи между объектами»; *лексический объект (lexical object)* — «простой лингвистический объект, обозначающий сущность»; *аксиома (axiom)* — «любое замкнутое предложение, утвержденное в качестве такового полномочным источником»; *предметная область (universe of discourse)* — «все сущности, которые были, есть или могут иметь место». Максимальную семантическую сложность имеют: *внешний уровень (external level)* — «все аспекты, относящиеся к ориентированному на пользователя представлению информации, видимые на уровне внешних интерфейсов информационной системы»; *концептуальная подсхема (conceptual subschema)* — «непротиворечивая совокупность предложений, которая выражает необходимые высказывания, истинные для некоторой предметной области, ограниченной представлением одного пользователя и, как таковой, являющейся частью концептуальной схемы, относящейся к информационной системе»; *схема баз данных (data base schema)* — «определение формы представления и структуры базы данных для возможных совокупностей всех предложений, имеющих в концептуальной схеме и информационной базе, включая аспекты манипулирования этими формами».

5. КОЛИЧЕСТВЕННЫЕ СЕТЕВЫЕ ПАРАМЕТРЫ СЕМАНТИЧЕСКОЙ СЛОЖНОСТИ И ГРАММАТИЧЕСКАЯ СТРУКТУРА ЛЕКСИЧЕСКОЙ ЕДИНИЦЫ

Гипотеза о зависимости между семантической и грамматической сложностью слова впервые была выдвинута и частично доказана в 1986 г [7]. В настоящей работе исследовалась зависимость между количеством СК, D-порядком и суммарным весом входных цепочек, с одной стороны, и числом корневых морфем, образующих слово (термин), — с другой. Результаты свидетельствуют о том, что с увеличением числа корневых морфем растет среднее значение всех параметров семантической сложности (таблица). Эта закономерность присуща всем без исключения рассмотренным терминологиям и выборкам общелитературной лексики

Зависимость между семантической и грамматической сложностью слова

Категория лексики	Язык	Число корневых морфем в лексич. единице	Среднее значение семантической сложности	
			Количество СК	D-порядок
Общелитературная	Русский — 83 слова	1	3,7	1,5
		2	5,2	2,6
	Английский — 280 слов	1	3,6	1,7
		2	5,7	2,6
Терминология	Украинский — 413 терминов	1	4,3	1,7
		2	6,0	2,4
		3	7,6	3,4
		4	8,7	3,1
	Русский — 819 терминов	1	4,2	1,1
		2	5,4	1,9
		3	8,9	2,9
		4	8,8	2,8
	Английский — 372 термина	1	5,1	2,2
		2	7,5	3,2
		3	9,5	4,1
		4		

В общелитературной лексике (русский и английский языки) среди семантически простейших слов 96% имеют одну корневую морфему, 4% — две. В этой категории среди слов с одной корневой морфемой минимальную семантическую сложность имеют 38% слов, среди слов с двумя корневыми морфемами — 9%. В терминологической лексике (в среднем для всех трех языков) среди семантически простейших терминов 47% имеют одну корневую морфему, 25% — две, 14% — три, 5% — четыре. Минимальную семантическую сложность здесь имеют 61% слов с одной корневой морфемой, 32% — с двумя, 6% — с тремя, 1% — с четырьмя.

6. ЗАКЛЮЧЕНИЕ

Корреляция между параметрами семантической сложности и другими свойствами слова подтверждает лингвистическую релевантность этих параметров. В то же время представляется небезынтересной дальнейшая интерпретация полученных результатов. Так, о чем мо-

жет свидетельствовать зависимость между степенью семантической сложности и соотношением числа знакомых и незнакомых слов? Все респонденты свободно владели языком (соответственно, украинским, русским, английским). Поэтому, если респондент признавал слово незнакомым, его можно считать мало употребляемым как активно, так и пассивно, т. е. малораспространенным.

Следовательно, можно допустить, что семантическая сложность представляет собой один из факторов, от которых зависит распространенность слова. Семантически простые слова являются наиболее используемыми и наиболее распространенными. Это согласуется с иным наблюдением: чем выше языковая семантическая значимость слова в данном лексико-семантическом объединении, тем больше его частотность [1].

Языковая семантическая значимость слова определяется, главным образом, количеством семантически зависимых от него слов, т. е. числом семантических дериватов. Применительно к терминологии языковую значимость термина можно интерпретировать как число понятий, истолкование которых невозможно без привлечения (прямого или опосредованного) данного термина. Языковая значимость слова — это значимость «в среднем», не связанная с конкретным словоупотреблением. Наибольшую языковую значимость имеют слова с более или менее общими (но не самыми общими!) значениями. Коррелятом языковой значимости является речевая значимость, которая определяется конкретным коммуникативным актом [2]. Речевая значимость слова не всегда совпадает с его языковой значимостью. Например, в терминологии информационных систем большую языковую значимость имеют такие термины, как *поиск информации, хранение данных, база данных*, однако в текстах акцент может смещаться в сторону более специфичных, хотя и имеющих меньшую языковую значимость, терминов (*пермутационный указатель, пертинентность, стратегия поиска*).

Между семантической сложностью и языковой семантической значимостью слова существует обратная (статистическая) зависимость. Таким образом, к наиболее распространенным относятся слова, имеющие большую языковую семантическую значимость и малую семантическую сложность.

Теперь о зависимости между семантической сложностью и трудностью понимания. Здесь причина кроется, по-видимому, в том, что увеличение степени семантической сложности слова отражает, в частности, рост числа семантических составляющих значения. А это означает, в свою очередь, что увеличивается число и удельный вес тех составляющих, которые соотносятся со все более и более специфичными понятиями, восприятие этих понятий требует все больших усилий и высокого профессионализма. Увеличивается и структурная сложность значения.

Наконец, о зависимости между семантической сложностью и морфемным строением слова. Здесь действуют несколько факторов. Во-первых, среди семантически простейших слов много таких, которые выражают определенные основополагающие понятия и принадлежат к ядру лексики. Часто оказывается, что эти слова утратили свою этимологию. Среди них много слов с большой частотой употребления в речи, а такие слова проявляют тенденцию к структурной простоте (*plant, piece, body*). Во-вторых, с увеличением степени семантической сложности, т. е. с ростом числа семантических составляющих значения, возникает необходимость привлекать дополнительный лексический материал (в форме слов, корневых или словообразовательных морфем) для того, чтобы выразить эти составляющие. Особенно ярко это проявляется в терминологической лексике.

В этих закономерностях проявляется универсальный закон оптимизации: в языке существует тенденция к

установлению оптимальных соответствий между разными, часто противоречащими друг другу, аспектами или свойствами языковой единицы. В нашем случае такими свойствами являются: семантическая сложность, языковая семантическая значимость, грамматическая сложность, понятность, распространенность (частотность).

Оптимальность механизма реализации коммуникативной функции языка обеспечивается благодаря тому, что факторы, затрудняющие использование слова (большая семантическая сложность, трудность понимания, громоздкость), уравновешиваются меньшей частотой обращения к соответствующим словам. Таким образом, в пространстве языка в одном «углу» сконцентрированы наиболее важные для среднего носителя языка слова, которые имеют максимальную языковую семантическую значимость и большую частоту обращения к ним, но в то же время семантически и грамматически более или менее просты, понятны и легко усваиваются, что облегчает их употребление в речи. В противоположном «углу» собраны слова (преимущественно термины), которые имеют небольшую общеязыковую семантическую значимость и сравнительно небольшую частоту обращения к ним, хотя необходимы и важны в отдельных коммуникативных сферах. Такие слова (термины) семантически и грамматически более сложны, трудны для понимания. Их использование ограничивается, главным образом, профессионально ориентированными средами. Здесь позволительно провести аналогию с хорошо организованным складом, где ближе ко входу хранятся сравнительно несложные, но важные в быту и необходимые практически каждому инструменты — молоток, клещи, отвертка. Далее нас ожидают более сложные, но реже востребуемые — электродрель, точило, еще дальше — токарный и сверлильный станки, а на заднем дворе — мощный бульдозер и экскаватор.

Полученные результаты имеют и практическое применение. Известно, что к научным и техническим терминам предъявляются противоречивые требования. В частности, термин должен быть точным и в то же время кратким. Требование точности влечет за собой стремление выразить по возможности все основные признаки содержания соответствующего понятия. Одна-

ко это ведет к увеличению длины термина. Поэтому создание термина — это поиски компромисса между требованиями точности и краткости. Вопрос о допустимой длине термина неминуемо возникает при создании и экспертизе терминологии. Решается он чаще всего либо волюнтаристским путем, либо на основании анализа эмпирических данных. Выявленные закономерные соответствия между степенью семантической сложности термина и его длиной позволяют по-новому подойти к определению оптимальной длины термина. Зная количественные параметры семантической сложности определенного термина, можно вычислить его рекомендуемую длину. Разумеется, для этого потребуются накопить больше статистических данных, чем сейчас имеется.

СПИСОК ЛИТЕРАТУРЫ

1. Скороходько Э. Ф. Семантические сети и автоматическая обработка текста. — Киев, 1983. — 218 с.
2. Скороходько Э. Ф. Сетевое моделирование лексики // Использование ЭВМ в лингвистических исследованиях. — Киев, 1990. — С. 129—166.
3. Скороходько Э. Ф. Компьютерный анализ и корректирование системы терминов на основе сетевого моделирования // НТИ. Сер. 2. — 1994. — № 2. — С. 1—9.
4. Короткий тлумачний словник української мови. — Київ: Рад. школа. — 1978. — 296 с.
5. Труб В. М. К проблеме функционального сходства нейтрального и контрастного отрицания // Логический анализ языка. — М., 1990. — С. 47—54.
6. Соболева П. А., Барченкова М. Д. О направлении производности (мотивации) // Актуальные вопросы дериватологии и дериватографии. — Владивосток, 1986. — С. 134—144.
7. Скороходько Э. Ф. Семантическая и структурная сложность существительного в современном английском языке // Там же. — С. 126—134.

Материал поступил в редакцию 20.02.95.