

© 1994 г. Л. ЛЁННГРЕН

**ОПЫТ ПОСТРОЕНИЯ АССОЦИАТИВНОГО СЛОВАРЯ  
ШВЕДСКОГО ЯЗЫКА**

**ВВЕДЕНИЕ**

В бывшем Центре вычислительной лингвистики при Уппсальском университете несколько лет велась работа по составлению ассоциативного словаря шведского языка (АСШЯ). Речь идет о словаре совершенно, насколько нам известно, нового типа. Очень кратко его можно охарактеризовать следующим образом.

Данный словарь имеет форму древовидной структуры, узлы которой заняты лексическими единицами – лексемами, а ветви соответствуют некоторым выбранным отношениям между лексемами. Эти отношения напоминают отношения родства, так что каждая лексема оказывается окруженной своей "семьей". Сами лексемы расположены традиционным образом – в алфавитном порядке, однако статья, отведенная каждой лексеме в настоящем словаре, резко отличается по своему содержанию от традиционной словарной статьи: в словарной статье АСШЯ перечисляются лексемы близлежащих узлов, т.е. именно "члены семьи" заглавной лексемы. Более точно, для каждой лексемы указываются все непосредственно семантически связанные с ней лексемы, и только они.

Большую помощь в работе над этим проектом оказали сотрудники Центра вычислительной лингвистики, в первую очередь Г. Фредрикссон, которая участвовала в работе около года; она помогала составлять исходный список лексем и снабжать каждую лексему соответствующим анализом, т.е. указывать ее "родителей". Кроме того, сотрудники-программисты, в первую очередь А. Килар, написали целый ряд программ, которые на основе упомянутого исходного анализа, т.е. отношения "ребенок : родитель", устанавливают другие отношения родства и выписывают словарные статьи.

Первоначальной целью настоящего исследования было найти эффективный метод автоматически – при помощи семантических дескрипторов – извлекать из текстов информацию. Особенно хотелось подвергнуть проверке гипотезу о том, что наилучшими дескрипторами являются слова, которые в данном тексте имеют высокую частотность, но вообще в языке отличаются низкой частотностью. Однако очевидная проблема состоит в том, что частотность для этих целей желательно установить на возможно более обобщенном уровне – так, чтобы никакая семантическая совместная встречаемость не была утрачена. Поэтому нужно подсчитывать частотность не словоформ и даже не лексем, а прямо значений. Первым шагом к такому обобщению является возведение словоформ к соответствующей лексеме, но обобщение может продолжаться и на уровне словообразования и даже чисто семантического поля. Можно утверждать, например, что лексема *локомотив* "скрывает" в себе семантический элемент "поезд". Нам представлялось, что, перенося частотность, в первую очередь, каждой отдельной словоформы на соответствующую лексему, а потом частотность каждой отдельной лексемы на лексему, представляющую более обобщенное семантическое содержание (нечто похожее на семантическое поле), можно будет выявить информацию, заложенную даже в сравнительно коротком тексте. Ведь нужно стремиться к тому, чтобы частотность именно тех

лексем, которые лучше всего отражают содержание данного текста, была как можно выше.

Данная гипотеза была проверена нами на очень небольшом материале. Оказалось, что первого шага обобщения достаточно для передачи содержания не слишком коротких текстов (порядка двух тысяч слов). Содержание таких текстов довольно хорошо выявляется уже на уровне лексемной частотности.

Однако для осуществления многоуровневой системы информационного поиска необходимо располагать очень богатым списком заранее установленных семантических связей. Этот список должен содержать множество соотношений типа *кровью* : *кровь*, *кровенной* : *кровь* и *гемоглобин* · *кровь*. В связи с этим интересно установить, до какой степени возможна автоматизация каждого шага обобщения. Обнаружение словоизменительных связей поддается автоматизации в силу их регулярного характера, в то время как гораздо менее регулярные словообразовательные связи нужно устанавливать, в основном, "вручную". То же в еще более высокой степени, конечно, относится к чисто семантическим, морфологически не выраженным, отношениям.

Исходя из названных соображений мы начали составлять список отношений на лексемном уровне. В этом списке каждая лексема возводилась к ближайшему, более простому семантическому соседу. Что в этом случае означает "более простой", объясним ниже. Таким образом, лексема *песчаный* была возведена к лексеме *песок*, лексема *книга* к лексеме *читать* и т.д. Вскоре, однако, оказалось во многих случаях целесообразным возводить лексему не только к одному, а сразу к двум другим, более простым, лексемам, с тем, чтобы ассоциативные связи были более полно отражены. Особенно ощущалась эта потребность применительно к сложным словам, которые очень типичны для шведского языка. Например, русскому словосочетанию *каменный дом* соответствует в шведском языке сложное слово *stenhus*, которое целесообразно возвести сразу к двум более простым лексемам *hus* "дом" и *sten* "камень".

На довольно раннем этапе стало очевидно, что образовывался своего рода тезаурус, где каждое слово было соотносено с одним или двумя более простыми словами. В том случае, когда невозможно было найти еще более простое слово, структура наверху "замыкалась" искусственной лексемой PRIM. Вся эта связанная структура приобретала иерархический характер благодаря тому, что вертикальное отношение естественно истолковать как отношение зависимости. На основе этих вертикальных отношений подчинения легко устанавливаются и горизонтальные – отношения соподчинения. Например, лексемы *желтый* и *красный* взаимосвязаны в силу того, что они оба зависят от лексемы *цвет*. Таким же образом, *поезд* и *машина* связаны общей подчиняющей лексемой *ехать* и т.д.

В качестве источников лексем были сначала использованы словники, полученные при автоматической обработке текстов, потом шведский частотный словарь, включающий 10.000 слов [1]. Кроме того, в ходе анализа приходилось добавлять многие слова, для того чтобы заполнять незанятые узлы. По мере того как словарь рос, становилось все более ясно, что лексика шведского языка будет отражена в нем довольно полно. [Тогда мы приняли радикальное решение: ввести в него все слова большого шведского толкового словаря [2], в результате чего в настоящее время АСШЯ включает свыше 70 000 слов.]

#### СТРУКТУРА СЛОВАРЯ

АСШЯ похож на тезаурус или даже является тезаурусом, только весьма необычного типа. Что такое тезаурус? Чем он отличается от других словарей? В традиционных словарях – двуязычных и толковых – дается значение слова, либо в виде эквивалента на другом языке, либо в виде толкования. Последнее можно рассматривать как эквивалент на том же языке, что и толкуемое слово. Толкование часто состоит из родового понятия и детерминанта, однако нередко в качестве толкования дается просто синоним, т.е. более знакомое слово, обозначающее то же

понятие. В словарях синонимов даются преимущественно только эквиваленты последнего типа.

Другими словами, традиционные словари основываются на принципе эквивалентности, в то время как тезаурусы основаны на более широком понятии семантического соотношения. Классическим примером и образцом является "Roget's Thesaurus of English Words and Phrases" [3], изданный впервые почти сто сорок лет тому назад. В 1930 г. была опубликована шведская версия этого словаря [4]. Тезаурусы этого типа обычно состоят из двух частей: основной части с понятийными классами, упорядоченными по какому-нибудь логическому принципу, и вспомогательной части, содержащей алфавитный список всех слов с указанием, к какому или каким понятийным классам они относятся. В одном тезаурусе подобного типа [5] слова, кроме того, снабжены короткими толкованиями. Число понятийных классов в названных тезаурусах – около 1.000. Подобную структуру имеет и "Лексическая основа русского языка" [6] (с педагогическим уклоном). Существуют также тезаурусы более сложного типа, например семантический словарь Ю.Н. Караулова [7], в котором применяется система дескрипторов и множителей, извлеченных из толкований традиционного словаря.

Стоит упомянуть еще один тип словаря, а именно словообразовательный словарь, в котором описываются деривационно-морфологические соотношения лексем, и одновременно те семантические соотношения, которые закреплены деривационной системой. Шведского словаря такого типа не существует, но что касается русского языка, то существует "Словообразовательный словарь русского языка" А.Н. Тихонова [8]. Заглавными словами статей этого словаря являются простые (=непроизводные) слова; под заглавным словом приводятся все непосредственно производные от этого слова; под каждым производным приводятся, в свою очередь, все производные от него и т.д., причем иерархическая глубина отмечается различными отступами вправо. Однако единицы одного уровня упорядочены традиционным образом, по алфавиту. Кроме того, словарь снабжен алфавитным указателем всех производных слов.

Упсальский словарь имеет характер тезауруса в том отношении, что он основан на семантических отношениях между словами, но в общем он резко отличается от всех описанных выше типов словарей. Он сходен с традиционными двуязычными и толковыми словарями в том отношении, что в нем имеется лишь один "вход", а именно, словарные статьи, упорядоченные по алфавиту. В этом словаре не нужен отдельный алфавитный указатель, т.к. каждая встречающаяся где-либо в словаре лексема выступает одновременно и в роли заглавного слова словарной статьи. Аббревиатуры тоже включены прямо в словарь, а не даны особым списком, как это обычно делается.

Понятие "семантического отношения" толкуются очень широко: включаются сюда и энциклопедические отношения, которые обычно не признаются языковыми. В словаре содержится около 3.000 слов, которые пишутся с заглавной буквы и чаще всего обозначают явления со специфической референцией. Речь идет, в частности, об именах и фамилиях политических деятелей, писателей, художников, о названиях городов, рек, гор, даже литературных произведений и художественных фильмов.

Со словообразовательным словарем настоящий словарь сходен в том отношении, что слова в нем образуют иерархическую структуру, которая, однако, отнюдь не основана только на степени морфологической сложности. Кроме того, как уже было сказано, иерархическая структура охватывает весь АСШЯ, в то время как в словообразовательном словаре она не простирается за пределы словообразовательных гнезд.

АСШЯ построен на двух, как мне кажется, вполне естественных положениях:

1) С точки зрения определенной лексемы L все остальные лексемы распадаются на семантически соотнесенные с L и несоотнесенные с L. Если представить себе, что лексемы расположены в семантическом пространстве, то соотнесенные с L лексемы можно назвать семантическими соседями лексемы L. Среди них есть более близкие и

более далекие, например, среди соседей лексемы *лошадь* можно считать лексемы *подкова* и *ржать* близкими соседями, а лексему *овес* – более далеким: понятия "подкова" и "ржать" однозначно связаны с понятием "лошадь", в то время как "овес" выступает как более независимое понятие (овес используется не только как корм для лошадей).

От констатации наличия семантической связи и определения ее силы можно, естественно, перейти к качественной характеристике данной связи. Однако для целей АСШЯ такой анализ не требуется. Что же касается направления связи между двумя соседями, то оно всегда должно быть установлено однозначно, но для этого я прибегаю не только к семантическим критериям (см. ниже).

Семантическое соседство напоминает то, что выявляется при помощи так называемого ассоциативного теста, однако последнее имеет более психологический или синтагматический характер. В [9] указаны такие связи, как, например, *книга: интересная, фамилия: лошадиная*. Подобного типа связи включать в наш словарь нецелесообразно, т.к. они не основаны на семантических межлексемных отношениях.

2) Лексика данного языка не представляет собой однородную массу; она имеет центр и периферию. Например, *вот* более центрально, чем *скот*, а это слово, в свою очередь, более центрально, чем *бегемот*. Большинство носителей языка сходятся во мнениях относительно иерархии "центральности", которая, по-видимому, определяется несколькими факторами. Главным фактором является общая частота слова; более часто употребляемые, и тем самым более известные слова, несомненно, воспринимаются как более центральные (ср. вышеприведенные примеры).

Что касается других факторов, то судить о них легче, если сравнивать семантически соотнесенные слова. Так, стилистически нейтральные слова воспринимаются как более центральные, чем их стилистически маркированные синонимы (*вкальывать : работать*). Большую роль играет и морфологическая сложность: производное слово более периферийно, чем слово, из которого (или – в случае сложных слов – из которых) оно произведено (*песчаный : песок*). Фразеологизмы возводятся к однословным синонимам (*сыграть в ящик : умереть*). Более или менее аксиоматически можно предполагать, что слова со специфической референцией менее центральны, чем соответствующие слова с генерической референцией (*Москва : город*).

Некоторые логико-семантические отношения тоже обладают зависимостной асимметрией и тем самым создают условия для иерархической структуры. Например, видовое понятие более периферийно, чем соответствующее родовое (*никель : металл*), название части, вообще говоря, более периферийно, чем название целого (*крыша : дом*). Направленность здесь устанавливается при помощи критерия, который вообще применим ко всем названным выше типам отношения: асимметрия следует из того факта, что число лексем, способных заполнять более центральные узлы, меньше, чем число лексем, способных заполнять более периферийные узлы. Другими словами, если "смотреть" в направлении центра, то однозначность, или детерминированность, больше, чем в противоположном направлении. Так, если исходить из лексемы *никель*, то однозначно попадаешь в лексему *металл*; если, с другой стороны, исходить из лексемы *металл* в пределах того же отношения, то попадешь в целый класс лексем: *никель, медь, железо* и т.д. С топологической точки зрения это и естественно: в периферии места больше, чем в центре.

Критерий детерминированности применим и в случаях отношения между названиями ситуаций и названиями объектов (партиципантов ситуаций), например, *ржать : лошадь, нож : резать*. Здесь играют роль соображения следующего типа: "ржать" неразрывно связано с "лошадью", в то время как "лошадь" выступает во многих других ситуациях, кроме "ржать"; "нож" специально предназначен для "реза-

ния", а "резать" можно и другими инструментами, нежели "ножом" (хотя для шведского языка это не так очевидно, как для русского).

Если дихотомия "центральный : периферийный" определяется многими факторами, то для того, чтобы можно было установить ее однозначно, требуется, чтобы эти факторы – хотя бы в большинстве случаев – сходились, т.е. были согласованы. К счастью, так и обстоит дело. Частота слова обратно пропорциональна его длине и морфологической сложности, сложное слово в типичном случае является гипонимом своего главного компонента.

Итак, если гипотезы 1) и 2) верны, то для каждой лексемы А в данном языке можно найти лексему Б, такую, что Б – близкий сосед лексемы А и одновременно занимает более центральную позицию, чем А. Эту лексему Б, которая лучше всех отвечает данным требованиям, мы называем "матерью" лексемы А, а исходную, более периферийную лексему А – "ребенком" лексемы Б. Например (сначала дается ребенок, потом – мать): *книга : читать; нос : лицо; крыша : дом; пыль : сухой*.

Отношение "ребенок : мать", конечно, асимметрично: у ребенка есть только одна мать, в то время как у матери может быть несколько детей. На этом факте и основана иерархическая структура словаря. Как уже было сказано, эта структура замыкается фиктивной лексемой PRIM. Дети одной матери образуют группу, связанную отношением семантического соседства: у них есть семантическое сходство, унаследованное от общей матери. Единственное исключение составляют дети лексемы PRIM (их около 50), которые могут и не быть семантически соотносительными.

Эта простая структура осложняется тем, что факультативно может быть указан и отец лексемы. Например, лексему *береста* желательно связать одновременно с лексемами *кора* и *береза*. *Ржать* желательно связать не только с лексемой *лошадь*, но и с лексемой *звук* для того, чтобы была образована группа *ржать, мычать, блеять, лаять, мурлыкать* и т.д. Особенно часто отцы указываются у сложных слов. Но тем не менее, большинство слов остаются без отца.

В принципе любая лексема может "вступить в брак" и при этом быть или матерью, или отцом. Исключение составляет PRIM, которая всегда выступает только в роли матери и в брак никогда не вступает. Одна и та же лексема может быть матерью по отношению к одной лексеме и отцом – по отношению к другой; однако никакая лексема не может вступить в брак сама с собой.

Исключительно важную роль при выборе родителей играют горизонтальные группы лексем. Именно они должны восприниматься языковой интуицией как естественные, даже "красивые" серии слов, и тем самым они служат при анализе важным коррективом. Можно даже сказать, что главная цель анализа заключается в создании оптимальных горизонтальных серий.

Поскольку для каждой лексемы указываются максимально два родителя (мать и – факультативно – отец), то образуются следующие семейные отношения (необходимо обратить внимание на то, что "пол" различается у родителей, но не у детей). Исходное слово называем "я", а его партнер – "ты":

- 11 ты моя мать
- 12 ты мой отец
- 21 я твоя мать
- 22 я твой отец
- 31 я твоя супруга
- 32 я твой супруг
- 40 у меня с тобой общие родители
- 41 у меня с тобой общая мать
- 42 у меня с тобой общий отец
- 43 моя мать – твой отец
- 44 мой отец – твоя мать.

Как уже было сказано выше, отношение ребенок : родитель, т.е. отношения 11 и 12, задаются вручную. Эти исходные данные представлены списком следующего типа:

книга : читать  
никель : металл  
золото : металл + желтый  
береста . кора + береза  
гараж : хранить + машина и т.д.

Компьютерные программы превращают этот исходный список в словарь со статьями, которые в распечатанном виде могут выглядеть по-разному. В более ранней версии они имели такой вид:

*хранить*: 12 иметь; 21 ангар, бумажник, гараж, кошелек, сейф, хранилище...; 31 деньги [3], машина, самолет...; 41 владеть, общий, собственный...

*ржать*: 11 звук; 12 лошадь; 21 ржание; 41 мычать, блеять, лаять...; 42 копыто, ехать верхом...; 43 издавать...; 44 лошадка, кобыла...

*Ленинград*: 11 город; 12 Россия; 22 петербуржец, ПГУ...; 32 житель, университет...; 40 Волгоград, Мурманск...; 41 Гетеборг, Мюнхен...; 42 Днепр, Горбачев...; 43 горсовет; 44 Сибирь...

П р и м е ч а н и е: после супруга *деньги* указывается в скобках, сколько раз заглавное слово вступает в брак с данным супругом: в результате трех "браков" между *хранить* и *деньги* образуются дети – *бумажник, кошелек, сейф*.

В более новой и упрощенной версии вторая и третья группы совмещены, а четвертая пропущена. Словарная статья выглядит так: *хранить*: 12 иметь; 21 хранилище, (машина:) гараж, (самолет:) ангар, (деньги:) бумажник, кошелек, сейф...

Несомненно, что вид статьи можно улучшить еще.

#### АНАЛОГИЯ С ГЕНЕАЛОГИЧЕСКИМ ДЕРЕВОМ

Для того чтобы сделать структуру АСШЯ более ясной, сравним ее с генеалогическим деревом и объясним, чем различаются эти структуры.

Как уже было сказано, в АСШЯ различие по "полу" проводится только у взрослых, но не у детей. В отличие от реальных отношений родства, очень распространен партеногенез: у большинства слов нет отца. Изобилует также многоженство/многомужество, причем одно и то же слово может вступать в брак как в качестве мужа, так и в качестве жены. Отношения типа 43 и 44 в реальной жизни не встречаются, так как вряд ли кто-нибудь рождает ребенка и до, и после смены пола.

Очень спорным является вопрос, следует ли в структуре АСШЯ разрешать "кровосмешение", т.е. допустимо ли больше, чем одно отношение родства между двумя лексемами. Иногда это кажется неизбежным и оправданным, например, в случае *парусник* : *судно* + *парус*; *парус* : *судно*; другими словами, *парус* выступает одновременно как супруг и ребенок лексемы *судно*. Подобное явление имеет место в шведском сложном слове *skogstrad* "лесное дерево", где *дерево* является одновременно супругой и матерью лексемы *лес*. Другой относительно распространенный и оправданный тип кровосмешения наблюдаем в случае шведского *tallskog* : *skog* + *tall* ("сосняк : лес + сосна"), где вступившие в брак лексемы *лес* и *сосна* имеют общую мать *дерево* (причем в таких случаях обязательно имеет место неодинаковое семантическое отношение между матерью и ребенком).

Однако иногда кровосмешения следует избегать, как в случае шведского сложного слова *hustak* "крыша дома", которое можно считать синонимом простого слова *tak* "крыша" (таких плеоназмов или почти плеоназмов в шведском языке довольно много); здесь мы просто возводим менее центральный член синонимичной пары к более центральному, и кровосмешения не возникает (ср. *tak* "крыша": *hus* "дом").

У всех лексем, кроме PRIM, есть один или два родителя. Детей у любой лексемы может быть сколько угодно, включая нуль (сухие ветви встречаются не только в периферии, но и в центральных позициях). Другими словами, по отношению к вертикальным связям лексемы распадаются на восемь типов, в отличие от людей,

которые распадаются лишь на три (у людей обязательно два родителя, и ни один человек не может быть одновременно матерью и отцом).

Структура генеалогического древа является динамической: она отражает течение времени, причем все единицы системы живут приблизительно одно и то же время и сменяют друг друга по вертикальной оси. Лексику языка мы, напротив, рассматриваем как статическую структуру, где положение каждой лексемы определяется исключительно ближайшими соседями и относительным расстоянием от центра. Топологически здесь более уместно говорить о круге или шаре, чем о дереве.

Если смотреть на лексику с хронологической точки зрения, то можно констатировать, что лексемы ведут себя следующим образом: а) живут крайне неодинаковое время; б) за время своего существования проникают в позиции, неодинаково близкие от центра; в) могут сменять родителей, т.е. превращаться в другие лексемы.

#### СРАВНЕНИЕ СО СТАНДАРТНЫМ ТОЛКОВЫМ СЛОВАРЕМ

В обычном толковом словаре заглавное слово снабжено морфологической информацией, одним толкованием для каждого значения слова и примерами, показывающими его употребление. В АСШЯ заглавное слово дается вместе с его семьей, т.е. с родителями, детьми, супругой/супругом и братьями/сестрами. Есть, конечно, определенное сходство между толкованием традиционного словаря и родителем или родителями АСШЯ. Они могут даже совпадать: это происходит в том случае, когда в качестве толкования (соответственно, в качестве матери) дается более знакомый, т.е. более центральный, синоним, ср., например, в [10] *оплеуха*: "То же, что пощечина". Последнее, в свою очередь, толкуется как "Удар по щеке ладонью". Соответствующий анализ в АСШЯ может опираться на это толкование и выбрать, например, в качестве матери слова *пощечина* лексему *удар* и в качестве отца лексему *щека* (или *ухо*). Информация, заложенная в синтаксической структуре толкования при этом, конечно, теряется; остается только указание на то, какой член пары главный и какой – детерминирующий.

Однако в общем разница между двумя типами словарных статей очень большая. Это касается, прежде всего, самого понятия лексемы. В традиционном словаре допускается и широко распространена полисемия, т.е. наличие нескольких означаемых у одного означающего. АСШЯ в этом отношении на той же позиции, что и Комбинаторно-толковый словарь И.А. Мельчука и А.К. Жолковского [11], т.е. различные означаемые являются поводом для расщепления данных единиц на разные лексемы. Эти лексемы оснащаются индексами, например *правило*<sup>1</sup>: *линейка* + *строить*; *правило*<sup>2</sup>: *весло* + *управлять*. Почти всегда они имеют различных родителей, но есть случаи, когда у них общий родитель, например, швед. *växt*<sup>1</sup> "растение" и *växt*<sup>2</sup> "рост" (с резко отличающимися друг от друга группами детей), которые оба возводятся к лексеме *växa* "расти".

Расщепление означающих проводится и в традиционном словаре в случаях, когда полностью отсутствует связь между соответствующими означаемыми, но интересно отметить, что в толкованиях это расщепление никогда не отмечается, так как подразумевается, что из контекста всегда ясно, какая лексема имеется в виду. В тезаурусе индексированные лексемы всегда оснащены своим индексом, независимо от того, выступают они в качестве заглавного слова или нет.

Как уже было указано, еще одна существенная разница между традиционным словарем и АСШЯ заключается в том, что все слова, встречающиеся где-либо в АСШЯ, обязательно выступают также как заглавные слова отдельных статей: этот словарь имеет полностью замкнутую структуру. Традиционные же словари часто содержат в своих толкованиях слова, которые сами как заглавные слова не выступают. Особенно это верно применительно к сложным словам в шведском языке. Эта лексическая единица отчасти находится на границе между лексической и

синтаксической областями языка. Новые сложные слова очень легко образуются в шведском языке, причем они часто ничем не поражают, никак не маркированы. Такие слова совершенно сознательно не даются как заглавные слова. Но обычно есть и много случаев, где в толкованиях употребляются слова, которые нужно бы дать в качестве заглавных, но которые просо упускаются из виду. В АСШЯ с его строгим контролем такие упущения исключены.

Уппсальский АСШЯ, благодаря тому, что на начальном этапе слова черпались непосредственно из словариков конкретных текстов, содержит довольно много слов, которые – по недосмотру или сознательно – не представлены в традиционных словарях. Речь идет прежде всего о двух следующих типах слов: сложные слова и собственные имена. Сложные слова, как уже было сказано выше, чаще всего имеют совершенно нормальный вид, но могут, конечно, иметь маркированный, окказиональный характер. Собственные имена, как известно, встречаются в толкованиях традиционных словарей, но как заглавные слова принципиально не даются.

Задумавшись над тем, что представляет собой толкование. Оно, в первую очередь, предназначено для того, чтобы объяснить значение незнакомого слова при помощи знакомых слов. На такую функцию толкования выполняют только применительно к довольно периферийным словам, которые легко толкуются через более центральные. Толкования же центральных слов имеют другой характер. Они уже не объясняют значение слова, а только определяют его, а именно частично или целиком при помощи менее знакомых слов. Практическая функция таких толкований, как, например, "орган обоняния, находящийся на лице человека, на морде животного" [10] скорее заключается в том, что а) выделяется одно из значений у многозначного слова; б) определяется объем данного значения (например, в шведском языке для обозначения носов у животных употребляется специальное слово).

С точки зрения характера толкования лексика, представленная в толковых словарях, распадается на два слоя: периферийный и центральный. Требование, чтобы каждое слово было снабжено определением, независимо от того, выполняет ли оно объяснительную функцию или нет, неизбежно приводит к порочным кругам, избегать которых всякий лексикограф старается, но, на мой взгляд, по сути дела, тщетно. Далеко не всегда было осознано, что определение не равняется толкованию (в смысле объяснения).

В АСШЯ никакого такого расслоения не происходит. То, что соответствует толкованию, т.е. родитель/родители, имеет совершенно одинаковый характер везде, так как всегда соблюдается требование, чтобы в качестве родителя выбирался более центральный семантический сосед. Объяснительная сила родителя остается неизменной. Единственная необходимая аксиома заключается в том, что самой центральной лексемой является PRIM, к которой возводятся в конечном счете все остальные лексемы.

С названным различием связано еще одно: с точки зрения определенной лексемы в традиционном словаре можно "смотреть" в принципе только в одном направлении, а именно или в направлении центра (для периферийных слов), или в направлении периферии (для центральных слов). В АСШЯ же всегда открывается вид в трех направлениях: в центр (вверх), в периферию (вниз) или в сторону (горизонтально).

#### **ЗАЧЕМ НУЖЕН АССОЦИАТИВНЫЙ СЛОВАРЬ?**

Наконец, я попытаюсь ответить на вполне оправданный вопрос: зачем нужен такой словарь, как АСШЯ? Пожалуй, слишком просто было бы указать только на тот факт, что организация лексики по содержанию, т.е. по семантическим принципам – старая мечта лексикографов и философов. Решение этой задачи, однако, имеет и вполне практическое значение. В первую очередь можно указать на следующие возможные применения:

1) При сочинении текста нередко случается, что знаешь, что хочешь сказать, но никак не находишь правильные слова. В таких случаях иногда пользуются синони-



мическим словарем. Богатый ассоциативный словарь, наверное, способен оказать еще более эффективную помощь. Разница между традиционным толковым словарем и ассоциативным словарем состоит в том, что в первом случае пользователь ищет значения, исходя из слов, а во втором – он ищет слова, исходя из значений, представленных словами-соседями. Первый тип словаря помогает при пассивном, рецептивном употреблении языка, второй – при активном, продуктивном (ср. здесь словарь [11], который тоже относится к "активному" типу);

2) Кроссворды и другие языковые игры являются популярным занятием многих людей. Издаются словари и энциклопедии, специально предназначенные для этой категории. АСШЯ содержит немало энциклопедической информации, прежде всего каскады ассоциаций, причем без ограничений, касающихся характера семантической связи или направления поиска;

3) Исследователи языка нуждаются в богатом эмпирическом материале. В последнее время наблюдается усиление внимания к лексическому компоненту описания языка. Растет интерес к машинным фондам и словарям нетрадиционного типа. АСШЯ является сокровищем для лексических исследований. Здесь предлагаются объемные серии слов, связанных по содержанию. Мы не исключаем даже, что сама структура словаря может пролить свет на некоторые трудные проблемы лексикологии, например, вопрос выделения и отграничения значений (в случаях полисемии).

4) Для многих других областей знания, кроме собственно лингвистики, очень важно, чтобы общая, лексически закрепленная, когнитивная сеть носителей определенного языка была исследована как можно более детально. Речь идет, в частности, о таких новых областях как искусственный интеллект, экспертные системы, информационный поиск и машинный перевод. Для шведского языка АСШЯ, благодаря богатству представленного в нем материала, сможет внести значительный вклад в дело этой инвентаризации<sup>1</sup>.

#### СПИСОК ЛИТЕРАТУРЫ

1. *Allén S* 10.000 i topp. Ordfrekvenser i tidningstext, Stockholm, 1972.
2. *Svensk ordbok* / Ed. by Allén S et al., Stockholm, 1986.
3. *Roget's Thesaurus of English Words and Phrases* / New edition prepared by Kirkpatrick B. L., 1987.
4. *Bring S.C.* Svenskt ordföråd ordnat i begreppsklasser. Uppsala, 1930.
5. *Lewis N.* A Modern Thesaurus of English Synonyms. N.Y., 1958 (1965).
6. *Лексическая основа русского языка* / Под ред. Морковкина В.В. М., 1984
7. *Караулов Ю.Н.* Русский семантический словарь М., 1982.
8. *Тихонов А.Н.* Словообразовательный словарь русского языка. I-II. М., 1985.
9. *Словарь ассоциативных норм русского языка* / Под ред. Леонтьева А.А., М., 1977.
10. *Ожегов С.И.* Словарь русского языка. Изд. 22-е. М., 1990.
11. *Мельчук И.А., Жолковский А.К.* Толково-комбинаторный словарь современного русского языка. Вена, 1984.

<sup>1</sup> Автор выражает сердечную благодарность И.А. Мельчуку, который прочитал предварительный вариант статьи и сделал ряд ценных замечаний.