

ГУМАНИТАРНЫЕ НАУКИ В ОБРАЗОВАНИИ

А. И. Глазырина

МЕТОДИКА СОСТАВЛЕНИЯ СВОДНОГО СПИСКА БАЗОВЫХ ТЕРМИНОВ РУССКОГО КОМПЬЮТЕРНОГО ПОДЪЯЗЫКА

В статье описывается методика составления списка базовых терминов на примере терминосистемы русского компьютерного подъязыка; обосновывается необходимость привлечения данных частотных, дескрипторных, толковых, специальных словарей, а также специальных толковых словарей смежных дисциплин.

Целью данной статьи является демонстрация впервые разработанной методики выделения базовых терминов русского компьютерного подъязыка (РКП).

Компьютерный подъязык находится в стадии формирования и содержит много предтерминов и окказионализмов. Современные словари русского компьютерного подъязыка насчитывают до 50–60 тыс. единиц, которые дополняются сотнями тысяч компьютерных новообразований в текстах. Возникла проблема определения границ русского компьютерного подъязыка. Следовательно, мы выделяем в данном массиве терминов ядро и периферию.

Исследователи данной предметной области сталкиваются с проблемой выделения именно базовой лексики, так как исследование всего массива компьютерных терминов практически нереально в силу, во-первых, многочисленности входящих в данный подъязык единиц, во-вторых, открытости данной терминосистемы инновациям и новым пополнениям.

При обучении языку для специальных целей (в англоязычной литературе – *LSP (Language For Special / Specific Purposes)*) обучающиеся вынуждены дважды преодолевать «языковой барьер»: совершенствовать уже полученные навыки и умения или осваивать новый для них иностранный язык (в нашем случае английский), а также новую предметную область, если они еще не имеют соответствующего профессионального образования. Именно поэтому важно начать обучение с того лексического минимума, который необходим будущему специалисту в его профессиональной деятельности. Данная методика позволяет решить поставленную задачу: составить список базовых терминов определенного подъязыка. Наша методика отличается универсальностью, так как позволяет составлять списки базовых терминов, во-первых, в любой предметной области, во-вторых, в любом языке.

Перейдем к описанию понадобившихся нам словарей, так как предметная область компьютерного подъязыка все еще «ищет себя». Более того, РВП основывается на старой, жесткой, стандартизированной классификации, уходящей в 70–80-е гг., когда информатика и документалистика были двумя сторонами одной медали.

При отборе материала мы использовали частотные словари¹: «Русско-английский частотный словарь по электронике» (И. И. Убин), «Частотный словарь индексирования» (Л. В. Сахарный – около 4200), «Английский частотный словарь по кибернетике» (Р. В. Адамов, Л. Н. Александрова, В. М. Андрющенко), «Частотный словарь английского подъязыка электронно-вычислительной техники» (В. А. Букович), «Частотный англо-русский словарь-минимум по электронике» (М. П. Алексеев), «Частотный словарь английского подъязыка радиоэлектроники» (А. П. Лясковская, Т. Н. Хомутова, Л. Б. Пацеля, И. Б. Самохвалова), «Актуальные проблемы терминологии по информатике и документации. Частотные словари терминов»; дескрипторные: «Дескрипторный словарь по информатике» (Э. Г. Азгальдов, О. А. Высочанская, М. И. Хаскина – 2241), «Дескрипторный словарь по информатике» (Н. А. Пашенко, Е. Б. Ксенофонтова, В. Ф. Скоробогатая); толковые: «Толковый словарь современной компьютерной лексики» (В. Л. Дорот, Ф. А. Новиков – более 2500), «Толковый англо-русский и русско-английский словарь компьютерных терминов» (А. Косцов, В. Косцов – более 5000), «Русско-английский толковый словарь по информатике» (В. И. Першиков, А. С. Марков, В. М. Савинков – 6000), «Большой толковый словарь компьютерных терминов» (А. Синклер), «Англо-русский словарь по программированию и информатике (с толкованиями)» (А. Б. Борковский), «Толковый словарь по вычислительной технике и программированию» (А. П. Заморин, А. С. Марков, М. Р. Шура-Бура), «*PC networking handbook*» (E. Tittel); русско-английские индексы в англо-русских словарях по компьютерному подъязыку: «Англо-русский толковый словарь терминов и сокращений по ВТ, Интернету и программированию (Э. М. Пройдаков, Л. А. Теплицкий – примерно 30000 статей), «Англо-русский словарь по вычислительной технике и программированию (Е. К. Масловский – 52000), «Термины современной информатики: программирование, вычислительная техника, Интернет (Е. Ю. Ваулина, В. Н. Рычков – более 6000), «Вычислительная техника и новые информационные технологии» (1400), «Англо-русский словарь по вычислительной технике» (В. К. Зейденберг, А. Н. Зимарев, А. М. Степанов – 37000), «Англо-русский словарь по вычислительной технике» (В. К. Зейденберг, Ю. Л. Зиман, А. Н. Зимарев. – 24000), «Англо-русский словарь пользователя персональной

¹ В скобках указано количество содержащихся в том или ином издании терминов (дескрипторов).

ЭВМ» (Е. К. Масловский – 10000), «Словарь терминов по информатике на русском и английском языках» (Г. С. Жданова, Е. С. Колобродова, В. А. Полушкин, А. И. Черный); многоязычные словари: «Словарь по электротехнике: английский, немецкий, французский, нидерландский, русский» Я. Н. Лучинский, Б. А. Алексеев, Б. Махлин), «Русско-англо-французский словарь по информационной теории и практике» (Г. С. Жданова, Е. С. Колобродова, В. А. Полушкин и др.), «Терминологический словарь по информатике (многоязычный)».

Кроме того, в ходе исследования потребовалось использовать и некоторые другие словари: «Комплексный частотный словарь русской научной и технической лексики» (П. Н. Денисов, В. В. Морковкин, Ю. А. Сафьян).

Вышеприведенные источники свидетельствуют о наличии огромного терминологического пласта лексики компьютерного подязыка. Для нашего исследования мы попытались выделить примерно 2 тыс. ед. Мы считаем, что выборка в 2 тыс. ед. соответствует целям и задачам нашего исследования, так как чрезмерно большой объем материала для исследования (20–30 тыс. терминов) может поставить под сомнение достижение каких бы то ни было результатов.

Как справедливо заметил В. Т. Титов, «познать лексико-семантическую систему языка – это прежде всего минимизировать ее, но не произвольно, а с сохранением ее своеобразия, заключенного в важнейших существенных параметрах». В. Т. Титов указывает на количество в 1–2 тыс. ед., которые являются ядерными единицами [4].

Общеизвестно, что «отобранная часть элементов совокупности (выборка) будет представлять всю совокупность с приемлемой точностью при двух условиях: она должна быть достаточно многочисленной, чтобы в ней могли проявиться закономерности, существующие в генеральной совокупности; элементы выборки должны быть отобраны объективно, независимо от воли исследователя, чтобы каждый из них имел одинаковые шансы быть отобранным или же чтобы шансы эти были известны исследователю». И далее: «представительность, или репрезентативность, выборки обеспечивается не только ее объемом, но и строгим соблюдением научно обоснованных правил отбора, гарантирующих его объективность» [2].

К традиционным словарям (изданным в «бумажном» виде), которыми мы воспользовались, мы добавили интерактивный «Русско-английский глоссарий по информационному обществу» (О. Н. Вершинская, Ю. Д. Вольнский, Т. В. Ершова, Н. В. Кривошеин, А. С. Мендкович, М. В. Моисеева, С. А. Нехаев, Г. Л. Смолян, Ю. Е. Хохлов, Д. С. Черешкин, С. Б. Шапошник). Данный глоссарий является совместным проектом Британского Совета в России, Института развития информационного общества и проекта «Российский портал

развития» (грант # CG 012 программы *infoDev* Всемирного Банка). Мы исключили из глоссария те термины, которых нет в «Англо-русском словаре по вычислительной технике и программированию» под редакцией Е. К. Масловского (2004) и «Англо-русском толковом словаре терминов и сокращений по ВТ, Интернету и программированию» под редакцией Э. М. Пройдакова, Л. А. Теплицкого (2005). Из 100 терминов в нашем сводном словнике оказалось 46 единиц.

Необходимость выделения базовой лексики (1–2 тыс. ед.) для нашего исследования заставила обратиться к лингвистическим принципам отбора единиц для списка базовых терминов русского компьютерного подъязыка. Из существующих шести принципов (по классификации И. В. Рахманова) мы, прежде всего, руководствовались принципом частотности, являющимся основополагающим в процедуре выделения базовых терминов. Использование частотных словарей компьютерного подъязыка позволяет выделить ядро русского компьютерного подъязыка, отделить его от периферии. За основу мы взяли принятый в статистической лексикографии порог надежности, равный 35.

Однако использование лишь критерия частотности не может считаться единственным и достоверным инструментом в подобного рода исследованиях. Дополнительным критерием, который позволяет выявить слова, быть может, малочастотные, но функционально значимые, является принцип выделения дескрипторов – смысловых доминант. Дескрипторные словари по информатике послужили вторым необходимым этапом в многоступенчатой методике создания списка базовых терминов русского компьютерного подъязыка.

Методика составления сводного списка базовых нормативных терминов русского компьютерного подъязыка наглядно демонстрирует 7 этапов, составляющих основу процедуры отбора базовых нормативных терминов русского компьютерного подъязыка.

Первым этапом в составлении сводного словника нашего исследования явилось сопоставление словников 11 отраслевых русскоязычных словарей друг с другом.

Для того чтобы определить, сколько пар словарей нам придется сравнивать, воспользуемся упрощенной формулой вычисления суммы арифметического ряда (Σ): $\Sigma = n(n-1)/2$, где n – количество словников словарей для попарного сравнения. Следовательно, для получения объективного списка базовых терминов нам понадобится сравнить 55 словарей.

Русско-английский словарь по электронике под редакцией И. И. Убина состоит из двух частей. Мы использовали материал второй части, включающей (в отличие от первой, содержащей все лексические единицы корпуса текстов русских переводов с частотой более 1) наиболее информативные для выбран-

ной тематики лексические единицы с частотой I. Данный словарь не подходит по временным рамкам, но он «отвечает» за сектор «электроника».

Сопоставив 55 словарей друг с другом, мы обнаружили количество совпадающих в той или иной паре единиц (существуют пары, в которых совпадает 1 слово или вообще нет совпадений). Данные сведения содержатся в табл. 1.

Процентное соотношение между общим количеством терминов в паре сравниваемых словарей и количеством совпавших терминов вычисляется путем деления количества совпавших терминов на количество терминов, содержащихся в 1% от общего числа терминов.

Таблица 1

Количество совпавших лексем при сопоставлении словарей друг с другом

	A	B	C	D	E	F	G	H	I	J	
B ¹	44										
C	119	77									
D	56	24	121								
E	3	0	1	0							
F	119	45	375	224	0						
G	108	45	570	271	1	705					
H	96	77	789	136	1	320	436				
I	20	19	122	12	0	28	41	85			
J	7	2	107	18	0	15	36	43	8		
K	89	36	369	204	0	349	607	319	37	25	
Итого словарей	10	9	8	7	6	5	4	3	2	1	55

Как видно из табл. 1, наибольшее количество совпавших терминов содержится в следующих словарях, представленных в виде множе-

¹ Буквенными индексами обозначены использованные русскоязычные терминологические словари: А – Азгальдов Э. Г. Дескрипторный словарь по информатике. – М., 1991; В – Актуальные проблемы терминологии по информатике и документации. Частотный словарь терминов. – М., 1982; С – Ваулина Е. Ю. Термины современной информатики: программирование, вычислительная техника, Интернет: англо-русский словарь: более 6000 слов, устойчивых словосочетаний и сокращений. – М., 2005; D – Вычислительная техника и новые информационные технологии. – М., 1992; E – Глоссарий по информационному обществу. – М., 2001; F – Дорот В. А. Толковый словарь современной компьютерной лексики. – М., 2004; G – Косцов А., Косцов В. Толковый англо-русский и русско-английский словарь компьютерных терминов. – М., 2004; H – Першиков В. И. Русско-английский толковый словарь по информатике. – М., 1999; I – Прохоров К. Я. Англо-русский словарь по микроэлектронике. – М., 1985; J – Русско-английский частотный словарь по электронике / Под ред. И. И. Убина. – М., 1977; K – Синклер А. Большой толковый словарь компьютерных терминов. – М., 1998;

лей: Н×С, F×G, K×G, G×С, G×Н, F×С, С×K, F×K, F×Н, Н×K. Итак, наиболее содержательными оказались словники словарей, имеющих буквенные индексы С, D, F, G, Н, K, то есть словари, датированные 1998–2005 гг.¹

Во-первых, наименьшее количество совпавших терминов содержится во множествах, образованных словарями по микроэлектронике и электронике. Данные словари вошли в исследуемую нами предметную область, так как в советское время предметная область компьютерного подъязыка еще не выделялась отдельно. Во-вторых, наименьшее количество совпавших терминов содержится во множествах, образованных глоссарием по информационному обществу. Данный факт можно объяснить тем, что смысловыми доминантами данного словника являются термины глобализация, информатизация, компьютеризация.

Процентное соотношение между общим количеством терминов в паре сравниваемых словарей и количеством совпавших терминов можно представить следующим образом (табл. 2):

Таблица 2

Количество совпавших лексем при сопоставлении словарей друг с другом (процентное соотношение), %

	A	B	C	D	E	F	G	H	I	J
B	6,6									
C	2,5	1,7								
D	3,3	1,7	2,2							
E	0,6	0	0,02	0						
F	4,9	2	6	7	0					
G	2,2	0,9	6,5	4,7	0,02	10,9				
H	5,2	4,85	14	5,2	0,06	9,5	4,2			
I	1,2	1,4	2,2	0,5	0	0,9	0,7	3,3		
J	0,9	0,4	2,3	1,2	0	0,6	0,7	2,5	0,5	
K	3,7	1,7	6	6,4	0	8,9	9,4	9,6	1,2	1,1

На втором этапе мы анализируем те списки, которые получились в результате сопоставления 55 пар словников словарей. Это позволяет выявить те термины, которые имеют хотя бы одно совпадение в этом сводном списке. Данный сводный список насчитывает 2037 единиц, т. е. 27,7% от общего числа терминов на первом этапе (7361 единица). На данном этапе мы также удаляем все повторяющиеся термины, например: слово *компьютер* повторялось 45 раз, *принтер* – 38, *дискета* – 30 и т. д.

¹ Сюда, однако, входит и словарь, изданный в 1992 г., который до сих пор не теряет своей актуальности.

На третьем этапе мы добавили к сводному списку базовых терминов сводный список терминов, полученный из текстовых источников. Для этого мы составили с помощью специальной компьютерной программы частотный словарь заголовков классификатора ISO исследуемой нами предметной области. Данный список насчитывает 189 ед. Сопоставив объединенный список терминов из вышеупомянутых словарей и объединенный список терминов из текстовых источников, мы обнаружили всего 18 дополнительных терминов (сюда вошли такие функционально значимые термины, как *компьютерная сеть, хеширование, портативный компьютер, электронный*).

На четвертом этапе мы посчитали целесообразным исключить все общенаучные и технические термины, так как они не являются отраслевыми терминами. Для этого мы специально провели сверку нашей редакции сводного словника и «Комплексного частотного словаря русской научной и технической лексики» (П. Н. Денисов, В. В. Морковкин, Ю. А. Сафьян) с целью исключения общенаучных и общетехнических терминов. Вышеуказанная процедура показала, что предварительная редакция сводного словника содержала 37 общенаучных и технических терминов.

На пятом этапе мы «вычли» из данного списка (2018 ед.) номенклатурные наименования, опираясь на признаки номенов, выделенных З. И. Комаровой [3, с. 8]. Всего в нашем списке базовых терминов оказалось 30 подобных единиц.

На шестом этапе мы «вычитаем» из сводного списка те лексемы, которые не содержат ни одного контактного элемента английского происхождения¹. Данная группа насчитывает 182 ед.: *автоматизированный, авторитетный, адаптация, адаптивная дифференциальная импульсно-кодированная модуляция, адресный, адресуемость, анализатор, аномалия, база, базирование, бесшумный, блокнот, бобина, буквица, возвращение, время, встраивать, выключать, выравнивать, вычисление, громкость, группировать, дата, двуправленный, действенный, держатель, деструктор, десятичный, детектор, дефис, дискретный, диффузонт, доставлять, доступность, завершение, зависимый, загрузчик, закомментировать, замещение, запоминание, запуск, запятая, звук, идентичность, издание, измеритель, изолятор, имитировать, инверсия, инвертирование, индикация, индустрия, инициализация, категоризация, квитанция, квитирование, классификация, когерентный, коллизия, компромисс, конфигурирование, концепция, копия, корректность, коррекция, кривая, кристалл, лексика, лингвистика, лицензирование, манипулятор, ма-*

¹ Под контактным элементом (контактемой) мы понимаем активированную и/или отображаемую на определенном языковом уровне единицу языка-источника в языке-адресате.

нипуляция, масштабировать, масштабируемость, мигание, многоязычный, многозадачность, многопоточность, множественный, модификатор, назначение, наращивание, неоднозначность, неполнота, непротиворечивость, несовместимый, обеспечение, обработка, обслуживание, обходить, оглавление, однонаправленный, оригинал, оснастка, откладывать, отсечка, оцифровка, оцифровывать, пара, пассивный, переадресация, перевод, перезапуск, перекрестие, перемена, переназначение, печать, подгонка, подзадача, подложка, подсеть, позиционирование, полнота, полноэкранный, полупроводник, полу-чатель (приемник), поставка, строитель, посылка, прогнозирование, программируемый, пульсировать, разбор, разборка, разделитель, разъединение, расписание, распознавание, распространение, растривание, рекомендация, реферат, рисование, робо (то) техника, свертка, свертывание, секретность, символ-заполнитель, система, системотехника, ситуация, скелет, скрайбирование, скрывать, слежение, собственность, совместимый, согласованность, сопоставлять, сортировать, срочный, станция, стратегия, структурирование, схема, считывание, твердотельный (твердотельный), техник, травление, тракт, тройка, тройной, удаленный, упаковка, упорядоченность, усечение, услуга, утилизация, фактор, факультативный, фиксатор, фильтрация, фонд, формализация, форматировать, форм-фактор, функционирование, четкость, штырек, эквивалент, экспертиза, эскиз.

Особую группу составляют 5 слов (*закомментировать, многопоточность, перезапуск, перекрестие, подложка*). В словари данные термины еще не включены и, соответственно, не обозначены пометой *информ*. Мы считаем, что, несомненно относясь к терминам информатики (ниже мы приводим их толкования из новейших специальных толковых словарей), они в скором времени будут включены в толковые словари русского языка. *Закомментировать* (разг.) – в программировании – поставить знаки комментария в начале и конце секции кода или в начале каждой строки [1]; *многопоточность* – метод, используемый при создании многопоточных *СОМ*-объектов, допускающий выполнение нескольких потоков одновременно внутри данного *СОМ*-объекта; является альтернативой *apartment model* [1]; *перезапуск* – повторный запуск программы [6, с. 362]; *перекрестие* – пересекающиеся линии, используемые в некоторых устройствах ввода информации в компьютер (например, планшет) для указания координат конкретной точки [1]; *подложка* – 1. Пластина, на которой формируется интегральная микросхема. 2. материал в печатных платах, на который наносятся элементы электронной схемы, или материал, на который наносится магнитный слой при изготовлении магнитного диска, ленты [5, с. 220].

По мере уровневого анализа контактом мы неизбежно сталкивались с синонимами и дублетами.

На седьмом этапе мы исключили 249 дублетов, перифраз и абсолютных синонимов из нашего сводного списка, выделив таким путем базовые нормативные термины русского компьютерного подъязыка, например: *режим разделения времени и разделение времени, зарезервированное слово и ключевое слово, носитель данных и носитель информации, прямой доступ и произвольный доступ*.

Итак, окончательная редакция сводного списка насчитывает 1557 ед., исключая 10 гибридов¹ (например, *архиватор, ассемблирование, однопроцессорный, подменю, полубайт*).

Использование при отборе терминологической лексики дополняющих друг друга лингвостатистического и качественного анализов обеспечивает правильность, объективность и методическую целесообразность отбора терминологических единиц.

Мы надеемся, что предлагаемая методика выделения базовых терминов компьютерного подъязыка может быть применена и в изучении других подъязыков.

Литература

1. Англо-русский толковый словарь терминов и сокращений по ВТ, Интернету и программированию, 1998–2004 гг., Э. М. Пройдаков, Л. А. Теплицкий. 12,5 тыс. статей.
2. Большая Советская Энциклопедия. – М.: Советская энциклопедия, 1974. – <http://slovari.yandex.ru>
3. Комарова З. И. Структура специального слова и ее лексикографическое описание: Моногр. – Свердловск; Изд-во Урал. гос. ун-та, 1991. – 156 с.
4. Титов В. Т. Общая квантитативная лексикология романских языков / В. Т. Титов. – Воронеж: Изд-во Воронежск. гос. ун-та, 2002. – 237 с.
5. Косцов А., Косцов В. Толковый англо-русский и русско-английский словарь компьютерных терминов. Более 5000 слов и словосочетаний. – М.: Мартин, 2004. – 304 с.
6. Дорот В. Толковый словарь современной компьютерной лексики: Свыше 2000 терминов и словосочетаний / В. Дорот, Ф. Новиков. – 2-е изд., перераб. и доп. – СПб.: БХВ – Петербург, 2002. – 512 с.

¹ Под гибридом мы понимаем «скрещенное» слово, составленное из разноязычных элементов.