

# Взаимодействие лексики и терминологии в общезначимой сфере языка<sup>1</sup>

Б.В. Добров, Н.В. Лукашевич

Научно-исследовательский вычислительный центр МГУ им. М.В. Ломоносова;

НО Центр информационных исследований

119992, Москва, Ленинские горы, НИВЦ МГУ, 339

[louk@mail.cir.ru](mailto:louk@mail.cir.ru), [dobroff@mail.cir.ru](mailto:dobroff@mail.cir.ru)

В работе утверждается, что существует широкий промежуточный слой между традиционно рассматриваемыми отдельно общей лексикой и терминологией. Этот слой мы называем «общественно-политической областью». Общественно-политическая область включает в себя лексику и терминологию, которая, с одной стороны, известна достаточно широкому слою населения, с другой стороны, соответствует понятиями профессиональных сфер деятельности. В статье рассматривается структура общественно-политической области языка, ее пересечение со специальной, научной лексикой. По результатам использования тезауруса РуТез делается вывод о важности указанной области для автоматической обработки текстов

## 1. Введение

подавляющее большинство текстов, хранимых в современных электронных коллекциях и нуждающихся в эффективной обработке и поиске, принадлежат к так называемой деловой прозе и содержат как общезначимую лексику, так и терминологию конкретных предметных областей. Между тем общезначимые слова и термины изучаются представителями совершенно различных научных дисциплин.

Основоположник Венской школы терминологии Вюстер (Wuster, 1979) подчеркивал, что одно из существенных различий между методами исследования, используемых лингвистами и терминологами, заключается в том, что терминологи начинают свое рассмотрение с понятия, которое должно быть точно определено и не зависит от своего наименования, а лексикологи начинают с языкового выражения. Поэтому традиционно терминологи говорят о понятиях, а лингвисты о значениях.

Вместе с тем в настоящее время в рамках исследований, связанных с построением для различных языков иерархических систем значений типа WordNet (Climent et al., 1996; Miller et al., 1990), происходит сближение двух подходов. Создавая wordnet'ы для своих языков, лингвисты выстраивают значения слов и языковых выражений в виде иерархических систем, пытаются найти схожие понятия для различных языков, выстраивают верхний независимый от языка уровень таких систем, пытаются использовать созданные структуры

---

<sup>1</sup> Эта работа частично выполняется при поддержке гранта # 03-01-00472 Российского фонда фундаментальных исследований.

для общезначимого языка как заготовку для автоматизированного выявления понятийных систем в конкретных предметных областях (Vossen, 2001; Buitellar, Sacalenu, 2001).

Для повышения качества автоматической обработки текстов деловой прозы необходимы исследования совместного функционирования лексико-терминологических понятийных систем, принципов разработки смешанных лексико-терминологических ресурсов, механизмов использования общезначимых лингвистических ресурсов для порождения предметно-ориентированных ресурсов

В статье мы утверждаем, что существует широкая предметная область, которая занимает промежуточное положение между общезначимой системой понятий и профессиональными понятийными системами, частично с ними пересекаясь. Мы называем эту область «общественно-политической», поскольку она содержит понятия, значимые для современного общества в целом, а именно понятия общезначимого языка, практически совпадающие с понятиями конкретных терминологических систем, и профессиональные понятия, понятные носителям языка.

## **2. Отличительные черты термина**

Большинство терминов рассматривают термин как слово или выражение, обозначающее понятие предметной области. Совокупность терминов предметной области образуют ее терминологическую систему. Отношение конкретного термина с другими терминами терминотерминологической системы предметной области задается посредством точной дефиниции. Более того, на практике критерием разграничения термина и нетермина часто служит наличие или, соответственно, отсутствие дефиниции (Шелов, 2001).

Во многих работах перечисляются следующие свойства идеального термина (Sager, 1990):

- термин должен относиться непосредственно к понятию, он должен выражать понятие ясно,
- термин не должен иметь синонимов,
- значение термина должно быть точным и не должно пересекаться по значению с другими терминами,
- значение термина не должно зависеть от контекста.

Это означает, что в идеальных случаях должно быть установлено взаимнооднозначное соответствие между понятиями и терминами в конкретной предметной области.

Для терминологической системы наилучшей ситуацией является отсутствие синонимии, отсутствие омонимии, отсутствие полисемии. Хотя такая идеальная ситуация встречается только в узких хорошо структурированных предметных областях, все вышеперечисленные свойства терминов подчеркивают их значительные отличия от общезначимой лексики.

## **3. Взаимопереходы терминов и общезначимой лексики. Общественно-политическая область**

Как отдельный способ формирования терминов рассматривается превращение в термин общезначимого слова - терминологизация, когда общезначимое слово получает новое терминологическое значение в конкретной предметной области. В то же время широко распространен и обратный процесс - детерминологизация, когда появившийся в некоторой специальной области термин становится словом общей лексики. В работе (Суперанская и др., 2003) отмечается, что при этом специальное значение в общей лексике редуцируется, термин приобретает прагматические свойства, которых он прежде был лишен, то есть возникает новое слово с терминологическим значением, требующее уже не дефиниции, а толкования. Породивший новое слово термин остается в своем терминологическом поле без изменений. Таким образом, считается, что фактически при процессе детерминологизации появляется омоним термина.

Однако в реальности значительное количество общезначимых слов русского языка имеет очень близкие по смыслу терминологические соответствия в одной или более областях без серьезного сдвига значения. Например, понятия, соответствующие общезначимому слову «здание», являются необходимым элементом терминологии по крайней мере в двух областях: в области строительства и в области коммунального обслуживания.

Если предположить, что всякое терминологическое значение должно быть отделено от общелексического значения, то это означает, что эти значения функционируют в различных контекстах, и совмещение их в одном и том же предложении приводит к такому явлению как «игра слов» (Cruse, 1980). Между тем если такое слово как «здание» употребляется в газетном тексте, то часто невозможно отличить, в каком значении общелексическом или терминологическом это слово употреблено, как, например, в следующем фрагменте текста публикации газеты «Известия» от 18.02.2004 под названием «Кто ответит за Трансвааль»:

*Допросы строителей, проектировщиков, людей которые отвечали за эксплуатацию здания, пока мало что дали, - заявили "Известиям" в прокуратуре. Все они, естественно, отрицают свою вину и кивают друг на друга. Будут проверены все, в том числе и поставщики строительства, они могли предоставить материалы не того качества, которые заявлялись. Объективно ответить на вопрос, кто виноват, можно будет только после завершения экспертиз.*

Несомненно, в профессиональной области границы понятия могут быть четче, строитель увереннее непрофессионального носителя языка должен отличать «здания» от других видов строений. При этом для непрофессионала такие «граничные» объекты относятся к области неопределенности - «vagueness» (Шелов, 1998), когда он уже не может четко классифицировать строение и охотно принимает профессиональную классификацию.

Поэтому по нашему мнению, понимание слова «здание» у строителя и обычного носителя языка отличается настолько незначительно, что сравнимо с индивидуальными различиями смыслов у разных людей, и этими различиями можно пренебречь.

Таким образом, подавляющее большинство общезначимых слов-артефактов должны иметь чрезвычайно близкие по смыслу терминологические аналоги по крайней мере в двух предметных областях: области производства этого артефакта и профессиональной области его обслуживания.

Кроме того, если человек соприкасается в своей повседневной жизни с профессиональными сферами деятельности, ему необходимо понимать и использовать те же самые значения, что и профессионалам. Поэтому неслучайно многие англоязычные общеязыковые толковые словари используют для толкования лексики права максимально точные дефиниции:

*arson - Law. the malicious burning of another's house or property, or in some statutes, the burning of one's own house or property, as to collect insurance (Random House Unabridged dictionary, 1999);*

*arson - Criminal law. The act of intentionally or recklessly setting fire to another's property or to one's own property for some improper reasons (Collins, 1991);*

можно сравнить с толкование в (Большой толковый словарь, 1998):

*Поджог - ... 2. Умышленное устройство пожара*

Таковыми же свойствами совместного лексического и терминологического употребления обладают и многие другие общеизвестные слова: названия транспортных средств, должности и профессии, технические устройства, произведения искусства, социальные и природные явления и многие другие. Перечислим лишь некоторые из таких слов, начинающихся на букву «а»: *аванс, авиабаза, автобус, автопилот, агроном, адвокат, аккредитив, алгебра, алгоритм, алебастр* и т.п.

Мы оцениваем, что до 40 процентов слов, содержащихся в общих толковых словарях, обладают похожими свойствами.

Кроме того, существует значительное количество многословных выражений, которые являясь терминами в специальных предметных областях, понятны носителям языка, например, *военная помощь, авиационная промышленность, внешняя миграция*. Это означает, что взаимопроникновение лексики и терминологии имеет значительно больший масштаб, чем это предполагалось ранее терминологами и лексикологами.

На существование промежуточного слоя между общей и специальной лексикой указывалось и ранее. На Рис. 1.1 воспроизведен рисунок из (Rondeau, 1980), на котором показан такой слой. На наш взгляд, эта пограничный лексико-терминологический слой «значительно шире» (Рис.1.2).

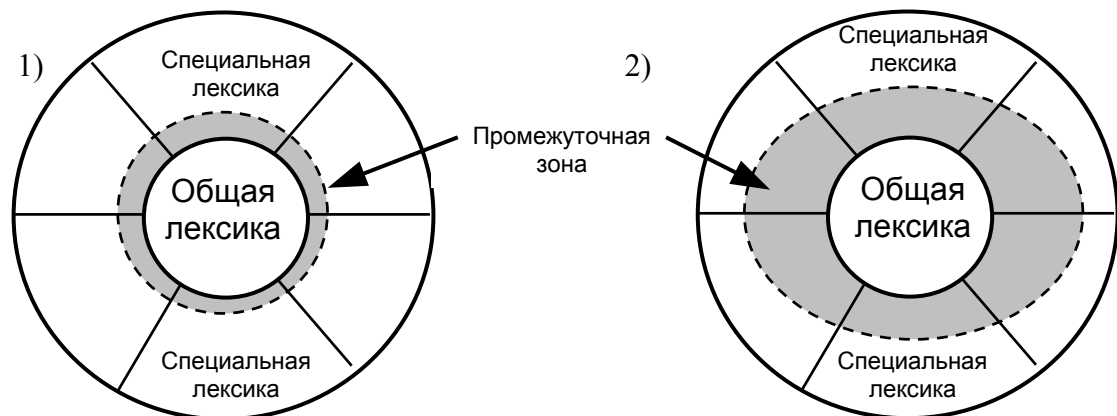


Рис.1

#### 4. Общественно-политический тезаурус как составная часть тезауруса РуТез

В 1994 году мы начали собирать терминологию экономики и права по официальным документам Российской Федерации, создавая тезаурус для автоматического индексирования. С 1996 года мы используем его также и для автоматической обработки новостных и газетных материалов, позже тезаурус применялся и для обработки таких «обращенных к человеку» текстов как социологические опросы. Мы назвали этот тезаурус Общественно-политическим и в результате многочисленных экспериментов и работ пришли к выводу о существовании смешанного лексико-терминологического слоя – общественно-политической области. С 1997 года на базе Общественно-политического тезауруса мы стали строить тезаурус русского языка РуТез, двигаясь не сверху вниз, как это обычно бывает при построении больших понятийных ресурсов, а снизу вверх, тем самым обеспечивая самые сложные, неопределенные уровни понятийной системы освоенным (а часто уже и работающим в конкретных приложениях) языковым и понятийным материалом (Лукашевич, Добров, 2002). В настоящее время объем тезауруса РуТез составляет 45 тысяч понятий, 107 тысяч слов и выражений, объем общественно-политического тезауруса 30 тысяч понятий, 75 тысячи слов и терминов. Общественно-политический тезаурус включает сейчас не только чисто общественно-политическую область, но также значительную часть профессиональной терминологии таких значительно пересекающихся с общественно-политической областью сфер деятельности как право, налоги, таможня и т.п., что дало возможность автоматической обработки законодательства и ведомственных актов.

## 5. Лексико-терминологический состав общественно-политической области и его свойства

Общественно-политическая область включает в себя лексику и терминологию, которая, с одной стороны, известна достаточно широкому слою населения, с другой стороны, соответствует понятиям профессиональных сфер деятельности. Рассмотрим состав и особенности общественно-политической области на примере реализации тезауруса РуТез как модели понятийной системы русского языка.

Если представить себе иерархию понятий от более общего к более частному, то наиболее верхние уровни занимает зона общей лексики – Общий лексикон, более нижние уровни занимает Общественно-политическая область (Рис. 2, стрелки на этом и следующих рисунках обозначают направления развития тезауруса РуТез). Одновременно именно в общественно-политической области находятся верхние уровни профессиональных понятийных систем (Рис. 3).

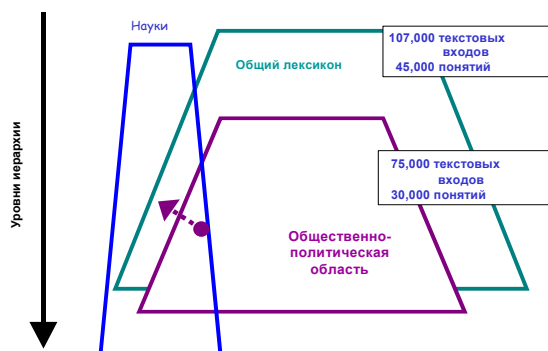


Рис.2. Общественно-политическая область vs. Общий лексикон

Разные предметные области имеют различные по величине перенесению с общественно-политической областью. Так, понятийная система предметной области «Выборы» практически полностью находится в Общественно-политической области, в то время как сферы различных промышленных производств пересекаются с общественно-политической областью лишь по небольшому числу понятий (Рис. 3).

Можно выделить совокупность непроизводственных регулирующих сфер деятельности, которые значимы в повседневной деятельности многих людей и, значит, в значительной степени пересекаются с общественно-политической областью, такие как Налоги, Бухгалтерия, Право, Таможня, Банковская сфера, образуя правовой и финансовый блоки областей. Они не содержат обширной специальной иерархии понятий (Рис. 4).

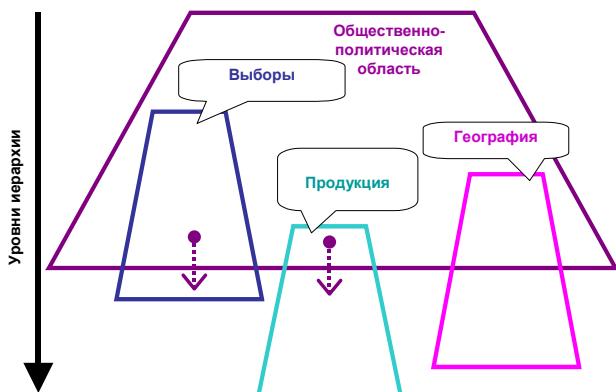


Рис.3. Специальная лексика vs. Общественно-политическая

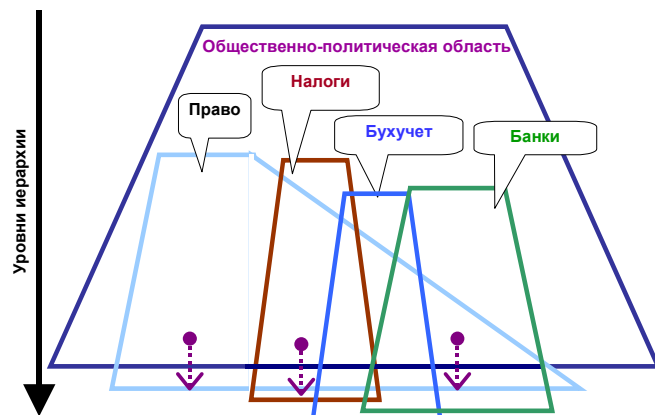


Рис.4. Взаимосвязь специальной лексики внутри Общественно-политической области

Научные понятийные системы пересекаются с общественно-политической областью сложнее. Общественно-политическая область включает основные виды наук, научных учреждений, общенаучные понятия. Однако каждая наука задает свою категоризацию изучаемых явлений, в связи с чем ее верхние уровни классификации могут значительно отличаться от классификации, на базе общей лексики (Рис. 5).

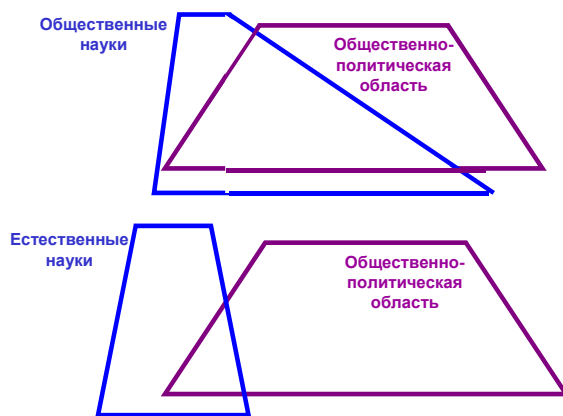


Рис.5. Научная лексика vs. Общественно-политическая область

Общественно-политическая область в значительной мере пересекается со сферой общественных наук, так как содержит понятия общественной жизни, которые изучаются общественными науками. При этом каждая общественная наука может иметь свою собственную классификацию рассматриваемых явлений.

## 6. Общественно-политическая область и автоматическая обработка текстов

Выделение такой области, а также выделение среди общезначимой лексики лексем, принадлежащих этой области, является чрезвычайно полезным для разработки

лингвистических ресурсов и технологий автоматической обработки больших электронных коллекций.

1. Прежде всего, терминология и лексика из этой области активно используется в самых разных по жанру, но значимых для жизни общества текстах, как законы, международные договора, другие официальные документы, газетные сообщения, экономические документы (Loukachevitch, Dobrov, 2002). Таким образом, создание лингвистического ресурса в общественно-политической области может значительно повысить эффективность и содержательность обработки всех этих видов документов.
2. Поскольку общественно-политическая область содержит наиболее общеизвестные понятия многих профессиональных предметных областей, то лингвистический ресурс, разработанный для общественно-политической области, может стать источником существенного понятийно-терминологического материала для создания лингвистических ресурсов в конкретных предметных областях.
3. Одновременно общественно-политическая область – это область общезначимая и содержит значительное количество общелексического материала, который относится к нижним и средним наиболее конкретным уровням языковой системы языка, поэтому понятийная структура общественно-политической области является и существенным базисом, на который можно опираться, например, выстраивая понятийную иерархическую систему языка типа WordNet (Miller и др. 1990).
4. Кроме того, если рассмотреть количество многозначных общезначимых слов внутри общественно-политической области и вне ее, то многозначных слов в общественно-политической области значительно меньше, а процедура автоматического разрешения многозначности работает эффективнее, поскольку часто значения относятся к различным подобластям общественной жизни, например, в подавляющем большинстве текстов контексты разных значений словоформы *судов* как средства водного транспорта и судебного органа существенно различаются. Это различие можно также эффективно использовать при автоматической обработке текстов, используя, например, комбинированную обработку текстов и запросов при решении информационно-поисковых задач, а именно пытаться разрешать многозначность для слов и терминов, относящихся к общественно-политической области, и использовать пословную обработку для остальной общезначимой лексики.

## Заключение

Граница между лексикой и терминологией представляет собой широкую промежуточную зону. Она содержит лексемы, значения которых совпадают с понятиями конкретных предметных областей и термины, понятные носителям языка. Эта зона включает в себя понятия, значимые для общества в целом, поэтому мы называем ее общественно-политической областью.

Лексико-терминологические ресурсы, разработанные для общественно-политической области, полезны для приложений по автоматической обработке разнообразных типов текстов.

Знания об общественно-политической области очень важны как для создания лингвистических ресурсов в конкретных предметных областях, так и как основа для описания абстрактной лексики языка.

## Литература

1. Buitellar P., Sacalenu B., Extending Synsets with Medical Terms In proceedings of the NAACL workshop on WordNet and Other Lexical Resources: Applications, Extensions and Customizations, Pittsburg, USA, 2001.
2. Climent S., Rodriguez H., Gonzalo J., Definitions of the links and subsets for nouns of the EuroWordNet project. - Deliverable D005, WP3.1, EuroWordNet, LE2-4003, 1996.
3. Collins English Dictionary: Third Edition. HarperCollins Publishers (1991).
4. Cruse D., Lexical Semantics. – Cambridge University Press. –1986.

5. Loukachevitch N.V., Dobrov B.V., Evaluation of Thesaurus on Sociopolitical Life as Information Retrieval Tool // Proceedings of Third International Conference on Language Resources and Evaluation (LREC2002) / M.Gonzalez Rodriguez, C. Paz Suarez Araujo (Eds.) – Vol.1 – 2002, Gran Canaria, Spain – p.115-121.
6. Miller G., Beckwith R., Fellbaum C., Gross D. and Miller K., Five papers on WordNet, *CSL Report, 43*, Cognitive Science Laboratory, Princeton University, 1990.
7. Random House Unabridged dictionary. Random house, Inc, 1999.
8. Rondeau G. Introduction a terminologie. Quebec, 1980.
9. Sager, J.C.: A Practical Course in Terminology Processing. Amsterdam: J. Benjamins, 1990.
10. Vossen P., Extending, Trimming and Fusing WordNet for Technical Documents. In: Proceedings of WordNet and Other Lexical Resources: Applications, Extensions and Customizations, Pittsburg, USA, 2001.
11. Wuster E., Einfurung in die Allgemeine Terminologielehre und terminologishe Lexicographie. - Vien; N.Y., 1979/Bd 1 - 2.
12. Большой толковый словарь русского языка. – Санкт-Петербург, изд-во Норинт, 1998.
13. Лукашевич Н.В., Добров Б.В., Тезаурус русского языка для автоматической обработки больших текстовых коллекций // Компьютерная лингвистика и интеллектуальные технологии: Труды Международного семинара Диалог'2002 / Под ред. А.С.Нариньяни – М.: Наука – 2002. – Т.2 - С.338-346.
14. Ожегов С.И., Шведова Н.Ю., Толковый словарь русского языка: 80000 слов и фразеологических выражений / Российская АН.; Российский фонд культуры;- 2-е изд., испр. и доп. - М.: АЗЪ, 1995.
15. Суперанская А.В., Подольская Н.В., Васильева Н.В., Общая терминология. Вопросы теории. М: УРСС, 2003.
16. Шелов С.Д., Определение терминов и понятийная структура терминологии. – Изд-во С.-Петербургского Университета, 1998.