

*О.Н. Ляшевская*

**ТОПОЛОГИЧЕСКИЕ КЛАССЫ ИМЕН  
В СЕМАНТИЧЕСКОЙ РАЗМЕТКЕ  
НАЦИОНАЛЬНОГО КОРПУСА РУССКОГО ЯЗЫКА<sup>1</sup>**

**1. Введение**

Лексико-семантические ресурсы играют все большую роль в корпусной лингвистике и в отрасли автоматической обработки текстов. Если взять в качестве примера английский язык, то можно указать такие известные банки данных, как WordNet<sup>2</sup>, EuroWordNet<sup>3</sup>, FrameNet<sup>4</sup>, PropBank<sup>5</sup>, NomBank<sup>6</sup>, TimeBank<sup>7</sup>,

---

<sup>1</sup> Работа выполнена при финансовой поддержке РФФИ, грант № 07-04-00240а.

<sup>2</sup> *Fellbaum C. (ed).* WordNet: An Electronic Lexical Database. Cambridge, Mass.: MIT Press, 1998.

<sup>3</sup> *Vossen P.* Introduction to EuroWordNet // N. Ide, D. Greenstein, P. Vossen (eds.). Special Issue on EuroWordNet. Computers and the Humanities. 1998. 32: 2–3. P. 73–89.

<sup>4</sup> *Baker C.F., Fillmore Ch.J., Lowe J.B.* The Berkeley FrameNet project // Proceedings of Framenet as a «Net» // Proceedings of COLING–ACL'98. University of Montreal, 1998. P. 86–90.

<sup>5</sup> *Kingsbury P., Palmer M.* Propbank: the next level of Treebank // Proceedings of Treebanks and Lexical Theories. 2003; *Palmer M., Gildea D., Kingsbury P.* The proposition bank: A corpus annotated with semantic roles // Computational Linguistics Journal. 2005. 31: 1.

<sup>6</sup> *Meyers A., Reeves R., Macleod C., Szekely R., Zielinska V., Young B., Grishman R.* Annotating Noun Argument Structure for NomBank. Proceedings of LREC–2004. Lisbon. 2004.

<sup>7</sup> *Pustejovsky J., Hanks P., Sauri R., See A., Gaizauskas R., Setzer A., Radev D., Sundheim B., Day D., Ferro L., Lazo M.* The TIMEBANK Corpus // Proceedings of Corpus Linguistics 2003.

Penn Discourse TreeBank<sup>1</sup>, VerbNet<sup>2</sup>, USAS<sup>3</sup>, не говоря уже о разных видах лексических онтологий. Многократное использование этих компьютерных ресурсов в разметке английских корпусов (ср. например последний масштабный проект семантической разметки Американского Национального корпуса<sup>4</sup>) позволило не только выявить их сильные и слабые стороны, но и связать эти системы друг с другом.

С началом развития корпусной лингвистики в России стали активно разрабатываться и русскоязычные семантические ресурсы, в частности, RussNet<sup>5</sup> и семантический словарь НКРЯ<sup>6</sup>. Разработка русского FrameNet и других аналогов – будем надеяться, дело ближайшего будущего.

В настоящем докладе речь пойдет о новом семантическом инструменте – **топологической классификации** предметных имен, которая составляет особый раздел многофасетной

---

<sup>1</sup> Prasad R., Dinesh N., Lee A., Miltsakaki E., Robaldo L., Joshi A., Webber B. The Penn Discourse TreeBank 2.0 // Proceedings of LREC-2008. Marrakech, Morocco.

<sup>2</sup> Kipper K., Korhonen A., Ryant N., Palmer M. Extending VerbNet with novel verb classes // Proceedings of LREC-2006.

<sup>3</sup> Piao S.S.L., Archer D., Mudraya O., Rayson P., Garside R., McEnery T., Wilson A. A Large Semantic Lexicon for Corpus Annotation // Proceedings of Corpus Linguistics 2005.

<sup>4</sup> Ide N., Baker C., Fellbaum C., Fillmore Ch., Passonneau R. MASC: The Manually Annotated Sub-Corpus of American English // Proceedings of LREC-2008. Marrakech, Morocco.

<sup>5</sup> Азарова И. В., Марина А. С. Автоматизированная классификация контекстов при подготовке данных для компьютерного тезауруса RussNet // Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог-2006». М., 2006. С. 13–17.

<sup>6</sup> Кустова Г.И., Ляшевская О.Н., Падучева Е.В., Рахилина Е.В. Опыт семантического расширения морфологической разметки: таксономическая классификация лексики в Национальном корпусе русского языка // Национальный корпус русского языка: 2003–2005. М., 2005.

лексической классификации, лежащей в основе семантической разметки НКРЯ. Топологические признаки лексемы – это информация о форме и пространственной ориентации физического предмета, обозначаемого данным именем.

Подобно тому, как перечисленные ранее программные продукты являются практическим воплощением теоретических идей лингвистики XX века (теория семантических ролей и фреймов Ч. Филлмора, классификация глаголов на основании трансформаций глагольного управления Б. Левин и др.), топологическая разметка базируется на двух постулатах: на постулате современной лексикографии о том, что описание предметного имени должно содержать информацию о существенных физических (telic) характеристиках обозначаемого объекта, и на когнитивной теории топологических (геометрических) типов Л. Талми<sup>1</sup>.

Опираясь на результаты психологических исследований, Л. Талми утверждает, что человек воспринимает и описывает окружающее пространство с помощью ограниченного набора пространственных схем, например, «плоскость», «линия», «выступ» и др. Данный способ категоризации проявляется в способах номинации окружающих объектов и в лексической сочетаемости – с предлогами, прилагательными, существительными и глаголами. Многие объективные характеристики объектов человек легко игнорирует, например, говоря *гусеница ползет по стволу*, он схематизирует ствол как линию, вдоль которой происходит движение, но «забывает» про контуры объекта и даже его объем.

---

<sup>1</sup> *Talmy L. How language structures space // Pick H.L., Jr., Acredolo L.P. (eds.) Spatial orientation: Theory, research and application. N.-Y.: Plenum Press, 1983. P. 225-282; Talmy L. The fundamental system of spatial schemas in language // Hamp B. (ed.), From perception to meaning: Image Schemas in Cognitive Linguistics. Mouton de Gruyter, 2006. P. 37-47.*

На русском материале теория топологических типов была развита Е.В. Рахилиной<sup>1</sup>. В частности, была разработана методика выделения топологических типов и обнаружены классы, релевантные для описания лексической сочетаемости предметных имен с прилагательными формы. В настоящее время работа продолжается в рамках проекта семантической разметки Национального корпуса русского языка. Ставится задача исследовать поведение предметных имен в разных видах пространственных конструкций и составить полный инвентарь топологических классов<sup>2</sup>.

## 2. Инвентарь топологических классов

Поскольку речь идет о пространственных характеристиках физических объектов, топологическая классификация охватывает предметную лексику. Материалом для классификации служат имена, входящие в семантический словарь НКРЯ (более 28 тыс. предметных значений). На этом массиве

На массиве лексики, входящей в семантический словарь НКРЯ, выделяется более 30 топологических классов, в том числе:

- «поверхности» разных видов:
- «горизонтальные поверхности» (*пол, поднос, этаж*),
- «вертикальные поверхности» (*стена, табло*),
- «наклонные поверхности» (*скат, лестница, стулья*);
- «пластины» (объекты, имеющие небольшую толщину и плоскую поверхность: *ладонь, подошва*);

---

<sup>1</sup> Рахилина Е. В. Когнитивный анализ предметных имен: семантика и сочетаемость. М.: Русские словари, 2000.

<sup>2</sup> Ляшевская О.Н., Рахилина Е.В. Топология в классификации русских предметных имен // Труды III Международного конгресса исследователей русского языка «Русский язык: исторические судьбы и современность». М., 2007. С. 469–470; Десятова А.В., Ляшевская О.Н., Махова А.А. Конструкция с творительным формы «X Y-ом» // Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог–2008». М., 2008. С. 133–139.

- «выступы» (*горб, грудь, колено; холм*);
- «вместилища» (*стакан, сундук, озеро*), с подклассом
- «углубления» в поверхности (*яма, впадина, выбоина*);
- «отверстия» (*дыра, горловина, окно*);
- «стержни» (вытянутые объекты жесткой формы: *шест, палка, игла, карандаш*);
- «столбы» (вертикально ориентированные стержни: *столб, башня, мачта*);
- «дуги» (непрямые стержни: *арка, мост, радуга*);
- «веревки» (*ремень, проволока, очередь, змея*);
- «линии» (*линия, пробор, черта*);
- «полосы» (вытянутые объекты, имеющие поверхность: *дорога, грядка*);
- «шары» (*ягода, мяч, комок*);
- «кольца и круги» (*кольцо, корона, колесо, пяточок*) и др., а также особый класс
- «объекты характерной формы» (*домик, сердечко, гармошка, колокол*).

Как видно, классы могут быть системно противопоставлены друг другу. Они различаются количеством измерений: объем – плоскость – вытянутый объект – точка; жесткостью/гибкостью объекта (так, жесткие «стержни» противопоставлены гибким «веревкам»); ориентацией в пространстве: ср. *глубокий карман* (углубление, вертикально ориентированное сверху вниз) – *глубокая ниша* (горизонтально ориентированное углубление). Еще один важный топологический признак – прикрепленность объекта к другим объектам. Так, можно сказать *концы веревки*, но не *\*концы мачты*, поскольку в последнем случае речи идет о «столбе», прикрепленном к опоре (у такого объекта может быть только один конец).

Несмотря на то, что мы можем проанализировать систему классификации в терминах различительных признаков, важно подчеркнуть, что топологический класс не сводим к их сумме

([+ трехмерный], [+ горизонтально ориентированный] и т. д.); напротив, каждый класс соотносится с простой зрительной схемой, которая мыслится как единое целое.

### 3. Разметка предметных имен по системе классов НКРЯ

Топологическая классификация является одним из фасетов лексической классификации НКРЯ. Другие фасеты включают:

- разряд: предметное имя, не предметное (абстрактное) имя, имя собственное;
- таксономический класс: лицо, животное, инструмент, транспортное средство и т. п.;
- мереология: часть предмета, квант или порция вещества, множество или совокупность и т. п.;
- оценка: положительная или отрицательная;
- словообразовательный класс: отглагольное или отадективное имя, диминутив, сингулятив, *nomina agentis*.

Соответственно, в семантическом словаре каждому предметному значению приписываются классы по всем шести независимым основаниям. Приведем в качестве примера разметку некоторых значений существительного *игла*:

- |                          |   |
|--------------------------|---|
| 1. <i>швейная игла</i>   | 2. <i>иглы дикобраза</i>                |
| разряд: «предметное имя» | разряд: «предметное имя»                |
| таксономия: «инструмент» | таксономия: —                           |
| мереология: —            | мереология: «часть тела:<br>животного»  |
| топология: «стержень»    | топология: «стержень:<br>прикрепленный» |
| оценка: —                | оценка: —                               |
| словообразование: —      | словообразование: —                     |

Характеристика имени по топологическому классу обладает той особенностью, что обычно она сохраняется при переходе от одного значения существительного к другому<sup>1</sup>. Здесь срабатывает достаточно распространенный принцип семантического переноса «по форме», ср. *кольцо с рубином на пальце* и *Садовое кольцо*; *угощать чаем с баранками* и *крутить баранку* (о шоферах).

В редких случаях топологический класс может несколько видоизменяться. Если сравнить значения существительного *ложе*, то можно увидеть, как плоскость «прогибается» при переходе от горизонтальной поверхности к углублению, ср. *на ложе из голых досок*, но *в ложе водохранилища* (обратный переход можно видеть у существительного *лоток*, ср. *лотки для стока воды* и *газеты на лотках*). Как правило, предпосылкой такой трансформации служит наличие целого класса функционально близких наименований, имеющих сдвоенный топологический статус (в данном случае, *вместилще/поверхность*, ср. *на кровати/в кровати*, *на постели/в постели*; *на блюде/в блюде*, *на тарелке/в тарелке*).

#### 4. Топологическая информация в корпусе

Укажем две основных области применения информации о топологических классах, которая содержится в разметке корпуса: это, с одной стороны, теоретические лингвистические и когнитивные исследования, а с другой стороны, практическое улучшение корпусной разметки.

Топологическая классификация является одновременно и результатом, и инструментом изучения языкового материала, связанного с восприятием пространства. Но насколько этот инстру-

---

<sup>1</sup> Конечно, с учетом того, что топологическая классификация не распространяется на абстрактные значения, а также значения веществ и множеств, ср. *номер*, *постройка*, *титан*, *чугун*, *класс*).

мент совершенен? Так как классификация лексическая, а именно, топологический класс определяется на словарном значении имени существительного, нельзя требовать, чтобы этот класс в точности обслуживал все употребления имени в корпусе. Напротив, в соответствии с принципами Грамматики Конструкций, любая конструкция (с предлогом, прилагательным, другим существительным и т. п.) может задавать свое пространственное видение обозначаемого объекта. Ср. здесь «могущественную» конструкцию с предлогом *в*, которая помогает нам представить практически любой предмет как вместилище (*в столе, в плече, в небо* и т.д.). Таким образом, в семантическом словаре указывается только прототипический топологический класс имени.

Как же тогда определить, какую пространственную интерпретацию имеет обозначаемый объект в конкретном примере из корпуса? Наша исследовательская гипотеза такова: зная топологический прототип, мы можем задать правила перехода в другой топологический класс (топологических трансформаций) в разных конструкциях.

Топологическая классификация может быть использована и в практических целях – для решения задач, связанных с разрешением многозначности в корпусе. В системе семантических фильтров, предназначенных для снятия омонимии в НКРЯ, топологическая информация находит применение наряду с другими лексико-семантическими и морфологическими признаками контекста.

В частности, топологический класс «выступ» является контекстным признаком, в котором прилагательное *острый* обозначает форму объекта («Суживающийся к концу» (МАС), ср. *острый кадык, носик, коленки, груди, выступы, борода*) – в сочетании с другими типами предметных имен прилагательное *острый* обозначает физическое качество (ср. *острый нож*) или вкусовые свойства (ср. *острый сыр*).