

ЛЕКСИЧЕСКАЯ ТИПОЛОГИЯ КАК ИНСТРУМЕНТ ДЛЯ СОЗДАНИЯ УНИВЕРСАЛЬНОГО СЛОВАРЯ: ПРИЗНАКИ «ТУГОЙ» И «УПРУГИЙ»

Е. А. Баскакова

Национальный исследовательский университет «Высшая школа экономики»

Поступила в редакцию 20 апреля 2015 г.

Аннотация: данная работа посвящена изучению семантического поля «тугой» и «упругий» в рамках лексической типологии на материале 9 языков (русского, английского, немецкого, французского, испанского, китайского, грузинского, сербского и иврита). Результатом анализа стал перечень минимальных ситуаций (фреймов), которые могут лексически противопоставляться в том или ином языке. Полученные фреймы легли в основу семантической карты, позволяющей визуализировать различия между отдельными языками в сфере лексикализации признаков «тугой» и «упругий». Эти же фреймы могут быть использованы при построении универсального автоматического словаря.

Ключевые слова: лексическая типология, автоматическая обработка языка, фреймовый подход, семантические поля «тугой» и «упругий».

Abstract: the research focuses on the semantic domains formed by the adjectives «tight» and «thick» within the framework of lexical typology. The research is based on the data from 9 languages: Russian, English, German, French, Spanish, Serbian, Chinese, Georgian and Hebrew. The result of the analysis consists of the list of minimal situations (frames) that form the basis of the semantic map which, in turn, visualizes the differences and similarities between the languages in the semantic domains «tight» and «thick». These frames can be used in the universal computerized dictionary for natural language processing.

Key words: lexical typology, natural language processing, frame-based approach, semantic domains «tight» and «thick».

Современное состояние развития компьютерных технологий, а также наличие электронных словарей и корпусных ресурсов для разных языков позволяют решать многие задачи, связанные с автоматической обработкой естественного языка. Вместе с тем, некоторые компьютерно-лингвистические задачи еще по-прежнему далеки от своего решения. В ряду таких нерешенных задач находится и машинный перевод. На сегодняшний день автоматические переводчики хорошо справляются с отдельными словами, но качество перевода целых предложений остается довольно низким. Это объясняется тем, что существующие двуязычные словари недостаточно подробно отражают особенности употребления каждого из приводимых переводных эквивалентов, т.е. не содержат полной информации о контексте, в котором может выступать тот или иной вариант перевода. Соответственно, по таким словарям невозможно адекватно предсказать переводной аналог для заданной лексемы в каждом конкретном контексте.

Прообраз словаря, который мог бы решать эту задачу, создается в настоящее время на материале качественных признаков участниками Московской

лексико-типологической группы (MlexT) [1; 2]. Он представляет собой базу данных. Каждая ее строка соответствует минимальной ситуации, к которой применим рассматриваемый признак. Так, для семантического поля «мокрый» к таким ситуациям относится, например, «одежда, намокшая под дождем», «одежда, не до конца высушенная после стирки», «воздух с повышенной влажностью», «помещение с повышенной влажностью воздуха» и т.д. (см. [3]). Как кажется, такие ситуации позволяют однозначно определять значение слова в каждом контексте и тем самым смогут стать надежной основой для перевода и построения типологических описаний.

Настоящее исследование выполнено в рамках проекта по созданию описанной базы данных и примыкает к активно развивающемуся в последние годы направлению – лексической типологии [4–7]. Объектом анализа стали семантические поля «тугой» и «упругий» на материале 9 языков (русского, английского, немецкого, французского, испанского, китайского, грузинского, сербского и иврита). Языковые данные были получены из словарей (толковых, двуязычных, тезаурусов и др.), корпусов (НКРЯ, DWDS,

CREA, COCA и др.)¹ и в ходе опроса информантов по специально созданной анкете. Отметим, что данные, собранные с помощью анкеты, позволили не только уточнить сведения, полученные из словарей и корпусов, но и выявить отрицательный языковой материал – случаи, когда употребление той или иной лексемы в определенном контексте невозможно.

Лексико-типологическое описание в рамках проектов MLexT подразумевает решение нескольких задач. К ним относятся, во-первых, выделение минимальных ситуаций, для которых релевантен рассматриваемый признак (такие ситуации в терминологии MLexT называются фреймами). Во-вторых, исследуется распределение этих фреймов между лексическими единицами, в частности, выявляются фреймы, которые типологически частотно совмещаются в одной лексеме².

Рассмотрим последовательно, какие результаты дает решение этих задач на материале полей «тугой» и «упругий».

Семантические поля «тугой» и «упругий»: фреймовая структура

Исследуемые семантические поля обладают довольно сложной структурой и охватывают разнообразные типы ситуаций.

Веревка/струна: данный фрейм подразумевает сильно натянутые объекты, как правило, закрепленные с обеих сторон.

Ткань (натянута): в этом случае также описывается натянутый объект, но этот объект относится к другому топологическому классу [9; 10]. Так может характеризоваться, например, ткань на барабане или на пальцах для вышивания.

Леска: фрейм описывает тянущиеся предметы, которые не теряют целостности при растяжении. Объект с такими свойствами необходим, например, при ловле рыбы.

Ткань (не рвется): эта ситуация аналогична фрейму «леска», но реализуется с объектами другого топологического класса.

Ветка: прилагательное характеризует свойство вытянутого объекта не ломаться при сгибании.

Резинка: топологически резинки представляют собой объекты, похожие на веревки, однако в прото-

типическом случае они воспринимаются экспериенциально (о тугой резинке обычно говорят в связи со сдавливающей одеждой).

Тесная одежда: этот фрейм непосредственно связан с предыдущим и во многих языках объединяется с ним в одной лексеме. Речь идет о предметах одежды, которые сдавливают человека, например, корсет, галстук или шляпа.

Облегающая одежда: здесь подразумевается одежда, которая сидит на человеке точно по фигуре.

Узел: прилагательные исследуемого поля могут описывать туго завязанный узел.

Кожа: данный фрейм сходен с ситуацией натянутой ткани, только он предполагает особый тип «ткани» – кожу человека.

Части тела: метонимически связаны с предыдущим фреймом – кожа на них как бы «натянута». Так могут описываться, например, бедра или ягодицы.

Мяч: здесь может подразумеваться, во-первых, зрительное свойство объекта (материал, из которого он изготовлен, туго натянут), а во-вторых, возникающие благодаря этому функциональные характеристики (способность отскакивать при ударе).

Подушка: в данном случае также, с одной стороны, выражается внешний эффект – как если бы внешний материал объекта был сильно натянут, а с другой – экспериенциальное свойство, а именно способность прогибаться при воздействии. В данном фрейме исследуемое семантическое поле сходится с прилагательными зоны «жесткого».

Кнопка: имеется в виду свойство объекта оказывать сопротивление при нажатии.

На основе сформированных фреймов и стратегий их объединения в изученных языках сначала был сконструирован каркас общей семантической карты (рис. 1), на который далее были нанесены конкретные языковые данные. В качестве иллюстрации приведем семантическую карту для семантических полей «тугой» и «упругий» в испанском языке (рис. 2). Объединенные фреймы на карте соответствуют областям употребления прилагательных.

Таким образом, выявленные фреймы включают достаточно разнородные ситуации («туго натянутая веревка» и «упругая ветка», или «тесная одежда» и «упругий мяч»), что может создать впечатление бессистемной организации исследуемых полей. Однако в самых разных языках представленные фреймы покрываются одними и теми же лексемами, а это значит, что за внешней разнородностью ситуаций должна стоять внутренняя логика, связывающая их между собой. И действительно, как уже можно было предположить исходя из приведенного перечня фреймов, некоторые из них связаны друг с другом имплицативными отношениями. В частности, если объект

¹ НКРЯ – Национальный корпус русского языка. URL: <http://www.ruscorpora.ru/DWDS> – Das Digitale Wörterbuch der Deutschen Sprache (корпус немецкого языка). URL: <http://www.dwds.de/> CREA – Corpus de Referencia del Español Actual (корпус испанского языка). URL: <http://corpus.rae.es/creanet.html>

COCA – Corpus of Contemporary American English (корпус английского языка). URL: <http://corpus.byu.edu/coca/>

² В рамках данного теоретического направления были уже проведены исследования в области глаголов движения в воде [5], глаголов боли [7], глаголов вращения [8], глаголов разделения объекта на части и ряд других.

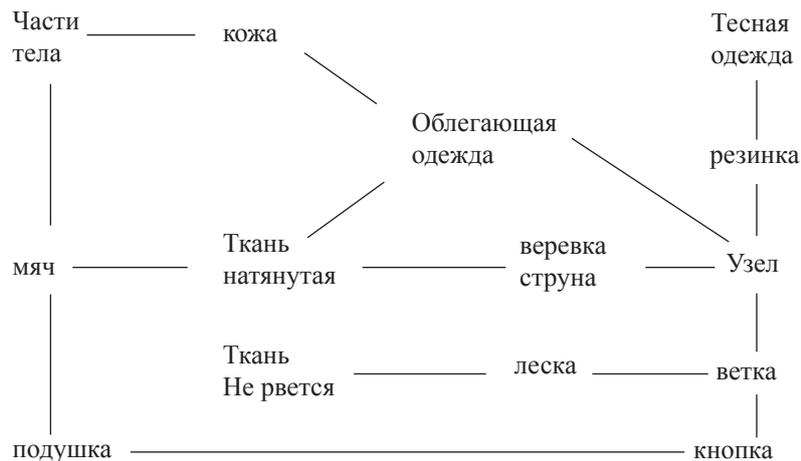


Рис. 1. Общая семантическая карта

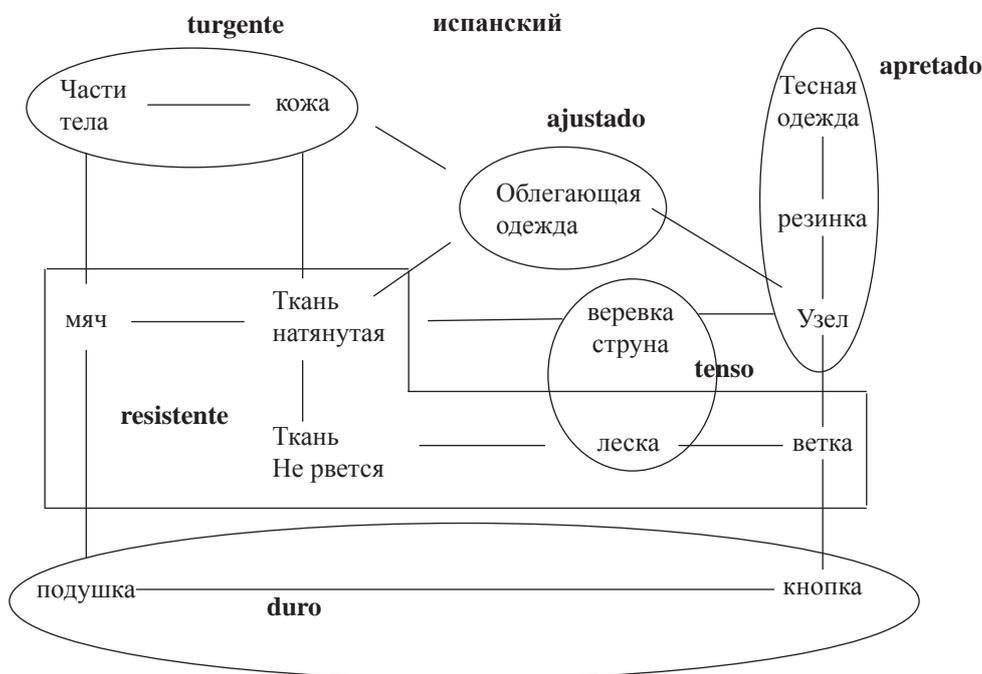


Рис. 2. Семантические поля «тугой» и «упругий» в испанском языке

сильно натянут (фрейм «веревка»), то это может вызывать представление об усилиях, которые требовались для его натягивания, т.е. о свойстве плохо поддаваться воздействию (фрейм «резинка»). В то же время, если объект можно сильно натянуть, то это означает, что он не рвется и не ломается при воздействии (отсюда ситуации «упругая ветка» или «нервущаяся ткань»).

Вместе с тем, некоторые фреймы, противопоставляемые в нашем описании, соответствуют по сути одной и той же экстралингвистической ситуации. Так, тесная одежда обычно облегает тело, однако мы выделяем здесь два разных фрейма – «тесная одежда» и «облегающая одежда». Основанием для такого решения является тот факт, что в ряде языков эти ситу-

ации выражаются разными лексемами, ср. в немецком:

(1) *Sie trug ihre übliche **straffe Hose** und das T-Shirt mit zwei roten Streifen.* (На ней были ее обычные облегающие брюки и футболка с двумя красными полосками.)

(2) *Ich könnte eine zu **enge Hose** keine 2 Stunden anlassen, da würde ich Bauchschmerzen bekommen!* (Я бы не мог и двух часов проносить тесные брюки, у меня бы заболел живот.)

Аналогичное противопоставление характерно для испанского языка, где имеется специальное прилагательное для «облегающей одежды», а «тесная одежда» совмещается с резинкой и узлом. Ср.:

(3) *Ella luce como siempre un elegante vestido largo, su **ajustado corsé** y un sombrero de paja.* (Она как

всегда блистает в элегантном длинном платье, **облегающем корсете** и соломенной шляпе.)

(4) *Durante 3 años, utilizó un **apretado corsé** que le dificultaba respirar con normalidad. (В течение трех лет она носила **тесный корсет**, который мешал ей нормально дышать.)*

В подобных случаях существенным параметром для лексикализации значений из поля «тугой» оказывается способ восприятия ситуации – переживается ли она изнутри, экспериенциально (ср. примеры 1 и 3) или же со стороны внешнего наблюдателя, зрительно (ср. примеры 2 и 4).

Для сравнения в русском, а также французском, английском, китайском и иврите данные фреймы объединяются в одной лексеме. Ср. для русского:

(5) *Чтобы попасть в раздевалку, нужно было обогнуть по периметру зал с тренажерами, которым распорядилась средних лет дама в **тугом трико**, обтягивавшем ее с развитыми зрелыми формами фигуру <зрительное восприятие>.*

(6) *Ведь после таких **тугих штанов** остаются синяки <внутреннее ощущение>.*

Стратегии совмещения фреймов: типологически частотные модели

В процессе работы ставилась задача не только собрать и проанализировать данные, но и понять, как лексический материал распределяется между выделенными нами фреймами, т.е. какие фреймы объединяются в одной лексеме. Рассмотрим некоторые характерные случаи совмещений.

Фреймы «веревка/струна» и «кожа» объединяются в одной лексеме во французском, немецком, английском, русском и сербском языках, ср. немецкое прилагательное *straff*:

(7) *Wenn eine **straffe Saite** vibriert, ist der ausgesandte Klang proportional zu der Länge der Saite. (Когда **тугая струна** вибрирует, издаваемое звучание пропорционально длине струны.)*

(8) ***Straffe Haut** ist eine Mischung aus Veranlagung und guter Pflege. (Упругая кожа – сочетание наследственной предрасположенности и хорошего ухода.)*

Еще одной частотной комбинацией фреймов в одной лексеме является группа «узел» + «резинка» + «тесная одежда» + «облегающая одежда». Эти фреймы покрываются одной лексемой в русском, французском, английском, китайском и иврите. Ср. во французском языке:

(9) *Tirer sur le bout peu à peu et serrer jusqu'à ce que vous obteniez un **noeud serré**. (Постепенно потянуть за кончик и тянуть, пока не получится «**тугой узел**».)*

(10) *Évitez de porter des tenues qui compressent l'abdomen: maillot trop moulant, porte-dossard trop serré, cuissard avec un **élastique serré** à la taille. (Ста-*

йтесь не носить одежду, которая сдавливает живот: слишком облегающую майку, слишком тугой номерной знак, трико с **тугой резинкой** на талии.)

(11) *Nous adorons toutes la silhouette extraordinaire que le **corset serré** à l'extrême nous donne. (Мы все любим ту великолепную фигуру, которую дает нам **тесный** до крайности **корсет**.)*

(12) *Avec des pantacourts très larges comme ceux-ci, on opte pour un **chandail serré** et un veston ajusté. (К широким капри, как эти, выбирают **обтягивающий свитер** и пиджак по фигуре.)*

В целом, языки нашей выборки прибегают к довольно разнообразным моделям совмещения фреймов. Это обусловлено тем, что некоторые из ситуаций, описываемых прилагательными поля «тугой» и «упругий», подразумевают комплексные ситуации, в которых лексема того или иного языка может «высвечивать» определенный аспект. Рассмотрим в качестве примера фрейм «узел». Так, «узел» может описываться как результат затягивания. В таком случае он совмещается со струной (концы веревки, из которой завязывается узел, натягиваются, и тем самым получается «тугой» узел) (ср. нем. *straff*). Узел может восприниматься визуально: его фрагменты тесно прижаты друг к другу, и это сближает его с фреймом «облегающая одежда» (одежда также тесно прижата к телу человека) – это характерно, например, для французского *serré*. Кроме того, при лексикализации данного фрейма может подчеркиваться следствие исходной ситуации затягивания узла – сложность его развязывания или прочность. При таком значении «узел» описывается тем же словом, что и «ветка», как в испанском или сербском языках. Возможность разного фокуса при описании объекта способствует не только межъязыковой, но и внутриязыковой вариативности: в одном языке могут существовать разные лексемные для одного фрейма. Ср. в испанском языке:

(13) *No hagas el **nudo tan apretado**. (Не затягивай слишком **тугой узел** (букв. **сжатый**)).*

(14) *Es un **nudo resistente**, utilizado para atar una cuerda a una argolla o bien para atar una cubeta que se va a sumergir en agua o se va a levantar. (Этот **прочный узел** используется для того, чтобы привязать веревку к кольцу или чтобы привязать ведро для погружения его в воду и подъема.)*

Таким образом, поле *тугой* подразумевает комплексную ситуацию, имеющую разные проявления, каждое из которых связано друг с другом семантическими отношениями различной природы. Семантические карты, построенные на основе полученных данных, отражают параметры, релевантные для лексических противопоставлений в том или ином языке. В частности, значимым для многих языков оказывается тип восприятия объекта: зрительный (внешний)

vs. экспериенциальный (внутренний). Многоаспектность ситуации обуславливает смежность изучаемой области с другими семантическими полями, в частности, с «тесным», «узким», «прочным», «твердым», «жестким», «плотным», «гибким», что задает перспективы дальнейшего исследования данного поля.

ЛИТЕРАТУРА

1. Рахилина Е. В. Фреймовый подход к лексической типологии / Е. В. Рахилина, Т. И. Резникова // Вопросы языкознания. – 2013. – № 2. – С. 3–31.

2. Кюсева М. В. Типологическая база данных адъективной лексики / М. В. Кюсева, Т. И. Резникова, Д. А. Рыжова // Компьютерная лингвистика и интеллектуальные технологии : по материалам ежегодной Международной конференции «Диалог» (Бекасово, 29 мая – 2 июня 2013 г.). Вып. 12 (19). – М. : Изд-во РГГУ, 2013. – С. 407–419.

3. Reznikova T. Wetness in a cross-linguistic perspective / T. Reznikova, V. Kruglyakova // The Typology of Physical Qualities. Benjamins / eds. E. Rakhilina, T. Reznikova.

Национальный исследовательский университет «Высшая школа экономики»

Баскакова Е. А., студентка факультета гуманитарных наук Школы Лингвистики

E-mail: baskakova.elen@gmail.com

Тел.: 8-963-778-38-64

4. Копчевская-Тамм М. С самыми теплыми чувствами (по горячим следам Стокгольмской экспедиции) / М. Копчевская-Тамм, Е. В. Рахилина // Типология и теория языка : от описания к объяснению. К 60-летию А. Е. Кибрика. – М. : ЯРК. – С. 462–487.

5. Глаголы движения в воде: лексическая типология / под ред. Т. А. Майсак, Е. В. Рахилиной. – М. : Индрик, 2007. – 752 с.

6. Cutting and breaking events : A crosslinguistic perspective. Cognitive linguistics (special issue) / eds. A. Majid, M. Bowerman. – 2007. – Vol. 18, № 2. – P. 133–152.

7. Брицын В. М. Концепт БОЛЬ в типологическом освещении / В. М. Брицын [и др.]. – Киев, 2009. – 424 с.

8. Круглякова В. А. Семантика глаголов вращения в типологической перспективе : дис. ... канд. филол. наук / В. А. Круглякова. – М. : РГГУ, 2010. – 350 с.

9. Talmy L. How language structures space / L. Talmy // Talmy L. Toward a cognitive semantics. V. I. / L. Talmy. – Cambridge, MA : MIT Press, 2000. – P. 177–254.

10. Рахилина Е. В. Когнитивный анализ предметных имен : семантика и сочетаемость / Е. В. Рахилина. – М. : Русские словари, 2000. – 416 с.

National Research University «Higher School of Economics»

Baskakova E. A., Student of the Linguistics School of the Humanities Faculty

E-mail: baskakova.elen@gmail.com

Tel.: 8-963-778-38-64