

Маркова Мария Владимировна

АВТОМАТИЧЕСКАЯ СЕМАНТИЧЕСКАЯ РАЗМЕТКА ПРЕДЛОЖЕНИЙ АНГЛИЙСКОГО ЯЗЫКА

Статья посвящена разработке метода автоматической семантической разметки предложений английского языка. Под семантической разметкой предложения подразумевается определение описанной в нем ситуации, а также разрешение лексической многозначности слов и словосочетаний, входящих в его состав. Рассматриваемый метод разметки опирается на такие лингвистические теории как "Семантика фреймов" Ч. Филлмора и "Классификация глаголов" Б. Левин. В основе его практической реализации – интеграция лингвистических ресурсов FrameNet, VerbNet, WordNet, синтаксического анализатора Stanford Syntax Parser.

Адрес статьи: www.gramota.net/materials/1/2013/6/31.html

Статья опубликована в авторской редакции и отражает точку зрения автора(ов) по рассматриваемому вопросу.

Источник

Альманах современной науки и образования

Тамбов: Грамота, 2013. № 6 (73). С. 107-109. ISSN 1993-5552.

Адрес журнала: www.gramota.net/editions/1.html

Содержание данного номера журнала: www.gramota.net/materials/1/2013/6/

© Издательство "Грамота"

Информация о возможности публикации статей в журнале размещена на Интернет сайте издательства: www.gramota.net

Вопросы, связанные с публикациями научных материалов, редакция просит направлять на адрес: almanac@gramota.net

УДК 004.89

Технические науки

Статья посвящена разработке метода автоматической семантической разметки предложений английского языка. Под семантической разметкой предложения подразумевается определение описанной в нем ситуации, а также разрешение лексической многозначности слов и словосочетаний, входящих в его состав. Рассматриваемый метод разметки опирается на такие лингвистические теории как «Семантика фреймов» Ч. Филлмора и «Классификация глаголов» Б. Левин. В основе его практической реализации – интеграция лингвистических ресурсов FrameNet, VerbNet, WordNet, синтаксического анализатора Stanford Syntax Parser.

Ключевые слова и фразы: семантическая разметка; ситуационный фрейм; семантический фрейм; актанты; семантические роли; классификация предикатов.

Маркова Мария Владимировна*Новосибирский национальный исследовательский государственный университет**(Новосибирский государственный университет)**markova.maria.v@gmail.com***АВТОМАТИЧЕСКАЯ СЕМАНТИЧЕСКАЯ РАЗМЕТКА ПРЕДЛОЖЕНИЙ АНГЛИЙСКОГО ЯЗЫКА[©]**

Задачу распознавания смысла текстов на естественном языке машиной научный мир решает уже давно. Заманчивой является идея о том, чтобы компьютерная система могла анализировать тексты, в частности, извлекать из них различные данные, факты, сведения, полезные при решении разных практических задач. Такие задачи могут возникнуть, например, в системах машинного перевода, в системах принятия решений, в вопросно-ответных системах, в задачах информационного поиска, в системах голосового управления и распознавания, список можно продолжать и далее.

В данной статье описана работа, целью которой являлась **разработка эффективного метода автоматической семантической разметки предложений английского языка на основе существующих лингвистических теорий и создание его практической реализации путем интеграции ресурсов, охватывающих большой объем языка, дополнения их слабых мест и добавления алгоритма означивания.**

Существенным плюсом данного метода можно назвать использование накопленного опыта прикладной лингвистики. Таким образом, в данной работе упор сделан не на составление словарных статей, наполнение лексических баз и другую объемную лингвистическую работу, требующую специальной подготовки, а на анализ существующих лингвистических теорий и их совместное использование, а с точки зрения программной реализации – на интеграцию лексических ресурсов.

Двумя основными теориями, лежащими в основе рассматриваемого метода автоматической семантической разметки предложений английского языка, являются *семантика фреймов* Ч. Филлмора [5] и *классификация глаголов* Б. Левин [6]. Поэтому для понимания метода разметки необходимо понимание данных теорий.

Первая из них – теория семантики фреймов – была озвучена Чарльзом Филлмором в 1976 году. В рамках данной теории язык рассматривается с точки зрения предикатной лексики, предполагает актантную структуру слов. Все слова в языке делятся на *предикатные* и *непредикатные*. Непредикатные называют предметы и их свойства, ими чаще всего являются существительные и прилагательные. Предикатными словами, как правило, являются глаголы и слова отглагольных частей речи. Они описывают ситуации, другими словами, некоторый фрагмент действительности, в котором можно выделить одного или нескольких участников либо отметить их значимое отсутствие. В теории семантики фреймов предикат образует *фрейм*, то самое описание ситуации с набором участников и атрибутов. Участники фрейма – *актанты*. Согласно данной теории, предикат накладывает логические ограничения на свои зависимые слова, выступающие актантами фрейма, в зависимости от той роли, которую они выполняют.

Например, предложение «*Carmen bought a dress for \$ 50 (Кармен купила платье за 50 долларов)*» отражает ситуацию **Покупки**. В данной ситуации всегда присутствуют (или подразумеваются) **Покупатель, Продавец, Товар, Плата** – они будут обязательными участниками ситуации. Также есть необязательные участники ситуации, не влияющие на ее понимание, такие как **Место, Время, Условие** и т.д. Причем, в роли **Покупателя** могут выступать **Человек** или **Организация**, но никак не может выступать **Животное** или **Предмет мебели**. Это логические ограничения, накладываемые предикатом на актанты.

Помимо подобных семантических ограничений, предикат диктует своим зависимым словам место в предложении и форму, в английском эти ограничения достаточно сильны, в отличие от русского, например. Такие синтаксические ограничения отражены в классификации глаголов Бет Левин. Это – вторая основная теория, лежащая в основе метода автоматической семантической разметки. Согласно исследованиям Б. Левин, существует корреляция между изменением смысла глагола и его синтаксическим поведением. По принципу схожести синтаксического поведения было классифицировано свыше 3000 английских глаголов и

выделено более 250 классов с оригинальными моделями управления, выраженными списком возможных синтаксических вариаций в каждом классе. Модель управления задает порядок слов в предложении, что не учтено в теории семантики фреймов.

Таким образом, влияние предиката на свои зависимые слова говорит о том, что синтаксис тесно связан с семантикой, и слово не имеет смысла без своего контекста. На данной идее базируется рассматриваемый метод автоматической семантической разметки. Описанные лингвистические теории представляют одни и те же признаки языка, но с разных сторон, и их совместное использование позволило создать более полную картину языка, а построенный на их основе метод дает возможность получить из внешнего представления предложения его смысловое содержание, именуемое семантической разметкой. Под семантической разметкой предложения подразумевается как описание ситуации, в нем скрытой, так и разрешение лексической многозначности слов, входящих в его состав.

Практическая реализация метода представляет собой интеграцию лингвистических ресурсов *FrameNet* [1], *VerbNet* [2] и *WordNet* [4]. Также при разметке задействованы синтаксический анализатор, в данной реализации им стал *Stanford Syntax Parser* [3], и отображения между ресурсами. *FrameNet* и *VerbNet* – это практические реализации описанных выше лингвистических теорий. *WordNet* – лексическая база данных, описывающая объекты реального мира; базовой сущностью которой является *синсет* (список синонимичных понятий). Часть используемых отображений между ресурсами была получена от сторонних разработчиков, часть – создана в рамках данной работы.

Процесс разметки условно можно разделить на два этапа. Первый – подготовительный, на данном этапе с помощью синтаксического парсера выделяется синтаксическая структура предложения, определяются ключевое слово, зависимости между словами, а также начальные формы слов и их части речи. Затем, с использованием полученной информации, всем словам предложения приписываются синсеты *WordNet*. Также на данном этапе по ключевому слову определяются подходящие семантические фреймы *FrameNet* и классы глаголов *VerbNet*. Итог подготовительного этапа – первоначальное означивание, без фильтрации «лишних смыслов».

В качестве примера возьмем предложение «*Carmen bought a dress for \$ 50*» и далее будем обращаться именно к нему. После подготовительного этапа ему будет приписана синтаксическая структура «*NP V NP PP*», где *NP* – *noun phrase*, *V* – *verb*, *PP* – *preposition phrase*, *PP* = {*for*} + *NP*; семантический фрейм – «*Commerce_buy*» из *FrameNet*, описывающий ситуацию покупки; класс глаголов «*get-13.5.1*» из *VerbNet*.

Второй этап – собственно означивание, в результате которого отсеиваются неподходящие предложению семантические фреймы и неверные (в контексте предложения) смыслы слов и словосочетаний.

В первую очередь синтаксическая структура предложения сравнивается с моделью управления, описанной в классе глаголов, определенном на прошлом этапе, и выделяются подходящие синтаксические вариации класса.

Для нашего примера из класса глаголов «*get 13.5.1*» будут выделены синтаксические вариации *NP.Agent V NP.Theme {for} NP.Asset* и *NP.Agent NP.Theme {for} NP.Beneficiary*. Они синтаксически эквивалентны, но последние элементы несут в себе семантические различия. Отследить эти различия и выбрать «правильную» вариацию можно с помощью отображения между ресурсами *FrameNet* и *VerbNet*.

Используя отображение между синтаксическими вариациями и семантическими фреймами, сопоставляющее их элементы, отбрасывается семантически неподходящая вариация.

Таким образом, в рассматриваемом примере останется только *NP.Agent V NP.Theme {for} NP.Asset*.

Так от конкретного предложения мы перешли к семантическому фрейму, определив при этом роли слов, входящих в его состав.

В примере «*Carmen bought a dress for \$ 50*»: «*Carmen*» – *NP.Agent* – **Buyer**, «*a dress*» – *NP.Theme* – **Goods**, «*\$ 50*» – *NP.Asset* – **Money**, где последние элементы последовательности соответствуют актантам семантического фрейма «*Commerce_buy*».

Далее, используя отображение между актантами синтаксических фреймов и синсетами *WordNet*, словам и словосочетаниям предложения сопоставляются конкретные смыслы. Таким образом решается проблема лексической многозначности.

Описанный метод автоматической семантической разметки универсален относительно используемых ресурсов, другими словами, лексическая база данных и словарь могут варьироваться в зависимости от предметной области и решаемой задачи, и дальнейшая разработка системы, построенной на базе метода, может осуществляться уже не за счет его модификации, а за счет обогащения используемых лингвистических ресурсов. Относительно языка он также универсален. Текущая реализация предполагает работу с английскими предложениями, поскольку были использованы английские версии *WordNet* и *FrameNet*, но оба ресурса многоязычны, соответственно, их замена позволит размечать тексты на других языках.

В заключение отметим, что описанный метод автоматической семантической разметки предложений имеет как теоретическую, так и прикладную ценность. Результаты тестирования показывают, что подход к семантической разметке предложения, опирающийся на семантику фреймов и классификацию глаголов, учитывающий синтаксическую структуру предложения, работает, дает положительные результаты, может развиваться и дополняться.

Список литературы

1. **Онлайн-версия ресурса *FrameNet*** [Электронный ресурс]. URL: <https://framenet.icsi.berkeley.edu> (дата обращения: 15.05.2013).
2. **Онлайн-версия ресурса *VerbNet*** [Электронный ресурс]. URL: <http://verbs.colorado.edu/~mpalmer/projects/verbnet.html> (дата обращения: 15.05.2013).
3. **Официальный информационный ресурс проекта *Stanford Syntax Parser*** [Электронный ресурс]. URL: <http://nlp.stanford.edu/software/lex-parser.shtml> (дата обращения: 15.05.2013).
4. **Официальный информационный ресурс проекта *WordNet*** [Электронный ресурс]. URL: <http://wordnet.princeton.edu> (дата обращения: 15.05.2013).
5. **Fillmore Ch. J.** Frame Semantics // *Linguistics in the Morning Calm*. Seoul, South Korea: Hanshin Publishing Co., 1982. P. 111-137.
6. **Levin B.** English Verb Classes and Alternations: a Preliminary Investigation. Chicago: University of Chicago Press, 1993.
7. **Ruppenhofer J. et al.** FrameNet II: Extended Theory and Practice. Berkeley, California: International Computer Science Institute, 2006.

УДК 008:930(07)

Культурология

Статья раскрывает значимость режиссерской работы, которую отличают смыслопорождение, созидание новой семиотической реальности. Режиссерская работа имеет свои пространственно-временные координаты, в рамках которых порождаются новая культурно-театральная реальность и уникальное культурно-историческое событие, действо, существующее в нераздельном единстве переживаний, чувственного мира зрителя, актера, режиссера. Основное внимание автор обоснованно акцентирует на культурологическом анализе художественного пространства и художественного времени, которые, являясь формой координации персонажей и их состояний, способствуют в конечном итоге показу относительности факторов человеческого бытия.

Ключевые слова и фразы: искусство; художественное пространство; художественное время; хронотоп; спектакль; режиссер.

Молтобарова Кумысбала Ишановна, д. филос. н., профессор

Академия пограничной службы Республики Казахстан

moltobarova@mail.ru

ВЗАИМОСВЯЗЬ ВРЕМЕННЫХ И ПРОСТРАНСТВЕННЫХ ОТНОШЕНИЙ В ИСКУССТВЕ®

Искусство испокон веков показывает относительность всех социальных ценностей, норм, отношений, институтов – нравственных, художественных, экономических, исторических, политических, правовых, научных и т.д. Бесконечность и неисчерпаемость бытия и жизни ведут художника к поискам социальной и моральной гармонии в искусстве. Ведь не всегда эта относительность социальных явлений обнаруживает себя на поверхности, она, как правило, скрыта, существует, как говорят, имплицитно.

А особенности отношения к окружающему миру, степень его осознания проявляются наиболее отчетливо в уровне и формах восприятия таких всеобщих понятий как пространство и время, которые в искусстве обозначаются как художественное пространство и художественное время.

В истории философии вопрос о познавательном статусе категорий пространства и времени рассматривался по-разному. Одни философы придерживались субстанциальной концепции, где отношение между пространством, временем и материей виделось как отношение между двумя видами самостоятельных субстанций. Другие, в реляционной концепции, пространство и время представляли как системы отношений, образуемых взаимодействующими материальными объектами, вне них несуществующими. А создание теории относительности привело к раскрытию диалектических связей в природе, опровергнув субстанциальную концепцию и подтвердив правильность реляционной концепции. Но теория относительности точку над і в этом вопросе не поставила, споры и вокруг нее еще продолжаются. Вообще, представления о времени и пространстве всегда вызывали полемику. Так, И. Ньютон и его сторонники, отстаивая объективность пространства и времени, для наглядности сравнивали их с резервуаром, вместилищем, с ящиком без стен, с наемной квартирой, которая может быть занятой или пустой. Кант резко критиковал Ньютона за его понимание пространства и времени как предметного бытия, а сам их истолковывал как априорные формы созерцания рассудка. В общем, идеалисты объясняли пространство и время как духовные сущности, домарксистский материализм – как предметные сущности, естествоиспытатели придерживались взглядов, сходных с точкой зрения Ньютона.

Насколько своеобразным, оригинальным было представление о времени и пространстве в искусстве можно представить, соприкоснувшись с творчеством народного артиста СССР и Казахской ССР, Лауреата