

Маркова Мария Владимировна

**СТАНДАРТИЗАЦИЯ ЛЕКСИКО-СЕМАНТИЧЕСКИХ РЕСУРСОВ В ТЕРМИНАХ СТАНДАРТА
МЕТАОБЪЕКТ FACILITY**

Статья посвящена модификации метода автоматической семантической разметки предложений английского языка. Описанная в статье модификация метода опирается на стандарт MetaObject Facility, существенные плюсы подобного видоизменения в том, что метод разметки становится независим от конкретики языка и степени наполнения ресурсов. Новое представление позволяет, не нарушая стандарта MOF, добавить онтологии, что повышает точность метода и расширяет область его применения.

Адрес статьи: www.gramota.net/materials/1/2015/3/18.html

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

Источник

Альманах современной науки и образования

Тамбов: Грамота, 2015. № 3 (93). С. 72-76. ISSN 1993-5552.

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

Содержание данного номера журнала: www.gramota.net/materials/1/2015/3/

© 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

пассионарный толчок мог дать жизнь новым этносам – китайскому, японскому, южноафриканскому, и они вскоре будут играть лидирующую роль на международной арене. Прошло 220-230 лет с момента толчка, и новые этносы находятся в фазе подъёма, которая сопровождается быстрым ростом пассионарности, постепенным расширением территориальных владений.

Таким образом, для украинского этноса будет более полезно связать свою историческую судьбу с российским, находящимся приблизительно в той же фазе этногенеза и положительно комплементарным, чем с западноевропейским суперэтносом – более старым и отрицательно комплементарным. Такая интеграция будет положительна и для российского этноса – это даст возможность скорейшего объединения российского (евразийского) суперэтноса в Евразийский Союз, который станет противовесом Европейскому. В евразийском пространстве украинский этнос сохранит свою национальную идентичность, в то время как в западноевропейском её потеряет. Исходя из теории этногенеза, западноевропейский суперэтнос в скором времени уступит лидерство более молодым этносам – ими могут стать японский, китайский, южноафриканский (если учесть, что пассионарный толчок XVIII в. действительно прошел по их линии).

Список литературы

1. Гумилев Л. Н. История – наука естественная, или Визит к профессору Гумилеву // Сельская молодежь. 1988. № 2. С. 44-49.
2. Гумилев Л. Н. От Руси к России. М.: АСТ; Астрель, 2012. 411+5 с.
3. Две Украины и одно решение [Электронный ресурс]. URL: <http://www.evrazia.org/article/1825> (дата обращения: 07.02.2015).
4. Ольховиков О. Будущее Украины и мира в свете пассионарной теории [Электронный ресурс]. URL: http://www.2000.ua/v-nomerc/svoboda-slova/ekspertiza_svoboda-slova/buduschec-ukrainy-i-mira-v-svete-passionarnoj-teorii_arhiv_art.htm (дата обращения: 07.02.2015).

GEOPOLITICAL SITUATION CONCERNING THE UKRAINIAN ETHNOS IN VIEW OF PASSIONARY THEORY OF ETHNOGENY

Lepinskaya Violetta Olegovna

Academy for the Humanities and Pedagogy (Branch) of Crimea Federal University named after V. I. Vernadsky in Yalta
violetta09@bk.ru

This article studies a geopolitical situation concerning the Ukrainian ethnos in terms of L. N. Gumilev's ethnogeny conception. The idea of creating the Eurasian Union in contrast to the European one is considered and a thesis that Ukraine will play its part in it is advanced. Moreover, the passionary impetus of the XVIII century, which was not described by Lev Nikolaevich in his main works, but was mentioned in one of his interviews, is studied.

Key words and phrases: the Ukrainian ethnos; passionarity; the Eurasian Union; complementarity; passionary impetus; super-ethnos.

УДК 004.89

Технические науки

Статья посвящена модификации метода автоматической семантической разметки предложений английского языка. Описанная в статье модификация метода опирается на стандарт MetaObject Facility, существенные плюсы подобного видоизменения в том, что метод разметки становится независим от конкретики языка и степени наполнения ресурсов. Новое представление позволяет, не нарушая стандарта MOF, добавить онтологии, что повышает точность метода и расширяет область его применения.

Ключевые слова и фразы: семантическая разметка; ситуационный фрейм; семантические роли; мета-объектное средство; MetaObject Facility; метамодель; онтология.

Маркова Мария Владимировна

Новосибирский национальный исследовательский государственный университет
markova.maria.v@gmail.com

СТАНДАРТИЗАЦИЯ ЛЕКСИКО-СЕМАНТИЧЕСКИХ РЕСУРСОВ В ТЕРМИНАХ СТАНДАРТА METAOBJECT FACILITY

Введение

Лексико-семантические ресурсы используются в большинстве задач, связанных с обработкой естественного языка, таких как: разрешение лексической многозначности, извлечение информации, а также определение семантических ролей и типов. В последние годы аспекты переиспользования и объединения лексико-

семантических ресурсов становятся более актуальными. Во многом это связано с трудозатратностью построения таких ресурсов с нуля. Стандартизация ресурсов играет важную роль в данном контексте, поскольку облегчает их интеграцию и делает переиспользование более простым. Системы обработки естественных языков, построенные согласно стандартам, могут легко подключать стандартизованные лексико-семантические ресурсы, и, таким образом, для них будет возможно легкое переключение между различными стандартизованными ресурсами. Другими словами, стандартизация лексико-семантических ресурсов делает их совместимыми.

В этой работе речь пойдет об описании метода автоматической семантической разметки (или ACP) [2], включающего в себя интеграцию лексико-семантических ресурсов *WordNet* [18], *FrameNet* [7], *VerbNet* [17] и реализованного для предложений английского языка. Предлагается рассмотреть метод в терминах стандарта *MetaObject Facility* [14], подобное изменение представления позволит видоизменить алгоритм так, что он станет независим от конкретики языка и степени наполнения ресурсов. Также будет предложена модификация метода через добавление онтологий, направленная на повышение точности семантической разметки.

Разработки по описанной тематике популярны и востребованы. С точки зрения схожести подхода к объединению ресурсов (выделения моделей и метамodelей) интересна разработка *Ubiquitous Knowledge Processing Lab* (Германия), описанная в статье [5].

MetaObject Facility

MetaObject Facility (MOF, переводится как метаобъектное средство) – один из стандартов OMG¹ для разработки, управляемой моделями. Это – закрытая архитектура метамоделирования; в ней определяется МЗ-модель, конформная самой себе. MOF задает строгую архитектуру метамоделирования, где элементы модели нижележащих слоев строго соответствуют элементам модели вышележащих слоев. MOF ограничивается обеспечением средств для определения структуры или абстрактного синтаксиса языка или данных. В качестве примера можно сравнить MOF с расширенной формой Бэкуса – Наура (РБНФ), а именно: MOF играет ту же роль для определения метамодели, которую РБНФ играет для определения грамматики языка программирования.



Рисунок 1. Текст как модель в архитектуре MOF

MOF реализовано как четырехслойная архитектура, где M0 – самый нижний слой, представляющий объекты реального мира, M1 – модель объектов реального мира, M2 – язык описания моделей слоя M1, или метамодель и M3 – язык описания моделей слоя 2, или мета-метамодель. Для текста, как описания реального мира, MOF-представление изображено на Рисунке 1. MOF задумывалось как переход между разными метамоделями, поскольку представляет собой мощную основу для их описания. Если две отличные друг от друга метамодели конформны MOF (другими словами, они могут быть описаны посредством M3), то все конкретные модели слоя M1, базирующиеся на них, могут располагаться в общем хранилище и совместно обрабатываться с использованием модельных трансформаций.

MOF является частью концепции *Model Driven Architecture* (MDA) [12]. Это – модельно-ориентированный подход к разработке программного обеспечения. Суть MDA состоит в построении абстрактной метамодели управления и обмена метаданными (или моделями) и определении способов трансформации метамодели в поддерживаемые технологии программирования, где MOF определяет создание метамодели.

Метод автоматической семантической разметки

Лексико-семантические ресурсы, предлагаемые для рассмотрения в данной статье – *WordNet*, *FrameNet*, *VerbNet*, а также семантический парсер были подробно описаны в работе [2], и с их помощью был реализован одноименный метод разметки.

Цель метода автоматической семантической разметки – сопоставить исходному неозначенному предложению семантическое означивание путем ряда трансформаций. Первая трансформация – получение синтаксической структуры текста. В практической реализации метода достигается с помощью *Stanford Syntax Parser* [16], который определяет ключевое слово, части речи и начальные формы всех слов в предложении и зависимости между ними.

Затем, с использованием полученной информации, всем словам предложения приписываются синсеты *WordNet*, по ключевому слову определяются подходящие семантические фреймы *FrameNet* и классы глаголов

¹ Консорциум *Object Management Group* занимается продвижением и разработкой объектно-ориентированных технологий и стандартов. Официальный сайт консорциума: <http://www.omg.org/>

VerbNet. Это – вторая трансформация. Также она включает в себя отсеивание лишних смыслов. Существующая на данный момент реализация метода производит разметку предложений английского языка.

В качестве третьей трансформации предлагается произвести переход к аннотации текста, опирающейся на термины подходящей онтологии. Эта трансформация – часть будущих работ по расширению метода автоматической семантической разметки. Трансформации представлены на Рисунке 2.

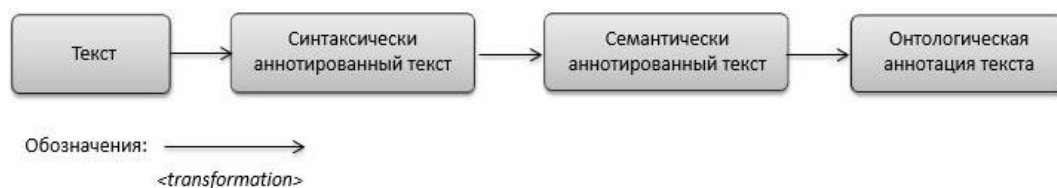


Рисунок 2. Трансформации в методе АСР

Применение стандарта MOF к методу автоматической семантической разметки. Добавление онтологий

Интеграция лексико-семантических ресурсов *WordNet*, *FrameNet* и *VerbNet* достигнута в методе семантической разметки за счет отображений между ресурсами.

В данном разделе предлагается описать элементы обобщенного метода семантической разметки в терминах стандарта *MetaObject Facility*. Для этого будут выделены объекты слоя M0 – “реального мира” – и определены их модели и метамодели.

Метод автоматической семантической разметки задает означивание предложения, таким образом, первый элемент слоя M0 – предложение. Предложение – частный случай обобщенного понятия текст, методы обработки построения и анализа текстов распространяются на предложения. Рассмотрим их.

Информация заключена в тексте, но текст строится и анализируется с использованием естественного языка (или сокращенно ЕЯ). Общая схема обработки текстов инвариантна по отношению к выбору ЕЯ: независимо от того, на каком языке написан исходный текст, его анализ проходит одни и те же стадии: разбиение текста на предложения и слова, определение характеристик отдельных слов и синтаксический анализ, семантический анализ. Причем, наиболее зависима от языка стадия характеристики отдельных слов и синтаксического анализа. Ввиду вышесказанного, на уровне M1, модели текста, помещаем модель языка. Модель языка описывает его основные характеристики (здесь мы не будем углубляться в лингвистику, скажем лишь, что на данном уровне английский и арабский языки будут отличаться как минимум направлением чтения символической последовательности и кодировкой в случае автоматизированной обработки текста). Таким образом, для элемента слоя M0 “Текст” на слое M1 будет располагаться модель конкретного ЕЯ, а на M2 – способ описания моделей языка, или метамодель естественного языка. Про метамодель стоит сказать, что это – сложный элемент, который включает в себя синтаксическую и семантическую метамодели среди прочих.

Синтаксически аннотированное предложение – следующий элемент слоя M0. Это – синтаксическая структура предложения, составляется на базе глубинной структуры текста [1] и теории формальных грамматик [3], в нее заложенной. По этой причине глубинная структура текста – модель синтаксической структуры текста – зависит от языка. В качестве метамодели выступает теория, лежащая в основе генеративной лингвистики, – формальная грамматика. Формальная грамматика, или просто грамматика в теории формальных языков, – способ описания формального языка, то есть выделения некоторого подмножества из множества всех слов некоторого конечного алфавита. Это – часть метамодели языка, описанной на слое M2.

Третий элемент слоя M0 – семантическая разметка текста, полученная в результате трансформации. В ней содержится информация о фреймах *FrameNet*, подходящих данному предложению, словам и словосочетаниям приписаны синсеты *WordNet*, определен соответствующий класс глаголов *VerbNet*, как результат – модель управления предложением. Над этим элементом слоя M0 стоит модель, представляющая собой объединение англоязычных ресурсов *WordNet*, *FrameNet* и *VerbNet*. Объединения удалось достигнуть путем добавления отображений между ресурсами, подробно описанных в работе [2].

Метамодель объединенной модели лексическо-семантических ресурсов *WordNet*, *FrameNet* и *VerbNet* – объединенное представление теорий, лежащих в основе данных ресурсов. Описание этой метамодели – тема отдельной статьи, но краткое описание теорий, лежащих в ее основе, приведено в следующих предложениях. Для англоязычного *WordNet* метамоделью является лексическая онтология *WordNet* [13]. Она активно используется для создания *WordNet* для различных языков, в том числе русского [4; 9]. Метамодель *FrameNet* – лексикон, основанный на семантике фреймов Чарльза Филмора [6]. На базе этой метамодели создано множество *FrameNet-style* лексических баз для различных языков, таких как: английский, японский, испанский, немецкий, польский и другие [8]. Для *VerbNet* также создаются адаптации для ряда языков помимо английского [11]. В их основе лежит классификация глаголов [10], которую можно считать метамоделью *VerbNet*.

Как и исходный метод АСР, реализованный для предложения английского языка, подход, представленный на Рисунке 3, содержит еще один уровень разметки предложения – онтологический. Добавление трансформации, порождающей онтологическую разметку, позволяет определить концептуальные связи между объектами предложения, что расширяет область применения метода.

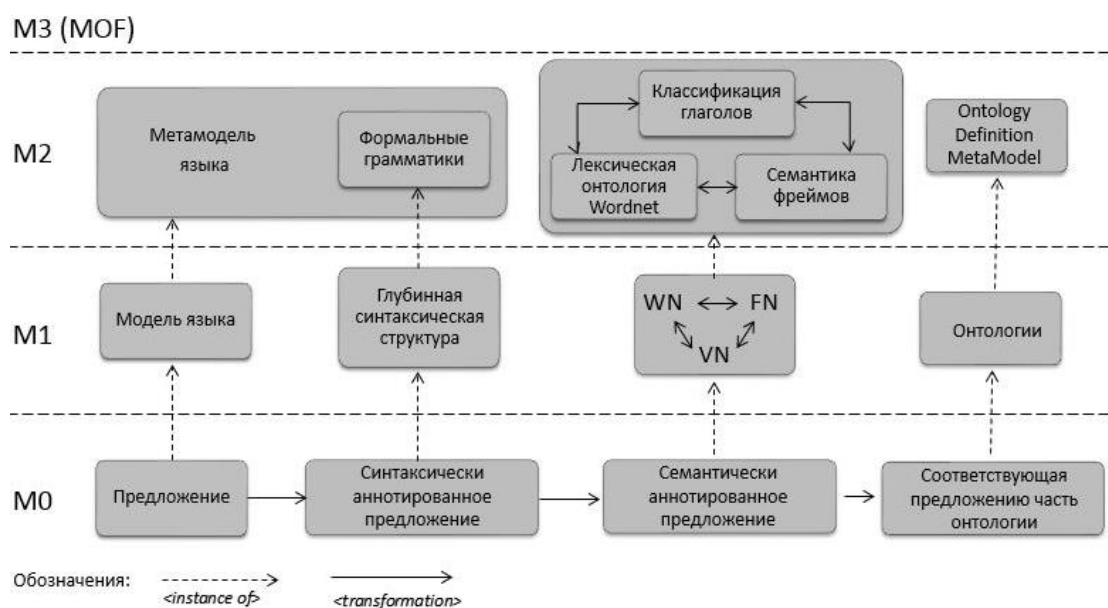


Рисунок 3. Метод АСР в MOF-представлении

Онтологическая разметка предложения

Онтология определяет понятия, их значения, а также отношения между ними в рамках задачи описания некоторой предметной области. Онтология включает:

- понятия из исследуемой предметной области;
- отношения между этими понятиями;
- свойства, или атрибуты, этих понятий;
- функции и процессы, в которые вовлечены понятия;
- ограничения и правила, накладываемые на понятия в рамках предметной области.

Использование онтологий при разметке предложения позволяет дополнить информацию, содержащуюся в нем непосредственно, информацией из онтологии. Наличие в онтологии связей, ограничений и правил, накладываемых на понятия, позволяет строить запросы для выборки информации из онтологии, делать логические выводы, если они не заданы в явном виде, производить оценки и заключения.

По своей сути онтологическая разметка уровня M0 – ограниченная часть онтологии, включающая в себя объекты, описанные в исходном предложении. Модель данного элемента – непосредственно подходящая по контексту онтология. Мета модель – еще одна спецификация OMG, *Ontology Definition Model* (ODM) [15]. Данная спецификация позволяет применять концепты MDA при разработке онтологий. В представленной схеме успешно вписывается в MOF-представление метода.

Заключение

Было показано, как можно дополнить существующий метод автоматической семантической трансформации, создающей онтологическую аннотацию. Также метод был описан в терминах спецификации *MetaObject Facility*.

Обобщенное до уровня моделей и метамodelей представление метода автоматической семантической разметки дает новые возможности его развития. Такая архитектура позволяет уйти от конкретики языка. А также наличие MOF-представления позволит уйти от ручной разработки к полуавтоматической, согласно методологии MDA (Model Driven Architecture) [12].

Добавление онтологий повышает качество семантической разметки и расширяет список возможностей при работе с текстом. Содержащиеся в онтологиях связи между понятиями, их ограничения, а также список правил позволяют строить запросы для выборки информации из онтологий, производить логические выводы, оценки и заключения.

Список литературы

1. **Глубинная структура** [Электронный ресурс] // Лингвистический энциклопедический словарь. URL: <http://tapemark.narod.ru/lcs/110a.html> (дата обращения: 13.01.2015).
2. **Маркова М. В.** Автоматическая семантическая разметка предложений английского языка: магистерская диссертация. Новосибирск: НГУ, 2013.
3. **Математическая лингвистика** [Электронный ресурс] // Лингвистический энциклопедический словарь. URL: <http://www.tapemark.narod.ru/lcs/287c.html> (дата обращения: 13.01.2015).
4. **Русский WordNet** [Электронный ресурс]. URL: <http://wordnet.ru/> (дата обращения: 13.01.2015).
5. **Eckle-Kohler J., Gurevych I.** Standardizing Lexical-Semantic Resources – Fleshing out the Abstract Standard LMF. Darmstadt, 2012. 10 p.

6. **Fillmore Ch. J.** Frame Semantics and the Nature of Language // Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech. Berkley, California, 1976. P. 21-23.
7. **FrameNet** [Электронный ресурс]. URL: <https://framenet.icsi.berkeley.edu> (дата обращения: 13.01.2015).
8. **FrameNets in Other Languages** [Электронный ресурс]. URL: https://framenet.icsi.berkeley.edu/frndrupal/framenets_in_other_languages (дата обращения: 13.01.2015).
9. **Global WordNet Association** [Электронный ресурс]. URL: <http://globalwordnet.org/> (дата обращения: 13.01.2015).
10. **Levin B.** English Verb Classes and Alternations: A Preliminary Investigation. Chicago, USA: University of Chicago Press, 1993. 348 p.
11. **Lin Sun, Korhonen A., Poibeau T., Messiant C.** Investigating the Cross-Linguistic Potential of VerbNet-Style Classification [Электронный ресурс]. URL: http://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CCIQFjAA&url=http%3A%2F%2Fwww.cl.cam.ac.uk%2F~alk23%2Fcoling-2010-french.pdf&ei=8L60VPO9K-ndywPR_YCYCw&usq=AFQjCNF4KHldXm-azHjLwl0he2D56R1tcg&bvm=bv.83339334,d.bGQ (дата обращения: 13.01.2015).
12. **MDA – The Architecture of Choice for a Changing World** [Электронный ресурс]. URL: <http://www.omg.org/mda/> (дата обращения: 13.01.2015).
13. **Miller G. A., Beckwith R., Fellbaum C. D., Gross D., Miller K.** Introduction to WordNet: An Online Lexical Database // International Journal of Lexicography. 1990. Vol. 3. № 4. P. 235-244.
14. **OMG's MetaObject Facility** [Электронный ресурс]. URL: <http://www.omg.org/mof/> (дата обращения: 13.01.2015).
15. **Ontology Definition Metamodel (ODM)** [Электронный ресурс]. URL: <http://www.omg.org/spec/ODM/> (дата обращения: 13.01.2015).
16. **The Stanford Parser: A Statistical Parser** [Электронный ресурс]. URL: <http://nlp.stanford.edu/software/lex-parser.shtml> (дата обращения: 13.01.2015).
17. **VerbNet. A Class-Based Verb Lexicon** [Электронный ресурс]. URL: <http://verbs.colorado.edu/~mpalmer/projects/verbnet.html> (дата обращения: 13.01.2015).
18. **WordNet. A Lexical Database for English** [Электронный ресурс]. URL: <http://wordnet.princeton.edu> (дата обращения: 13.01.2015).

STANDARDIZATION OF LEXICAL-SEMANTIC RESOURCES IN THE TERMS OF METAOBJECT FACILITY STANDARD

Markova Mariya Vladimirovna
Novosibirsk State University
markova.maria.v@gmail.com

The article is devoted to the modification of the method of the automatic semantic marking of the English sentences. The modification of the method described in the paper is based on *MetaObject Facility* (MOF) standard, the significant advantages of this modification is that the method of marking becomes independent on the specifics of language and the degree of resources fullness. New representation enables to add ontology without violating the standard of MOF that increases the accuracy of the method and extends the field of its application.

Key words and phrases: semantic markup; situational frame; semantic roles; *MetaObject Facility*; metamodel; ontology.

УДК 323

Социологические науки

Статья посвящена улучшению механизма противодействия коррупции. На примере передового опыта борьбы с коррупцией предложены некоторые рекомендации по совершенствованию и ужесточению законодательной базы РФ, регулирующей антикоррупционную деятельность. Выявлена и обоснована необходимость общественного контроля и влияния на процесс принятия общественно важных решений.

Ключевые слова и фразы: противодействие коррупции; государственная и муниципальная служба; общественный контроль; антикоррупционная политика; ротация кадров; совершенствование механизма противодействия коррупции.

Михолап Яна Владиславовна

Комсомольский-на-Амуре государственный технический университет
pinacolada-girl@mail.ru

О СОВЕРШЕНСТВОВАНИИ МЕХАНИЗМА ПРОТИВОДЕЙСТВИЯ КОРРУПЦИИ

Проблема коррупции в той или иной степени существует во всех государствах, и наша страна не является исключением. Коррупция в России проникла во все сферы деятельности общества, но наиболее подвержены