



«УТВЕРЖДАЮ»

Проректор по научной работе СПбГУ

С.В. Аплонov

2018г.

ЗАКЛЮЧЕНИЕ

федерального государственного бюджетного образовательного учреждения высшего образования "Санкт-Петербургский государственный университет"

Диссертация «Анализ параметров семантической связности с помощью дистрибутивных семантических моделей (на материале русского языка)» выполнена в федеральном государственном бюджетном образовательном учреждении высшего образования «Санкт-Петербургский государственный университет» на кафедре математической лингвистики.

В период подготовки диссертации соискатель **Паничева Полина Вадимовна** проходила обучение в очной аспирантуре федерального государственного бюджетного образовательного учреждения высшего образования «Санкт-Петербургский государственный университет» (с 2015 года по 2018 год).

В 2008 году окончила отделение прикладной лингвистики при кафедре математической лингвистики факультета филологии и искусств федерального государственного бюджетного образовательного учреждения высшего образования «Санкт-Петербургский государственный университет».

В 2018 году окончила очную аспирантуру при кафедре математической лингвистики филологического факультета федерального государственного бюджетного образовательного учреждения высшего образования «Санкт-Петербургский государственный университет» по специальности 10.02.21 – Прикладная и математическая лингвистика.

Документы о сдаче кандидатских экзаменов выданы в 2018 году федеральным государственным бюджетным образовательным учреждением высшего образования «Санкт-Петербургский государственный университет».

Научный руководитель – кандидат филологических наук, доцент Ольга Александровна Митрофанова работает в должности доцента на кафедре математической лингвистики филологического факультета федерального государственного бюджетного образовательного учреждения высшего образования «Санкт-Петербургский государственный университет».

По итогам обсуждения диссертации «Анализ параметров семантической связности с помощью дистрибутивных семантических моделей (на материале русского языка)» принято следующее заключение.

Оценка выполненной соискателем работы

В диссертационном исследовании Паничевой Полины Вадимовны решена важная научная задача, связанная с моделированием параметров семантической связности на основе дистрибутивных семантических моделей.

В результате исследования дано теоретическое обоснование взаимодействия парадигматического и синтагматического компонентов в сочетании семантически связанных лексем; разработан и практически реализован алгоритм анализа семантической связности в синтаксически связанных словосочетаниях и линейных контекстах; алгоритм успешно применен для автоматической идентификации метафорических контекстов русских глаголов.

Теоретическое лингвистическое обоснование работы опирается на концепции Р.Якобсона, В.Г.Гака, Д.Н.Шмелева, Ю.Д.Апресяна как представление о соотношении парадигматического и синтагматического в сочетании связанных семантически лексем, а также концепции Дж.Лакоффа и Й.Уилкса как интерпретации когнитивных оснований для переноса (видоизменения) значения слов в лингвистическом контексте, отражающем экстралингвистический. Параметры семантической связности определены на основе расстояния между контекстными векторами в дистрибутивно-семантической модели.

В ходе работы получены и проанализированы эмпирические данные о статистико-комбинаторных параметрах «правильности» применительно к тематически разным текстам и корпусам. На основе этих данных был определен диапазон семантической связности, заданы границы собственно аномальности (неправильности). Сформулировано и проверено предположение о том, что метафорические употребления русских глаголов характеризуются определенным местом в диапазоне семантической связности, а именно, представляют собой явления, лежащие на нижнем отрезке диапазона семантической связности. Это предположение было проверено на материале 110 контекстов, выбранных случайным образом из веб-корпуса русского языка RuTenTen для каждого из девяти многозначных переходных русских глаголов (*бомбардировать, очертить, пилить, распылять, разбавлять, выкраивать, взорвать, взвесить, зажигать*). Сделан вывод о том, что диапазон семантической связности характеризует метафорические употребления глаголов как отличные от прямых употреблений. Соответственно, диапазон семантической связности признан полезным параметром в автоматической идентификации метафоры. На основе этих результатов предложено применение параметров семантической связности не только в исследовании метафоры, но и для выявления лексических ошибок и клинических нарушений речи.

Диссертация является законченной научной квалификационной работой, соответствующей требованиям, предъявляемым к кандидатским диссертациям.

Личное участие соискателя в получении результатов, изложенных в диссертации

В процессе работы над диссертацией Паничева Полина Вадимовна провела самостоятельное экспериментальное исследование, в ходе которого получены следующие результаты:

1. Разработан и реализован алгоритм анализа параметров семантической связности. Осуществлено применение алгоритма анализа параметров семантической связности для автоматической идентификации метафоры.
2. Выявлен диапазон показательных параметров семантической связности для метафорических употреблений набора глаголов русского языка.
3. Проведена оценка работы алгоритма анализа параметров семантической связности и применимости аппарата анализа семантической связности в задаче автоматической идентификации метафоры:
 - а. Реализована трехкратная кросс-валидация данных о контекстах метафорических и прямых употреблений глаголов, получено соответствующее разбиение данных на обучающую и тестовую выборки;
 - б. Проведен эксперимент по машинному обучению при решении задачи бинарной классификации на метафорические/прямые употребления глаголов.
4. Доказано, что алгоритм анализа параметров семантической связности является адекватным инструментом для решения задачи автоматической идентификации метафоры;
5. Определено, что глагольная метафора в русском языке характеризуется низкими значениями параметров семантической связности.

Степень достоверности результатов проведенного исследования

Достоверность научных результатов исследования определяется следующими особенностями процедуры подбора и анализа данных: 1) использованием процедуры разметки метафорических/прямых словоупотреблений Metaphor Identification Procedure Vrije Universiteit te Amsterdam (MIPVU), для которой в предшествующих работах была проведена оценка согласованности экспертов; в результате была показана высокая степень согласованности экспертов при разметке русскоязычных текстов в соответствии с MIPVU; 2) использованием данных, извлеченных из веб-корпуса русского языка RuTenTen (общий объем свыше 18 млн. словоупотреблений) и представленных в виде контекстных выборок, характеризующих сочетаемость девяти различных глаголов; при этом установлено, что метафорические употребления большинства глаголов характеризуются единообразной тенденцией к низким значениям параметров семантической связности, что говорит о достоверности полученной модели метафорического значения глаголов; 3) проведением эксперимента по машинному обучению на основе кросс-валидации данных и получением высоких результатов для автоматической

идентификации метафоры, подтверждающих значимость параметров семантической связности для данной задачи.

Новизна результатов проведенного исследования

Новизна исследования состоит в следующем:

- 1) впервые применены принципы дистрибутивно-семантического моделирования для операционализации понятия семантической связности;
- 2) описаны различные диапазоны параметров семантической связности на уровне конструкций, ошибок, метафоры, нарушения связности текста;
- 3) впервые использованы данные о параметрах семантической связности для решения лингвистических и междисциплинарных задач на материале русского языка.

Практическая значимость результатов проведенного исследования

Проведённое исследование вносит серьезный вклад в понимание семантической связности и ее роли в контексте глагольных употреблений. Реализован конструкт семантической связности и алгоритм оценки его диапазонов в рамках дистрибутивной семантики, который может быть использован для выявления метафорических словоупотреблений, лексических ошибок, нарушений речи. Выявлен характерный для метафорических употреблений глаголов диапазон семантической связности, что позволяет формализовать идентификацию глагольной метафоры, в том числе, распространить алгоритм анализа параметров семантической связности на более широкий круг глаголов. В перспективе понятие семантической связности может быть успешно применено к автоматическому выявлению и исправлению лексических ошибок, а также автоматической идентификации нарушений связности речи в клинической лингвистике.

Результаты исследования могут быть использованы в курсах по автоматическому анализу русскоязычных корпусов текстов, моделированию семантического уровня естественного языка.

Полнота изложения материалов диссертации в работах, опубликованных соискателем

Основные результаты выполненного исследования изложены в 7 печатных работах, из которых 6 опубликованы в ведущих рецензируемых изданиях, проиндексированных в международной базе данных Scopus и ведущих рецензируемых изданиях, рекомендуемых ВАК, что подтверждает достаточную степень апробации исследования.

1. Panicheva, P. A Multi-Feature Classifier for Verbal Metaphor Identification in Russian Texts / Y. Badryzlova, P. Panicheva // Conference on Artificial Intelligence and Natural Language: 7th International Conference, AINL 2018, St. Petersburg, Russia, October 17-19, 2018, Proceedings / ed. by

D. Ustalov, A. Filchenkov, L. Pivovarova, J. Žižka. – CCIS, vol. 930. – Springer, Cham, 2018. – P. 23-34 (0.73 п.л., доля авт. 0.2 п.л.). – ISBN (online) 978-3-030-01204-5, ISSN (online) 1865-0937, DOI: <https://doi.org/10.1007/978-3-030-01204-5>.

2. Panicheva, P.V. Distributional semantic features in Russian verbal metaphor identification / P.V. Panicheva, Y.G. Badryzlova // Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference «Dialogue» (2017). – Issue 16. – Vol. 1.: Computational Linguistics: Practical Applications. – Moscow: RSUH, 2017. – P. 179-190 (0.7 п.л., доля авт. 0.5 п.л.). – ISSN (online) 2075-7182.

3. Panicheva, P. Semantic Feature Aggregation for Gender Identification in Russian Facebook / P.Panicheva, A.Mirzagitova, Y.Ledovaya // Artificial Intelligence and Natural Language: AINL 2017 / ed. by A. Filchenkov, L. Pivovarova, J. Žižka. – CCIS, vol. 789. – Springer, Cham, 2017. – P. 3-15 (0.81 п.л., доля авт. 0.5 п.л.). – ISBN (online) 978-3-319-71746-3, DOI: https://doi.org/10.1007/978-3-319-71746-3_1.

4. Panicheva, P.V. Estimating Syntagmatic Association Strength Using Distributional Word Representations / G.T. Bukia, E.V. Protopopova, P.V. Panicheva, O.A. Mitrofanova // Computational Linguistics and Intellectual Technologies: Proceedings of the Annual International Conference «Dialogue» (2016). – Issue 15. – Moscow: RSUH, 2016. – P. 112-122. (0.72 п.л., доля авт. 0.15 п.л.) – ISBN (online) 2075-7182.

5. Panicheva, P. Developing a Toolkit for Distributional Analysis of Abnormal Collocations in Russian / P.Panicheva, O.Mitrofanova // Proceedings of the 13th Conference on Natural Language Processing (KONVENS). – Bochum: Bochumer Linguistische Arbeitsberichte, 2016. – P. 203-208 (0.59 п.л., доля авт. 0.39 п.л.). – ISSN: 2190-0949.

6. Panicheva, P. Distributional analysis of Russian lexical errors / P. Panicheva // ExLing 2016: Proceedings of 7th Tutorial and Research Workshop on Experimental Linguistics, 27 June – 2 July 2016, Saint Petersburg, Russia / ed. by A. Botinis. – Saint Petersburg: International Speech Communication Association, 2016. – P. 123-127 (0.22 п.л.). – ISBN (Ebook) 978-960-466-161-9, ISSN (Ebook) 2529-1092.

7. Panicheva, P. Evaluating Distributional Semantic Models with Russian Noun-Adjective Compositions / P. Panicheva, E. Protopopova, G. Bukia, O. Mitrofanova // Analysis of Images, Social Networks and Texts: 5th International Conference, AIST 2016: Yekaterinburg, Russia, April 7-9, 2016: Revised Selected Papers / ed. by D.I. Ignatov, M.Yu. Khachay, V.G. Labunets, N. Loukachevitch, S.I. Nikolenko, A. Panchenko, A.V. Savchenko, K. Vorontsov. – CCIS, vol. 661. – Springer, Cham, 2017. – P. 236-247 (0.71 п.л., доля авт. 0.5 п.л.) – ISBN (online) 978-3-319-52920-2. – DOI: https://doi.org/10.1007/978-3-319-52920-2_22.

Ценность научных работ соискателя

В опубликованных работах представлены основные результаты диссертационного исследования: сформулированы теоретические основания анализа семантической связности, разработан исследовательский инструментарий для анализа семантической связности, опирающийся на дистрибутивно-семантические модели, осуществлено практическое применение данного инструментария к задачам оценки допустимости синтагматической связи, автоматического исправления лексических ошибок, автоматической идентификации метафоры.

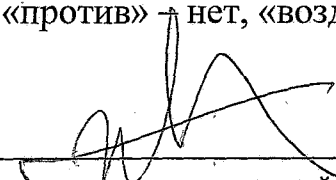
Соответствие содержания диссертации научной специальности

Представленная Полиной Вадимовной Паничевой диссертация «Анализ параметров семантической связности с помощью дистрибутивных семантических моделей (на материале русского языка)» соответствует специальности 10.02.21 – Прикладная и математическая лингвистика.

Диссертация «Анализ параметров семантической связности с помощью дистрибутивных семантических моделей (на материале русского языка)», выполненная Паничевой Полиной Вадимовной, рекомендуется к защите на соискание ученой степени кандидата филологических наук по специальности 10.02.21 – Прикладная и математическая лингвистика.

Заключение принято на заседании экспертной группы, сформированной решением декана факультета от 31.08.2018 № 89-106 в составе четырех человек.

Присутствовало на заседании 4 человека. Результаты голосования: «за» – 4 чел., «против» – нет, «воздержались» – нет, протокол № 2 от 03.11.2018 г.


(подпись председателя экспертной группы)

(Мартыненко Григорий Яковлевич,
д.филол.н., профессор,
профессор кафедры математической
лингвистики)