

Гуманитарная терминология как сеть:
теория графов
о закономерностях научного стиля *

Б. В. Орехов

ИНСТИТУТ ЯЗЫКОЗНАНИЯ РАН
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ –
ВЫСШАЯ ШКОЛА ЭКОНОМИКИ, МОСКВА

Аннотация. Сети (графы) активно используются современной наукой для моделирования и анализа самых разнообразных явлений, в том числе активно применяются и в области изучения естественного языка. Сети позволяют отразить в модели то, что обычно теряется при традиционном количественном подходе к предмету: взаимосвязи между элементами. Построение сети на определенном образом обработанном тексте научного стиля позволяет лучше представить себе терминологическую систему автора, а также применить некоторые количественные метрики, явно показывающие наиболее важные фрагменты общего целого. Исследование проводится на материале научных текстов О. М. Фрейденберг, А. Ф. Лосева и Е. М. Мелетинского. Результаты показывают различия в их подходах к выстраиванию терминологической системы и разницу в стиле изложения.

* Исследование выполнено за счет гранта РФФИ № 14-28-00130 «Лингвистические технологии во взаимодействии гуманитарных наук» в Институте языкознания РАН.

Орехов Б. В. Гуманитарная терминология как сеть: теория графов о закономерностях научного стиля // Критика и семиотика. 2016. № 2. С. 94–101.

ISSN 2307-1737. Критика и семиотика. 2016. № 2
© Б. В. Орехов, 2016

Ключевые слова: теория графов, терминология, научный стиль, центральность графа.

УДК 81

Контактная информация: Орехов Борис Валерьевич, кандидат филологических наук, доцент Национального исследовательского университета – Высшая школа экономики (ул. Мясницкая, 20, Москва, 101000, Россия; nevmenandr@gmail.com)

Графы, или сети (в нашей работе эти термины рассматриваются как синонимы), все активнее применяются в последнее время для научного анализа языкового материала. В отличие от уже ставших традиционными количественных подходов исследование с помощью сетей позволяет не просто посчитать частотность того или иного факта в выборке, но и учесть его связи с другими элементами системы. Графы – это математические объекты, которые состоят из вершин (или узлов) и связей между ними, именуемых ребрами. Такой объект легко визуализировать, т. е. представить в виде изображения точек и связывающих их линий. Для неспециалистов зачастую именно такой картинкой и ограничивается эвристический потенциал графа. Характерный пример – статья Франко Моретти «Теория сетей, анализ сюжета», вошедшая в его книгу «Дальнее чтение» [2016, с. 288–323]. Показав на примере шекспировского «Гамлета», какие сети персонажей может построить исследователь, литературовед останавливается на демонстрации изображений и не использует мощный математический аппарат теории сетей. Если же говорить про подход к подобному материалу с позиций специалистов, то их прежде всего будут интересовать именно численные метрики графа (см.: [Beveridge, Shan, 2016]).

В этой статье мы попытаемся с помощью графа описать основные закономерности терминологической системы некоторых важных для отечественной традиции ученых-гуманитариев. Мы уже предпринимали попытку такого рода, сосредоточившись на проблематике мифа [Орехов, 2015]. Сейчас перед нами стоит более объемная задача общей характеристики терминологической системы.

Но прежде чем, собственно, строить сеть, необходимо определиться с тем, что и как мы будем моделировать с ее помощью. Если для установления связей-ребер между персонажами-узлами в художественном тексте можно выработать конвенциональные процедуры (например, совместное появление на сцене в пьесе), то в случае с терминологией, основания для построения модели менее очевидны.

Мы подошли к вопросу с идеологических позиций компьютерной лингвистики, которая в общем случае считает значимой совместную встречаемость элементов в тексте. Так, важным основанием для многих алгоритмов машинной обработки естественного языка стал принцип фор-

мирования выборки bag of words («мешок слов»), который можно сформулировать следующим образом: выборка представляет собой слова текста, сваленные в мешок, т. е. при этом не учитывается порядок их следования, рассматривается только факт нахождения одного с другим в мешке. Этот принцип лег в основу и нашей работы. В качестве исходных данных использовались научные тексты О. М. Фрейденберг объемом 291 057 словоформ (до подсчетов из текста были удалены служебные слова и местоимения), А. Ф. Лосева (641 444 словоформы) и Е. М. Мелетинского (202 630 словоформ), авторов, принадлежавших к разным поколениям, но так или иначе занимавшихся важной для всей гуманитарной науки проблематикой мифа. Каждое слово из представленных текстов считалось термином, что кажется оправданным для гуманитарной науки. При этом ребро между узлами графа мы рисовали только в том случае, если два слова оказывались в текстах рядом (т. е. на расстоянии не более чем 5 словоформ одно от другого) не менее чем $N/5\,000$ раз, где N – это общий объем корпуса данного автора в словоформах (т. е. 58, 128 и 40 соответственно для Фрейденберг, Лосева и Мелетинского). Такой подход оправдал себя в [Орехов, 2015], показав неплохие, т. е. интуитивно приемлемые, результаты.

Рассмотрим сначала сеть О. М. Фрейденберг (рис. 1). Если посчитать в этой сети такую важную характеристику графа, как степенная центральность, то мы получим список наиболее важных узлов графа, т. е. таких, у которых больше всего «соседей». Эта метрика позволяет узнать, какие слова и в каком порядке являются наиболее важными элементами построенной системы:

образ	0.154
смерть	0.089
форма	0.071
новый	0.059
трагедия	0.053
бог	0.053
сюжет	0.047
хор	0.041
божество	0.041
греческий	0.041
понятие	0.035
царь	0.035
язык	0.035
герой	0.035
драма	0.029

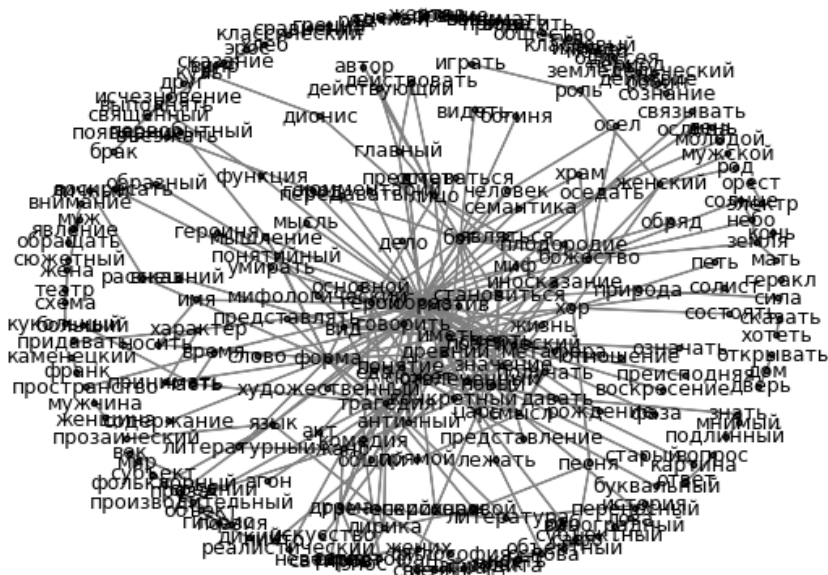


Рис. 1. Сеть понятий О. М. Фрейденберг

Верхушка списка довольно подробно и хорошо описывает опорные для Фрейденберг понятия, выводя в лидеры термин «образ», вошедший и в заглавие одной из важнейших для ученого работ: «Образ и понятие». Примечательно, что среди вполне привычных слов, так или иначе соотносимых с основным для Фрейденберг материалом античной культуры, появляется экзистенциальное слово «смерть», которое показало себя важным элементом терминологической системы Фрейденберг еще в цитированном выше прошлом исследовании.

Сеть понятий А. Ф. Лосева (рис. 2) выглядит гуще и плотнее, кроме того, насыщеннее смотрится ее центральная часть. Сопротивление сети Фрейденберг централизации понятий было отмечено нами ранее на более узком материале – часть графа, связанная с понятием «миф». Здесь идея о децентрализации терминологического графа Фрейденберг подтверждается низким показателем степенной центральности для наиболее важных узлов графа.

Посчитаем степенную центральность для терминологической сети А. Ф. Лосева:

вещь	0.262
число	0.247
смысл	0.214

иметь	0.132
бытие	0.123
время	0.120
понятие	0.111
становление	0.099
категория	0.084
точка	0.072
сущность	0.072
говорить	0.072
эйдос	0.069

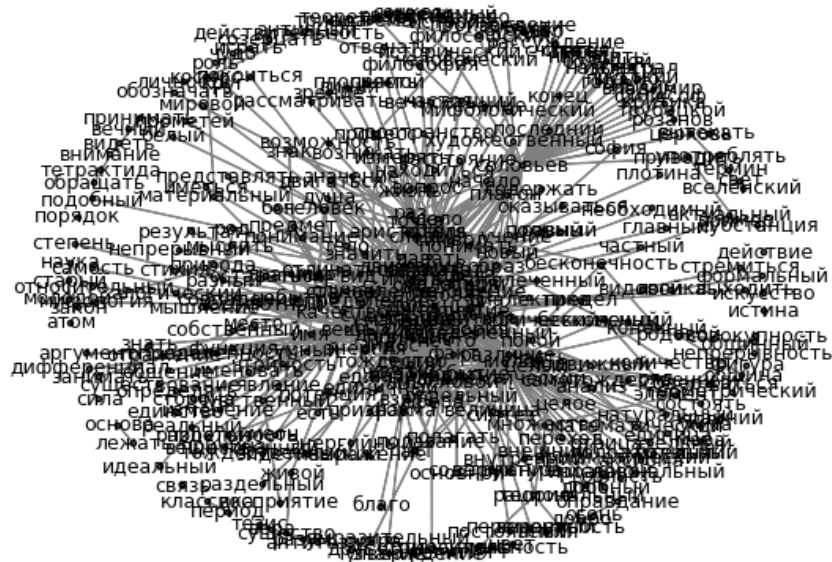


Рис. 2. Сеть понятий А. Ф. Лосева

Довольно логично, что в философских трудах Лосева наиболее важными оказываются именно философские термины: «число» и «вещь». Благодаря сети мы можем посмотреть и на их ближайших соседей, т. е. слова, которые связаны со словом «число» непосредственно (их полный список – 81 единица): *полагание, число, сущность, вопрос, время, диалектический, переход, должный, становиться, внешний* и т. д. Отметим, что числовые показатели степенной центральности у Лосева для наиболее важных терминов выше, чем для наиболее важных узлов у Фрейденберг.

Визуально граф терминов Мелетинского (рис. 3), скорее, похож на аналогичный граф Лосева, чем на граф Фрейденберг. По всей видимости, такая структура терминологического графа вообще естественна для гуманитарной науки.

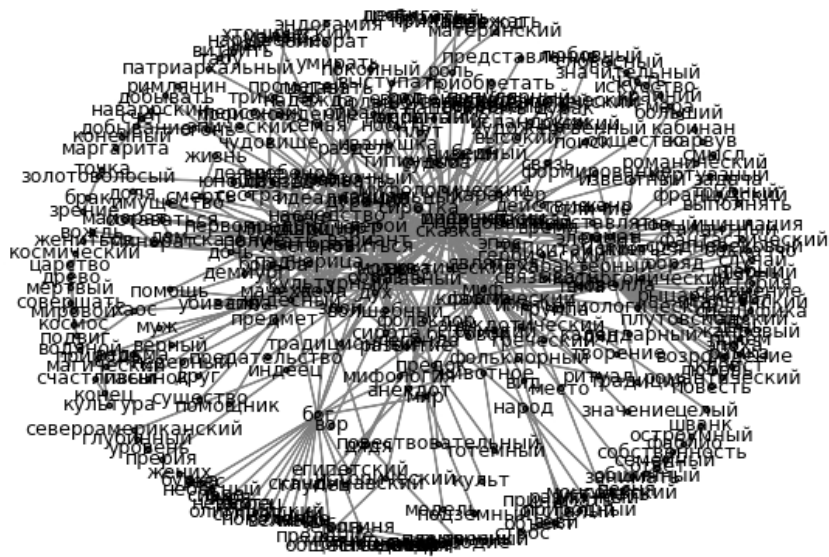


Рис. 3. Сеть понятий Е. М. Мелетинского

Рассмотрим на этом материале уже опробованную метрику степенной центральности:

сказка	0.352
герой	0.207
новелла	0.122
бог	0.079
миф	0.072
жена	0.043
роман	0.034
известный	0.0308
сила	0.0303
сын	0.028

Любопытны здесь пересечения со списком слов-узлов графа Фрейденберг (бог, герой), хотя эти термины находятся на иных позициях с точки зрения их важности в общей системе. Кроме того, следует обратить вни-

мание на сравнительно большие значения метрики: 0.352 и 0.207. По всей видимости, они говорят о большей однородности тематического поля автора, о его тяготении к описанию одних и тех же предметов похожим набором терминов. Такую речевую стратегию вообще естественно связать с традициями академической науки, образцовым представителем которой можно считать Е. М. Мелетинского. В то же время О. М. Фрейденберг всегда отличалась своеобразием стиля (и современниками воспринималась как «непонятная»), а А. Ф. Лосев – многообразием тематики, к которой прилагалась его философская система. В то же время и при визуальном анализе, и при сопоставлении числовых значений степенной центральности графы терминов Лосева и Мелетинского похожи больше, чем граф Фрейденберг сходен с любым из представленных. Очевидно, это означает, что самое сильное влияние на формирование графа, построенного по описанным принципам, оказывает именно авторский стиль.

Любопытна параллель нашим наблюдениям над децентрализацией графа понятий, которая обнаруживается при изучении литературы, посвященной О. М. Фрейденберг. Н. В. Брагинская пишет: «Выделение “центральных мыслей” О. М. Фрейденберг так же делает их бедными и условными, как и обнажение семантики мифа уплощает миф и лишает его глубины» [Брагинская, 1998, с. 741]. Такова природа стиля и мышления О. М. Фрейденберг, по всей видимости, использовавшей различные термины для выражения своей идеи, что, с одной стороны, усложняло стиль и форму изложения, с другой – делало ее более гибкой и способной к выражению более сложных смыслов.

Таким образом, в «обычной» терминологической системе мы находим ряд важных для автора понятий, которые выделяются в графе с помощью степенной центральности. Особняком стоит сеть понятий, построенная на текстах О. М. Фрейденберг, терминология которой избегает централизованности (как сама Фрейденберг избегала традиционных дискурсивных установок академического сообщества) и стремится к полиморфности выражения.

Список литературы

Брагинская Н. В. Послесловие к 1-му изданию // Фрейденберг О. М. Миф и литература древности. М.: Вост. лит., 1998. С. 734–743.

Моретти Ф. Дальнее чтение. М.: Изд-во Института Гайдара, 2016.

Орехов Б. Моделирование терминологического тезауруса работ Р. Г. Назирова о мифологии и истории фольклорных сюжетов // Назировский архив. 2015. № 2. С. 118–131.

Beveridge A., Shan J. Network of Thrones // Math Horizons Magazine. 2016. Vol. 23. No. 4. P. 18–22. URL: <http://dx.doi.org/10.4169/mathhorizons.23.4.18>

Article metadata

Title: Humanitarian terminology as a network: graph theory about the regularities of scientific style

Author: B. V. Orekhov

Author's e-mail: nevmenandr@gmail.com

Author affiliation: The Institute of Linguistics, Russian Academy of Sciences; National Research University – Higher School of Economics

Abstract. Complex networks (graphs) are widely used by modern science to model and analyze a variety of phenomena, including natural language. Networks allow you to reflect in the model that is usually lost in the traditional quantitative approach to the subject: the relations between the elements. Building a network on a processed scientific text allows you to better visualize the terminological system of the author, and to apply the some quantitative metrics, most clearly showing fragments of the whole. The study is carried out on the material of scientific texts by O. M. Freudenberg, A. F. Losev and E. M. Meletinskiy. The results show differences in their approaches to terminological system and the difference in style of presentation.

Key terms: graph theory, terminology, scientific style, the centrality of the graph.

Reference literature (in transliteration):

Braginskaja N. V. Posleslovie k 1-mu izdaniju [Afterword to the first edition] // Freudenberg O. M. Mif i literatura drevnosti [Myth and literature of ancient times]. M.: Vost. lit., 1998. S. 734–743.

Moretti F. Dal'nee chtenie [Further reading]. M.: Izd-vo Instituta Gajdara, 2016.

Orehov B. Modelirovanie terminologicheskogo tezaurusa rabot R. G. Nazirova o mifologii i istorii fol'klornyh sjuzhetov [Modelling of thesaurus of the works by R.G. Nazirov about mythology and history of folk plots] // Nazirovskij arhiv. 2015. № 2. S. 118–131.

Beveridge A., Shan J. Network of Thrones // Math Horizons Magazine. 2016. Vol. 23. No. 4. P. 18–22. URL: <http://dx.doi.org/10.4169/mathhorizons.23.4.18>