

Что может рассказать теория графов
о терминологической системе
О.М. Фрейденберг?

В статье рассматривается способ автоматического извлечения из научных текстов терминов и объединения их в единую сеть, которую можно проанализировать с помощью количественных методов. В качестве центрального объекта анализа выступает корпус текстов О.М. Фрейденберг, который сравнивается с аналогичным – А.Ф. Лосева, Е.М. Мелетинского, А.А. Потебни, А.Н. Веселовского, Я.Э. Голосовкера и Р.Г. Назирова. Сеть понятий Фрейденберг демонстрирует своеобразие и ряд специфических особенностей.

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

Интуитивно ощущаемая читателем (даже не очень внимательным) особость стиля О.М. Фрейденберг хорошо известна¹. Риторика (в плане *elocutio*) ее работ не похожа ни на академическое литературоведение предшественников, ни на советскую научную продукцию младших современников и потомков. В целом понятны и истоки этого своеобразия: специфическая в языковом и интеллектуальном отношении эпоха, на которую приходится формирование ее научного инструментария, соединяется с биографически обусловленным нахождением в творчески насыщенной среде; дело усугубляется последующей научной изоляцией. При этом сама специфика ее особого стиля, пожалуй что, еще не описана, что не так уж удивительно: Фрейденберг пользуется вниманием тех, кто в научных работах ищет план выражения и не смешивает труды коллег по цеху с предметом исследования (в качестве такого предмета выступают философия и литература чаще всего).

Между тем, если следовать идеям самой Фрейденберг, именно план выражения требует пристального и кропотливого изучения,

ведь «вопреки тому, что говорит диалектический материализм, форма не может быть оторвана от содержания и содержание не может предшествовать форме и быть ведущим его началом; вопреки тому, что говорил Марр, морфология не есть семантика»².

Мы решимся предложить не вполне привычный для традиционной филологии цифровой способ описания особенностей стиля О.М. Фрейденберг, основанный на автоматической обработке текстов, извлечении из них терминов и последующем построении из этого материала графа (или – сети). Само обращение к компьютерному инструментарию при работе с текстом уже некоторое время не является ни чем-то новым, ни экзотическим. Мощная и довольно хорошо зарекомендовавшая себя область компьютерной лингвистики помогает создавать коммерчески прибыльные продукты в области информационного поиска, искусственного интеллекта, машинного перевода, а в приложении к культурно значимым объектам та же компьютерная лингвистика помогает находить решения для того, что обозначается рамочным термином *distant reading*³, т. е. комплекс подходов, противопоставленных *close reading* (медленному чтению). В рамках этого комплекса постулируется ограниченность человеческих возможностей в работе с большими массивами текстуального материала, и компьютер привлекается к анализу с тем, чтобы помочь исследователю охватить как можно больший объем исходных данных. При этом ученый вынужден работать с текстом не напрямую (отсюда – *distant reading*), а через посредство того, что из них удалось извлечь средствами машинной обработки.

В нашем случае исследовательский объект, с которым мы будем работать непосредственно, это граф терминов, который, как мы утверждаем, напрямую связан со стилевым уровнем организации научного текста. 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»⁴. Показав на примере шекспировского «Гамлета», какие сети персонажей может построить исследователь, ли-

тературовед останавливается на демонстрации изображений и не использует мощный математический аппарат теории сетей. В то же время численные метрики графа не следует упускать из вида⁵.

Если для установления связей-ребер между персонажами-узлами в художественном тексте можно с легкостью выработать конвенциальные 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 уже упоминавшейся 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 формирования выборки *bag of words* («мешок слов»), который можно сформулировать следующим образом: «выборка представляет из себя слова текста, сваленные в мешок, т. е. при этом не учитывается их порядок следования, рассматривается только факт нахождения одного с другим в мешке». Этот принцип с некоторыми коррективами 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)⁶, А.Ф. Лосева (641 444 словоформы), Е.М. Мелетинского (202 630 словоформ), Я.Э. Голосовкера (150 082 словоформы), А.А. Потебни (118 519 словоформ), А.Н. Веселовского (149 031 словоформа), Р.Г. Назирова (122 556 словоформ) – авторов, принадлежащих 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 именно гуманитарной специфики. 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 пять словоформ одно от другого) не менее чем $N/5000$ 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 нашей предыдущей работе⁷, показав неплохие, т. е. интуитивно приемлемые, результаты. И именно здесь кроются те самые коррективы принципа *bag of words*: с одной стороны, важной оказывается дистанция между словами (если расстояние между словами оказывалось больше, чем пять, то они не считались связанными), что не вполне характерно для компьютерной работы с текстами. С другой стороны, слова, охваченные этой дистанцией, признавались равноправными, т. е. попавшими в «мешок слов» на общих основаниях с другими.

Визуальный анализ, которым часто ограничиваются, позволяет увидеть несколько вершин, к которым сходится большинство линий-ребер соседей. Они находятся в центре рисунка. В терминах теории графов принято говорить, что такие вершины обладают наибольшей *центральностью*, а сами вершины признаются наиболее важными для системы, которую моделирует граф. Извлечем из графа список узлов с наибольшей центральностью (см. табл. 1).

Таблица 1

Термины, обладающие в графе понятий О.М. Фрейденберг наибольшей центральностью

Термин-узел	Коэффициент центральности
образ	0,154
смерть	0,089
форма	0,071
иметь	0,059
новый	0,059
трагедия	0,053
бог	0,053
сюжет	0,047
хор	0,041
божество	0,041

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 ожидаемых слов, так или иначе соотносимых с основным для Фрейденберг материалом античной культуры, появляется экзистенциальное слово «смерть», приобретающее особенное терминологическое значение.

Вспомним слова Н.В. Брагинской: «По сути дела, в работах Фрейденберг нет разворачивания мысли, нет дискурсивности, но нет и описательства. У ее работ есть центр – центральная в прямом смысле этого слова мысль... Для работ О.М. Фрейденберг нужна какая-то другая пространственная организация текста, при которой эта “центральная мысль” так бы и помещалась в центре, подобно

источнику света, а материал, который она “освещает”, располагался “кругом”»⁸. Мы не уверены, что было бы справедливо уравнивать «мысль», о которой говорит исследователь, и термин, являющийся единицей нашей модели, но если пытаться искать и формализовывать «мысль», то, видимо, именно термины будут ближайшими ее выразителями. Если принять такое допущение, то «образ», «смерть» и «форма» будут самыми центральными (в терминологическом смысле) репрезентациями главной «мысли», а освещаемый ею материал – это ближайшие соседи перечисленных узлов. Таких соседей у узла «образ» в нашей сети получилось 52, среди них (выберем существительные): *комментарий, время, жизнь, функция, сюжет, предмет, значение, семантика, мышление, еда, мотив, бог, характер, метафора, форма, город, смерть, смысл, божество, миф, мысль, понятие, природа, вид, язык, иносказание, царь, герой, трагедия*. Этот перечень «ключевых слов» действительно некоторым образом отражает содержание текстов Фрейденберг, но, разумеется, в примитивизированном виде, как это уже было сформулировано: «Выделение “центральных мыслей” О.М. Фрейденберг так же делает их бедными и условными, как и обнажение семантики мифа, уплощает миф и лишает его глубины»⁹.

Можно отметить и полное доминирование в приведенном перечне существительных (глагол «иметь», возможно, следовало бы совсем исключить из выборки, так как часто он выполняет в текстах вполне служебную роль). В оставшемся за пределами таблицы перечне первый полноценный глагол возникает только на 17-м месте, и это нетривиальная и мало предсказуемая в таком списке лексема «оседать». Ср. «Законченной религии еще не было, но уже создавалась особая область, где оседали в законсервированном виде отработанные веками представления»¹⁰. Действительно, слово «оседать» становится для Фрейденберг значимым термином, выражающим процесс формирования культурных наслоений, реконструкцией которых она и занимается. Эта явная физико-химическая метафора, безусловно, индивидуальна, по крайней мере в смысле тех масштабов, в которых она проявлена в текстах. В то же время в связи с глаголами нужно вспомнить и еще об одном наблюдении над стилем анализируемых текстов: «О.М. Фрейденберг признавалась, что глаголы, связывающие временем, она насильно вводит в текст: “Обобщающая мысль не мыслит в конкретном прошлом, уж лучше тогда praesens, но и он сковывает, к земле гнет мысль” (архив)»¹¹. Таким образом, мы, кажется, находим в нашей distant-reading модели подтверждение, пришедшее со стороны close reading и авторефлексии.

В графе понятий А.Ф. Лосева благодаря размерам корпуса оказалось гораздо больше узлов – 1 008. Визуальный анализ этой сети был бы крайне затруднен, так что мы сразу перейдем к выявлению наиболее центральных терминов (см. табл. 2).

Таблица 2

Термины, обладающие в графе понятий А.Ф. Лосева наибольшей центральностью

Термин-узел	Коэффициент центральности
вещь	0,262
число	0,247
смысл	0,214
соловьев	0,141
бытие	0,123
время	0,120
понятие	0,111
становиться	0,099
становление	0,099
категория	0,084

Как можно заключить из табл. 2, в корпусе Лосева тексты философской тематики полностью поглотили те, которые были ориентированы на специфику мифа и материи, близкие Фрейденобергу. «Вещь», «число» и «смысл» стали наиболее важными понятиями в реконструируемой терминологической системе, и их принадлежность к другому семантическому полю, нежели у Фрейденоберга, очевидна. Среди центральных терминов Лосева не прощупывается никакой специфически античной тематики (у Фрейденоберга она была на первом плане), а их укорененность именно в философском дискурсе контрастирует с литературоведческим дискурсом Фрейденоберга («бытие» vs. «сюжет», «категория» vs. «трагедия»).

В графе понятий Е.М. Мелетинского оказалось 584 узла. Среди центральных терминов ориентация на литературоведческий тезаурус еще более заметна, что создает гораздо больше пересечений с графом Фрейденоберга. Лексемы «мотив» (коэффициент центральности – 0,073, далее у всех терминов приводится в скобках), «образ» (0,057), «сюжет» (0,050) и др. являются пересечением в списках наиболее центральных терминов. Таким образом, жанровая специфика оказала влияние на наш материал гораздо большее, чем поколенческая.

Любопытно, что и у Мелетинского, и у Фрейденберг гораздо большей центральностью, чем у Лосева, обладает термин «бог» (0,066 у Мелетинского и 0,053 у Фрейденберг; у Лосева – 0,015). По всей видимости, это обусловлено спецификой древнего материала, которому посвящены основные труды ученых в их корпусах. Иными словами, и Лосев, и Фрейденберг, и Мелетинский мало говорят о христианском боге, а Лосеву при этом не требуется упоминание и других богов.

Несмотря на то что корпус Фрейденберг приблизительно в два раза превышает корпус Я.Э. Голосовкера, сеть понятий последнего насчитывает 1 487 узлов, т. е. почти в четыре раза объемнее сети Фрейденберг. Это может означать большее, чем у Фрейденберг, разнообразие сочетаемости (вернее, совместной встречаемости) терминов. Иными словами, у Фрейденберг большее число слов чаще оказывается в текстах в одном и том же окружении, в то время как у Голосовкера слова чаще образуют новые связи и контексты. Те же показатели для корпусов А.А. Потебни, А.Н. Веселовского и Р.Г. Назирова составляют 1 350, 263 и 232 соответственно. Таким образом, Лосев, Мелетинский и Потебня больше похожи друг на друга по вариативности контекстов словоупотребления, в то время как Фрейденберг, Веселовский и Назиров стараются (сознательно или бессознательно) минимизировать вариативность контекстов употребления своих терминов. Наиболее центральные термины упомянутых авторов приведены в таблице 3.

Таблица 3

Термины, обладающие в графах понятий Я.Э. Голосовкера, А.А. Потебни, А.Н. Веселовского и Р.Г. Назирова наибольшей центральностью.

Полужирным выделены совпадения с наиболее центральными терминами О.М. Фрейденберг

Голосовкер	Потебня	Веселовский	Назиров
мир	слово	песня	миф
образ	мысль	царь	бог
смысл	язык	соломон	сказка
иван	образ	сын	герой
воображение	человек	легенда	культ
человек	иметь	жена	становиться
логика	говорить	народный	сюжет
бог	звук	образ	смерть
миф	басня	новый	русский
черт	известный	поэзия	религия

Таблица 3 представляет нам любопытнейший материал для комментирования терминологических систем. С одной стороны, именно у философски ориентированных Голосовкера и Лосева на первых позициях оказывается слово «смысл». С другой стороны, глагол «становиться», естественный для философа Лосева (вспомним важное для философии XX в. понятие «Dasein»), оказывается среди наиболее центральных у филолога Назирова. У последнего же среди самых центральных узлов оказывается типично Фрейденовское (и, как нам казалось поначалу, сугубо индивидуальное в плоскости терминологии) понятие «смерть». Если исключить момент прямого влияния (у Назирова есть ссылки на Фрейденовский, так что совсем отрицать генетические отношения нельзя), то, возможно, следует сказать, что появление «смерти» в ряду наиболее центральных понятий продиктовано спецификой мифологического материала. В то же время все терминологические графы в основном избегают опираться на глаголы как на центральные узлы, так что наше первоначальное предположение, что это именно специфика стиля Фрейденовский, было не вполне справедливым.

Итак, предложенная модель терминологической системы О.М. Фрейденовский включает в себя 382 единицы терминов, что, как мы могли убедиться при сравнении с другими подобными моделями, скорее, не так много (средний показатель) и не зависит от объемов исходного корпуса. Наиболее важными понятиями выступают «образ», «смерть» и «форма», что в целом подтверждается читательской интроспекцией. Серьезное влияние на такую модель оказывают жанровые особенности авторских текстов: весьма сходные и по жанру, и по материалу труды Е.М. Мелетинского и Р.Г. Назирова приводят к построению во многом повторяющих друг друга графовых терминологических моделей (в перечне 10 наиболее центральных терминов совпадают «сказка», «герой», «бог»). Из-за этого же понятие «образ», главное для Фрейденовский, с меньшим коэффициентом центральности, но все же фигурирует в 4 из 6 «топовых» перечней других моделей. По всей видимости, для лучшего уяснения специфики терминологии (и, соответственно, стиля) Фрейденовский потребуются взять в качестве контрастного материала корпус русского антиковедения.

Примечания

¹ Беглое замечание Н.В. Брагинской: «Стиль автора, его речевая манера, с одной стороны, индивидуальная, а с другой – несущая на себе отпечаток времени и устаревшей сегодня нормы» (*Фрейденовский О.М.* Миф и литература древности.

- М.: Восточная литература, 1998. С. 763). Ср. тенденциозное, но не менее от этого показательное замечание современника: «Это заумь. Может, вы и что-то правильное хотите сказать, но ведь это абсолютно непонятно...» (цит. по: *Фрейдберг О.М.* Поэтика сюжета и жанра. М.: Лабиринт, 1997. С. 431. Примеч. 24).
- ² *Фрейдберг О.М.* Миф и литература древности. С. 27.
- ³ *Moretti F.* Distant Reading. Verso, 2013. По-русски – «дальнее чтение», мы предпочитаем перевод «отвлеченное чтение».
- ⁴ *Moretti Ф.* Дальнее чтение. М.: Изд-во Института Гайдара, 2016. С. 288–323.
- ⁵ См.: *Beveridge A., Shan J.* Network of Thrones // *Math Horizons Magazine.* 2016. Vol. 23. № 4. P. 18–22.
- ⁶ Полный перечень вошедших в выборку текстов занял бы слишком много места. Отметим только, что в нее попали все крупные опубликованные работы О.М. Фрейдберг, а также соответствующая часть сайта – <http://freidenberg.ru/>
- ⁷ *Орехов Б.* Моделирование терминологического тезауруса работ Р.Г. Назирова о мифологии и истории фольклорных сюжетов // Назировский архив. 2015. № 2. С. 118–131.
- ⁸ *Фрейдберг О.М.* Миф и литература древности. С. 741.
- ⁹ Там же.
- ¹⁰ Там же. С. 221.
- ¹¹ Там же. С. 741–742.