

АВТОМАТИЗАЦИЯ ОБРАБОТКИ ТЕКСТА

УДК (088.84):81'322.2

Г. Я. Мартыненко

Золотое сечение формулы изобретения

Обсуждаются результаты измерения соотношения словесных масс в ограничительной и отличительной частях формулы изобретения с выделенной новизной. Установлено, что это соотношение тяготеет к золотому сечению, которое может рассматриваться как толистическая характеристика, безразличная к физическому объему формулы, но чутко реагирующая на гармоническое соотношение объемов ее структурно-значимых частей. В более широком контексте выявленная тенденция связана с числовыми закономерностями, отражающими соотношение между известной и новой информацией, в частности при тема-рематическом членении предложения и текста.

1. ВВЕДЕНИЕ

В ряде письменных жанров в условиях жестко регламентированной коммуникации порождаются тексты с высоким уровнем стандартизации внешней, зримой формы, отражающей распределение семантической информации в линейной развертке текста. К таким регламентированным текстам в жанре поэзии относятся так называемые "твердые формы" [1], в первую очередь разные варианты сонета. Что касается научно-технической литературы, то здесь к твердым формам можно отнести патентное описание и его основную часть — формулу изобретения. Предельно жесткая организация последней интенсивно обсуждается в патентоведении и прикладной лингвистике [2–5].

Несмотря на то, что сонет и формула изобретения не имеют на первый взгляд серьезных оснований для сопоставления, в этих жанрах можно все-таки разглядеть некоторые родственные черты, позволяющие считать формулу изобретения "сонетом техники" [6]:

1. И для сонета, и для формулы изобретения характерны лаконичность, обозримость, лингво-полиграфическая "картинность" текста. Это способствует быстрому и целостному восприятию текста и заключенной в нем идеи.

2. И для формулы изобретения, и для сонета (классического и его вариантов) характерна "жесткая" структурированность формы, проявляющаяся, например, в многозвенности формулы и в разбиении сонета на tercеты и катрены, следующие друг за другом в фиксированном порядке.

3. И сонет, и формула являются средоточием максимально концентрированной информации, порции которой размещаются в линейной цепи текста в соответствии с формированием различных вариантов семантической доминанции путем противопоставления основной и сопутствующих идей, движения от зачина, экспозиции, завязки к кульминации, концовке, развязке.

В последнее время при исследовании сонета и сонетоподобных форм активизируется стремление

к разысканию числовых констант (или почти констант), управляющих динамикой распределения семантической информации в тексте. Наибольшей популярностью среди них пользуется так называемое золотое сечение. Например, в работе [7] обсуждается связь между ритмической структурой русского классического сонета и его семантической организацией, а в работе [8] на материале стихотворений А. С. Пушкина изучается семантическая и эстетическая нагрузка симметричных паттернов, присущих золотому сечению и зеркальной симметрии.

На основании сказанного, нами было выдвинуто предположение, что правило золотого сечения действует не только в пределах твердых форм поэтической речи, но и в иных жестко организованных текстах, например, в формуле изобретения.

2. ПОСТАНОВКА ЗАДАЧИ

Каждый пункт отечественной формулы изобретения строится в виде одного грамматического предложения, имеющего вид дефиниции через ближайший род и видовое отличие. Это предложение состоит из двух частей: ограничительной и отличительной, разделенных словом *отличающийся*. Первая часть содержит признаки, характеризующие прототип (известное знание), а вторая — признаки, составляющие новизну данного технического решения.

На логико-семантическое и количественное соотношение между признаками, включаемыми в ограничительную и отличительную части в патентоведении, обращается весьма серьезное внимание. В частности, важной проблемой считается поиск оптимального соотношения между числом признаков, характеризующим степень известности данного технического решения, и количеством признаков, формирующих его новизну [9; 10]. Степень такой оптимальности, соразмерности, "гармоничности" может быть грубо оценена путем измерения

объема словесных масс, относящихся к упомянутым частям формулы. Нами была выдвинута гипотеза, что эти объемы, выраженные в количестве словоупотреблений, в оптимальных формулах регулируются правилом золотого сечения.

Воспользуемся классической формулировкой этого правила, представляющего собой систему двух уравнений:

$$\begin{cases} a + b = c; \\ \frac{a}{c} = \frac{b}{a}, \end{cases}$$

где c — “целое”, состоящее из a — “большого” и b — “меньшего”.

Золотое сечение может быть истолковано и как статистическое распределение по альтернативному признаку. В такой трактовке система уравнений будет переписана в виде:

$$\begin{cases} a + b = 1; \\ \frac{a}{a+b} = \frac{b}{a}, \end{cases}$$

где a — доля “большого”, b — доля “меньшего”. В итоге получаем:

$$a^2 = b; a = \sqrt{b}.$$

А после решения квадратного уравнения

$$a^2 + a - 1 = 0$$

относительно a получаем:

$$a = 0,618; b = \sqrt{0,618} = 0,382.$$

Таким образом, устанавливается константное доленое участие “меньшего” и “большого” в “целом”.

Для дисперсии и стандартного отклонения получаем следующие выражения:

$$\sigma^2 = a \cdot b = a^3 = 0,236.$$

$$\sigma = \sqrt{ab} = a\sqrt{a} = 0,486.$$

3. СТАТИСТИЧЕСКОЕ НАБЛЮДЕНИЕ И РЕЗУЛЬТАТЫ ОБРАБОТКИ ДАННЫХ

Путем случайного механического отбора из отечественного бюллетеня “Открытия, изобретения, промышленные образцы, товарные знаки” были извлечены 500 формул изобретения. Из выборки были исключены формулы, относящиеся к изобретениям на применение, поскольку они имеют структуру без выделенной новизны, а также переводы зарубежных изобретений, тексты которых находятся под сильным воздействием оригинала, построенного в соответствии с национальным патентным правом и национальными традициями. Для приблизительной оценки статистической устойчивости золотой пропорции выборочная совокупность была разбита на две подвыборки по 250 формул каждая. При этом статистической обработке были подвергнуты только однозначные формулы и первые пункты многозначных формул, поскольку другие части формулы и соотношение их объемов (заголовок,

цель и другие пункты кроме первого) заслуживают самостоятельного исследования.

В каждой формуле нами подсчитывалось число словоупотреблений, относящихся к ограничительной и отличительной частям, а также к формуле в целом. Далее подсчеты осуществлялись двумя способами: недифференцированным и дифференцированным. При первом способе суммировались словесные объемы всех выборочных формул и их частей, а затем на массивах всех словоупотреблений, относящихся к формуле в целом и ее частям, вычислялись пропорции эмпирического золотого сечения. При таком подходе внимание концентрируется исключительно на соотношении объемов отличительной и ограничительной частей, при полном отвлечении от весьма мощного экстенсивного фактора — абсолютного “физического” (лингвополиграфического) объема формулы и ее частей. Второй способ основывается на измерении долей отличительной и ограничительной частей в каждой формуле с последующим построением распределения формул по величине этого параметра.

Остановимся сначала на недифференциальном способе измерения. Нами были получены следующие данные, приведенные в табл. 1.

Таблица 1

Соотношение словесных объемов ограничительной и отличительной частей формулы изобретения

	Огранич. часть m	Отличит. часть k	Вся формула $m + k$	$p = \frac{m}{m+k}$	$q = \frac{k}{m+k}$
Выборка 1	6048	9702	15750	0,384	0,616
Выборка 2	5975	9926	15901	0,376	0,624
Итого	12023	19628	31651	0,380	0,620

Полученные результаты говорят о весьма большой близости полученных соотношений (в особенности итоговой сумме выборок) к золотому сечению. Измерим предельную ошибку итоговой выборки, например, для словесной доли отличительной части (q). Воспользуемся формулой:

$$\Delta = t \cdot \sqrt{\frac{p \cdot q}{N}},$$

где $p \cdot q$ — дисперсия альтернативного признака, N — объем выборки (общее число словоупотреблений в выборочных формулах), t — коэффициент доверия, равный трем при доверительной вероятности, близкой к единице (0,997).

Произведем необходимые вычисления:

$$\Delta = 3 \cdot \sqrt{\frac{0,62 \cdot 0,38}{31651}} = 0,0082.$$

Итак, с вероятностью близкой к единице, можно утверждать, что отношение словесной массы отличительной части к словесной массе формулы в целом имеет доверительный интервал:

$$0,6118 < q < 0,6282,$$

в который попадает теоретическое значение золотого сечения, равное 0,618.

Теперь приступим к дифференцированному способу измерения.

Естественно полагать, что все многочисленные "соучастники" построения текста формулы изобретения (изобретатели, работники патентных служб, эксперты патентного ведомства и др.) руководствуются множеством факторов, управляющих процессом текстообразования. Среди сильных, постоянно действующих факторов можно отметить следующие:

- 1) комплекс нормативных требований к тексту формулы;
- 2) категориальность изобретения, т. е. его отнесенность к способам, устройствам, веществам и др.;
- 3) многозвенность формулы: ее разбиение на иерархически или ассоциативно упорядоченные пункты;
- 4) классификационный фактор: область техники, класс Международной классификации изобретений (МКИ).

К постоянно действующим факторам следует также отнести и неосознанное, инстинктивное тяготение составителей формулы к соразмерному, гармоничному распределению словесных масс между ограничительной и отличительной частями в соответствии с правилом золотого сечения.

Кроме постоянно действующих факторов, на соотношение объемов отличительной и ограничительной частей влияет множество случайных причин, с трудом поддающихся систематическому учету.

В связи с этим значительный интерес представляет исследование закономерностей колеблемости упомянутых соотношений относительно золотого сечения.

Для исследования этих закономерностей в каждой формуле изобретения были вычислены доли словесных масс отличительной и ограничительной частей и построены соответствующие статистические распределения. В табл. 2 показано одно из таких распределений. Соответствующая гистограмма приведена на рисунке (график построен на основе суммы для выборок).

Таблица 2

Распределение долей словесных масс отличительной части формулы изобретения

Доля отличительной части	Выборка 1	Выборка 2	Выборка 1 + Выборка 2
0,2-0,3	1	2	3
0,3-0,4	9	12	21
0,4-0,5	40	29	69
0,5-0,6	59	59	118
0,6-0,7	71	78	149
0,7-0,8	53	54	107
0,8-0,9	17	16	33
Итого	250	250	500

Распределение, показанное на рисунке, имеет вид куполообразной кривой с умеренной левосторонней асимметрией.

Среднее значение данного распределения равно 0,6184, практически совпадая с золотым сечением, и незначительно отличается от величины 0,6200, полученной первым способом. Также близки к золотому сечению значения, вычисленные вторым способом для обеих подвыборок: 0,6168 и 0,6200 соответственно. В целом в этом собрании средних величин царит удивительное постоянство, диктуемое правилом золотого сечения.

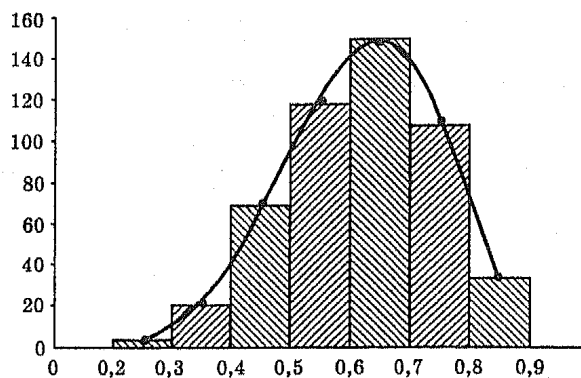


График распределения долей словесных масс отличительной части формулы изобретения

Тем не менее, обратим внимание на одну интересную особенность.

Все меры центральной тенденции данного распределения (среднее значение, мода и медиана) попадают в один и тот же модальный интервал [0,6-0,7]. Известно, что в распределениях с типом асимметрии, подобным тому, который присущ нашему материалу, меры центральной тенденции располагаются в таком порядке:

$$\text{мода} > \text{медиана} > \text{среднее арифметическое} [10].$$

Такая упорядоченность присуща и нашему распределению; при этом медиана, равная 0,621, превышает золотое сечение на очень малую величину. Из этого следует, что вопрос о том, какая из двух мер центральной тенденции (среднее арифметическое или медиана) обнаруживает большее согласие с золотым сечением, не может считаться безоговорочно решенным в ситуации, когда речь идет о разнице в несколько тысячных. Все же отметим, что "наши симпатии" находятся на стороне медианы, поскольку она характеризует симметричные свойства формулы изобретения в двух измерениях: в частотной шкале (зеркальная симметрия) и в шкале, отображающей системно-структурные свойства ("золотая" симметрия).

Что касается показателей вариации данного распределения, то они существенно меньше тех, которые были получены первым (недифференцированным) способом; для итоговой выборки (см. табл. 2) дисперсия равна 0,0164, стандартное отклонение — 0,128, а коэффициент вариации — 0,207. Это объясняется тем, что вычисленная вторым способом дисперсия является так называемой межгрупповой дисперсией, т. е. дисперсией выборочных долей относительно средней доли, близкой к золотому сечению. Общая дисперсия в данном случае подчиняется правилу сложения дисперсий (правилу прямоугольного треугольника) и вычисляется по формуле:

$$\sigma^2 = \delta^2 + \overline{\sigma^2},$$

где σ^2 — общая дисперсия, δ^2 — межгрупповая дисперсия, $\overline{\sigma^2}$ — средняя групповых дисперсий.

Что касается нашего эмпирического материала, то здесь:

$$\delta^2 = 0,0164.$$

$$\sigma^2 = 0,2199.$$

$$\sigma^2 = 0,2199 + 0,0164 = 0,2363.$$

На основании полученных данных можно измерить степень влияния группировочного признака (разброс долей словесных масс отличительной части около золотого сечения) на общую вариацию. Для этого может быть использовано эмпирическое корреляционное отношение, вычисляемое по формуле:

$$\eta = \sqrt{\frac{\delta^2}{\sigma^2}}.$$

Для наших данных η равно 0,267. Эта величина говорит о том, что общая вариация лишь на 26,8% зависит от факторного признака. Такое заключение подтверждается также внешним видом графика на рисунке, где доли довольно кучно концентрируются в районе золотого сечения.

О высокой концентрации долей около золотого сечения свидетельствуют также результаты измерения предельной ошибки выборки, вычисляемой по формуле:

$$\Delta = t \cdot \sqrt{\frac{\delta^2}{n}},$$

где δ^2 — межгрупповая дисперсия, n — объем выборки (число выборочных формул изобретения), t — коэффициент доверия, равный трем при доверительной вероятности 0,997. Для нашего материала Δ равна 0,0161. Следовательно, доверительный интервал средней доли отличительной части составляет:

$$0,6023 < \text{средняя доля} = 0,6184 < 0,6245,$$

в который попадает теоретическое значение золотого сечения.

В заключение обратим внимание на один весьма принципиальный момент.

Формула изобретения является целостным образованием как на технико-правовом, так и на языковых уровнях. Целостность формулы в существенной мере поддерживается ее краткостью, компактностью, «портативностью». Это свойство регулируется рядом нормативных требований:

- 1) необходимостью использования для характеристики идеи изобретения минимального количества существенных признаков (требование общности);
- 2) необходимостью изложения каждого пункта формулы в виде одного грамматического предложения (требование «лаконичности»);
- 3) необходимостью структуризации идеи изобретения в виде основной идеи (первый пункт) и подчиненных ей сопутствующих идей, т. е. второй и последующие пункты (требование многозначности).

Несмотря на эти требования, объем формулы варьирует в довольно широких границах. Золотое сечение проявляет при этом практически полное безразличие к колебанию объема формулы, фиксируя устойчивое соотношение между словесными массами ограничительной и отличительной частей. Эту закономерность зафиксировали наши вычисления: коэффициент корреляции между «эмпирическим золотым сечением» (т. е. долей словесной массы отличительной и ограничительной частей) и словесным объемом формулы в целом близок к нулю (0,0138). Это означает, что золотое сечение может рассматриваться как холистическая

(или синергетическая) характеристика, безразличная к физическому объему формулы изобретения, но чутко реагирующая на гармоническое соотношение объемов ее структурно значимых частей.

4. ВЫВОДЫ

Итак, выдвинутая нами гипотеза о близости к золотому сечению соотношения словесных масс ограничительной и отличительной частей формулы изобретения не противоречит опытным данным. Для окончательного подтверждения этой гипотезы необходимо произвести несколько дополнительных выборок большого объема с учетом фактора многозначности формулы, категориального разнообразия технических решений, фактора комплексности изобретений и их классификационной принадлежности.

Конечно, в каждой отдельной формуле соотношение известного и нового знания может существенно отклоняться от золотого сечения, однако оно довольно откровенно «прорисовывается» уже в выборках умеренного объема, материализуя некоторую скрытую тенденцию, управляющую равновесным «перетеканием» словесных масс из ограничительной части формулы в отличительную, и обратно.

Представляется, что в более широком контексте выявленная тенденция возможно связана с числовыми закономерностями, отражающими соотношение между известной и новой информацией, в частности при тема-рематическом членении предложения и текста.

СПИСОК ЛИТЕРАТУРЫ

1. Гаспаров М. Л. Избранные статьи. — М.: Наука, 1995.
2. Киселев О. М. Лингвистические особенности формулы изобретения // Вопр. изобретательства. — 1979, № 10. — С. 26–28.
3. Мартыненко Г. Я. Некоторые лингвистические аспекты двойственной природы патентной документации // НТИ. Сер. 2. — 1979. — № 10. — С. 10–16.
4. Могилевский В. Н. Лингвистические особенности формулы изобретения на устройства // Вопр. изобретательства. — 1978. — № 12. — С. 26–28.
5. Чебанов С. В., Мартыненко Г. Я. Семантика описательных текстов. Типологический аспект. — СПб: Изд-во С.-Петербургского ун-та, 1999.
6. Карпунин Г. Сонеты техники // Литературная газета. — 1968. — № 4. — С. 10.
7. Гринбаум О. Н. Гармония строфического ритма в эстетико-формальном измерении. — СПб: Изд-во С.-Петербургского ун-та, 2000.
8. Абрамов М. А., Волошинов А. В. Статистический анализ симметричных отношений в стихотворениях А. С. Пушкина. — <http://www.biophys.msu.ru/awse/confer/NLW99/010.htm>.
9. Бокастов В. Н. Признаки объектов изобретений // Вопр. изобретательства. — 1968. — № 9. — С. 14–17.
10. Новожилов А. Г. Составление формулы изобретения в странах с германской системой патентования // Вопр. изобретательства. — 1965. — № 11. — С. 13–17.
11. Пасхавер И. С. Средние величины в статистике. — М.: Статистика, 1979.

Материал поступил в редакцию 13.03.02