

УДК 81'33

UDC 81'33

Сидняев Николай Иванович, Бутенко Юлия Ивановна,**Гаража Владислав Владленович****Московский государственный технический университет им. Н. Э. Баумана****г. Москва, Российская Федерация****Nikolai I. Sidnyaev, Juliia I. Butenko, Vladislav V. Garazha****Bauman Moscow State Technical University (BMSTU)****Moscow, Russian Federation****iuliabutenko2015@yandex.ru, sidn_ni@mail.ru, vgaradzha@gmail.com**

**СТАТИСТИЧЕСКАЯ ОЦЕНКА АССОЦИАТИВНОЙ СИЛЫ
НЕОСМЫСЛЕННЫХ БУКВОСОЧЕТАНИЙ
STATISTICAL ASSESSMENT OF MEANINGLESS LETTER STRINGS
ASSOCIATIVE POWER**

Аннотация

В статье предложена методика составления статистики самых часто встречающихся триграмм в текстах различной длины, сравнение нескольких небольших отрывков с общей статистикой. На основе полученных данных предложена минимальная адекватная выборка. Предложена методика проверки выдвигаемых гипотез по проверке законов распределения с использованием различных критериев. Представлена статистическая обработка результатов количественного анализа триграмм. Проведён расчёт метрологических параметров оценки неизвестных параметров распределения триграмм. В количественном анализе проведено не бесконечно большое число определений, а несколько независимых определений, то есть имеющие выборку (выборочную совокупность) объёмом 5–6 вариантов. Описаны условия выбора лингвистических моделей, а так же виды лингво-математических моделей: идеальные и воспроизводящие. Приведены методологические функции прикладной лингвистики. Рассмотрены специальные разделы математики, используемые в лингвистической теории и практике. Проведена статистическая проверка на предмет, может ли выборка быть извлечена из логнормальной генеральной совокупности – сложная не-параметрическая гипотеза. Проверка проведена с помощью критерия Колмогорова.

Abstract

The article proposes a method of compiling statistics of the most common trigrams in texts of different lengths, comparing several small passages with general statistics, and on the basis of the obtained data, a minimum adequate sample is proposed. The method for verification of hypotheses is proposed to test the distribution laws by using different criteria. The statistical processing of the results of the quantitative analysis of trigrams is presented. Calculation of metrological parameters for estimation of unknown parameters of the trigram distribution is performed. In the quantitative analysis, not an infinitely large number of definitions but several independent definitions is made, that is, having a sample (total sample) of 5–6 options. The conditions for the choice of linguistic models, as well as the following types of linguistic-mathematical models are described: ideal and reproducing. The methodological functions of applied linguistics are reviewed. The special sections of mathematics used in linguistic theory and practice are reviewed. The possibility of extracting the sample from the log-normal general population is statistically tested as a complex non-parametric hypothesis. The test was carried out using Kolmogorov's criterion.

Ключевые слова: инженерная лингвистика, математика, лингво-математическая модель, языкознание, теория вероятностей, статистика.

Keywords: engineering linguistics, mathematics, linguistic-mathematical model, linguistics, probability theory, statistics.

1. Введение

Под ассоциативной силой (далее – АС) неосмысленного буквосочетания принято понимать степень трудности, с которой человек преобразует данный стимул в привычные для него единицы языкового кода – слова и словосочетания. Например, носители русского языка обычно легко связывают неосмысленное буквосочетание *ЗВО* со словом *звон*, а сочетание *ТБЛ* – со словами *таблица* или *Тбилиси*, однако преобразование сочетаний типа *ГУВ* или *ОУА* в слова русского языка неизменно вызывает у них затруднения. В этом смысле мы будем говорить, что буквосочетания *ЗВО* и *ТБЛ* обладают большей ассоциативной силой, чем сочетания *ГУВ* и *ОУА*. Дальнейший интерес к изучению ассоциативной силы вербальных стимулов был связан с тем, что, как было обнаружено, эта характеристика прогнозирует ряд важных аспектов речевого поведения человека [Лесохин, 1982, с. 22], в том числе результаты запоминания в опытах по парно-ассоциативному обучению, пороги зрительного распознавания неосмысленных буквосочетаний и т. п. [Иванов, 2004, с. 25].

Так, например, если составить из букв порядка 2300 легко произносимых неосмысленных буквосочетаний и из них выбрать по запоминанию наиболее подходящие по требованию эксперимента, то, оказывается, испытуемые запоминают некоторые стимулы быстрее чем другие. В связи с этим можно сделать вывод о возможности существования ещё одной характеристики неосмысленных буквосочетаний, способной влиять на их запоминаемость. Эту характеристику, как правило, называют значением неосмысленного буквосочетания. Здесь предполагается, что буквосочетание обладает тем большим значением, чем оно легче ассоциируется с привычными словами. Именно это свойство в дальнейшем было названо ассоциативной силой неосмысленных буквосочетаний.

В процессе эксперимента обнаружено, что ассоциативная сила вербальных стимулов (в данном случае – неосмысленных буквосочетаний) является важной характеристикой вербального материала, необходимой для исследования не только процессов запоминания, но и некоторых других существенных аспектов речевого поведения [Люгер, 2003, с. 48]. Именно этим объясняется стремление исследователей получить надёжные количественные оценки данного параметра [Пиотровский, 2008, с. 51]. Со времени первых попыток измерения АС неосмысленных слогов появились работы, в которых исследовались разные аспекты этого вопроса, предлагались более совершенные методы проведения ассоциативного эксперимента, публиковались таблицы неосмысленных буквосочетаний с соответствующими количественными оценками АС. Если составить всевозможные трехбуквенные комбинации типа «согласный – гласный – согласный» (СГС) так, что первая и последняя согласные должны были быть разными ($C_i C_j$; $i \neq j$), а затем исключить все комбинации, являющиеся словами, то можно получить окончательный список из 2019 неосмысленных слогов-триграмм типа СГС. Полученные триграммы удобно представить в виде последовательных серий в тахистоскопе, по 252 триграммы в каждой. Максимальное время экспозиции отдельного стимула составляло 3 сек., но, если испытуемые давали ассоциацию за более короткое время, экспериментатор сразу показывал следующий стимул [Волошин, 2004, с. 115 ; Селиванова, 2010, с. 86].

2. Анализ экспериментальных данных ассоциативной силы неосмысленных буквосочетаний

Рассмотрим следующий эксперимент. Триграммы были предъявлены 34 испытуемым в виде буклетов по 100 страниц, где на каждой странице была записана только одна триграмма. Порядок стимулов в каждом буклете был различным. Максимальное

время работы с одним стимулом равнялось 30 сек. Испытуемые сообщили, что эксперимент проводится с целью выяснения того, какие ассоциации возникают в ответ на предъявление серии триграмм. Прочитав стимул, испытуемые должны были записывать вокруг него на данной странице все слова, которые приходят им в голову. Инструкция побуждала испытуемых написать как можно больше разных слов за 30-секундный период, в конце которого испытуемые получали сигнал перейти к следующему стимулу. При обработке результатов были подсчитаны следующие показатели:

1. Среднее число ассоциаций f , порождённых за фиксированный промежуток времени (в данном случае – за 30 сек). Это число подсчитывалось по формуле [Кандрашина, 1989, с. 22 ; Сидняев, 2014, с. 34]:

$$f = \frac{1}{N} \sum_{i=1}^N R_{S_i},$$

где S_i – предъявленный стимул; R_{S_i} – число ответов, данных i -м, испытуемым на стимул S_i ; N – число испытуемых.

2. Оценка свойства данного стимула порождать одни и те же ассоциации у разных испытуемых, которая определялась следующим образом. Для каждой триграммы был сделан частотный словарь порождённых слов-ассоциаций. Была подсчитана доля испытуемых, давших самую частую ассоциацию. Трансформация этой величины представляет собой оценку вероятности p . В работах приводится соответствующая таблица значений f и p [Ильин, 1985, с. 36 ; Кутюра, 2010, с. 15].

При разработке модели возникает ряд вариантов решения одной и той же задачи. Если разрабатываемая система отличается большой сложностью, проверить, к чему приведёт выбор того или иного варианта, очень трудно, а иногда и практически невозможно до тех пор, пока система не будет полностью реализована. В этом случае «дешевле» обратиться к исследованию живого прототипа и попытаться выяснить, какой из вариантов лучше согласуется с полученными экспериментальными фактами [Физиология речи..., 1976, с. 4].

Обсуждая полученные результаты, можно выдвинуть следующие две гипотезы.

1. Триграммы, которые порождают больше ассоциаций, будут легче связываться между собой в парно-ассоциативном эксперименте, чем триграммы, которые порождают меньше ассоциаций.

2. Испытуемые, которые дают много ассоциаций на неосмысленные буквосочетания, будут быстрее заучивать новые неосмысленные буквосочетания, чем те испытуемые, которые дают мало ассоциаций.

Для получения более надёжных оценок АС для всех возможных комбинаций типа СГС из букв латинского алфавита (включая и те, которые являются осмысленными английскими словами), проведён эксперимент следующим образом.

2480 триграмм типа СГС разбивались на 9 групп и затем записывались на диктофон, которые предъявлялись так, чтобы каждые две секунды появлялся новый стимул. Результатом является перестроенная таблица, в которой триграммы располагались в том порядке, в котором они объединялись в кластеры. Алгоритм кластеризации включает вычисление и построение исходной матрицы 9×9 [Миринова, 2015, с. 155–157].

Опыт проводился в три сеанса. После проведения второго сеанса испытуемым предлагался тест для проверки словарного запаса: требовалось записать все слова, начинающиеся с буквы S (за исключением имён собственных), которые придут испытуемым в голову в течение 5 минут.

Для окончательного анализа результатов были рассмотрены ответы 216 испытуемых. Коэффициент корреляции между числом слов, порождённых испытуемыми в

тесте по проверке словарного запаса, и числом ответов «да» в ассоциативном эксперименте, свидетельствует о наличии корреляции при уровне значимости $\alpha = 0,05$ (95%): ($p = 0,142$; $N = 312$). Однако при подсчёте того же коэффициента в отдельности для мужчин и женщин оказалось, что он не достигает 95% уровня значимости. Таким образом, можно сделать заключение о том, что связь оценок АС и словарного запаса является весьма слабой.

Эксперимент был спланирован и проведён следующим образом. Триграммы набора были произвольно разделены на 28 списков по 75 триграмм, предъявляемых испытуемыми в такой форме, чтобы результаты можно было обработать на ЭВМ.

В эксперименте участвовало 261 испытуемых (158 мужчин и 103 женщины), которые были разбиты на 6 групп. Стимулы предъявлялись в течение трёх сеансов с интервалом в неделю [Фрумкина, 1971, с. 17]. Эксперимент проводится с целью узнать, сколько слов-ассоциаций могли бы служить в качестве ответов на некоторые буквосочетания. Испытуемым давались списки трёхбуквенных сочетаний (триграмм), которые должны были оценить каждую триграмму в зависимости от того, с каким числом слов или предметов она связывается. Всего имелось пять возможных оценок: 1) плохо; 2) неудовлетворительно; 3) удовлетворительно; 4) хорошо; 5) отлично.

Требовалось оценить предъявленные триграммы с помощью шкалы, проставляя у каждой триграммы галочку в соответствующей графе. Поясним это примером. Возьмём триграмму *TEX*. Сначала надо произнести её по буквам (*t-e-x*), потом прочесть её как слог (*tex*). Триграмма *TEX* связывается с большим числом слов и словосочетаний, поэтому *TEX* получит самую высокую оценку, т. е. необходимо поставить триграмме *TEX* галочку в в соответствующей графе «отлично».

На основе анализа ответов 200 испытуемых получено три вида оценок АС триграмм.

Оценка a . Испытуемый должен был помещать в первую категорию шкалы («*none*») только те триграммы, которые не вызывают у него никаких ассоциаций. Если 100 испытуемых из 200 поместили бы данную триграмму в первую категорию, это означало бы, что половина испытуемых не нашла для них ассоциаций. Предлагается здесь оценивать АС триграмм по формуле: $a = (N - f_1) / N$, где N – число испытуемых (в данном случае 200); f_1 – число испытуемых, поместивших данный стимул в категорию 1.

Величина a , подсчитываемая по данной формуле сопоставима, с величиной m . В приведённом примере $a = 0,5$; $m = 50\%$. Таким образом, в общем случае $m = a \cdot 100\%$. Проведённое сопоставление полученных оценок a с оценками тех же триграмм [Белянин, 2005, с. 30], приводит к коэффициенту корреляции $r = 0,88$.

II. Субъективные представления о числе ассоциаций на данный стимул a' – надёжность оценок a .

Среднее число ассоциаций, которые, по мнению испытуемых, могут порождаться стимулом i , подсчитывается на основе распределения оценок стимула i по пятиразрядной шкале с помощью формулы

$$a' = \frac{f_1(1) + f_2(2) + f_3(3) + f_4(4) + f_5(5)}{N},$$

где 1, 2, ..., 5 – значения категорий шкалы; $f_1, f_2, \dots, f_5$ – число испытуемых, поместивших стимул i соответственно в 1, 2, ..., 5 категории шкалы.

Например, пусть 50 испытуемых, положили триграмму *ROV* в категорию 1, а 10, 70, 20 и 50 испытуемых – соответственно во вторую, третью, четвёртую и пятую категории. Тогда для триграммы *ROV* будем иметь:

$$a'_{rov} = \frac{1 \times 50 + 2 \times 10 + 3 \times 70 + 4 \times 20 + 5 \times 50}{50 + 10 + 70 + 20 + 50} = \frac{610}{200} = 3,05$$

Необходимо отметить, что формула имеет ряд преимуществ по сравнению с предыдущей, потому что формула даёт возможность использовать все данные пятиразрядной шкалы, а в случае применения первой формулы часть информации теряется из-за объединения разрядов 2, 3, 4, 5 – в один. Это означает, что оценки a' являются более чувствительными и точными, чем оценки a . Надёжность оценок a определялась с помощью известной методики [Ковшиков, 2007, с. 18; Гридина, 2012, с. 32]. 200 испытуемых были произвольно разделены на две равные группы. Коэффициент корреляции результатов опыта между двумя группами оказался весьма высок, что свидетельствует о достаточной степени надёжности оценок a' .

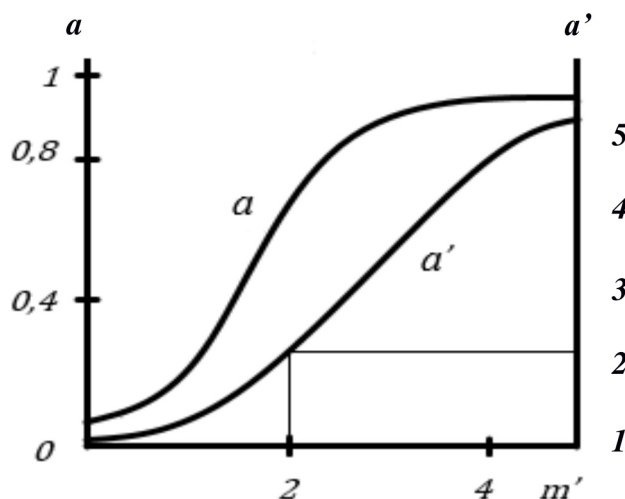
Имея в виду оценки a' , предложено подсчитывать дисперсию оценок АС триграммы i по формуле [Потапова, 1986, с. 37]:

$$\sigma^2 = \frac{f_1(1)^2 + f_2(2)^2 + f_3(3)^2 + f_4(4)^2 + f_5(5)^2}{N} - a'^2,$$

где 1, 2, ..., 5 – категории шкалы; $f_1, f_2, \dots, f_5$ – число испытуемых, поместивших стимул i соответственно в категории 1, 2, ..., 5; a' – оценка АС, полученная по формуле.

Рассмотрим зависимость между σ^2 и a' . Очевидно, что $\sigma^2_i = 0$ для всех случаев, когда все оценки стимула i находятся в одной и той же категории шкалы. Если оценки стимула i распределены равномерно по всем категориям шкалы (равномерное распределение), то $a_i = 3$ и $\sigma_i^2 = 2$.

Максимальное значение $\sigma^2 = 4$ соответствует случаю бимодального распределения, когда половина испытуемых помещает оценки стимула i в категорию 1, другая половина испытуемых – в категорию 5.



Р и с у н о к 1. Зависимость надёжность оценки a' от вероятности

Проведём анализ распределения σ^2 для произвольно выбранной группы стимулов объёмом в 120 триграмм от a' этих триграмм (см. рис. 1). Отметим, что σ^2 изменяется в зависимости от a' по простому параболическому закону, т. е. a возрастает от 0 около категории 1 (левого конца шкалы) до нестабильного максимума около категории 3 (середины шкалы) и снова убывает до 0 около категории 5 (правого конца шкалы). Это соответствует обычной картине разброса оценок, наблюдаемой в

психометрических экспериментах [Потапова, 1986, с. 109]. Оценки АС триграмм на интервальной шкале – m' , a' представляет собой числа на порядковой шкале. Как известно, порядковая шкала накладывает определённые ограничения на операции, производимые с числами. Более удобными для расчётов и анализа являются численные оценки на интервальных шкалах. В данном случае, удобно нормализовать оценки a' , трансформируя их в числа на интервальной шкале. С этой целью произвольно отбирались 21 триграмма и для них были получены АС в единицах интервальной шкалы, с использованием формулы:

$$m' = l + k \frac{\frac{N}{2} - N_1}{N_2},$$

где N – число испытуемых, участвующих в опыте; l – нижняя граница категории, содержащей медиану; N_1 – число испытуемых, поместивших данный стимул ниже l ; N_2 – число испытуемых, поместивших стимул в ту категорию, которая содержит медиану; k – ширина категории, содержащей медиану. Средняя ширина интервалов категорий 2, 3, 4, выраженных в единицах, следующая: категория 2–0,84; 3–0,61; 4–0,63.

Поскольку вычисление m' всех триграмм является очень трудоёмкой работой, то, пользуясь оценками a , a' и m' для 21 случайно отобранной триграммы, построим номограмму для трансформации a и a' в m всех 2100 триграмм типа СГС (см. рис. 1).

Посредством этой номограммы можно подсчитать m' любой триграммы при условии, что a или a' этой триграммы заранее известны. Например, пусть a' триграммы *BOF* будет 2,0 (по таблице значений a'). От точки 2,0 шкалы a' мы восстанавливаем перпендикуляр до пересечения с кривой a' (см. рис. 1), потом от точки пересечения с кривой a' опускаем перпендикуляр к шкале m' . Точка пересечения с $m'=2$ при $a'=1$ представляет значение a триграммы *BOF*.

Проведённый анализ показывает, что оценки a , a' и m' обладают большой устойчивостью и надёжностью. Сопоставление результатов ассоциативного эксперимента, проведённого по методу порождения и оценок a' , приводит к постановке новой задачи.

Инструкция сформулирована так, что каждый испытуемый отвечает на вопрос о том, как много разных слов-ассоциаций порождает у него данная триграмма. Если стимул порождает много разных слов-ассоциаций, то по методу шкалирования он должен быть помещён в один из последних разрядов шкалы (4 или 5). Если стимул порождает мало разных слов-ассоциаций, то он должен быть помещён в первые разряды шкалы – 2 или 3.

Состав ответов с качественной стороны просто не рассматривается. При попытке подойти к ответам испытуемых не с точки зрения наличия – отсутствия ответа на данную триграмму, а с точки зрения количества разных ассоциативных ответов обнаруживается, что одна триграмма обладает свойством порождать у всех испытуемых одно и то же слово-ассоциацию, а другая триграмма у разных испытуемых порождает разные ассоциации. Например, в качестве ответа на триграмму *ЗВО* абсолютное большинство испытуемых дают одно и то же слово – звон, а на триграмму *НАВ* – даётся большее число разных слов. АС обеих триграмм (m'), будет примерно одной и той же, поскольку в протоколах не будет прочерков; практически все испытуемые дадут какой-либо ответ. Очевидно, однако, что оценки АС обеих триграмм, полученные методом шкалирования, должны существенно расходиться. Более того, триграмма *ЗВО* должна получить очень низкую оценку m' (или a') – ведь она порождает у индивидов мало разных слов, а триграмма *НАВ* – очень высокую оценку, поскольку она порождает у индивидов много разных слов. Возникает вопрос о том, в какой мере сопоставимы

оценки типа (1) m' и оценки типа (2) m' – точнее говоря, характеризуют ли эти оценки один и тот же аспект механизма порождения ассоциаций или разные аспекты.

В результате эксперимента для каждой триграммы получено четыре типа оценок АС:

F_1 – частота доминантной ассоциации;

F_2 – число слов-ответов, встречающихся по одному разу;

N – число разных слов-ответов;

H – теоретико-информационная мера неопределённости ответов, подсчитанная по формуле

$$H = \log_2 n - \frac{1}{n} \sum n_i \log_2 n_i ,$$

где n – число всех испытуемых; n_i – число испытуемых, давших в качестве ответа на данный стимул i -ю ассоциацию.

В таблице 1 представлены значения m' (типа 1) и N (типа 2) для каждой из четырёх групп стимулов (значения N в группах осреднены по 10 триграммам).

Т а б л и ц а 1. Результаты наблюдений триграмм

Группа	$m'(тип 1)$	$N, (тип 2)$	Группа	$m'(тип 1)$	$N(тип 2)$
I	3,70	31,10	III	2,10	41,10
II	2,85	31,80	IV	1,35	50,90

Здесь триграммы первой группы порождают много ассоциаций, а триграммы четвёртой группы порождают мало ассоциаций (ср. m' для групп I и 4 в табл. 1). Однако это находится в полном противоречии с результатами, полученными методом порождения. Из таблицы 1 видно, что на триграммы первой, группы в опыте типа 2 порождалось в среднем меньше разных слов ($N = 31,10$), чем на триграммы четвёртой группы ($N = 50,90$). Подсчитывая коэффициент корреляции между m' и N , обнаружена существенная отрицательная корреляция между ними ($r = -0,71$, $p < 0,001$; $v = 38$). Это означает, что чем меньше разных слов-ассоциаций порождает данная триграмма, тем больше будет её m' .

3. Методы получения количественных оценок АС триграмм

Сущность метода порождения заключается в следующем. От испытуемых требуется в ответ на предъявленную триграмму написать одно или несколько слов-ассоциаций, вызываемых в его сознании данной триграммой. Этот метод применяется в двух вариантах.

1. От испытуемых требуется записать только одно слово-ассоциацию, а именно, то, которое первым придёт им в голову. АС триграмм в этом случае вычисляется по формуле

$$m_i = \frac{N_1}{N} ,$$

где N – число испытуемых, участвовавших в опыте; N_1 – число испытуемых, давших ассоциацию на стимул i .

2. От испытуемых требуется записать все слова-ассоциации, которые порождаются у них данным стимулом за определённое время (30 сек, 60 сек и т. п.). При этом варианте метода порождения АС подсчитывается по формуле $a_i = S / N$, где S – общее число ответов, приписанных стимулу всеми испытуемыми. Данный вариант применялся в опыте Мандлера [Фрумкина, 1955, с. 155].

Метод шкалирования для получения АС триграмм применялся, например, Ноблом. Сущность этого метода заключается в том, что испытуемые оценивают с помощью шкалы, много ли ассоциации порождает у них данная триграмма. Так, если им представляется, что триграмма вообще не порождает никаких ассоциаций, они помещают её в категорию шкалы № 1, имеющей наименование «никаких ассоциаций»; если ассоциаций порождается много, то триграмма может быть помещена в категорию № 5 («очень много») или в категорию № 4 («довольно много»), и т. д. АС триграммы в этом случае может быть подсчитана по формуле

$$a_i = \frac{1(f_1) + 2(f_2) + 3(f_3) + 4(f_4) + 5(f_5)}{N},$$

где 1, 2, 3, 4 и 5 – номера категорий шкалы; $f_1, f_2, \dots, f_5$ – число испытуемых, поместивших стимул соответственно в 1, 2, ..., 5 категории шкалы.

Характерной особенностью методов многомерного шкалирования является то, что они базируются на введении так называемой функции несоответствия (критерия расхождения), с помощью которой производится оценка того, насколько хорошо полученное представление сохраняет информацию, имеющуюся в исходных данных. Так как в данном случае отношениями эмпирической системы являются некоторые соотношения между оценками сходства пар, то функция несоответствия должна оценивать степень сохранения этих соотношений между расстояниями в соответствующей числовой системе [Клигер, 1978, с. 81].

В качестве стимулов для данного эксперимента было использовано 200 русских неосмысленных триграмм, для которых были ранее получены сведения о частоте их встречаемости в текстах ($F_{об}$). Субъективные оценки их частоты ($F_{суб}$) и оценки их произносительной трудности (Pr). Последние две характеристики триграмм были получены в результате проведения психометрических экспериментов.

4. Условия проведения эксперимента

В нашем наборе были представлены все основные типы трёхбуквенных сочетаний гласных и согласных графем русского языка: СГС (40 триграмм); ГСГ (27); ССГ (50); ГСС (63) и ССС (19). Кроме того, в набор включена одна триграмма типа ГГГ (ОУА), служившая «якорным» стимулом в эксперименте по получению $F_{суб}$ триграмм.

Эксперимент проводился в два этапа. Весь набор был произвольно разделен на списки А и Б, включавшие соответственно 117 и 83 триграммы. Эксперимент по порождению ассоциаций на триграммы списка А проводился одновременно в двух группах испытуемых (38 и 32 человек); список Б был предъявлен другим группам испытуемых (30 и 20 человек); кроме того, этот список предъявлялся 30 испытуемых индивидуально. Методика проведения группового опыта во всех случаях была идентичной.

Эксперимент проводился в двух группах старшеклассников. 117 триграмм набора были предварительно случайным образом разбиты на 3 списка по 39 триграмм в каждом (списки № 1, 2 и 3). После объяснения инструкции (в каждой группе на это потребовалось около 7 мин.) учащимся было роздано по 6 бланков, перенумерованных от 1 до 6. На каждом бланке стояли цифры от 1 до 39. Затем испытуемые приступали к работе. По окончании первого списка (соответственно бланка № 1) давалась команда переходить ко второму бланку (давался следующий список из 39 триграмм). Первый бланк при этом отбирался. Между списками делалась минутная пауза. Порядок следования стимулов в списке был фиксированным; порядок же следования списков в группах варьировал: 1–2–3–2–1–3 для одной группы испытуемых и 3–1–2–1–2 для

другой группы испытуемых. Таким образом, весь набор предъявлялся испытуемым дважды. Испытуемые не были предупреждены, что списки из 39 триграмм будут повторяться. Однако им говорили, что могут повторяться отдельные триграммы. Продолжительность опыта (включая объяснение инструкции) составила 40–48 мин. Таким образом, в среднем на работу с каждой триграммой (включая её предъявление экспериментатором и запись испытуемым ответа) отводилось 8–10 сек. В результате опыта для каждой триграммы списка А было получено по 140 ответов (прочерк, т. е. отсутствие ассоциации на триграмму, также считался ответом).

Аналогичным образом проводился эксперимент с триграммами списка Б. В групповых опытах было получено (60+40) 100 ответов на каждую триграмму. В индивидуальных опытах от испытуемых требовалось записать по одной ассоциации на данную триграмму. Таким образом, всего на каждую триграмму списка Б получено по 130 ответов.

5. Результаты и обсуждение

Анализ данных, полученных нами в различных опытах, показал, что результаты не зависят от того, в какой группе испытуемых проводился опыт, а также был ли опыт групповым или индивидуальным. Об устойчивости результатов ассоциативных экспериментов такого рода свидетельствуют и многочисленные литературные данные [Горелов, 2001, с. 10]. Это позволяет объединить все полученные данные.

Количественный анализ. По результатам опытов для каждой триграммы набора были выписаны все полученные ответы и подсчитано число прочерков, т. е. число, показывающее, сколько раз на данную триграмму не было порождено ни одного слова-ассоциации. Затем по этим результатам по формуле (1) для каждой триграммы была подсчитана величина m – показатель ассоциативной силы данной триграммы. Полученные значения m для 200 триграмм набора приведены в таблице 2. Триграммы в таблице расположены в порядке убывания m . Таблица 2 может быть использована в качестве источника сведений об АС обследованной группы русских триграмм.

Качественный анализ. Представляло интерес прежде всего рассмотреть вопрос о разнообразии слов-ответов, выдаваемых на триграммы-стимулы. С этой целью необходимо было решить, какие ответы будут считаться разными, а какие – одинаковыми. В рамках данной задачи считали целесообразным рассматривать как «одинаковые» ответы, являющиеся однокоренными словами или грамматическими формами одного слова (ср. *звон* – *звонить* – *звонок* – *звонка*). Все слова такого рода объединяли в одно «гнездо». «Гнездо» могло содержать одно или несколько слов. Выбранные нами 4 триграммы имеют примерно одинаковое m – 95,4–95,7%, однако степень разнообразия ответов на эти триграммы весьма различна; если на триграмму *НИВ* 108 чел. из 140 дали слово *нива*, то, например, наиболее частый ответ на триграмму *СУН* соответствует показаниям всего 19 испытуемых (из 130). Каким же образом можно было бы численно выразить эту разницу в степени разнообразия ответов? С этой целью воспользовались теоретико-информационным подходом [Мурзин, 1991, с. 17]. Совокупность всех ответов на данную триграмму T рассматривается, как опыт с k возможными исходами, где k – число реально полученных слов-ответов (прочерки не учитываются). В качестве численной оценки разнообразия исходов такого опыта естественно использовать его энтропию H . Согласно формуле:

$$H_T = - \sum_{i=1}^n p_i \log_2 p_i ,$$

где n – число различных слов-ответов, полученных на триграмму T ; p_i – доля слов-ответов данного типа среди всех полученных слов-ответов.

Однако непосредственно сравнить значения величины H_T для различных триграмм нельзя, так как число слов-ответов k для каждого T различно.

Т а б л и ц а 2. Результаты наблюдения различных триграмм

Стимулы		нив	олн	нав	сун
Хар-ка	m H I	95,7 0,185 1,954	95,7 0,398 0	95,7 0,581 1,950	95,4 0,696 1,755
Распределение ответов		1. нива 108 2. вино 11 3. ленивый 2 4. нева 2 5. винт 1 6. Иван 1 15. ни в чем 1	1. солнце 45 2. волна 35 3. полный 19 4. олень 14 5. волновался 3 6. подсолнух 2 7. салон 1 8. луна 1 и т. д.	1. навык 34 2. навстречу 16 3. наводнение 12 4. наврать 8 5. наводить 6 6. ванна 5 7. Ваня 5 8. наверное 5 9. навар 4 10. навзничь 4 11. навеки 3 12. навигация 3 13. автобус 3 14. нарва 3 15. канава 2 16. навалиться 2 17. навывлет 2 18. новатор 2 19. на вечере 1 и т. д.	1. сундук 19 2. сунуть 13 3. сукно 8 4. суп 7 5. сын 7 6. сук 5 7. сон 5 8. Сун-Ят-Сен 5 9. уснуть 4 10. суслик 4 11. судно 4 12. сумка 3 13. сухой 2 14. сосунок 2 15. сумерки 2 16. сустав 1 17. султан 1 и т. д.

Для сравнения энтропий над величиной H_T проводятся некоторые преобразования, а именно; предлагается вычислять приведённую энтропию

$$H' = \frac{H_T}{\log_2 k}$$

Как видно из табл. 2, H' триграммы НИВ составляет 0,185. H' триграмм ОЛН, НАВ и СУН приводятся во второй строке таблицы 3. Как и следовало ожидать, численная оценка степени разнообразия ответов у данных триграмм весьма различна. Указанная процедура была применена ко всем триграммам набора, в результате чего каждая из 200 триграмм получила оценку H' .

Т а б л и ц а 3. Статистические результаты обработки триграмм

Абсолютная частота слова-ответа в эксперименте (f_i)	Число слов с данной абсолютной частотой (n_{f_i})	$f_i \cdot n_{f_i}$	Относительная частота, $P_i = \frac{f_i}{k}$	$P_i \cdot \log_2 P_i$	$n_{f_i} \cdot P_i \cdot \log_2 P_i$
1	11	11	0,0075	0,0529	0,5819
2	2	4	0,0149	0,0908	0,1816
11	1	11	0,0821	0,2959	0,2959
108	1	108	0,8060	0,2508	0,2508

Анализ данных таблицы 3 показывает, что множества ответов на триграммы *НИВ*, *ОЛН*, *НАВ* и *СУН* различаются еще в одном качественном отношении: если среди ответов на триграммы *НИВ* и *НАВ* преобладают слова, начинающиеся с данных триграмм (более 80%), то доля таких слов в ответах на триграмму *СУН* значительно меньше (ответы, начинающиеся с *СУН*, составляют всего 30,6% от общего числа ответов). Что же касается триграммы *ОЛН*, то в множество полученных на неё ассоциаций вообще не входит ни одного слова, начинающегося с *ОЛН*. Способность триграмм порождать ассоциации, включающие данные триграммы в качестве начала, тесно коррелирует с АС этих триграмм. В частности, что среди триграмм, имеющих высокие оценки АС, преобладают те, которые порождают преимущественно слова-ассоциации, начинающиеся с этих триграмм [Мурзин, 1991, с. 17]. Вместе с тем, как показывает анализ, имеются триграммы, одинаковые по АС, но различающиеся по способности порождать ассоциации, начинающиеся с них. В силу этого представлялось желательным иметь возможность получить ещё одну численную оценку свойств триграмм – а именно, оценку того, насколько легко порождаются в ответ на данную триграмму слова, содержащие её в качестве начала. В качестве такой оценки была предложена величина I , которая подсчитывается по формуле

$$I = 1 + \frac{\ln l}{\ln k'},$$

где l – число слов-ответов («гнёзд»), начинающихся с данной триграммы; k – общее число слов-ответов («гнёзд»), без учёта прочерков.

Как видно из приведённой формулы, она не определена для случая $l = 0$, т. е. для случая, когда в составе слов-ответов нет ни одной ассоциации, начинающейся с данной триграммы. В этом случае здесь условно принимали $I = 0$. Таким образом, I триграммы *ОЛН* будет равно 0. В случае, когда $l > 0$, величина I изменяется от 1 ($I = 1$ при $l \sim 1$, т. е. в случае, когда имеется всего один ответ, включающий данную триграмму в качестве начала) до 2 ($I = 2$ при $l = k'$, т. е. в случае, когда все слова-ответы начинаются с данной триграммы). I для триграмм *НИВ* и *НАВ* равно соответственно 1,954 и 1,950; I триграммы *СУН* значительно ниже – 1,755. Значение I для всех 200 триграмм набора приводится в таблице 5 (см. также [Опыт..., 1971, с. 180]).

6. Статистический анализ экспериментальных исследований

После предварительного тренировочного сеанса была дана следующая инструкция:

Выяснить, какие слова или словосочетания ассоциируются с неосмысленными слогами или порождаются ими. На каждом листе, есть номера (с 1 до 50). Читается по буквам каждый слог дважды. Например, В – А – К ... В – А – К. После первого чтения слога, записываем его строчными буквами, а после повторного чтения запишем рядом с данным слогом слово или словосочетание, порождаемое в представлении этим слогом.

Для каждого слога будет в распоряжении не более 25 секунд.

Далее проводится проверка, извлечена ли данная выборка из нормальной генеральной совокупности – сложная непараметрическая гипотеза. Проверку провести с помощью χ^2 – критерия Пирсона с количеством интервалов k , равным 4 и 6. Результаты измерений распознавания (время (мс)). Выборка по времени на чтение не более 25 секунд: вариационный ряд (см. табл. 4).

Т а б л и ц а 4. Вариационный ряд выборки по времени на чтение

10,036	10,195	10,897	11,254	11,576	12,547	12,608	13,700	13,834
13,837	13,962	14,024	14,040	14,095	14,116	14,150	14,175	14,365
14,379	14,441	14,470	14,546	14,841	15,967	16,005	16,185	16,274
16,489	16,799	16,898	17,217	17,566	17,843	17,893	17,928	17,934
18,127	18,201	18,447	18,554	18,593	18,705	19,249	19,374	19,700
21,333	22,250	23,129	23,676					

Обработка результатов наблюдений:

$n = 49$;

Размах выборки: $R_n = x_{max} - x_{min} = 23,676 - 10,036 = 13,64$

Для числа интервалов $k = 4$:

Длина интервала

$$d = \frac{R}{k} = \frac{13,64}{4} = 3,41$$

Интервалы в общем виде:

$$\Delta_1 = (-\infty; x_{min} + d); \Delta_2 = (x_{min} + d; x_{min} + 2d); \Delta_3 = (x_{min} + 2d; x_{min} + 3d); \Delta_4 = (x_{min} + 3d; +\infty)$$

После подстановки:

$$\Delta_1 = (-\infty; 13,446); \Delta_2 = (13,446; 16,856); \Delta_3 = (16,856; 20,266); \Delta_4 = (20,266; +\infty)$$

Количество n_i элементов выборки, попавших в каждый интервал:

$$n_1 = 7; n_2 = 22; n_3 = 16; n_4 = 4$$

Поскольку $n_4 < 5$ объединяем 3-й и 4-й интервалы, тогда:

$$\Delta_1 = (-\infty; 13,446); \Delta_2 = (13,446; 16,856); \Delta_3 = (16,856; +\infty)$$

$$n'_1 = 7; n'_2 = 22; n'_3 = 20$$

Вероятность попадания в каждый интервал:

$$P_i = \Phi\left(\frac{h_i - \bar{x}}{S}\right) - \Phi\left(\frac{h_{i-1} - \bar{x}}{S}\right)$$

$$h_0 = -\infty; h_1 = 13,446; h_2 = 16,856; h_3 = +\infty$$

найдем среднее выборочное:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{49} \sum_{i=1}^{49} x_i = 16,05$$

найдем СКО:

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = \sqrt{\frac{1}{48} \sum_{i=1}^{49} (x_i - \bar{x})^2} = 3,14$$

$$P_1 = \Phi\left(\frac{13,446 - 16,05}{3,14}\right) - \Phi\left(\frac{-\infty - 16,05}{3,14}\right) = \Phi(-0,83) - \Phi(-\infty) = 0,467$$

$$P_2 = \Phi\left(\frac{16,856 - 16,05}{3,14}\right) - \Phi\left(\frac{13,446 - 16,05}{3,14}\right) = \Phi(0,256) - \Phi(-0,83) = 0,134$$

$$P_3 = \Phi\left(\frac{\infty - 16,05}{3,14}\right) - \Phi\left(\frac{16,856 - 16,05}{3,14}\right) = 1 - \Phi(0,256) = 0,399$$

Величины $C_i = n p_i$

$$C_1 = 49 \cdot 0,467 = 22,883; C_2 = 49 \cdot 0,134 = 6,566; C_3 = 49 \cdot 0,399 = 19,551$$

Значение статистики:

$$\chi^2 = \sum_{i=1}^3 \frac{(n_i - n_{pi})^2}{n_{pi}} = \frac{(7 - 22,883)^2}{22,883} + \frac{(22 - 6,566)^2}{6,566} + \frac{(20 - 19,551)^2}{19,551} = 11,02 + 36,28 + 0,01 = 47,31$$

Число степеней свободы: $v = k - l - 1$; $k = 3$; $l = 2 \Rightarrow v = 0$

Для распределения χ^2 видим, что для $v = 1$ и $\alpha = 0,001$ $\chi_{0,999}^2(1) = 10,83$

Поскольку $\chi^2 > \chi_{0,999}^2(1)$, гипотезу о том, что данная выборка извлечена из нормальной генеральной совокупности нужно отклонить.

б) Для числа интервалов $k = 6$. Длина интервала $d = R/k = 13,64 \div 6 = 2,27$.
Интервалы в общем виде:

$$\Delta_1 = (-\infty; x_{\min} + d); \Delta_2 = (x_{\min} + d; x_{\min} + 2d); \Delta_3 = (x_{\min} + 2d; x_{\min} + 3d); \\ \Delta_4 = (x_{\min} + 3d; x_{\min} + 4d); \Delta_5 = (x_{\min} + 4d; x_{\min} + 5d); \Delta_6 = (x_{\min} + 5d; +\infty)$$

После подстановки:

$$\Delta_1 = (-\infty; 12,306); \Delta_2 = (12,306; 14,576); \Delta_3 = (14,576; 16,846); \\ \Delta_4 = (16,846; 19,116); \Delta_5 = (19,116; 21,386); \Delta_6 = (21,386; +\infty)$$

Количество n_i элементов выборки, попавших в каждый интервал:

$$n_1 = 5 \quad n_2 = 17 \quad n_3 = 7 \quad n_4 = 13 \quad n_5 = 4 \quad n_6 = 3$$

Поскольку $n_5 < 5$ и $n_6 < 5$, объединяем 5-й и 6-й интервалы, тогда:

$$\Delta'_1 = (-\infty; 12,306); \Delta'_2 = (12,306; 14,576); \Delta'_3 = (14,576; 16,846); \\ \Delta'_4 = (16,846; 19,116); \Delta'_5 = (19,116; +\infty)$$

$$n_1 = 5 \quad n_2 = 17 \quad n_3 = 7 \quad n_4 = 13 \quad n_5 = 7$$

Вероятность попадания в каждый интервал:

$$P_i = \Phi\left(\frac{h_i - \bar{x}}{S}\right) - \Phi\left(\frac{h_{i-1} - \bar{x}}{S}\right)$$

$$h_0 = -\infty; h_1 = 12,306; h_2 = 14,576; h_3 = 16,846; h_4 = 19,116; h_5 = +\infty$$

$$P_1 = \Phi\left(\frac{12,306 - 16,05}{3,14}\right) - \Phi\left(\frac{-\infty - 16,05}{3,14}\right) = \Phi(-1,19) - \Phi(-\infty) = 0,116$$

$$P_2 = \Phi\left(\frac{14,576 - 16,05}{3,14}\right) - \Phi\left(\frac{12,306 - 16,05}{3,14}\right) = \Phi(-0,469) - \Phi(-1,19) = 0,204$$

$$P_3 = \Phi\left(\frac{16,846 - 16,05}{3,14}\right) - \Phi\left(\frac{14,576 - 16,05}{3,14}\right) = \Phi(0,254) - \Phi(-0,469) = 0,280$$

$$P_4 = \Phi\left(\frac{19,116 - 16,05}{3,14}\right) - \Phi\left(\frac{16,846 - 16,05}{3,14}\right) = \Phi(0,976) - \Phi(0,254) = 0,235$$

$$P_5 = \Phi\left(\frac{+\infty - 16,05}{3,14}\right) - \Phi\left(\frac{19,116 - 16,05}{3,14}\right) = 1 - \Phi(0,976) = 0,165$$

Величины $C_i = n \cdot p_i$

$$C_1 = 49 \cdot 0,116 = 5,684; C_2 = 49 \cdot 0,204 = 9,996; C_3 = 49 \cdot 0,280 = 13,720;$$

$$C_4 = 49 \cdot 0,235 = 11,515; C_5 = 49 \cdot 0,165 = 8,085$$

Значение статистики:

$$\chi^2 = \sum_{i=1}^5 \frac{(n_i - n_{pi})^2}{n_{pi}} = \frac{(5 - 5,684)^2}{5,684} + \frac{(17 - 9,996)^2}{9,996} + \frac{(7 - 13,72)^2}{13,72} + \frac{(13 - 11,515)^2}{11,515} + \frac{(7 - 8,085)^2}{8,085} = 0,082 + 4,907 + 3,29 + 0,19 + 0,145 = 8,614$$

Число степеней свободы: $\nu = k - l - 1$; $k = 5$; $l = 2 \Rightarrow \nu = 2$

Для распределения χ^2 при уровне значимости $\alpha = 0,01$. $\chi_{0,999}^2(2) = 9,21$

Поскольку $\chi^2 > \chi_{0,999}^2(2)$, гипотезу о том, что данная выборка извлечена из нормальной генеральной совокупности можно принять на уровне значимости $\alpha = 0,01$.

Далее проверим, может ли та же самая выборка быть извлечена из логнормальной генеральной совокупности – сложная непараметрическая гипотеза. Проверку проведём с помощью критерия Колмогорова.

Для применения критерия Колмогорова необходимо произвести расчёт:

$$c_i = \max \left(\left| \frac{i-1}{n} - F_0(x_i) \right|, \left| \frac{i}{n} - F_0(x_i) \right| \right),$$

$$\text{где } F_0(x_i) = \Phi \left(\ln \left(\frac{x_i - \bar{x}}{S} \right) \right)$$

Для расчёта функции Лапласа (F_0) для логнормального распределения используется программа MathCAD и функция $\text{plnorm}(\chi, \mu, \sigma)$. Здесь в качестве μ берётся его оценка $\bar{x} = 16,05$ и вместо σ также берётся его оценка $S = 3,14$, т.е. $\text{plnorm}(\chi; 16,05; 3,14)$. Произведём расчёт c_1 :

$$F_0(x_1) = \Phi \left(\ln \left(\frac{10,036 + 16,05}{3,14} \right) \right) = \Phi(-1,915) = 1 - 0,973 = 0,027$$

$$c_1 = \max \left(\left| \frac{1-1}{49} - 0,027 \right|, \left| \frac{1}{49} - 0,027 \right| \right) = \max(0,027, 6,59 \cdot 10^{-3}) = 0,027$$

Расчёт остальных производим аналогично. Результаты расчёта приведены в таблицах 5–6. Из таблицы видно, что $\max c_i = 0,209$, $T = \max c_i \cdot \sqrt{n} = 0,209 \cdot \sqrt{49} = 1,463$.

Для критерия согласия Колмогорова критическое значение $t_{0,01} = 1,63$. Поскольку $T < t_{0,01}$, гипотезу о том, что выборка была извлечена из логнормальной генеральной совокупности следует принять на уровне значимости $\alpha = 0,01$. Далее проверяется факт того, может ли выборка быть извлечена из экспоненциальной совокупности – сложная непараметрическая гипотеза. Проверка проводится с помощью критерия ω^2 . Прежде чем приступить к проверке гипотезы по критерию ω^2 , необходимо рассмотреть функцию

$$f(\chi, \lambda) = 1 - e^{-\lambda \chi}, \chi > 0$$

По методу максимального правдоподобия точечная оценка параметра (табл. 6).

$$\hat{\lambda} = 1/\bar{x}; \hat{\lambda} = 1/16,05 = 0,062$$

Т а б л и ц а 5. Результаты статистических вычислений

i	c_i	i	c_i	i	c_i	i	c_i	i	c_i
1	0,027	11	0,063	21	0,201	31	0,033	41	0,036
2	0,012	12	0,045	22	0,209	32	0,052	42	0,032
3	0,011	13	0,028	23	0,119	33	0,063	43	0,031
4	0,018	14	0,011	24	0,079	34	0,047	44	0,043
5	0,025	15	0,011	25	0,015	35	0,031	45	0,041
6	0,03	16	0,022	26	0,014	36	0,032	46	0,034
7	0,013	17	0,039	27	0,023	37	0,018	47	0,036
8	0,097	18	0,051	28	0,014	38	0,022	48	0,028
9	0,089	19	0,127	29	0,022	39	0,011	49	0,011
10	0,076	20	0,112	30	0,014	40	0,025	50	

Т а б л и ц а 6. Результаты вычислений по функции Лапласа

i	$F_0(x_i)$	i	$F_0(x_i)$	i	$F_0(x_i)$	i	$F_0(x_i)$	i	$F_0(x_i)$
1	0,4632	11	0,5792	21	0,5923	31	0,6561	41	0,6842
2	0,4685	12	0,5808	22	0,5942	32	0,6635	42	0,6864
3	0,4911	13	0,5812	23	0,6015	33	0,6692	43	0,6968
4	0,5022	14	0,5827	24	0,6284	34	0,6702	44	0,6992
5	0,5121	15	0,5832	25	0,6293	35	0,6709	45	0,7052
6	0,5406	16	0,5841	26	0,6334	36	0,6711	46	0,7336
7	0,5423	17	0,5847	27	0,6354	37	0,675	47	0,7483
8	0,5723	18	0,5896	28	0,6402	38	0,6765	48	0,7616
9	0,5759	19	0,59	29	0,6471	39	0,6814	49	0,7696
10	0,5759	20	0,5915	30	0,6492	40	0,6835	50	

Для $\chi_{(1)} = 10,036$; $F_0(\chi_1) = 0,4632$. Тогда $F_0(\chi_1) - \frac{2 \cdot 1 - 1}{2 \cdot 49} = 0,4632 - 0,0102 = 0,4530$.

Расчёт остальных слагаемых проводится с использованием программы MathCAD $m\omega^2 = 0,1877$ по таблице для экспоненциального закона распределения критическое значение при $\alpha = 0,01$ $\omega = 0,2299$ (см. табл. 6). Поскольку $m\omega^2 < \omega_{0,01}$, то гипотезу о том, что выборка извлечена из экспоненциальной совокупности можно принять на уровне значимости $\alpha = 0,01$.

Очевидно, что в действительности оценки АС не основаны на представлении испытуемых о числе разных слов-ассоциаций, порождаемых данной триграммой-стимулом, хотя инструкция ориентировала испытуемых именно на этот признак. Можно предположить, что в процессе шкалирования испытуемых преобразует признак «триграмма, которая порождает много разных слов», в какой-то иной признак, например, в признак «триграмма, которая легко порождает слово-ассоциацию» или «триграмма, которая быстро порождает слово-ассоциацию». Это предположение нуждается в экспериментальной проверке выдвигаемых гипотез по полученным законам распределения, которая покажет, в какой мере сопоставимы данные об АС, получаемые методом статистической оценки.

7. Заключение

Таким образом, в настоящей работе были предложены характеристики, позволяющие проводить качественный анализ результатов ассоциативного эксперимента. Интересующий стимул преобразовывался в привычные для него единицы языкового кода – слова и словосочетания. В качестве примера показано, как носители русского языка связывают неосмысленное буквосочетание *ЗВО* со словом *звон*, а сочетание *ТБЛ* – со словами *таблица* или *Тбилиси*, однако выявлено, что преобразование сочетаний типа *ГУВ* или *ОУА* в слова русского языка неизменно вызывает затруднения. В таком случае делается вывод, что буквосочетания *ЗВО* и *ТБЛ* обладают большей ассоциативной силой, чем сочетания *ГУВ* и *ОУА*. Изучение ассоциативной силы вербальных стимулов было вызвано тем, что, было замечено влияние этого значимого параметра на прогнозирование ряда важных аспектов речевого поведения человека, в том числе на результаты запоминания в опытах по парно-ассоциативному обучению, на пороги зрительного распознавания неосмысленных буквосочетаний и т. п.

В статье предложены методика определения ассоциативной силы для группы неосмысленных русских триграмм. Авторы не ставят своей задачей получение более или менее полной таблицы ассоциативной силы русских триграмм. Задача состояла прежде всего в том, чтобы отработать методику проведения эксперимента и проведения статистической обработки по получению, ассоциативной силы неосмысленных буквосочетаний и найти подходящие способы описания закона распределения получаемых результатов. Учитывая относительную новизну подхода, ставилась задача разработать методику обработки результатов ассоциативного эксперимента этого типа и, в частности, ввести некоторые характеристики, позволяющие проводить качественный анализ получаемых результатов.

Для постановки опытов с контролируемыми параметрами наилучшими стимулами являются неосмысленные буквосочетания. Испытуемые запоминают некоторые стимулы значительно быстрее, чем другие. В связи с этим есть возможность существования ещё одной характеристики неосмысленных буквосочетаний, способной влиять на их запоминаемость. Буквосочетание обладает тем большим значением, чем оно легче ассоциируется с привычными словами языка. Именно это свойство позволило назвать ассоциативной силой неосмысленные буквосочетания.

Список литературы

- Белянин, В. П. Психолингвистика [Текст] / В. П. Белянин. – М. : Флинта, 2005. – 226 с.
- Волошин, В. Г. Компьютерная лингвистика [Текст] / В. Г. Волошин. – Сумы : Университетская книга, 2004. – 382 с.
- Горелов, И. Н. Основы психолингвистики [Текст] / И. Н. Горелов, К. Ф. Седов. – М. : Лабиринт, 2001. – 304 с.
- Гридина, Т. А. Методика завершения высказывания: стратегии текstopорождения [Текст] / Т. А. Гридина // Психолингвистические аспекты изучения речевой деятельности. – Вып. 10. – Екатеринбург, 2012. – С. 18–34.
- Иванов, В. В. Лингвистика третьего тысячелетия: Вопросы к будущему [Текст] / В. В. Иванов. – М. : Языки славянской культуры. 2004. – 208 с.
- Ильин, В. В. Природа науки: Гносеологический анализ [Текст] / В. В. Ильин, А. Т. Калинин. – М. : Высшая школа, 1985. – 230 с.
- Канрашина, Е. Ю. Представление знаний о времени и пространстве в интеллектуальных системах [Текст] / Е. Ю. Канрашина, Л. В. Литвинцева, Д. А. Поспелов. – М. : Наука, 1989. – 328 с.
- Клигер, С. А. Шкалирование при сборе и анализе социологической информации [Текст] / С. А. Клигер, М. С. Косолапов, Ю. Н. Толстова. – М. : Наука, 1978. – 107 с.

- Ковшиков, В. А. Психолингвистика. Теория речевой деятельности [Текст] / В. А. Ковшиков, В. П. Глухов. – М. : АСТ, Астрель, 2007. – 224 с.
- Кутюра, Л. Философские принципы математики [Текст] / Л. Кутюра. – М. : Изд-во «ЛКИ», 2010. – 272 с.
- Миронова, Д. М. Применение кластерного анализа в текстологии. Структурная и прикладная лингвистика : межвуз. сб. / под ред. А. С. Герда, И. С. Николаева. – СПб. : Изд-во С.-Петерб. гос. ун-та. – 2015. – Вып. 11. – С. 155–160.
- Мурзин, Л. Н. Текст и его восприятие [Текст] / Л. Н. Мурзин, А. С. Штерн. – Свердловск : Изд-во Урал. ун-та, 1991. – 172 с.
- Опыт экспериментального исследования ассоциативной силы неосмысленных русских триграмм [Текст] / А. П. Василевич, Е. Н. Герганов, О. Е. Загустина [и др.] // Вероятностное прогнозирование в речи / отв. ред. Р. М. Фрумкина, – М. : Наука 1971. – 199 с.
- Потапова, Р. К. Средства фонетического членения речевого потока в немецком и русском языках : учеб. пособие [Текст] / Р. К. Потапова, Л. П. Блохина. – М. : МГПИ им. Мориса Тереза, 1986. – 115 с.
- Рассел, С. Искусственный интеллект: современный подход, 3-е изд. : Пер. с англ. [Текст] / С. Рассел, П. Норвиг. – М. : Издательский дом «Вильямс», 2006. – 1408 с.
- Селиванова, Е. А. Лингвистическая энциклопедия [Текст] / Е. А. Селиванова. – Полтава : Довкиля-К, 2010. – 844 с.
- Сидняев, Н. И. Методологические аспекты преподавания высшей математики в контексте модернизации школьного математического образования [Текст] / Н. И. Сидняев // AlmaMater (Вестник высшей школы). – 2014. – № 5. – С. 33–40.
- Фрумкина, Р. М. Вероятность элементов текста и речевое поведение [Текст] / Р. М. Фрумкина. – М. : Наука, 1971. – 168 с.
- Физиология речи. Восприятие речи человеком [Текст] / Л. А. Чистович, А. В. Венцов, М. П. Гранстрем [и др.]. – Л. : Наука, 1976. – 388 с.

References

- Belyanin, V. P. (2005). *Psikholingvistika* [Psycholinguistics]. Moscow : Flinta Press.
- Voloshin, V. G. (2004). *Komp'yuternaya lingvistika* [Computer Linguistics]. Sumy : Universitetskaya kniga Press.
- Gorelov, I. N., Sedov, K. F. (2001). *Osnovy psikholingvistiki* [Fundamentals of psycholinguistics]. Moscow : Labirint Press.
- Gridina, T. A. (2012). Metodika zaversheniya vyskazyvaniya: strategii tekstoporozhdeniya [The technique of finishing the utterance: Text production strategy]. *Psikholingvisticheskie aspekty izucheniya rechevoy deyatel'nosti* [Psycholinguistic aspects of speech activity] (pp. 18–34). Ekaterinburg.
- Ivanov, V. V. (2004). *Lingvistika tret'ego tysyacheletiya: Voprosy k budushchemu* [Linguistics of the Third Century: Questions to the Future]. Moscow : Yazyki slavyanskoy kul'tury Press.
- Il'in, V. V., Kalinkin, A. T. (1985). *Priroda nauki: Gnoseologicheskii analiz* [The Nature of Science: Epistemological Analysis]. Moscow : Vysshaya shkola Press.
- Kandrashina, E. Yu., Litvintseva, L. V., Pospelov, D. A. (1989). *Predstavlenie znaniy o vremeni i prostranstve v intellektual'nykh sistemakh* [Representation of Knowledge about Time and Space in Intelligent Systems]. Moscow : Nauka Press.
- Kliger, S. A., Kosolapov, M. S., Tolstova, Yu. N. (1978). *Shkalirovanie pri sbore i analize sociologicheskoy informatsii* [Scaling method while collecting and analyzing sociological information] Moscow : Nauka Press.
- Kovshikov, V. A., Glukhov, V. P. (2007). *Psikholingvistika. Teoriya rechevoy deyatel'nosti* [Psycholinguistics. Theory of Speech Activity]. Moscow : Astrel' Press.
- Kutyura, L. (2010). *Filosofskie printsipy matematiki* [Philosophical principles of mathematics]. Moscow : LKI Press.

- Mironova, D. M. (2015). Primenenie klaster'nogo analiza v tekstologii [Cluster analysis in textual criticism]. *Strukturnaya i prikladnaya lingvistika* [Structural and Applied Linguistics] (Vol. 11, pp. 155–160). St Peterburg : St Peterburg State University Press.
- Murzin, L. N., Shtern, A. S. (1991). *Tekst i ego vospriyatie* [Text and its Perception]. Sverdlovsk : Ural University Press.
- Vasilevich, A. P., Gerganov, E. N., Zagustina, O. E. et al. (1971). Opyt eksperimental'nogo issledovaniya asociativnoy sily neosmyslennykh russkikh trigramm. In R. M. Frumkina, *Veroyatnostnoe prognozirovanie v rechi* [Probabilistic approach to speech]. Moscow : Nauka Press.
- Potapova, R. K., Blokhina, L. P. (1986). *Sredstva foneticheskogo chleneniya rechevogo potoka v nemetskom i russkom yazykakh* [Means of Sound Division of Speech in German and Russian] : A coursebook. Moscow : The Maurice Thorez Institute of Foreign Languages.
- Russell, S., Norvig, P. (2010). *Artificial Intelligence: A Modern Approach*. 3d edition. Moscow : Vil'yams Publishing House.
- Selivanova, E. A. (2010). *Lingvisticheskaya entsiklopediya* [Linguistic Encyclopedia]. Poltava : Dovkily –K Press.
- Sidnyaev, N. I. (2014). Metodologicheskie aspekty prepodavaniya vysshey matematiki v kontekste modernizatsii shkol'nogo matematicheskogo obrazovaniya [Methodological Aspects of Lecturing of Higher Mathematics in the Context of Modernization of School Mathematical Education]. *AlmaMater (Vestnik vysshey shkoly)* [AlmaMater (Higher School Herald)], 5, 33–40.
- Frumkina, R. M. (1971). *Veroyatnost elementov teksta i rechevoe povedenie* [The Probability of Elements of Text and Speech Behavior]. Moscow : Nauka Press.
- Chistovich, L. A., Ventsov, A. V., Granstrem, M. P. et al. (1976) *Fiziologiya rechi. Vospriyatie rechi chelovekom* [Physiology of Speech. Human Speech Perception]. Saint Petersburg : Nauka Press.