

УДК 81'443:[811.161.1+811.111]

Н.В. Шульга,
магістр педагогічних наук,
преподаватель кафедры иностранных языков БГПУ

СТАТИСТИЧЕСКИЙ АНАЛИЗ РЕДУПЛИКАТИВНЫХ ОБРАЗОВАНИЙ БЕЛОРУССКОГО, РУССКОГО И АНГЛИЙСКОГО ЯЗЫКОВ

Явление редупликации (от позднелат. *reduplicatio* 'удвоение') представляет собой непродуктивный способ слово – и формообразования, заключающийся в двукратном контактном повторении фонологически схожих компонентов (корней, основ и слов), в результате которого образуются редупликативные единицы, характеризующиеся полуслитным написанием.

В белорусском, русском и английском языках выделяются 3 универсальных типа редупликативных образований (далее – РО): 1) полные – компоненты которых фонологически идентичны (бел. *беры-беры* (ад сінг. *berі* 'слабасць'), рус. *хороший-хороший*, англ. *goo-goo* 'агу-агу'), 2) аблаутивные – компоненты которых различаются ударным гласным при аллитерации начальных согласных (бел. *піф-паф*, рус. *буки-бяки*, англ. *criss-cross* 'крест-накрест') и 3) рифмованные – компоненты которых различаются начальными согласными при рифмовании остальных звуков (бел. *мульці-пульці*, рус. *муси-пуси*, англ. *hanky-panky* 'шурь-мурь').

На сегодняшний день должное теоретическое осмысление получили следующие аспекты редупликации рассматриваемых языков: особенности фономорфологической и деривационной структуры РО [1–3], сущность их экспрессивных функций [4–5] и способы перевода [6].

Накопленный в результате проведенных исследований фактический материал¹ позволяет предположить, что в белорусском, русском и английском языках активность полных, аблаутивных и рифмованных РО, составляющих выборки разных объемов, различна. Однако лишь применение элементов математической статистики даст возможность объективно оценить полученные количественные результаты, увеличивая тем самым достоверность выводов по нашему диссертационному исследованию, посвященному выявлению типологических сходств и различий в системе редуплика-

тивных единиц сопоставляемых языков.

Таким образом, цель настоящей статьи состоит в оценивании с помощью статистических методов неизвестной вероятности распределения РО в белорусском, русском и английском языках.

Методика исследования и полученные результаты. Для проведения соответствующих расчетов необходимо знать: 1) общий размер выборки, то есть количество всех проанализированных словоупотреблений² в данном языке (с учетом повторяемости лексических единиц размер выборки составил: 13 146 673 словоупотребления в белорусском языке, 13 641 661 – в русском и 13 139 340 – в английском); 2) общее количество всех редупликативов, которые встречаются в выборке (с учетом повторяемости выделено: а) 547 полных РО в белорусском языке, 695 – в русском и 259 – в английском; б) 32 аблаутивных РО в белорусском языке, 72 – в русском и 390 – в английском; в) 176 рифмованных РО в белорусском языке, 295 – в русском и 535 – в английском языке).

Сравним вероятности распределения РО различных типов. Для этого рассмотрим два языка. Пусть p_1 и p_2 – неизвестные вероятности появления полных, аблаутивных и рифмованных редупликативов в первом и во втором языках соответственно. При уровне значимости $\alpha = 0,05$ проверяется нулевая гипотеза H_0 : $p_1 = p_2$ о равной вероятности появления редупликативов разных типов в двух генеральных совокупностях при конкурирующей гипотезе H_1 : а) $p_1 < p_2$, б) $p_1 > p_2$, согласно которой РО того или иного структурного типа в одном из сопоставляемых языков появляются с большей или меньшей вероятностью по сравнению с другим языком³ (таблица 1).

¹ Лексические единицы, составляющие редупликативный корпус, распределились следующим образом: 1) полные РО: 325 единиц в белорусском языке, 331 – в русском, 197 – в английском; 2) аблаутивные РО: 20 единиц в белорусском языке, 33 – в русском и 291 – в английском; 3) рифмованные РО: 100 единиц в белорусском языке, 182 – в русском и 392 – в английском.

² Под словоупотреблением в данном случае понимается «цепочка букв, заключенная в тексте между двумя пробелами» [7, с. 11]. Общее количество словоупотреблений удобно подсчитывать с помощью функции «Статистика» в программе MS Word.

³ Гипотезы выбираются исходя из количественных показателей РО, составляющих редупликативный корпус современного белорусского, русского и английского языков.

Таблица 1 – Гипотезы, применяемые к редупликативам трех сопоставляемых языков

ГИПОТЕЗЫ			
Языковая пара	белорусский vs. русский	белорусский vs. английский	русский vs. английский
Тип РО			
полные	$H_0: p_1 = p_2$ $H_1: p_1 < p_2$		$H_0: p_1 = p_2$ $H_1: p_1 > p_2$
аблаутивные		$H_0: p_1 = p_2$ $H_1: p_1 < p_2$	
рифмованные			

В качестве статистики критерия для проверки гипотез H_0 и H_1 выбирается случайная величина $E_{\text{набл}}$, которая вычисляется для каждой пары языков по следующей формуле [8, с. 237–239]:

$$E_{\text{набл}} = \frac{\frac{b_1}{c_1} - \frac{b_2}{c_2}}{\sqrt{\frac{b_1 + b_2}{c_1 + c_2} \times \left(1 - \frac{b_1 + b_2}{c_1 + c_2}\right) \times \left(\frac{1}{c_1} + \frac{1}{c_2}\right)}}, \quad (1)$$

где b_1 и b_2 – частота появления РО в первом и во втором языках из сопоставляемой пары; c_1 и c_2 – размер выборок в первом и во втором языках соответственно.

Основные числовые показатели, найденные по выборкам, извлеченным из печатных текстов трех сопоставляемых языков и применяемые в ходе вычисления значения статистики критерия, приведены в таблице 2.

Для получения выводов касательно соотношения вероятностей распределения РО разных типов в генеральных совокупностях сопоставляемых языков вычисленное значение статистики ($E_{\text{набл}}$) необходимо сравнить с крити-

ческой областью ($e_{\text{кр}}$), критическая точка которой находится из следующего равенства:

$$\Phi(e_{\text{кр}}) = \frac{1 - 2\alpha}{2} = \frac{1 - 2 \times 0,05}{2} = 0,45, \quad (2)$$

где $\Phi(e_{\text{кр}})$ – функция Лапласа.

Вычислив значение $\Phi(e_{\text{кр}}) = 0,45$, находим по таблице функции Лапласа [8, с. 390–391] значение $e_{\text{кр}} = 1,645$.

При конкурирующей гипотезе $H_1: p_1 > p_2$, проверяемой при сравнении полных РО белорусского и русского языков с английским, необходимо найти критическую точку правосторонней критической области. Если $E_{\text{набл}} < e_{\text{кр}}$, оснований отвергнуть нулевую гипотезу не имеется. Если $E_{\text{набл}} > e_{\text{кр}}$, нулевая гипотеза отвергается и принимается альтернативная гипотеза [8, с. 238]. При конкурирующей гипотезе $H_1: p_1 < p_2$ критическая точка располагается в левосторонней критической области. Если $E_{\text{набл}} > -e_{\text{кр}}$, оснований отвергнуть нулевую гипотезу не имеется. Если $E_{\text{набл}} < -e_{\text{кр}}$, нулевая гипотеза отвергается и принимается альтернативная гипотеза [8, с. 238].

Таблица 2 – Значение статистики критерия по выборке

Языковая пара	Тип РО	Частота РО	Размер выборки	Значение статистики ($\alpha = 0,05$)
белорусский ₁ vs. русский ₂	полные	$b_1 = 547$ $b_2 = 695$	$c_1 = 13\ 146\ 673$ $c_2 = 13\ 641\ 661$	$E_{\text{набл}} \approx -3,549$
	аблаутивные	$b_1 = 32$ $b_2 = 72$		$E_{\text{набл}} \approx -3,735$
	рифмованные	$b_1 = 176$ $b_2 = 295$		$E_{\text{набл}} \approx -5,083$
белорусский ₁ vs. английский ₂	полные	$b_1 = 547$ $b_2 = 259$	$c_1 = 13\ 146\ 673$ $c_2 = 13\ 139\ 340$	$E_{\text{набл}} \approx 10,137$
	аблаутивные	$b_1 = 32$ $b_2 = 390$		$E_{\text{набл}} \approx -17,433$
	рифмованные	$b_1 = 176$ $b_2 = 535$		$E_{\text{набл}} \approx -13,471$
русский ₁ vs. английский ₂	полные	$b_1 = 695$ $b_2 = 259$	$c_1 = 13\ 641\ 661$ $c_2 = 13\ 139\ 340$	$E_{\text{набл}} \approx 13,539$
	аблаутивные	$b_1 = 72$ $b_2 = 390$		$E_{\text{набл}} \approx -15,201$
	рифмованные	$b_1 = 295$ $b_2 = 535$		$E_{\text{набл}} \approx -8,873$

Таблица 3 – Статистическое решение

Языковая пара		Тип РО	Гипотезы	$E_{\text{набл}} \text{ vs. } e_{\text{кр}}$ ($\alpha = 0,05$)	Выводы
белорусский vs. русский	полные		$H_0: p_1 = p_2$ $H_1: p_1 < p_2$	-3,549 < -1,645	отвергнуть H_0
	аблаутивные			-3,735 < -1,645	отвергнуть H_0
	рифмованные			-5,083 < -1,645	отвергнуть H_0
белорусский vs. английский	полные		$H_0: p_1 = p_2$ $H_1: p_1 > p_2$	10,137 > 1,645	отвергнуть H_0
	аблаутивные			-17,433 < -1,645	отвергнуть H_0
	рифмованные			-13,471 < -1,645	отвергнуть H_0
русский vs. английский	полные		$H_0: p_1 = p_2$ $H_1: p_1 > p_2$	13,539 > 1,645	отвергнуть H_0
	аблаутивные			-15,201 < -1,645	отвергнуть H_0
	рифмованные			-8,873 < -1,645	отвергнуть H_0

Выводы касательно принятия или отклонения основной гипотезы, основанные на сравнении наблюдаемого значения критерия с критической областью, представлены в таблице 3.

Анализ результатов сравнения вероятностей распределения РО различных типов в сопоставляемых языках, представленный в таблице 3, позволяет установить следующее: вероятность появления полных редупликативов в белорусском и русском языках выше, нежели в английском языке. Статистически значимые отличия наблюдаются и в распределении аблаутивных и рифмованных редупликативов трех сопоставляемых языков, которые с большей вероятностью появляются в английском языке по сравнению с рассматриваемыми восточнославянскими языками. В белорусском языке вероятность проявления РО любых структурных типов ниже, чем в русском языке.

Определим относительную частоту (\bar{b}) проявлений редупликативов различных типов в трех сопоставляемых языках¹, а также рассчитаем приближенные границы доверительного интервала (b' ; b'') для оценки с надежностью $\gamma = 0,95$ неизвестной вероятности распределения РО в каждом из языков по следующим равенствам [8, с. 174–175]:

$$b' = \frac{b}{c} - t \cdot \sqrt{\frac{\frac{b}{c} \cdot \left(1 - \frac{b}{c}\right)}{c}}; \quad (3)$$

$$b'' = \frac{b}{c} + t \cdot \sqrt{\frac{\frac{b}{c} \cdot \left(1 - \frac{b}{c}\right)}{c}}; \quad (4)$$

где b – частота проявлений РО в языке; c – размер выборки в соответствующем языке; $t = 1,96$ –

¹ Относительная частота проявлений редупликативов (\bar{b}) представляет собой отношение частоты РО (b) к размеру выборки (c) и определяется по следующему равенству [8, с. 175]: $\bar{b} = \frac{b}{c}$.

значение аргумента функции Лапласа².

Основные числовые показатели, используемые в ходе вычисления относительной частоты (\bar{b}) и определения левых (b') и правых (b'') границ доверительного интервала, приведены в таблице 4.

В результате статистического оценивания неизвестной вероятности распределения и частоты РО различных типов в трех сопоставляемых языках наблюдаются следующие закономерности (таблица 4): наибольшую вероятность появления в белорусском и русском языках имеют полные редупликативы, относительная частота которых составила соответственно 0,0042 % (при доверительном интервале от 0,0038 % до 0,0045 %) и 0,0051 % (при доверительном интервале от 0,0047 % до 0,0055 %). В английском языке полные РО используются с относительной частотой в 0,00197 % (при доверительном интервале от 0,0017 % до 0,0022 %), что в 2,1 раз ниже, нежели в белорусском языке, и в 2,6 раз реже по сравнению с русским языком.

Наименьшей вероятностью проявлений в сопоставляемых восточнославянских языках характеризуются аблаутивные редупликативы: оценка вероятности того, что наугад взятое слово в белорусском языке окажется аблаутивным РО, составляет всего 0,00024 % (при доверительном интервале от 0,00016 % до 0,00033 %), что в 2,2 раза меньше, чем в русском языке, где относительная частота аблаутивных РО составляет 0,00053 % (при доверительном интервале от 0,00041 % до 0,00065 %). В английском языке вероятность проявлений

² Значение аргумента $t = 1,96$ было найдено по таблице функции Лапласа [8, с. 390–391], после того как мы вычислили значение функции Лапласа $\Phi(t) = 0,475$ по формуле

$$\Phi(t) = \frac{\gamma}{2} = \frac{0,95}{2} = 0,475,$$

где γ – заданная надежность.

Таблица 4 – Доверительный интервал для оценки неизвестной вероятности распределения редупликативов

Язык Тип РО		Частота РО (b)	Размер выборки (c)	Относит. частота (\bar{b}), $\cdot 10^{-5}$	Доверит. интервал ($\gamma = 0,95$)	
					$b', \cdot 10^{-5}$	$b'', \cdot 10^{-5}$
Бел.	полные	547	13 146 673	4,2	3,8	4,5
	аблаутивные	32		0,24	0,16	0,33
	рифмованные	176		1,3	1,1	1,5
Рус.	полные	695	13 641 661	5,1	4,7	5,5
	аблаутивные	72		0,53	0,41	0,65
	рифмованные	295		2,2	1,9	2,4
Англ.	полные	259	13 139 340	1,97	1,7	2,2
	аблаутивные	390		2,97	2,7	3,3
	рифмованные	535		4,1	3,7	4,4



Рисунок – Частота редупликативных образований

аблаутивных РО, относительная частота которых составляет 0,00297 % (при доверительном интервале от 0,0027 % до 0,0033 %), в 5,6 раз выше, нежели в русском языке, и в 12,2 раза больше, чем в белорусском языке.

Что касается рифмованных РО, то они наиболее часто используются в английском языке, характеризуясь относительной частотой в 0,0041 % (при доверительном интервале от 0,0037 % до 0,0044 %). В свою очередь, в белорусском языке вероятность использования рифмованных РО в 3 раза меньше по сравнению с английским языком (относительная частота составляет 0,0013 % при доверительном интервале от 0,0011 % до 0,0015 %) и в 1,6 раз меньше, нежели в русском языке, где относительная частота рифмованных РО составляет 0,0022 % (при доверительном интервале от 0,0019 % до 0,0024 %).

Графически соотношение частотности полных, аблаутивных и рифмованных РО в белорусском, русском и английском языках может быть представлено следующим образом (рисунок).

Таким образом, полученные опытным путем данные позволяют говорить о зависимости между типом сопоставляемых языков и оценкой неизвестной вероятности распределения РО. В белорусском и русском языках явление редупликации выступает не столько словообразовательным, сколько формообразовательным способом, поэтому значительно преобладают полные РО разных частей речи. В английском языке редупликация выполняет преимущественно словообразовательную роль и большинство редупликативов представлено рифмованными и аблаутивными типами.

ЛИТЕРАТУРА

1. Шульга, Н.В. Фономорфологическая структура редупликативных слов белорусского, русского и английского языков / Н.В. Шульга // Весці БДПУ. Сер. 1. Педагогіка. Псіхалогія. Філалогія. – 2010. – № 3. – С. 62–66.
2. Шульга, Н.В. Образование редупликативных слов и словоформ в современном белорусском, русском и английском языках / Н.В. Шульга // Научные стремления: молодеж. сб. науч. ст. / Совет

- молодых ученых НАН Беларуси, ООО «Лаборатория интеллекта»; редкол.: А.Н. Волченко [и др.]. – Минск, 2012. – Вып. 1. – С. 81–86.
3. Шульга, Н.В. Словообразовательные и формообразовательные модели редупликативных единиц: на материале белорусского, русского и английского языков / Н.В. Шульга // Весті БДПУ. Сер. 1. Педагогіка. Псіхалогія. Філалогія. – 2011. – № 4. – С. 34–38.
 4. Шульга, Н.В. Стилистический эффект редупликативных образований в пространстве художественных текстов современных русских писателей / Н.В. Шульга // Молодежь в науке – 2011: прил. к журн. «Весті Нацыянальнай акадэміі навук Беларусі»: материалы Междунар. науч. конф. молодых ученых, Минск, 25–29 апр. 2011 г.: в 5 ч. / Нац. акад. наук Беларуси; Совет молодых ученых НАН Беларуси; редкол.: А.А. Коваленя (гл. ред.) [и др.]. – Минск, 2012. – Ч. 2: Серия гуманитарных наук. – С. 225–228.
 5. Шульга, Н.В. Функциональная характеристика некоторых подтипов рифмованных редупликативных образований / Н.В. Шульга // Весті БДПУ імя І.П. Шамякіна. – 2011. – № 4. – С. 108–112.
 6. Шульга, Н.В. Алгоритм перевода редупликативных образований: на материале белорусского, русского и английского языков / Н.В. Шульга // Весті БДПУ імя І.П. Шамякіна. – 2012. – № 3. – С. 119–125.
 7. Пиотровский, Р.Г. Математическая лингвистика: учеб. пособие / Р.Г. Пиотровский, К.Б. Бектаев, А.А. Пиотровская. – М.: Высш. школа, 1977. – 383 с.
 8. Гмурман, В.Е. Руководство к решению задач по теории вероятностей и математической статистике: учеб. пособие / В.Е. Гмурман. – 9-е изд., стереотип. – М.: Высш. шк., 2004. – 404 с.

SUMMARY

The article describes the procedure and results of statistical estimation derived from unknown probability distribution of full, ablaut and rhymed reduplicative formations in Belarusian, Russian and English. It is concluded that there is a correlation between the type of language, the nature and frequency of reduplicative words and word forms.

Поступила в редакцию 27.06.2013 г.