

4. Брусенцев В. О. Алгебро-логічні моделі формалізації семантики речень та їх застосування в інформаційних системах штучного інтелекту : дис. ... канд. технічних наук: 05.13.23 – системи та засоби штучного інтелекту / В. О. Брусенцев. – Харків. – 2003. – 228 с.
5. Дарчук Н. П. Комп'ютерна лінгвістика (автоматичне опрацювання тексту) : підручник. – К. : Видавничо-поліграфічний центр "Київський університет", 2008. – 351 с.
6. Волошин В. Г. Комп'ютерна лінгвістика : Навчальний посібник / В. Г. Волошин. – Суми : Університетська книга, 2004. – 382 с.
7. Карпіловська Є. А. Вступ до прикладної лінгвістики: комп'ютерна лінгвістика : Підручник / Є. А. Карпіловська. – Донецьк : ТОВ "Юго-Восток, ЛТД", 2006. – 188 с.
8. Карпіловська Є. Корпуси кореневих морфем в українській та словацькій мовах: можливості альтернативних рішень // Є. Карпіловська // Morfematický výskum slovenčiny (možnosti jeho štatistického, elektronického a didaktického spracovania), Prešov, Filozofická fakulta Prešovskej university v Prešove, 2006. – С. 249–262.
9. Клименко Н. Ф. Морфемно-словотвірний фонд української мови як дослідницька та інформаційно-довідкова система / Н. Ф. Клименко, Є. А. Карпіловська, Л. І. Карпіловська та ін. // Мовознавство. – 1990. – № 6. – С. 41–50.
10. Марченко О. О. Алгоритми семантичного аналізу природномовних текстів: дис. ... канд. фіз.-мат. наук: 01.05.01 / О. О. Марченко. – К., 2005. – 150 с.
11. Перебийніс В. С. Напрямки і школи в сучасній структурній лінгвістиці / В. С. Перебийніс // Проблеми та методи структурної лінгвістики. – К., 1969. – Вип. 2. – С. 5–25.
12. Партико З. В. Прикладна і комп'ютерна лінгвістика : Вступ до спеціальності / З. В. Партико. – Львів : Афіша, 2008. – 224 с.
13. Феноменологія лексикографічних систем: Монографія / В. А. Широков; НАН України. Український мовно-інформаційний фонд. – К. : Наук. думка, 2004. – 327 с.

Borys Onyschenko, Tetyana Scherbina

Use of virtual keyboard in the process of typesetting of dialectic texts

The article demonstrates the possibilities of a developed virtual keyboard for the facilitation of researcher's work in the process of use of special symbols for text demonstration of dialectic speech.

Key words: system print, symbolic information, virtual keyboard.

**Татьяна Бодрова,
Надежда Тукмакова**

ОПРЕДЕЛЕНИЕ КОЭФФИЦИЕНТА РАНГОВОЙ КОРРЕЛЯЦИИ ЧАСТЕЙ РЕЧИ В РУССКИХ И ЧУВАШСКИХ ГАЗЕТНЫХ ТЕКСТАХ

У статті визначена міра статистичної залежності частин мови в російських і чувашських газетних текстах за допомогою коефіцієнта рангової кореляції, величина якого змінюється залежно від жанру газетної мови. Матеріалом дослідження стали популярні місцеві номери газет Чувашської Республіки "Хыпар" і "Советская Чувашия".

Ключові слова: частини мови, статистична залежність, коефіцієнт рангової кореляції, російські й чувашські газетні тексти.

Коэффициент ранговой корреляции в лингвостатистике означает вычисление статистической связи между лингвистическими объектами. Статистической же связью называется такая зависимость, при которой каждому значению одной величины может соответствовать несколько значений другой

зависимой величины. Эти зависимости выражаются в том, что при увеличении активности (росте частотности) явления А увеличивается или уменьшается активность явления В. Следует заметить, что лингвисты нередко могут сталкиваться с необходимостью “предвидения”, прогнозирования поведения какого-либо лингвистического события. При этом, во-первых, изучаемые признаки могут оказаться независимыми друг от друга, когда поведение одного признака никоим образом не влияет на поведение другого. Например, применение прошедшего времени никак не влияет на структуру предложений. Во-вторых, изучаемые признаки могут находиться в функциональной зависимости друг от друга. Тогда знание поведения одного признака дает точные сведения поведения другого.

В последнее время особый интерес лингвистов вызывает статистический анализ частей речи в различных функциональных стилях языка. В свое время А. А. Потебня заметил, что “в русском языке, как и в других сродных, по направлению к нашему времени увеличивается противоположность имени и глагола, возникают новые разграничения существительного и прилагательного, наблюдаются обособление прилагательного и наречия [8, с. 516]. Одним словом, по его мнению, в современном русском языке происходит все большая дифференциация частей речи.

Использование в лингвистике статистических методов безусловно способствует познанию языка на более высоком уровне и получению объективных знаний о нем. Это подчеркивали такие видные ученые, как И. А. Бодуэн де Куртене, А. М. Пешковский, М. Н. Петерсон, Е. Д. Поливанов, В. В. Виноградов.

При всем существующем интересе к данной тематике статистическое исследование частей речи еще несмело применяется в отечественном языкознании. Большая заслуга в этом направлении языкознания безусловно принадлежит Головину Борису Николаевичу и Носенко Ивану Андреевичу, которые впервые указали на необходимость не просто подсчета языковых единиц, но и определение статистической связи между изучаемыми языковыми явлениями, степень их взаимозависимости. Б. Н. Головин выделяет ряд конкретных исследовательских вопросов в лингвистике.

1. Имя существительное грамматически господствует над именем прилагательным. Означает ли это, что увеличение активности имен и существительного в речи всегда влечет за собой и рост активности имени прилагательного?

2. Имя существительное и глагол нередко воспринимаются как взаимосвязанные и конкурирующие в истории языка части речи. Существует ли между ними корреляционная зависимость в процессе их речевого функционирования и, если существует, какова она – прямая или обратная (т. е. возрастает или убывает частотность глагола при возрастании существительного)?

3. Предлоги в русском языке обслуживают нужды падежного управления, а шире – нужды морфологической сочетаемости имен с именами и имен с

глаголами. Есть ли корреляционная связь между функционированием имен существительных и предлогов, глаголов и предлогов?

4. Меняется ли сила корреляционной связи между наблюдаемыми явлениями языка в зависимости от языковых и речевых стилей, а также и других условий языкового функционирования и развития?

Очевидно, что перечень подобных вопросов здесь далеко не исчерпан. Актуальность настоящего исследования состоит в том, что впервые в сравнительно-сопоставительном аспекте изучено взаимодействие частей речи на материале русских и чувашских публицистических текстов. Выбор указанных языков не случаен. Общеизвестно: русский язык характеризуется лексико-фразеологическим богатством, разветвленной системой синтаксических структур, обширными стилистическими возможностями. Это язык межнационального общения, оказывающий значительное влияние на контактирующие с ним системы. Он является достаточно полно изученным в количественном аспекте и имеет хорошо наработанную литературу. В интенсивных и продолжительных контактах с русским – флективным языком – находится чувашский, агглютинативный по своей природе, также имеющий иные социолингвистические свойства, как правило, отличные от функциональных особенностей русского языка.

Статистический и сопоставительный аспекты в чувашском языкознании находятся в состоянии, которое требует уточняющего определения – «фрагментарный». Статистический аспект чувашской речи наиболее ярко представлен в диссертационных работах В. П. Зайцевой [5] и П. В. Желтова [4]. С сожалением приходится констатировать, что введение компаративных разысканий с равным привлечением чувашского материала было и, скорее всего, останется прерогативой чувашских ученых. Части речи в чувашском языкознании еще не изучались ни в статистическом, ни в сопоставительном аспектах.

Как в русской, так и в чувашской лингвистике отсутствуют специальные исследования, посвященные определению статистической связи между частями речи публицистического стиля. Между тем средства массовой информации занимают достаточно большое место в жизни человека. Важнейшей их функцией является формирование общественного сознания. Для определения нормы, выяснения степени употребительности и стилистической характеристики лексических, морфологических и синтаксических единиц исследователи все чаще обращаются к газетной речи.

В данной статье попытаемся решить вопрос: существует ли между частями речи корреляционная зависимость в процессе их речевого функционирования и, если существует, какова она – прямая или обратная?

Представим следующую гипотезу: во-первых, между частями речи как в русской, так и в чувашской газетной речи существует корреляционная связь. Эта статистическая связь меняется в зависимости от жанра публицистической речи; во-вторых, между сравниваемыми газетными текстами существуют как черты сходства, связанные с особенностями жанра, так и отличия в силу структурных особенностей сравниваемых языков. Достижение цели возможно

при решении следующих задач: определить репрезентативные выборки газет, найти величины коэффициентов ранговой корреляции и провести сравнительно-сопоставительный анализ полученных результатов.

Генеральную совокупность составили местные номера газет “Советская Чувашия” и “Хыпар”. Выбор газет обусловлен прежде всего тем, что названные издания являются основными газетами Чувашской Республики. В каждой из газет были исследованы по 2 жанра – информационный и аналитический. Способ изложения материала – традиционный стиль статей в местных газетах. Учитывался и размер выборок. Информационный жанр в газетах занимает небольшой объем (100–250 словоупотреблений), поэтому в качестве выборок послужили тексты статей в их полном, законченном виде. В аналитическом жанре была установлена граница в 300 словоупотреблений. Соответственно, выборки этого жанра явились одинаковыми и между собой, и между сравниваемыми газетами. Из каждого жанра мы исследовали по 10 выборок, так как такого количества достаточно для проведения статистических исследований: оно обеспечивает 95%-ую надежность в показателях, что позволяет увидеть закономерную тенденцию функционирования языковых единиц [2, с. 27].

Силу статистической связи между двумя изучаемыми явлениями можно измерить несколькими способами. Самые распространенные из них – выборочный коэффициент корреляции, ранговый коэффициент корреляции, коэффициент взаимной сопряженности, коэффициент линейной корреляции. Каждый из методов имеет свои преимущества перед другим. Считается, что первый и последний способы более строгие и требуют кропотливой работы. Вычисление рангового коэффициента корреляции является работой не менее трудоемкой, но зато не требует глубоких математических знаний. Именно поэтому мы воспользовались формулой рангового коэффициента корреляции.

Коэффициент ранговой корреляции был определен по формуле Спирмена:

$$\rho = 1 - \frac{1 - 6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

[7, с. 118–119].

Для того чтобы применить эту формулу, все выборочные частоты одного ряда ранжируют, то есть располагают в порядке их возрастания или убывания. Затем находят порядковые места (ранги) соответствующих частот другого ряда. Для получения нужного результата нужно найти величину d . Что такое d ? Это разность порядковых мест, занимаемых в двух ранжированных частотных рядах частотами двух изучаемых явлений, между которыми предполагается корреляционная связь. И, наконец, вычисляется величина коэффициента корреляции при условии, что n равно числу выборок.

Ранговый коэффициент может приобретать значения от -1 до $+1$. Величина ранговой корреляции может быть положительной, что свидетельствует о

прямой зависимости друг от друга частей речи: возрастание или убывание одной части речи влечет за собой возрастание или убывание другой, или отрицательной, при которой возрастание одной части речи сопровождается убыванием другой. И наоборот, убывание одной – возрастанием другой. Если полученное число коэффициента корреляции приближается к нулю, это будет свидетельствовать об отсутствии корреляционной связи или о наличии очень слабой связи.

Проиллюстрируем действие этой формулы в таблицах на примере информационного жанра в исследуемых нами газетных текстах.

Таблица 1

**Статистические данные
для вычисления коэффициента корреляции рангов
между именами существительными и глаголами в “Советской Чувашии”**

Существительное	<i>R1</i>	Глагол	<i>R2</i>	$d = R1 - R2$	d^2
90	9,5	20	3	6	36
82	7	29	9	-2	4
90	9,5	14	1	9	81
72	5	24	6	-1	1
83	8	20	4	4	16
64	3	26	8	-5	25
77	6	24	7	-1	1
69	4	21	5	-1	1
59	2	33	10	-8	64
43	1	17	2	-1	1

Таблица 2

**Статистические данные
для вычисления коэффициента корреляции рангов
между именами существительными и глаголами в “Хыпаре”**

Существительное	<i>R1</i>	Глагол	<i>R2</i>	$d = R1 - R2$	d^2
101	7,5	17	5,5	2	4
108	8,5	27	7	1,5	2,25
82	4	14	4	0	0
108	8,5	10	2,5	6	36
101	7,5	7	1	6,5	42,25
76	3	21	6	-3	9
91	5	17	5,5	-0,5	0,25
74	2	13	3,5	-1,5	2,25
94	6	13	3,5	2,5	6,25
67	1	10	2,5	-1,5	2,25

R1 – порядковый номер первой исследуемой части речи;

R2 – порядковые места частот другой исследуемой части речи;

n – число наблюдений (выборок).

Результаты вычислений представлены в таблице 3:

Таблица 3

**Величины коэффициентов
ранговой корреляции между частями речи
в русских и чувашских газетных текстах¹**

Коррелирующие части речи	Информационный жанр		Аналитический жанр	
	“Советская Чувашия”	“Хыпар”	“Советская Чувашия”	“Хыпар”
Существительное–глагол	- 0.39	-	-	- 0.49
Существительное–причастие	+ 0.38	+ 0.42	-	+ 0.47
Существительное–прилагательное	+ 0.76	+ 0.22	+ 0.42	+ 0.22
Существительное–предлог /послелог	+ 0.64	-	+ 0.41	-
Существительное–союз	+ 0.32	- 0.43	+ 0.24	- 0.9
Глагол- деепричастие	-	+ 0.19	-	+ 0.59
Прилагательное-наречие	-	+ 0.81	-	- 0.82
Причастие-деепричастие	-	+ 0.78	-	+ 0.36
Глагол-предлог/послелог	+ 0.16	-	+ 0.67	-
Глагол-союз	+ 0.38	- 0.58	+ 0.72	- 0.06

Сводные данные о величинах ранговой корреляции между частями речи свидетельствуют о том, что:

– между существительными и глаголами в информационном жанре (ИЖ) русской газеты и аналитическом жанре (АЖ) чувашской обнаружилась отрицательная статистическая связь, проявляющаяся в том, что при увеличении частоты имен существительных частота глаголов уменьшается и наоборот: при уменьшении существительных возрастает частота глаголов. В ИЖ “Хыпар” (“Х”) и АЖ “Советской Чувашии” (“СЧ”) обнаружилась другая тенденция – функциональная статистическая независимость указанных частей речи²;

– в ИЖ русской и чувашской газет, а также в АЖ газеты “Х” наблюдается прямая статистическая связь между существительными и причастиями. Наличие указанной статистической связи для русской газеты может быть интерпретировано ее ориентированностью на книжную речь: *Между тем*

¹ Результаты в таблицу внесены выборочно. При этом соблюдался основной принцип – цифры должны быть значимыми.

² Б. Н. Головин, исследуя корреляционную связь между существительными и глаголами в текстах А. И. Герцена, И. С. Тургенева и А. И. Гончарова, высказал предположение, что чем независимее друг от друга указанные части речи, тем более речь писателя рационалистичнее и ближе к публицистике [Головин Б. Н. Опыт применения корреляционного анализа в изучении языка // Вопросы статистической стилистики. – Киев, 1974. – С. 13]. Опыт применения корреляционного анализа в изучении языка // Вопросы статистической стилистики. – Киев, 1974. – С. 13].

только с батыревским хозяйством “Труд” заключен договор на хранение в “Пахча” с правом продажи 600 тонн картофеля и 300 тонн моркови [Нужен ли столице Большой ОПТ // СЧ. – 2001. – 4 авг.]; Подведены итоги экономического соревнования между крупными и средними предприятиями (организациями) промышленности республики за первое полугодие. За стабильную работу и достижение высоких экономических показателей предприятиям присвоены призовые места [Кто лучший? // Там же. – 2001. – 16 авг.]. Для чувашской газетной речи употребление причастий не является стилистически маркированным, т. к. глагольные формы в чувашском языке имеют большой удельный вес независимо от функциональной принадлежности;

– во всех жанрах газет выявилось тесное функциональное притяжение существительных с прилагательными. Полученные результаты дают положительный ответ на интересующий многих лингвистов вопрос: действительно ли прилагательные являются лишь покорными спутниками существительных? Цифры показывают: активность прилагательных зависит от активности существительных. Такая зависимость приводит к образованию устойчивых сочетаний, что является характерной чертой газетной публицистики. В русских газетах в силу своей высокочастотности устойчивыми можно назвать такие сочетания, как *активное участие, основная функция, реальные доходы, высокая награда, торжественная церемония* и т. д. В чувашских газетах – *пысăк уяв (большой праздник), пысăк зал (большой зал), пысăк ёс (большая работа), чаплă церемони (торжественная церемония), хаклă тăвансем (дорогие родственники), ырă хăнасем (добрые гости), полномочиллĕ представителĕ (полномочный представитель), суйлав кампанийĕ (предвыборная кампания)* и т. д.;

– степень статистической спаянности между существительными и предлогами в русской газете демонстрирует их положительную связь в обоих жанрах. Это объясняется особенностями языка газеты, характерными для него конструкциями “существительное + существительное с предлогом”, такими как *договор на хранение, плата за услуги, тариф за пользование* и др. Для чувашской газетной речи свойственна функциональная независимость послелогов от существительных, отсюда преобладание конструкции “существительное + существительное”: *бюджетна бюджет (бюджет с бюджетом), саккунна саккун (закон с законом), республикăна республика (республика с республикой), патшалăхна патшалăх (государство с государством)*;

– между глаголами и деепричастиями, а так же между причастиями и деепричастиями в обоих жанрах “Х” выявилась положительная статистическая связь. Данный факт определяется структурными особенностями чувашского языка, для которого характерно обилие аналитических конструкций: *Президент СССР халăх артисткин В.К. Кузьминан кăкăрĕ çине орден турсе хучĕ* – Президент прикрепил орден к груди заслуженной артистки СССР В. К. Кузьминой [Клементьев И. Ёсрĕ палăрнисене чысларĕс // Х. – 1999. – 30 дек.]; *2001çулта продукци туса кăларасси хăвăрт чакса ларнă. Укçа–тенкĕ тĕлĕшĕнчен пысăк йывăрлăхсем тухса тăнă* [Павлова И., Алексеев Л.

Сынсемпе – куҫа-куҫан // Там же. – 2002. – 30 апр.] – В 2001 году выпуск продукции резко снизился. С финансовой стороны возникли большие трудности.

– величины корреляции между прилагательными и наречиями в “Х” получились контрастными – прямая в ИЖ и обратная в АЖ. ИЖ является номинативным, в уточнении существительных важную роль играют прилагательные. В АЖ признак, скорее всего, достигается наличием более тесной статистической спаянности глаголов с деепричастиями. В чувашском языке вспомогательные глаголы придают дополнительный оттенок деепричастию, благодаря чему составные глаголы получают новые семантические качества. Они как бы живописуют действие, характеризуя его наглядно, образно, картинно;

– отличительные черты, связанные со структурными особенностями языков, отразились в употребляемости служебных частей речи – союзов и предлогов (в русском), послелогов (в чувашском). Для русской публицистики характерна прямая статистическая связь предлогов практически со всеми частями речи. В чувашской публицистике она в большинстве случаев близка к нейтральной, что объясняется особенностями этого языка: соединительную функцию между частями речи в нем могут выполнять различные аффиксы. В частотном соотношении количество послелогов в чувашских текстах значительно уступает функционально равноправным русским предлогам: в ИЖ “СЧ” – 245, “Х” – 59, в АЖ “СЧ” – 363, “Х” – 91. В русских текстах союзы употребляются гораздо чаще, чем в чувашских: в ИЖ “СЧ” – 165, АЖ – 250; в ИЖ “Х” – 35, АЖ – 93. В чувашском языке в определенных случаях функцию союза выполняет аффикс творительного падежа -па (-пе) и аффикс деепричастия -са (-се).

Таким образом, между частями речи как в русской, так и в чувашской газетной речи существует корреляционная связь – прямая (положительная) или обратная (отрицательная). Эта статистическая связь меняется в зависимости от жанра газетной речи, т. е. части речи в обеих газетах стилистически маркированы. Между русскими и чувашскими газетными текстами существуют как черты сходства, связанные с особенностями жанра, так и отличия в силу структурных особенностей сравниваемых языков и их типологического своеобразия. Количественно-качественный подход создает перспективу превращения обычной структурной модели языка в структурно-вероятностную, отличающуюся объективностью и реалистичностью. Такого рода исследования, необходимые для дальнейшего развития теории журналистики и совершенствования газетной практики и, безусловно, должны быть продолжены на выборках большего объема с применением компьютерных технологий.

ЛИТЕРАТУРА

1. Головин Б. Н. Опыт применения корреляционного анализа в изучении языка / Б. Н. Головин // Вопросы статистической стилистики. – Киев : Наукова думка, 1974. – С. 5–16.

2. Головин Б. Н. Язык и статистика / Б. Н. Головин. – М. : Просвещение, 1971. – 191 с.
3. Данилов А. П. Чăваш публицистики: Вĕренĕ пособиё / А. П. Данилов. – Ш. : Чăв. ун-чĕн изд-ви. – 1991. – 80 с..
4. Желтов П. В. Сопоставительно-сравнительное исследование морфем чувашского языка с применением формальных методов : Дис. ... канд. филол. наук: 10.02.20 / П. В. Желтов. – Чебоксары: Чуваш. гос. ун-т, 2010. – 194 с.
5. Зайцева В. П. Исследования частотности употребления слов в различных типах чувашских текстов: Дис. ... канд. филол. наук: 10.02.06. / В. П. Зайцева. – Чебоксары : Чуваш. гос. ун-т, 2000. – 138 с.
6. Квантитативные методы отбора учебного материала по иностранному языку для неязыкового вуза: Сб. науч. трудов / Урал. гос. ун-т, Свердловск, 1990. – 138 с.
7. Носенко И. А. Начало статистики для лингвистов / И. А. Носенко. – М. : Высш. шк., 1981. – 157 с.
8. Потебня А. А. Из записок по русской грамматике / А. А. Потебня. – М. : Учпедгиз, 1958. – Т. 1–2. – 536 с.
9. Сергеев В. И. Именные части речи: Учеб. пос. / В. И. Сергеев. – Чебоксары : Изд-во Чуваш. ун-та, 1994. – 152 с.

Bodrova Tatyana, Tukmakova Nadezhda

***Determination of the coefficient of rank correlation of parts of speech
in russian and chuvash newspaper texts***

The article deals with degree of statistical dependence of parts of speech in Russian and chuvash newspaper texts with the help of the coefficient of rank correlation coefficient, the value of which varies depending on the genre of the newspaper speech. This research served as the popular local newspapers of the Chuvash Republic "Хыпар" and "Sovetskaya Chuvashiya".

Key words: parts of speech, the rank correlation coefficient, Russian and chuvash newspaper texts, statistical relationship.