



О влиянии социокультурных процессов на изменение языка художественной прозы

Шерстинова Татьяна Юрьевна, НИУ ВШЭ СПб, tsherstinova@hse.ru

Аннотация/Abstract

Доклад посвящен результатам междисциплинарного проекта, целью которого являлось статистическое изучение изменений, происходящих в языке в переломные моменты истории, на примере одного из наиболее драматических периодов в истории нашей страны – первых 30-ти лет XX века. Исследование проведено на материале специально созданного Корпуса русского рассказа 1900-1930-х гг., в который включены тексты 300 писателей.

В результате проведенного исследования получены уникальные статистические данные, отражающие изменения языка и стиля художественной прозы за исследуемый период времени: в частности, построены частотные словари и сформированы списки «ключевых слов» для разных периодов рассматриваемой эпохи (дореволюционного, революционного, советского), выявлены основные тенденции в распределении частей речи и использовании синтаксических структур. Полученные результаты показывают, что действительно можно проследить корреляцию между социокультурными процессами, происходящими в обществе и определенными языковыми и стилистическими изменениями прозы.

Введение/Introduction

Доклад посвящен результатам междисциплинарного проекта, поддержанного грантом РФФИ, № 17-29-09173 офи_м «Русский язык на рубеже радикальных исторических перемен: исследование языка и стиля предреволюционной, революционной и постреволюционной художественной прозы методами математической и компьютерной лингвистики (на материале русского рассказа)», целью которого являлось статистическое изучение изменений, происходящих в языке в переломные моменты истории, на примере одного из наиболее драматических периодов для нашей страны – первых 30-ти лет XX века. Для оценки масштабов произошедших в языке трансформаций в проекте используются статистические методы анализа, применяемые к результатам компьютерной и экспертной обработки текстов художественной литературы.

Исследование проведено на материале специально созданного Корпуса русского рассказа 1900-1930-х гг. При разработке корпуса ставилась задача включения в него текстов максимального количества писателей, работавших и публиковавшихся в рассматриваемый период.

Методы и материал исследования/ Methods & Materials

Для проведения автоматического лингвистического анализа было отобрано 310 рассказов, представляющих творчество 300 русских писателей, среди которых есть всемирно известные (А.П.Чехов, Л.Н.Толстой, И.А.Бунин, М.А.Шолохов и др.), относительно известные (А.Белый, В.В.Вересаев, В.Г.Короленко, З.Н.Гиппиус, Б.А.Пильняк, и др.), а также малоизвестные или почти забытые в наше время авторы (Б.А.Верхоустинский, В.Я.Ленский, И.Ф.Колотовкин, Е.Н.Опочинин, О.П.Снегина и др.). Эти тексты составили аннотированный корпус, состоящий из трех хронологических периодов: 1) Рассказ начала XX века до Первой мировой войны (1900–1913), 2) Рассказ эпохи острых социальных потрясений — Первой мировой войны и последующей Революционной эпохи, включая Гражданскую войну (1914–1922), и 3) Рассказ постреволюционной эпохи (1923–1930). Общий объем текстов аннотированного корпуса превышает 1 млн. словоупотреблений. Все рассказы аннотированного корпуса прошли автоматическую обработку с помощью целого комплекса программных средств (mystem, UDPipe, Antconc, stylo, собственно-разработанных утилит).

Результаты/Results

В результате были получены уникальные статистические данные, отражающие языковые характеристики как отдельных текстов, так и обобщенных хронологических срезов. Для всех текстов были построены частотные словари, отражающие разнообразие используемой лексики и выявляющие наиболее характерные слова языка эпохи. Ключевые слова переломного, революционного периода (1917-1922) продолжают военную тематику (*винтовка, русский, (рас)стрелять, солдат, поручик, пулемет, казарма*) и отражают основные революционные реалии (*толпа, совет, коммунист, кандидат, комитет*). На постреволюционном этапе ключевыми становятся *комиссар, лагерь, товарищ, сельсовет, отряд, революция, военком, бандит, гражданин, советский, комсомолец, заявление*. Сравнение дистрибуции частей речи в разные временные периоды не показывает статистически значимых отличий, что позволяет говорить об устойчивости языковой системы на этом уровне, но можно говорить о некоторой тенденции к увеличению доли существительных и глаголов за счет снижения доли наречий, личных местоимений и союзов. Что же касается синтаксических структур, в среднем в революционный и постреволюционный периоды наблюдается их упрощение, в том числе уменьшение средней длины предложения. При этом максимально короткие фразы характерны для языка революционной эпохи.

Актуальность и значимость (результатов) исследования

Значимость осуществленного проекта для развития междисциплинарных направлений определяется апробацией современных средств обработки естественного языка на материале художественных текстов, а также совмещения гуманитарных и «точных» методов исследования литературного материала. В рамках данного проекта проблема системного анализа литературы и ее языка тесно переплетается с проблемой возрождения и сохранения литературного наследия, существенная часть которого до недавнего времени была вычеркнута из памяти народа. В этом также состоит важность и новизна полученных результатов.

Выводы/Conclusions

Обобщенные статистические данные, полученные в результате выполнения проекта, позволяют выявить наиболее значимые языковые трансформации, а также оценить масштабность и глубину произошедших в результате революционных событий изменений языка и стиля. Они показывают, что действительно можно проследить корреляцию между социокультурными процессами, происходящими в обществе, и определенными языковыми и стилистическими изменениями прозы: 1) на лексическом уровне наблюдается изменение частот значимой лексики, появляются новые реалии, 2) на уровне синтаксиса наблюдается упрощение синтаксических структур в революционный и постреволюционный периоды, 3) на уровне пунктуации тоже произошли изменения: если в довоенной прозе изобиловали многоточия, то в революционную эпоху, по понятным причинам, наступает время восклицательных знаков, 4) в то же время дистрибуция частей не показывает статистически значимых отличий, что говорит об устойчивости языка в этом отношении.

Литература/References

1. Мартыненко Г. Я. Методы математической лингвистики в стилистических исследованиях. СПб.: Нестор-История, 2019. 296 с.
2. Мартыненко Г. Я. Основы стилеметрии. СПб.: Изд-во СПбГУ, 1988. 176 с.
3. Martynenko G., Sherstinova T. Linguistic and Stylistic Parameters for the Study of Literary Language in the Corpus of Russian Short Stories of the First Third of the 20th Century. CEUR Workshop Proc. Vol. 2552, 2020, pp. 105–120.
4. Гребенников А. О., Скребцова Т. Г. Языковая картина мира в русском рассказе начала XX века. Философия и гуманитарные науки в информационном обществе, 2019, 3(25). С. 141-143
5. Sherstinova T., Ushakova E., and Melnik A. Measures of Syntactic Complexity and their Change over Time (the Case of Russian). In: 27th Conference of Open Innovations Association FRUCT, Univ. of Trento, Italy: 2020, pp. 221–229.
6. Sherstinova T., Grebennikov A., Skrebtsova T., Guseva A., Gukasian M., Egoshina I., and Turygina M. Frequency Word Lists and Their Variability (the Case of Russian Fiction in 1900-1930), In: 27th Conf. of Open Innovations Association FRUCT, University of Trento, Italy: 2020. pp. 366–373.