

**КОРПУСНОЕ ИССЛЕДОВАНИЕ РЕЧЕВОГО ЖАНРА
«ПОЗДРАВЛЕНИЕ»: ЛИНГВОКОГНИТИВНЫЙ И
СОЦИОЛИНГВИСТИЧЕСКИЙ АСПЕКТЫ**

Юлдашева Сабина Юсуфовна

преподаватель кафедры русского языка и литературы

Кокандский государственный университет

Аннотация. В тезисах обосновываются принципы составления специализированного корпуса поздравительных текстов на узбекском и русском языках. Описываются источники материала, этапы построения корпуса, система метаданных и методы корпусного анализа, позволяющие исследовать жанр в лингвокогнитивном и социолингвистическом аспектах.

Ключевые слова: корпусная лингвистика, специализированный корпус, речевой жанр «Поздравление», метаданные, частотный анализ, коллокации, ключевые слова, речевой этикет.

Корпусная лингвистика позволяет перейти от описания отдельных примеров к анализу больших массивов реальных текстов. Т. МакЭнери и Э. Харди подчёркивают, что корпусный подход даёт исследователю не только иллюстрации, но и количественные данные о частотности и распределении языковых единиц. Для этикетных жанров это особенно важно: их формулы кажутся очевидными носителю языка, но реальная частотность, вариативность и распределение формул по ситуациям общения остаются малоизученными.

Национальные корпуса содержат поздравления лишь фрагментарно, без указания на социальные параметры ситуации. Корпусные ресурсы узбекского языка активно развиваются, однако этикетные жанры в них представлены недостаточно. Поэтому для исследования жанра «Поздравление» необходим **специализированный корпус**, составленный по следующим принципам.

– **Репрезентативность.** Корпус включает тексты из разных каналов: поздравительные открытки, официальные поздравления организаций и государственных органов, публикации в социальных сетях, сообщения в мессенджерах, тосты и устные поздравления в транскрипции.

– **Сопоставимость.** Узбекский и русский подкорпусы сбалансированы по каналам и поводам и имеют одинаковую структуру метаданных, что делает возможным сопоставительный анализ.

– **Разметка метаданных.** Для каждого текста фиксируются повод, канал, степень официальности, пол, возраст и статус адресанта и адресата, характер их отношений.

– **Этическая корректность.** Персональные данные обезличиваются; используются только открытые источники или тексты, полученные с согласия авторов.

Построение корпуса проходит несколько этапов. На первом этапе осуществляется сбор текстов и их первичный отбор по жанровому признаку. На втором — очистка: удаление дубликатов, рекламных шаблонов, текстов, не являющихся поздравлениями. На третьем этапе выполняются токенизация и морфологическая разметка; для узбекского языка, как агглютинативного, особое значение имеет корректное выделение аффиксов. На четвёртом этапе тексты получают метаданные,

а на пятом — жанровую и семантическую разметку: выделяются речевые ходы и концепты пожеланий.

Корпус открывает широкие возможности для лингвокогнитивного анализа. Частотные списки и анализ ключевых слов выявляют ядро и периферию концептов, вербализуемых в пожеланиях. Анализ коллокаций показывает устойчивые сочетания, например *sihat-salomatlik tilayman, uzoq umr, крепкого здоровья, желаю счастья*. Конкордансы позволяют увидеть, в каких контекстах употребляется та или иная формула и как она модифицируется авторами.

Сравнение подкорпусов позволяет установить, какие концепты являются универсальными для обеих лингвокультур, а какие — национально специфичными. Можно предположить, что концепты «здоровье» и «счастье» будут частотными в обоих подкорпусах, тогда как концепты «долголетие» и «вера» окажутся более характерными для узбекских поздравлений. Корпусный подход позволяет проверить эту гипотезу на количественных данных, а не на отдельных примерах.

Социолингвистический анализ опирается на метаданные. Распределение формул по подкорпусам показывает, как степень официальности, возраст и статус участников влияют на длину текста, выбор обращения, форм вежливости и набора пожеланий. Например, можно сравнить долю обращений с формулами уважения (*Hurmatli, Muhtaram, Уважаемый*) в поздравлениях, адресованных старшим и ровесникам. Статистические методы, в частности критерий χ^2 , позволяют проверить, являются ли наблюдаемые различия значимыми.

Отдельное направление анализа связано с каналом коммуникации. Поздравления в мессенджерах и социальных сетях отличаются краткостью, использованием эмодзи, графических средств и смешением

языков. Корпус позволяет проследить, как цифровая среда трансформирует традиционный жанр и какие элементы его композиции сохраняются, а какие исчезают.

Важной частью корпусного исследования является жанровая разметка. Каждый текст делится на речевые ходы: обращение, собственно поздравление, оценку события или адресата, пожелание, завершение и подпись. Такая разметка позволяет установить, какие ходы обязательны для жанра, а какие факультативны, и как их набор зависит от канала и степени официальности. Например, официальное поздравление организации, как правило, содержит развёрнутую оценку заслуг адресата, тогда как поздравление в мессенджере нередко ограничивается одной формулой.

Сопоставительный анализ подкорпусов позволяет выявить и различия в способах выражения одного и того же концепта. Концепт «здоровье» в узбекском языке выражается устойчивыми парными сочетаниями (*sihat-salomatlik, sog'lik va omonlik*), а в русском — сочетаниями прилагательного с существительным (*крепкого здоровья, доброго здоровья*). Концепт «долголетие» в узбекских текстах нередко соединяется с религиозными формулами (*Alloh umringizni uzoq qilsin*), тогда как в русских он выражается светскими формулами (*долгих лет жизни*). Корпус позволяет не только зафиксировать такие различия, но и оценить их частотность.

Корпусные данные имеют и лингводидактическое значение. Частотные списки поздравительных формул, распределённые по ситуациям общения, могут служить основой учебных материалов по речевому этикету для студентов, изучающих русский язык в узбекской аудитории. Студент получает не абстрактный набор формул, а сведения о том, какая формула уместна в официальной, полуофициальной и

неофициальной ситуации и по отношению к адресату какого возраста и статуса.

Для обработки корпуса могут использоваться как специализированные программы (AntConc, Sketch Engine), так и средства языка программирования Python. Это делает исследование воспроизводимым: другой исследователь может повторить анализ на тех же данных и проверить выводы.

Результатом исследования станет размеченный двуязычный корпус поздравлений и описание жанра, основанное на количественных данных. Такой корпус может использоваться в лексикографии при составлении словарей речевого этикета, в преподавании узбекского и русского языков как иностранных, а также при разработке систем автоматической обработки текстов — классификаторов, чат-ботов и генераторов этикетных сообщений.

Литература

1. Бахтин М. М. Проблема речевых жанров // Бахтин М. М. Эстетика словесного творчества. — М.: Искусство, 1979. — С. 237–280.
2. Захаров В. П., Богданова С. Ю. Корпусная лингвистика. — 3-е изд. — СПб.: Изд-во СПбГУ, 2020.
3. Формановская Н. И. Речевой этикет и культура общения. — М.: Высшая школа, 1989.
4. Ҳақимов М. Ўзбек прагматингвистикаси асослари. — Тошкент: Akademnashr, 2013.
5. Biber D., Conrad S., Reppen R. Corpus Linguistics: Investigating Language Structure and Use. — Cambridge: Cambridge University Press, 1998.

6. Brown P., Levinson S. C. Politeness: Some Universals in Language Usage.
— Cambridge: Cambridge University Press, 1987.
7. McEnery T., Hardie A. Corpus Linguistics: Method, Theory and Practice.
— Cambridge: Cambridge University Press, 2012.