

Министерство науки и высшего образования РФ
Федеральное государственное автономное
образовательное учреждение высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Институт филологии и языковой коммуникации
Кафедра теории германских и романских языков и прикладной лингвистики

УТВЕРЖДАЮ
Заведующий кафедрой
_____ О.В. Магировская
« ____ » _____ 2023 г.

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

**ПОСТРОЕНИЕ И ОПТИМИЗАЦИЯ МУЛЬТИМОДАЛЬНЫХ
КОРПУСОВ ДИАЛОГИЧЕСКОЙ СПОНТАННОЙ РЕЧИ ДЛЯ
МАШИННОГО ОБУЧЕНИЯ (НА МАТЕРИАЛЕ РУССКОГО И
КИТАЙСКОГО СТУДЕНЧЕСКИХ ДИСКУРСОВ)**

45.04.02 Лингвистика

45.04.02.01 Межкультурная коммуникация и перевод

Научный руководитель	_____	канд. филол. наук, доц. О.А. Прохорова
Магистрант	_____	С.В. Ухоботова
Нормоконтролер	_____	А.А. Струзик

Красноярск 2023

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	4
ГЛАВА 1. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ПОСТРОЕНИЯ МУЛЬТИМОДАЛЬНЫХ КОРПУСОВ ДИАЛОГИЧЕСКОЙ СПОНТАННОЙ РЕЧИ	11
1.1. Становление мультимодальной лингвистики как продолжения теории дискурс-анализа.....	11
1.2. Функциональные характеристики устной коммуникации	15
1.2.1. Особенности речевых сбоев в устной спонтанной речи	22
1.2.2. Особенности невербального компонента в устной спонтанной речи	33
1.3. Общие принципы аннотирования и сегментирования данных при создании мультимодальных корпусов диалогической спонтанной речи	41
1.4. Мультимодальные корпусы как ресурсы для создания моделей машинного обучения	53
ВЫВОДЫ ПО ГЛАВЕ 1.....	60
ГЛАВА 2. ОПЫТ ПОСТРОЕНИЯ И ОПТИМИЗАЦИИ МУЛЬТИМОДАЛЬНЫХ КОРПУСОВ ДИАЛОГИЧЕСКОЙ СПОНТАННОЙ РЕЧИ ДЛЯ МАШИННОГО ОБУЧЕНИЯ.....	63
2.1. Дизайн эксперимента по сбору данных и принципы представления его результатов.....	63
2.2. Определение основных характеристик экспериментальных переменных и наблюдаемых параметров при предварительной статистической обработке мультимодальных данных на русском и китайском языках.....	72
2.3. Лингвистическое описание закономерностей невербальной компенсации речевых сбоев в русской спонтанной диалогической речи	82
2.4. Лингвистическое описание закономерностей невербальной компенсации речевых сбоев в китайской спонтанной диалогической речи	90
2.5. Оптимизация мультимодальных корпусов диалогической спонтанной речи на русском и китайском языках для машинного обучения	95
ВЫВОДЫ ПО ГЛАВЕ 2.....	115
ЗАКЛЮЧЕНИЕ	120
СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ.....	125

ПРИЛОЖЕНИЕ А.....	141
ПРИЛОЖЕНИЕ Б.....	142
ПРИЛОЖЕНИЕ В.....	144
ПРИЛОЖЕНИЕ Г.....	150

ВВЕДЕНИЕ

Появление и развитие электронно-вычислительных машин стимулировало новый виток сбора и обработки данных в электронном формате, возвестив о становлении самостоятельного лингвистического направления – корпусной лингвистики.

Основным фокусом работы корпусной лингвистики является разработка принципов создания и эффективного использования лингвистических корпусов в различных дисциплинах, включая лексикографию, переводоведение, лингводидактику, прикладную математику, компьютерные науки и пр. Последняя научная область, применяя результаты корпусных исследований, специализируется на системах распознавания речи, развитии автоматического перевода, генерации текстов, определении тональности в тексте и многом другом.

Быстрый темп развития и совершенствования современных технологий сегодня происходит в том числе по причине применения мультимодальных компьютерных архитектур, которые в отличие от одномодальных демонстрируют более продуктивные результаты, вследствие представления более разнородной, но вместе с тем согласованной информации. Переход к мультимодальным исследованиям свидетельствует об усложнении задач, стоящих перед учеными, и их стремлении приблизить машинные системы к выполнению работы, ранее доступной только человеку.

Вышеизложенное обуславливает актуальность мультимодального исследования и необходимость решения **научной проблемы**, которая состоит в построении мультимодальных корпусов и их адаптации под нужды машинного обучения.

Решение стоящей перед нами проблемы осложнено прежде всего тем, что на сегодняшний день практически отсутствует единый подход к представлению мультимодальных данных и их аннотированию, под которым подразумевается процесс расстановки меток/тегов к интересующим объектам.

В области искусственного интеллекта особое внимание уделяется двум направлениям, оба из которых сфокусированы на оптимизации, т.е. целенаправленной деятельности по получению наилучших из возможных результатов при работе над человеко-машинным взаимодействием и повышением общей «интеллектуальности» электронной системы.

Необходимость оптимизации мультимодального массива данных для определенных задач машинного обучения продиктована целым спектром ограничений в работе архитектуры интеллектуальных компьютерных систем. В рамках данной работы под одним из возможных ограничений мы подразумеваем проблемы автоматического распознавания речи, возникающие по причине присутствия помех в вокальной модальности. Спонтанная диалогическая речь изобилует подобными помехами, именуемыми лингвистами речевыми сбоями.

Мы полагаем, что продвинуться в решении поставленной научной проблемы можно, с одной стороны, путем анализа научных достижений специалистов по теории диалога, мультимодальной, жестовой, корпусной лингвистике, мультимодальному машинному обучению и комбинирования их подходов и методов для использования полученного теоретического конструкта при создании мультимодальных корпусов. С другой стороны, изучение взаимосвязи между вокальной и кинетической модальностями и выявление закономерностей в их использовании при реальном человеческом взаимодействии позволит не только выровнять мультимодальные данные (иными словами, согласовать информацию из разных модусов коммуникации), подготавливая их к машинному обучению, но и предложить модель обхода нежелательных помех в вербальном канале с лингвистической точки зрения.

Научная гипотеза настоящей работы заключается в том, что повышенная активация каналов кинетической модальности приходится на моменты возникновения речевых сбояв. При подтверждении предположения о том, что типы речевых сбояв имеют универсальную природу и стимулируют воспроизводимое жестовое поведение, как мы предполагаем,

возможна систематизация наблюдаемых закономерностей при невербальной компенсации речевых сбоев в виде мультимодальных схем, находящихся в составе отдельного коммуникативного события.

Целью работы является разработка мультимодальных корпусов диалогической спонтанной речи на русском и китайском языках и их оптимизация для машинного обучения.

Объектом исследования является диалогическая спонтанная речь в мультимодальном аспекте.

Предметом являются закономерности функционирования вокальной и кинетической модальностей в диалогической спонтанной речи.

В соответствии с поставленной целью необходимо решить ряд **задач**:

1) определить причины становления нового лингвистического направления – мультимодальной лингвистики и дать определение смежных понятий;

2) рассмотреть функциональные характеристики устной коммуникации, раскрыв особенности речевых сбоев и невербального компонента как важных факторов устной спонтанной речи;

3) исследовать ключевые подходы к аннотированию устной спонтанной речи и выявить прикладную ценность создания мультимодальных корпусов;

4) спроектировать дизайн эксперимента и обработать мультимодальные данные путем сплошной дискурсивной разметки;

5) систематизировать основные характеристики экспериментальных переменных, полученных при предварительной статистической обработке мультимодальных данных на двух языках с целью обнаружения ведущих параметров для оптимизации мультимодальных корпусов;

6) описать закономерности невербальной компенсации речевых сбоев в русской и китайской диалогической спонтанной речи;

7) разработать структуру и описать характеристики мультимодальных корпусов диалогической спонтанной речи; оптимизировать

корпусы путем сегментирования коммуникативных событий компенсации речевых сбоев и предложить вариант модели мультимодальной автоматизируемой последовательности обучения.

Материалом исследования послужили 16 аудиовидеозаписей на русском языке и 10 аудиовидеозаписей на китайском языке спонтанных диалогов, примерной длительностью 44 мин. и 21 мин., в объеме 159405 и 54386 аннотаций соответственно.

В эксперименте приняли участие 26 носителей русского языка и 20 носителей китайского языка.

Степень разработанности научной проблемы. Методология исследования сложилась под влиянием общей теории текста и дискурса, описанного в трудах Н.Д. Арутюновой, И.Н. Борисовой, В.Е. Гольдина, Г.В. Колшанского, М. Сэвилл-Труака и др.

Анализ отечественных и зарубежных авторов по теме магистерской диссертации показал, что ключевые положения формирующегося мультимодального подхода к анализу дискурса изложены в трудах Е.Е. Анисимовой, Ю.И. Детинко, А.А. Кибрика, Г. Кресса, Л.В. Куликовой, Т. ван Левена, Д. Макнейла, Л. Мондада, К. Мюллера, К. О'Халлорана, В.И. Подлесской, Х. Сакса, О.В. Фёдорова, Е.В. Чистовой и др.

В настоящее время мультимодальные исследования сосредоточиваются на изучении невербальной коммуникации, что также нашло отражение в трудах таких ученых, как Е.А. Гришиной, А. Кендона, Г.Е. Крейдлина, А.О. Литвиненко, Ю.В. Николаевой, Дж. Оберландера, М. Фостера, У. Фризена, П. Экмана и пр. Отметим также работы О.В. Магировской, Е.С. Привалихиной, Е.В. Прозоровой, Л.В. Куликовой, С.А. Шатохиной, выполненные в рамках невербальной коммуникации и посвященные одному из ее проявлений – жестовой речи глухих.

Значительный вклад в диалогическую лингвистику внесли работы М.М. Бахтина, Е.А. Земской, А.В. Михеева, Л.П. Якубинского. В том числе одному из проявлений разговорной речи – речевым сбоям – уделяют внимание

такие ученые, как Д. Журафски, Е.А. Земская, А.А. Кибрик, С. Нутебум, В.И. Подлеская, М. Столт, Ю. Чжао, Э. Щеглофф, Э.Б. Яковлева и др.

Разработке принципов аннотирования и построения корпусов посвящены труды С.Ю. Богдановой, Е.А. Гришиной, В.П. Захарова, А.А. Кибрика, М.В. Копотева, А.О. Литвиненко, Дж. Оберландера, В.И. Подлеской, В.А. Плунгяна, Т.И. Резниковой, О.В. Фёдоровой, М. Фостера и пр.

На исследовании и изложении основ мультимодального машинного обучения сосредоточивают свое внимание такие ученые, как А. Заде, П. Лян, М. Ма, С. Мерити, Л.Ф. Моренси, Р. Сочер, К. Сюн и др.

В настоящей работе использованы такие **методы**, как 1) общенаучные методы: количественный метод для выявления частотности проявлений в разных модусах коммуникации, качественный метод для интерпретации результатов статистического подсчета; 2) лингвистические методы: мультимодальный дискурс-анализ, разговорный анализ, лингвостилистический анализ дискурса, сопоставительный анализ, метод лингвистического аннотирования мультимодальных данных.

Магистерская диссертация состоит из введения, двух глав, заключения, списка использованной литературы и двух приложений.

Во Введении обосновывается актуальность и новизна работы, определяются цель, задачи, объект, предмет, методы и материалы исследования.

В Главе 1 рассматривается история становления нового лингвистического направления – мультимодальной лингвистики, приводятся основные функциональные характеристики устной спонтанной речи, исследуются подходы к аннотированию мультимодальных данных, выясняется прикладная польза создания мультимодальных корпусов и пути их оптимизации для машинного обучения.

В Главе 2 прописан принцип организации лингвистического эксперимента, проанализированы результаты базовых статистических

подсчетов проаннотированного массива данных русской и китайской спонтанной диалогической речи, описаны закономерности невербальной компенсации речевых сбоев в спонтанной коммуникации, заданы основные характеристики мультимодальных корпусов диалогической спонтанной речи на двух языках, представлен подход к оптимизации корпусов для машинного обучения путем сегментирования на коммуникативные события и предложена модель автоматизированной последовательности мультимодального машинного обучения при возникновении помех в потоке вербальной информации.

Заключение содержит в себе основные выводы и перспективы дальнейшего научного исследования.

Приложение А представляет техническое задание, выданное компаний Promobot.

Приложение Б включает образец согласия на участие в лингвистическом эксперименте на русском и китайском языках.

Приложение В представляет образцы опорных карточек на русском и китайском языках, использованных для лингвистического эксперимента.

Приложение Г содержит список актуализированных в ходе разметки жестов, представленных в соответствии с классификацией П. Экмана и У. Фризена. В приложении также хранится перечень симптоматических жестов и жестов нейтрального положения, не вошедших в классификацию, но упоминаемых в данной работе.

Магистерская выполнена по заказу и при активном участии компании Promobot.

Апробация работы. Теоретические положения и результаты практического исследования представлены в докладах на научных конференциях XIV Международной научно-практической конференции молодых исследователей «Язык, дискурс, (интер)культура в коммуникативном пространстве человека» (Красноярск, 2022) и XV Международной научно-практической конференции молодых

исследователей «Язык, дискурс, (интер)культура в коммуникативном пространстве человека» (Красноярск, 2023).

Отдельные этапы исследования были отражены в статье ВАК: Ухоботова С.В. Мультимодальные паттерны компенсации речевых сбоев (на материале студенческого диалогического дискурса) // Вестник Томского государственного педагогического университета. 2023. Вып. 1 (225). С. 9–18.

ГЛАВА 1. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ПОСТРОЕНИЯ МУЛЬТИМОДАЛЬНЫХ КОРПУСОВ ДИАЛОГИЧЕСКОЙ СПОНТАННОЙ РЕЧИ

1.1. Становление мультимодальной лингвистики как продолжения теории дискурс-анализа

Зародившись в русле философии, лингвистика, начиная с системно-структурной парадигмы знаний, стремилась «освободиться от уз» смежных с ней дисциплин, попутно продолжая заимствовать их методологический инструментарий (например, методы статистического подсчета, элементы теории вероятности, анкетирование, опрос и т.д.).

С провозглашением антропоцентрической парадигмы научных знаний, согласно которой «<мерой> всех вещей» является человек, начался масштабный экспансионизм, предполагающий проведение исследований на стыке научных дисциплин [Кубрякова, 1995: 212]. Так, в лингвистике сформировались новые научные направления, а именно социолингвистика, этнолингвистика, психолингвистика, когнитивная лингвистика, компьютерная лингвистика и т.д.

Благодаря междисциплинарному подходу лингвистика постепенно стала выходить за рамки анализа лишь текстовых произведений. Сформированный в русле лингвистики текста дискурсивный подход приобрел самостоятельный статус. Интерес ученых сосредоточился на дискурсе, который представляет «сложное коммуникативное явление, включающее, кроме текста, еще и экстралингвистические факторы (знания о мире, установки, цели адресанта), необходимые для понимания текста» [ван Дейк, 1989: 7]. В это же время ученые стали прописывать первые принципы порождения устного дискурса, познакомив научный мир с теорией речевых актов [Остин, 1986; Серль, 1986]. Так, акцент исследований постепенно перемещается с анализа текстовых компонентов, на толкование интенций

(скрытых смыслов) коммуникантов в устной речи. Несмотря на такой междисциплинарный характер лингвистики, исследования, сосредоточившиеся на рассмотрении одной модальности – вербальной, имели ограничения при конструировании более полной модели человеческой коммуникации.

На сегодняшний день существует множество дисциплин, теоретические подходы которых раскрывают разные аспекты мультимодальной природы человека и окружающего его мира [Bezemer, Jewitt, 2010]. Например, впервые о мультимодальности заговорили психологи и психотерапевты, обнаружив, что комплексный подход, включающий использование когнитивной, телесно-ориентированной, семейной и других типов психотерапии, позволяет лучше проработать конфликты, когнитивные и поведенческие проблемы и сохранить более продолжительный эффект. Подобные исследования проводились Д.Н. Оудсхоорном [Оудсхоорн, 1993], А.А. Лазарусом [Lazarus, 1981], В.А. Ананьевым [Ананьев, 1995]. Такой многомерный подход в психологии стимулировал в том числе лингвистические исследования, т.к. было очевидно, что язык является важным элементом в конструировании картины мира, а именно «<отражения> в психике человека предметной окружающей действительности» [Леонтьев, 1993: 18; Красных, 2003: 38].

Своим развитием в лингвистике мультимодальность обязана социологам и социолингвистам. Именно они были первыми, кто, приняв во внимание тот факт, что речь и письмо являются центральными модусами в коммуникации, сосредоточились на социальном и ситуативном использовании языка. Под влиянием социолингвистических исследований сложились такие новые направления, как конверсационный анализ [Sacks, 1972; Schegloff, 1996], интерактивная социология [Goffman, 1981], интерактивная социолингвистика [Gumperz, 1999], лингвистическая антропология [Duranti, 2001], микроэтнография [Erickson, 2004], лингвистическая этнография [Creese, 2008]. Впервые популяризируется метод

видео и фотофиксации человеческих интеракций для последующего разбора и анализа [Knoblauch et al., 2006].

С лингвистической точки зрения изучение мультимодальности прошло свой путь от более традиционных направлений, таких как креализованные тексты, невербальная семиотика [Крейдлин, 2002], [Анисимова, 2003], [Детинко, Куликова, 2017], [Chernyavskaya, Kulikova, 2020] к рассмотрению менее исследуемых областей науки, например, жестовых языков [Прозорова, 2007], [Магировская, Привалихина, 2019], [Kulikova, Schatokhina, 2020].

Сегодня потребность в мультимодальных исследованиях носит в том числе прикладной характер. Как подчеркивает отечественный ученый А.А. Кибрик в своей статье «Мультимодальная лингвистика», представители прикладных дисциплин уже давно осознали практическую значимость мультимодальных моделей коммуникации и готовы применять разработки в собственных технических устройствах, хотя теоретический интерес к мультимодальным исследованиям в лингвистике только набирает оборот [Кибрик, 2010]. Современный мультимодальный подход предполагает выявление закономерностей работы разных модусов (каналов) в процессе устной коммуникации для последующей «алгоритмизации и компьютерной реализации» [Крейдлин, 2014: 101].

В основе мультимодальных исследований лежит идея взаимосвязи всех известных модусов коммуникации – вербального, жестового и фонетического (просодического). Если исследователь ставит своей задачей изучение устной коммуникации, то прежде всего необходимо учитывать, что поступление информации и ее преобразование в знания будет проходить по разным каналам восприятия. Именно в этом заключается суть мультимодальности в узком смысле этого слова [Гришина, 2017].

Современная лингвистика не может предложить общепринятого универсального определения понятия «мультимодальность», вкладывая в него разные характеристики в зависимости от научной области. В рамках нашего исследования мы принимаем за основу определение, данное отечественным

ученым Г.Е. Крейдлиным, согласно которому под мультимодальностью понимается «описание общих законов и конкретных правил взаимодействия в устном коммуникативном акте вербальных и невербальных знаков» [Крейдлин, 2014: 101].

Отсутствие единообразия в применении понятия «мультимодальность» также порождает терминологическую омонимию. Например, в российском научном пространстве можно наблюдать отождествление понятий «поликодовость» («поликодовый текст»), «креолизация» («креализованный текст») и «мультимодальность» («мультимодальный текст»). Все перечисленные термины включают рассмотрение вербальных и невербальных компонентов, однако у каждого из них акценты расставлены по-разному. Так, креолизация сосредоточивается на интеграции разных систем, поликодовость – на языке как коде. Сам же феномен мультимодальности акцентирует внимание на способах восприятия информации, полученной коммуникантом из разных каналов (модусов). При таком изучении человеческой коммуникации, речи не всегда принадлежит абсолютно главенствующая роль. Несмотря на сравнительную совместимость трех терминов, принципиальной является мысль о том, что «креолизация», «поликодовость» и «мультимодальность» противопоставлены на основе «принципа вербальной центрированности/децентрированности в коммуникации» [Загидуллина, 2019: 183].

С развитием дисплейной коммуникации и передачей всех возможных элементов сообщения через Интернет, ученые начинают обращаться еще к одному термину – «мультимедийность», подразумевая под медиумом текст, изображение, звук и приравнивая его к понятию мультимодальности. Суть обоих терминов состоит в изучении комплекса способов передачи и восприятия информации, с той разницей, что «мультимедийность» сфокусирована на визуальной составляющей (например, «визуальная грамматика» [Kress, 1996]) и применима к интернет-пространству.

Безусловно, мультимодальность должна включать разные модусы, в том числе тактильные, вкусовые и пр. Однако в связи с ограниченностью технологий современного мира и отсутствием всесторонне разработанного методологического аппарата, лингвистам доступны к рассмотрению две модальности: вокальная (вербально-просодический канал) и кинетическая (окуломоторный канал, канал мануальных жестов, канал цефалических жестов, канал жестов лица (мимика), канал движения корпуса) [Кибрик, 2010: 135].

В нашем исследовании мы изучаем как собственно вербальный компонент, так и невербальный компонент, реализованные в двух доступных типах модальностей – *вокальной и кинетической*, что позволит нам наиболее полно проанализировать принципы построения устной спонтанной коммуникации.

Таким образом, мы можем говорить о том, что лингвистика, благодаря своей полипарадигмальной природе, постепенно выходит на другой уровень, включая в себя новые методы смежных научных дисциплин. Современное развитие технологий в том числе мотивирует лингвистов искать инновационные пути развития, расширять круг научных интересов. Становится очевиден факт качественной проработки методов при изучении вербального компонента, вместе с тем приведший к узконаправленному подходу в отношении устной коммуникации, основным характеристикам которой как раз и посвящен следующий параграф.

1.2. Функциональные характеристики устной коммуникации

На сегодняшний день количество исследований, посвященных письменной форме, значительно больше, чем устной форме естественного языка [Подлеская, Кибрик, 2007]. Причина маргинальности проблем устной коммуникации очевидно складывается из эфемерности и трудности обработки речи, что в свою очередь противопоставляется относительной логичности, плавности изложения и воспроизводимости письменных текстов. Лингвистика

текста задумывалась как направление, изучающее литературные тексты и критерии связности текста в разных его ипостасях. Вместе с тем, предполагалось, что сформированный в русле лингвистики текста дискурс-анализ позволит рассматривать устную форму коммуникации.

Классическое разделение языка и речи времен Ф. де Соссюра, появление терминов «устная/письменная речь», «устный/письменный текст», отсутствие четких критериев разграничения приведенных понятий – все это повлияло на фокус языковых исследований. Лингвисты стали больше обращаться к анализу фиксированных на бумаге произведений, сознательно или бессознательно оставляя устную форму общения в стороне. С переходом к антропоцентрической парадигме научного знания, ученые стали расширять и прорабатывать определение понятия «дискурс», понимая, что именно он смог бы связать устную и письменную формы коммуникации и по-прежнему бы вписывался в традиционные лингвистические методы исследований. Предметом дискурсивных исследований стали произведения речетворческого процесса – те же самые тексты, но уже с учетом экстралингвистических факторов (см. определение «дискурс – речь, погруженная в жизнь» [Арутюнова, 1990: 136–137]).

Стоит отметить, что устная форма естественного языка эволюционно появилась раньше, чем письменная, о чем свидетельствует наличие в третьем десятилетии XXI века бесписьменных малочисленных народов. Только в РФ насчитывается порядка 36 языков коренных малочисленных народов Севера, Сибири и Дальнего Востока [Бигуаа, Семенова, 2020: 36].

На сегодняшний день также заметен рост интереса к анализу устной речи, продиктованный стремлением создать новые или усовершенствовать уже имеющиеся системы автоматического распознавания и синтеза речи. Устная форма коммуникации представляется в данном случае удобным способом взаимодействия человека и компьютера. Использование такого типа коммуникации также обладает важным преимуществом над письменным, т.к. предполагается, что в будущем люди не будут задействовать дополнительные

инструменты и каналы связи, такие как мышь, клавиатура, экран монитора, освещение и пр. Именно по этой причине внимание специалистов по искусственному интеллекту, программистов и лингвистов постепенно переключается в область прикладных фонетических исследований, которые бы способствовали созданию речевых технологий с устным вводом и выводом информации, применяемых в машинных интерфейсах, информационных и банковских службах, а также устройствах, предназначенных для работы в экстремальных или опасных для человека условиях [Кодзасов, Кривнова, 2001: 514–515].

Рассмотрением особенностей речевой деятельности, описанием её основных видов и характеристик уже давно заняты ученые таких областей, как психология, лингвистика, социология, философия и пр. Ученые, проводящие исследования на стыке двух или нескольких дисциплин (например, психолингвистика, социолингвистика и т.д.), также достигли продуктивных результатов. Современные ученые и специалисты сходятся во мнении о разделении речи на монологическую и диалогическую.

С точки зрения лингвистики, речь является организованной формой «<словесного воспроизведения> знаний», которое «требует всякий раз словесной импровизации, форма и объем которой меняются в значительных пределах в зависимости от ситуативных условий и стратегии говорящего» [Кацнельсон, 1972: 111].

Психолингвисты определяют устную монологическую речь как тип речевой деятельности, опирающийся на «единый замысел, смысловую программу, которая заставляет говорящего контролировать формирование высказывания, не позволяя ему по ассоциативному признаку соскользнуть в сторону» [Горелов, Седов, 2001: 72].

В отличие от монологической речи, которая является объектом исследований самых разных направлений, идея диалогичности языка, восходящая к положениям М.М. Бахтина, и изучение ее главного проявления – диалогической речи – главным образом развивается лингводидактиками с

позиции организации процесса интеракции учеников и студентов при изучении иностранных языков. Диалог, с точки зрения, диалогической лингвистики, представляет «аспект в исследовании языка как системы, как языкового материала (как текста), как языковой способности и как речевой деятельности» [Богачанова и др., 2019: 5].

В соответствии с задачами данной работы нам подходит более общее определение, данное Л.П. Якубинским. Ученый относит диалогическую речь к «перемежающимся формам взаимодействий, подразумевающим сравнительно быструю смену акций и реакций взаимодействующих индивидов» [Якубинский, 1986: 25]. Ученый также неоднократно настаивает на рассмотрении непосредственной формы взаимодействия, т.е. «лицом к лицу», при которой речь характеризуется «непосредственным, в зрительном и слуховом отношении, восприятием высказывающего лица» [Якубинский, 1986: 25].

Подчеркивая важность разговорного, т.е. устного воспроизведения в диалоге (ср. внутренний диалог в психологии), отечественный ученый М.В. Матвеева приводит термин «разговорный диалог», который «представляет собой функциональный объект, состоящий из ряда реплик коммуникантов, каждая из которых обусловлена ситуативно и влияет на следующую реплику речевого партнера» [Матвеева, 2018: 92].

Е.А. Земская также подчеркивает, что разговорная речь предполагает вовлечение в коммуникацию как говорящего, так и слушающего (признак диалогичности), преобладание устного конструирования содержания сообщения. Разговорная речь характеризуется спонтанностью речевых актов, «сплавленностью речи с ситуацией общения», а также неофициальностью отношений коммуникантов [Земская, 2004: 290–291]. Неформальность отношений говорящего и слушающего, по нашему мнению, всё-таки является спорным критерием. Главным аргументом в данном случае является фактор симметричности-несимметричности, который включает «тождество-нетождество социальных признаков партнеров по коммуникации вне

оценочной их характеристики по шкале “выше-ниже”» [Китайгородская, Розанова, 2003: 106–107]. Предполагаем, что с лингвистической точки зрения официальность и неофициальность также определяется использованием участниками коммуникации особых речевых клише, выбором речевого жанра, количеством речевых сбоев.

Если отвлечься от социолингвистических факторов, то стоит также отметить некоторые другие характеристики ведения диалога или полилога. При анализе диалогической и полилогической речи Е.А. Гришина предлагает использовать критерий равноправность-неравноправность. Равноправный диалог предполагает попеременную смену ролей в разговоре. Неравноправным считается такой диалог, в котором одной из сторон коммуникации задается тема разговора, провоцирующая остальных на ее дальнейшее обсуждение. Такой стороной коммуникации, как правило, выступают модераторы (интервьюеры). Другим важным критерием при анализе является сопоставимость-несопоставимость, который как раз коррелирует с симметричностью и несимметричностью по принципу социальной значимости участников коммуникации [Гришина, 2005].

На основе критериев равноправности-неравноправности и сопоставимости-несопоставимости можно прийти к следующим выводам: монолог – один автор, равноправный диалог – два автора, неравноправный диалог с несопоставимым модератором – один автор, неравноправный диалог с сопоставимым модератором – два автора [Там же, 2005]. Фокус настоящей работы представляют диалоги равноправного типа.

Важно отметить, что при диалогическом взаимодействии адресат не является простым пассивным наблюдателем процесса коммуникации, а также наряду с адресантом активно принимает участие. В ходе восприятия и интерпретации адресатом речи адресанта может настать такой момент, когда сказанная вторым реплика будет противоречить пресуппозициям первого реагирующего партнера по коммуникации или требовать определенной вербальной/невербальной реакции, что может стать причиной использования

инициирующей реплики в стремлении поддержать/направить коммуникацию. В связи с чем, вслед за Е.В. Борисовой мы считаем необходимым ввести для дальнейшего лингвистического описания и систематизации материала два понятия – «инициирующий» и «реагирующий» коммуниканты. Под «инициирующим коммуникантом» понимается активный участник, непосредственно произносящий реплики с невербальным сопровождением. «Реагирующий коммуникант» представляет собой активного партнера, реагирующего на вербальном или невербальном уровнях на реплики инициирующего [Борисова, 2003: 5].

Согласно области применения устная речь также подразделяется на два больших класса речевых жанров: публичная и непубличная речь. Принципиальным фактором, отличающим один жанр от другого, является наличие слушателей и фиксация на звуко- или видеоносителях [Гришина, 2005: 103]. Согласившись с первым критерием, мы тем не менее считаем, что запись сама по себе не всегда может служить надежным ориентиром при выделении публичной и непубличной форм устной речи. Создание мультимодального корпуса предполагает ведение аудио и видеозаписей для дальнейшего аннотирования, но сам факт съемки не говорит о публичности мероприятия. Предполагаем, что Е.А. Гришина добавила фиксацию речи как критерий, основываясь на материале, используемом при создании мультимедийного русского корпуса, а именно студийные записи новостей, выступлений, кадры фильмов, сериалов и пр.

С публичностью/непубличностью устной речи также связан критерий подготовленности. Публичная устная речь обычно ассоциируется с подготовленным выступлением перед аудиторией, а непубличная речь – с неподготовленным. Хотя у этого умозаключения есть и свои исключения, например, спортивное интервью при онлайн-трансляции репортажа вряд ли можно назвать подготовленным с точки зрения выбора языковых средств и микротем диалогов.

Во всем многообразии трактовок понятия «спонтанная речь», мы придерживаемся определения, данного А.В. Михеевым. По мнению ученого, спонтанная речь соотносится с «естественной формой выражения мысли, обусловленной исключительно внутренней мотивацией говорящего, т.е. человеком, свободно владеющим языком» [Михеев, 1986: 40].

Н.Н. Рудык выделяет следующие характеристики спонтанной речи как самостоятельного вида коммуникативной деятельности:

- 1) неподготовленность, которая приводит к формированию новых комбинаций смыслов при использовании известных языковых единиц;
- 2) мотивированность, проявляющаяся в стимуляции участников коммуникации к началу вербального общения в ситуации неожиданности;
- 3) инициативность, предполагающая готовность и желание выразить собственные мысли;
- 4) эмоциональность, которая проявляется в умении оценивать объекты внеязыковой действительности на основе чувственного опыта;
- 5) экспрессивность как стремление говорящего воздействовать на слушающего с помощью речи, сопровождаемой жестами и мимикой [Рудык, 2010: 188–190].

Несмотря на достижения в области анализа устной формы коммуникации, неподготовленная, т.е. спонтанная диалогическая речь (*non task-oriented spontaneous dialogue*) по-прежнему вызывает трудности, т.к. в отличие от подготовленной речи первая не поддается четкому структурированию [Lopes, Hemmingsson, Astrand, 2018].

Как видно из приведенных выше критериев, такой многомерный феномен как устная речь не укладывается в строго упорядоченную систему, что также представляет сложность для ученых и специалистов. Устная форма является первичной по отношению к письменной, однако, в связи со сложностями ее упорядочения и фиксирования, а также ее представленности в научных публикациях, все еще уступает в пользу лингвистического анализа письменных текстов. Вследствие чего, устная спонтанная диалогическая речь

представляет собой, по нашему мнению, весьма актуальный предмет исследования.

1.2.1. Особенности речевых сбоев в устной спонтанной речи

В начале п. 1.2. мы упоминали, что устная коммуникация предстает в двух основных формах – подготовленная и неподготовленная (спонтанная) речь. Ученые установили, что, даже будучи подготовленной, речь выступающего всё равно содержит правки, паузы, колебания, хотя их количество значительно ниже, чем в неподготовленной [Подлесская, Кибрик, 2007]. Спонтанная речь не линейна, нить повествования не может быть заведомо известна ни говорящему, ни слушающему. По причине отсутствия конкретных прогнозов развития речевой коммуникации, человеческий мозг вынужден выдавать информацию порционно, анализируя ход коммуникативных событий в режиме реального времени. Такая мозговая деятельность требует больших когнитивных затрат на поиск и вербализацию необходимых языковых единиц, вследствие чего нередко происходят ситуации потери беглости речи (speech disfluencies) или речевые сбои.

На сегодняшний день принципы порождения и восприятия сбоев в речевом потоке, возрастные и гендерные характеристики участников коммуникации, особенности сбоев у больных афазией, методика обучения спонтанной речи на иностранном языке, создание систем распознавания речи и многое другое становится предметной областью исследований, результаты которых публикуются в сборниках Международного семинара по речевым сбоям в спонтанной речи (The Disfluency in Spontaneous Speech, DISS). Работа международного симпозиума, уже на протяжении 23 лет посвященная одному объекту – речевым сбоям, что указывает на актуальность и важность изучения данного феномена.

Под влиянием сложившихся критериев оценки связности и правильности изложения письменных текстов некоторые ученые, деятели

культуры и другие специалисты отождествляют речевые сбои с непростительными промахами и погрешностями, нарушающими благозвучие речи [Подлеская, Кибрик, 2007]. По этой причине некоторые из них даже призывают «очистить» речь и избавиться от пауз, хезитаций, дискурсивных слов (более известных как «слова-паразиты») и пр. Если в подготовленной речи проблема с обеспечением чистоты воспроизведения языковых единиц более-менее решается, то спонтанная речь неизбежно будет содержать речевые сбои, т.к. именно они дают возможность «выиграть время», пока происходит упаковка смыслов и их вербализация.

Произнесение речи и мыслеобразование тесно взаимосвязаны, но сами процессы отнюдь не синхронны. Между формированием мысли и облечением ее в звуковую оболочку должно пройти время, которое в потоке речи компенсируется использованием пауз, хезитаций, удлинений и др. Как пишет Э.Б. Яковлева, «хезитации – естественный феномен живой устной речи, обусловленный ментальными процессами. Избавляться от них так же нелепо, как и избавляться от способности мыслить» [Яковлева, 2021: 5]. От себя также добавим, что речевые сбои не ограничены лишь хезитациями и включают другие внеязыковые/языковые маркеры, однако идея об их влиянии на когнитивные принципы речепорождения и речепонимания является одной из ведущих в психолингвистических исследованиях конца XX начала XXI веков (см. работы Ф.Г. Эйслер [Eisler, 1968], Ю.В. Красикова, [Красиков, 1980], В. Левелта [Levelt, 1983], Г. Делла [Dell, 1986], Т.В. Ахутиной [Ахутина, 1989], А. Постма [Postma, 2000], А.А. Леонтьева [Леонтьев, 2003] и т.д.).

Хотя появление речевых сбоев в спонтанной речи неизбежно, как слушающие мы подсознательно понимаем, что количество помех в потоке речи здорового человека должно быть оптимальным, в противном случае их переизбыток может привести к коммуникативным неудачам. Тем не менее вплоть до сегодняшнего дня не представлены точные данные о допустимом объеме перестроений в речи без серьезных нарушений понимания общего смысла. Подтверждением стремления специалистов и простых обывателей

избавиться от нежелательных сбоев в речи являются научные работы и журнальные статьи в области риторики, лингвистики, социолингвистики, когнитивистики и т.д.

Если в психо- и социолингвистических исследованиях речевые сбои – ключ к пониманию мыслительных процессов индивида и / или социума, то для компьютерных лингвистов данный феномен представляет серьезное препятствие в создании автоматических систем распознавания звучащей речи (automatic speech recognition – ASR). Традиционно принято выделять три крупные сферы применения систем распознавания речи: 1) голосовое управление – распознавание отдельных команд (коммерческое применение), 2) поисковые системы (выделение ключевых слов в речевом потоке) и 3) дикторозависимое / дикторонезависимое распознавание слитной речи на базе большого словаря (speaker (in)dependent large vocabulary continuous speech recognition) [Верходанова, Карпов, 2012].

Согласно проведенным исследованиям, подготовленная речь распознается современными системами с точностью более 85%, хотя для достижения подобных высоких показателей требуется предварительное обучение системы на большом массиве данных корпусов звучащей речи. Распознавание спонтанной речи вместе с тем осложнено наличием речевых сбоев и паралингвистическими явлениями. К последним можно отнести кашель, причмокивание, хрипы, вздохи, выдохи, смех и пр. По результатам анализа точности обработки живой спонтанной речи системой, разработанной в лаборатории LIMSI, речевые сбои стали причиной ошибок в почти 13% случаев [Adda-Decker et al., 2003]. Ранее другие западные ученые провели исследование, которое показало, что оговорки в 100% случаев повлияли на ошибку в автоматическом распознавании звучащей речи, удлинения отдельных звуков – в 81% случаев, а паузы – в 92% соответственно [Butzberger et al., 1992].

Пристальное внимание к речевым сбоям ученых разных областей научного знания тем не менее пока не повлияло на формирование

общепринятого терминологического аппарата. Так, в зарубежных исследованиях встречаются следующие термины, описывающие данное лингвистическое явление: “disfluency”, “discontinuity”, “flustered speech”, “hesitation”, “speech management”, “turnholding devices”, “non-fluency” и т.д. [Верходанова, Карпов, 2012: 10]. В отечественных трудах по лингвистике также нет единства в понимании данного феномена, вследствие чего встречаются такие понятия, как «коммуникативная неудача» [Воронина, 2003], «речевое затруднение» [Александрова, 2010], «речевая ошибка» [Степанова, 2012], «речевой сбой» [Подлеская, Кибрик, 2007], [Ахапкина, 2013] и т.д. Так как в определениях большинства из перечисленных понятий существует поясняющее слово «сбой», мы склонны считать «речевой сбой» более универсальным термином и в рамках данной работы будем использовать именно его.

Речевые сбои включают целый спектр нарушений в процессе построения спонтанной речи, а именно заполненные и незаполненные паузы, оговорки, самоперебивания (самокоррекции), удлинения, повторы слов или отдельных звуков и пр. Прерывание сообщения происходит по классическому сценарию. Так, при возникновении речевого сбоя говорящий приостанавливает поток речи в так называемом «моменте прерывания» (interruption point) для исправления того, что было сказано [Nooteboom, Quene, 2008: 838]. Помимо описания функциональных характеристик сбоев, нередко обращаются к временным показателям. Репарандумом (reparandum) называется отрезок речи, который предстоит отредактировать, а исправление (repair) – это уже внесенные корректировки. В некоторых случаях исправления могут не содержать редактирующего компонента, как в случае с обрывом слова и его достраиванием (restart) [Gósy, 2019: 3].

Речевые сбои могут присутствовать в дискурсе человека по ряду лингвистических и экстралингвистических причин. С точки зрения языковой организации, они связаны с проблемами в планировании самого дискурса, включая продумывание очередного коммуникативного хода. Немаловажным

также является психоэмоциональное состояние коммуниканта, его готовность/неготовность к интеракции.

В широком смысле речевые сбои можно разделить на два больших класса. К первому классу относятся перерывы, именуемые паузами колебаний, паузами хезитаций или просто хезитациями. Наличие пауз свидетельствует о необходимости говорящим времени на обдумывание, планирование или перепланирование следующей «порции» информации. Основной функцией пауз колебаний является сегментирование потока речи. «Паузы на границах отрезков, образующих интонационное и семантическое единство, могут быть связаны как с членением потока речи, так и с хезитацией, а паузы внутри таких отрезков имеют, как правило, хезитационную природу» [Подлеская, Кибрик, 2007: 3]. Хезитации могут быть воплощены в абсолютных паузах, заполненных паузах, удлинении отдельных звуков, а также в лексических маркерах.

Среди лексических маркеров также стоит упомянуть особый тип хезитации – «маркеры препаративной подстановки» (встречается термин «препаративная замена») или placeholders [Там же, 2007: 9]. Суть данного явления заключается в использовании говорящим слова-заместителя до тех пор, пока нужное слово не «придет на ум». В зависимости от морфологической и грамматической структуры языка существуют разные виды маркеров препаративной замены, однако более универсальными являются указательные и вопросительные местоимения (*этот, тот, 这个 (это), 那个 (то)*), обобщенные слова (*вещь, 玩意 (штуковина), 东西 (вещь)*) и местоименные выражения (*это самое как его*). Было установлено, что в русском языке, принадлежащем к флективному типу, встречаются как согласуемые, так и несогласуемые маркеры препаративной замены. Согласуемый тип предполагает замещение с использованием слов-заместителей в одном роде, падеже, числе, что и необходимое задумываемое слово. В.И. Подлеская приводит следующий пример: «...мне дали...(0.6) \э́тот || ‘’(0.4) типа /медали,».

Предположительно, говорящая задумывала слово мужского рода (орден), но так и не вспомнив его, заменила на синонимичную конструкцию (типа медали) [Podlesskaya, 2010]. Несогласуемый тип маркеров включает как именные, так и глагольные группы. В ходе исследования было также обнаружено, что существуют примеры с маркерами указательных местоимений, к которым присоединяется глагольный префикс, как например в следующей клаузе «...(0,8) И он || он **приэто**= || не /привязан» [Подлеская, Кибрик, 2007: 11].

Второй класс речевых сбоев, в отличие от первого, подразумевает перестроения или изменения речевых отрезков и включает самоисправления (в других работах явление может именоваться «самокоррекциями»). Исправления говорящим самого себя происходят в тот момент, когда он решил, что порожденный им отрезок информации не отвечает коммуникативным задачам (неточная или неверная номинация, необходимые уточнения, изменение цепочки событий и т.д.). В этом случае говорящий заменяет полностью забракованный фрагмент на другой или повторяет при необходимости отдельные части из забракованного отрезка речи, добавляя новые фрагменты [Там же, 2007].

В отечественной лингвистике предложено выделять два основных режима самокоррекций – собственно коррекция (онлайн коррекция) и ретроспективная коррекция (офлайн коррекция) или редактирование. В онлайн режиме самоисправление говорящим происходит непосредственно в момент обнаружения проблемы. Как правило, при онлайн коррекциях появляется заполненная или незаполненная пауза, после которой следует замена забракованного фрагмента на новый или корректировка последовательности упоминания событий. Онлайн типу коррекций в большей степени свойственны обрывы слов (сильный или слабый фальстарт). Офлайн коррекция предполагает редактирование постфактум, т.е. работу с уже вербализированным отрезком. Чаще всего корректировки происходят в плане уточнения или исправления определенных фактов. Было установлено, что

в русском языке ретроспективные коррекции происходят обычно с использованием синтаксической конструкции «не X, (а) Y» [Подлеская, Кибрик, 2007].

Из примеров, приведенных выше видно, что инициирующий говорящий осознает моменты нарушений речевого потока и стремится восполнить, компенсировать образовавшийся недостаток путем проведения разного рода репаратур. Тем не менее, подобное восполнение в диалогическом взаимодействии может проводиться не только на внутреннем (индивидуальном), но и на внешнем (межличностном) уровне. Последний в свою очередь выражается в виде «наложений» речи (overlap, overlapping, overlapping talk или overlapping speech).

Несмотря на очевидное присутствие наложений в процессе интеракции, в отечественной научной литературе они, как объект исследования, представлены крайне редко. В зарубежных трудах данный феномен рассматривается в основном для формирования механизмов и принципов эффективного коммуникативного общения, но не дается четкое определение этому понятию. Так, в работах (см. [Schegloff, 2000], [Stolt, 2008], [Abbas, 2020] и т.д.) приводится следующее объяснение: наложение речи подразумевает, что второй собеседник активизируется в тот момент, когда инициирующий собеседник еще не закончил свой коммуникативный ход.

Вступление второго коммуниканта в контакт, не дожидаясь окончания коммуникативного хода, может свидетельствовать о «погружении» в тему, реальном проживании человеком ситуации, его раздражении, недовольстве по отношению к сказанному. Подобное поведение реагирующего коммуниканта также может быть связано с предугадыванием/дистраиванием мыслей своего партнера по коммуникации, в особенности в моменты возникших речевых затруднений. Такая активная ответная реакция представляет *соперничающий тип наложения речи* (competitive overlap) [Stolt, 2008].

Изначально, явление наложения речи рассматривалось лингвистами как неэффективное или даже деструктивное с точки зрения организации процесса

коммуникации. Разговаривать поочередно и не перебивать собеседника считается одним из правил коммуникации. Ему стараются придерживаться во всех смыслах, «вплоть до того, что одной из первых вещей, которой учат маленьких детей, является ожидание своей очереди в процессе разговора» [Stolt, 2008: 6]. Позднее в оппозицию к соперничающему типу наложений речи стали приводить несоперничающий тип (*non-competitive overlap*), направленный на кооперативное общение участников коммуникации [Там же, 2008]. Для поддержания лаконичности научной работы мы предлагаем именовать его *кооперативным типом наложения речи*.

Э. Щеглофф выделяет в своей работе четыре вида наложений речи. По нашему мнению, представленная выше классификация представляет оба описанных выше типа – соперничающий и кооперативный:

1) пограничное прерывание (*terminal overlap*) – говорящий начинает с того места, где как ему показалось, закончил свой коммуникативный ход предыдущий коммуникант;

2) продолжение (*continuers*) – соревновательный стиль ведения диалога, т.к. реагирующий коммуникант осознает, что иницирующий еще не окончил свою мысль, находится в середине повествования, но тем не менее вставляет в свою речь долексические маркеры типа “uh”, “huh”, “mm hm”;

3) условная возможность передачи слова (*conditional access to the turn*) – ситуация, при которой говорящий приглашает другого участника общения вступить в вербальную коммуникацию. Примерами такой смены инициативы является приглашение к совместному поиску искомого слова, которое ведущий коммуникант не может вербализовать, или к совместному построению высказывания, начатого одним коммуникантом и продолженного другим;

4) хоровая речь (*chordal / choral talk*) – пример кооперативного ведения диалога, который предполагает одновременность произнесения высказываний, а не один за другим. К такому типу относятся коллективные приветствия, прощания, поздравления и пр. [Schegloff, 2000: 5–6].

Если Э. Щеглофф исследовал само восприятие наложений речи и способы преодоления коммуникативного затруднения участниками диалогов/полилогов, то группа французских ученых (см. [Adda et al., 2007]) сосредоточилась на выявлении закономерностей между наличием наложений речи и речевых сбоев в политическом дискурсе. В широком смысле наложение представлено двумя крайностями: «*обратная связь*» (backchannel) и «*смена ролей*» (turn-taking) [Adda et al., 2007].

Отметим, что смена ролей (turn-taking) при наложении представлена очевидной «*инициативной сменой*» (turn request) и «*вовлеченной сменой*» (anticipated turn taking). В отличие от инициативы смены ролей, подразумевающей попытку захватить коммуникативный ход путем прерывания говорящего по ряду определенных причин, вовлеченная смена наступает в тот момент, когда реагирующим коммуникантом были замечены признаки завершения речевого акта (например, нисходящая интонация) [Adda et al., 2007]. По нашим наблюдениям вовлеченная смена может сформироваться и по инициативе самого иницилирующего говорящего, когда тот не может восполнить образовавшийся пробел в собственном речевом потоке и обращается напрямую или косвенно за помощью к реагирующему.

Согласно проведенному литературному обзору, в отечественном языкознании практически отсутствуют исследования и/или описания приведенных выше явлений наложения речи, в связи с чем остро встает вопрос об адекватности и лаконичности перевода этих понятий. Предложенные нами варианты их перевода выполнены с применением приема модуляции, заключающегося в использовании т.н. «контекстуальных синонимов» [Рыбин, 2007: 164].

Обратная связь, или как переводит в одной из статей О.В. Федорова «маркер обратной связи», реагирующего коммуниканта представляет другую крайность наложений речи [Федорова, 2022]. Ее также невозможно считать пассивным типом взаимодействия, т.к., как уже было упомянуто в п.1.2., все коммуниканты являются активными участниками. Она не только не нарушает

коммуникативную реплику иницирующего коммуниканта, но и является для него неким ожидаемым ориентиром, демонстрирующим получение адресантом информации [Борисова, 2018: 17]. По результатам проведенного исследования на французском языке было выяснено, что реагирующий коммуникант, дающий обратную связь, не мешает иницирующему [Adda et al., 2007]. Маркерами обратной связи являются как вокальные (например, поддакивания), так и кинетические сигналы (улыбка, кивки, наклоны головы, качания головой и т.д.) [Федорова, 2022]. По данным нашей выборки в плане коммуникативного взаимодействия двух собеседников при компенсации речевых сбоев такой вид реакции может формироваться, если иницирующий собеседник самостоятельно совершил репаратуру или продолжил развертывание коммуникативной реплики после хезитации.

В рамках данного исследования мы систематизируем материал, рассматривая его с позиции внутреннего и внешнего уровня компенсации речевых сбоев. Нарушения в вербальном канале являются стимулом к реакции второго собеседника, выражающиеся в трех основных сценариях: инициатива смены как соперничающий тип, вовлеченная смена как кооперативный тип и обратная связь.

При изучении влияния наложений речи на речевые сбои ученые пришли к выводу о том, что большинство нарушений свойственны для иницирующего говорящего, чем для реагирующего [Adda et al., 2007]. Речь первого содержит намного больше речевых сбоев в самом начале речевого потока, нежели в середине или конце, что доказывается правилом «линии отклонения» (declension line) [Shriberg et al., 2001].

В рамках изучения феномена нарушения беглости речи, ученые выяснили, что вербализация новой, прежде неупотребляемой говорящим информации порождает много речевых сбоев по причине низкого уровня предугадывания развития коммуникативных событий. Впрочем, количество речевых помех не становится больше, если сам говорящий ожидает появление новой информации и готов к ее упоминанию [Arnold et al., 2003].

Ученые предполагают, что пол и возраст участников коммуникации влияет на количественные показатели речевых сбоев, однако результаты исследований, проведенных на базе разных языков, не дают универсальных выводов. Согласно работам западных ученых, в английском языке мужчины произносят больше лексических маркеров хезитации, чем женщины [Shriberg, 1996]. Исследование на материале телефонных диалогов носителей китайского языка тем не менее таких результатов не показало. Ученые установили, что объем заполненных пауз в китайском языке не влияет в значительной мере от пола человека. По полученным демографическим данным больше всего заполненных пауз в речи женщин на юге Китая, меньше всего – в речи мужчин на севере. Уровень образования, как предполагалось изначально, не повлиял кардинально на использование заполненных пауз в китайском языке [Zhao, Jurafsky, 2005].

На основе вышесказанного можно отметить, что изучение речевых сбоев в спонтанной речи – сложная, но необходимая часть анализа особенностей порождения устной спонтанной речи. Речевые сбои как вербализованный ключ к пониманию когнитивных процессов речепорождения составляет вместе с тем проблему при создании универсальных систем автоматического распознавания речи. Несмотря на невозпроизводимость устной речи, учеными были установлены и классифицированы основные закономерности появления нарушений в процессе построения и развертывания устной речи, что свидетельствует о повторяемости и воспроизводимости речевых сбоев.

Наряду с речевыми сбоями, в устной коммуникации существует и другой показательный механизм мыслительной деятельности – жестикуляция в разных каналах кинетической модальности, чему и посвящен следующий подпараграф.

1.2.2. Особенности невербального компонента в устной спонтанной речи

Еще во времена античности великие философы, риторы, актеры театра подчеркивали важность невербальной коммуникации. В древних произведениях, посвященных риторике, есть упоминания о жестах, которые «приличествуют» оратору, и жестах, которых ему стоит остерегаться [Гаспаров, 1972: 26]. В настоящий момент исследования о жестике человека приобрели небывалый рост. Выявленные учеными-антропологами и специалистами по межкультурной коммуникации универсалии невербального общения на материале полевых и лабораторных исследований создали основу для углубления и расширения фокуса трудов в данной научной области. Изучение жестового аспекта коммуникации стало настолько актуальным, что им стали заниматься ученые из разных научных направлений – лингвисты, социологи, нейробиологи, историки, психологи, специалисты по компьютерным технологиям и многие другие. Знаменательным также стало основание в 2002 году Международной научной ассоциации по изучению жестов (The International Society for Gesture Studies – ISGS), публикующей на постоянной основе сборники с результатами исследований ученых в сфере невербалики.

В лингвистике, сосредоточившейся на анализе преимущественно письменных текстов, жест долгое время находился на периферии научного интереса. Ситуация изменилась лишь с развитием когнитивной науки, основой понимания которой стало рассмотрение языка человека в непосредственной связи с когнитивными процессами, происходящими в головном мозге: память, мышление, эмоции, процессы концептуализации и категоризации и т.д. Британский ученый и специалист по жестам А. Кендон заявил, что «жестикация настолько взаимосвязана с речью», что они «тем самым должны рассматриваться как два аспекта одного процесса» [Kendon, 2007: 6].

По причине многоаспектности предмета исследования на сегодняшний день существует целое многообразие определений термина «жест». Для сужения круга задач научной работы, в начале XXI века жест воспринимался преимущественно как движение рук (см. [McNeil, 1992], [Kendon, 2004]), т.к. основная доля информации из невербального канала относится к мануальным жестам. В широком смысле, под жестом понимается визуально считываемое движение человеческого тела или таких его частей, как движение рук, головы, корпуса, направление взгляда и пр. Вслед за Г.Е. Крейдлиным мы определяем «жест» как носитель информации, который выступает в виде «знаковых кинетических единиц выражения и передачи информации» [Крейдлин, 2002: 10].

При трактовке понятия «жест» как знака или символа возникает вопрос о границе между коммуникативными телодвижениями и неосознанными физиологическими движениями. Последние остаются за областью исследования, т.к. считается, что не несут важной смысловой нагрузки. Тем не менее, движения, продиктованные физиологией, могут быть особо актуализированы в моменты затруднений при развертывании устного дискурса, репрезентуя процесс мыслепорождения и принципы восприятия информации реагирующим говорящим. В контексте данной работы мы учитываем все движения человеческого тела, сопровождающие речевые сбои и влияющие на вербально-жестикуляционное поведение реагирующего коммуниканта.

При рассмотрении жестов в их связи с речью, важно упомянуть функциональную метрику, названную «континуумом Кендона» (Kendon's continuum). На основании предложенной Д. Макниллом классификации можно выделить следующие типы жестов:

- 1) жестикуляция (gesticulation), сопровождающаяся в обязательном порядке речевым потоком;

2) квазиязыковые жесты (language-like gestures), которые могут заменять отдельные языковые единицы (например, кивание в знак согласия вместо слова «да»);

3) пантомимы (pantomimes) – жесты, передающие информацию вместо речи;

4) эмблемы (emblems) – жесты с немотивированной формой;

5) собственно жестовые языки (sing languages) с целой системой оторванных от языка, но вместе с тем социально закрепленных движений, полностью заменяющих естественный язык [McNeill, 1992: 37].

В приведенной классификации именно первый класс жестов представляет особый интерес для лингвистов по причине прослеживаемой связи с речью человека. Жестикуляция позднее получила название речевых жестов (co-speech gestures) или жестов, синхронизированных с речью (speech-synchronized gesticulation). Однако, ученые затем выяснили, что в зависимости от ситуации функционирования, один и тот же жест может быть интерпретирован и как речевой жест, и как квазиязыковой жест, и как значимый элемент пантомимы, или как составная часть жеста жестового языка глухонемых [Гришина, 2017].

Важным обстоятельством при анализе жестов является описание их основных свойств. По мере развития научных направлений о невербальном поведении – кинесики, мимики, жестики, проксемики и пр. – выводились новые или опровергались выявленные характеристики жестов. Так, А. Кендон в своем труде подчеркивает универсальность следующих черт жестов: осознанность, бескорыстие, намеренность, информативность и контролируемость [Kendon, 2004]. Не все ученые позднее согласились с перечисленными свойствами. Е.А. Гришина отмечает, что из всех свойств достоверными являются только бескорыстие и контролируемость жеста. Если под бескорыстием понимается отсутствие практической цели телодвижения (например, ситуация, когда человек уворачивается от летящего на него объекта, не является жестом), то контроль предполагает желание и готовность

говорящего физически совершать или не совершать определенный жест [Гришина, 2017].

Впрочем, не все ученые придерживаются позиции контролируемой организации жестовой деятельности. И.А. Крым определяет три состояния использования жестов человеком:

- 1) сознательное (осознанное) употребление, которое предполагает аналитическую работу сознания человека;
- 2) бессознательное (неосознанное) употребление, при котором выполнение жеста происходит автоматически, т.е. произвольно;
- 3) сверхсознательное употребление, происходящее на уровне интуиции [Крым, 2004: 13].

Возвращаясь к упомянутому выше континууму А. Кендона, мы также не можем не упомянуть одно из основных свойств, отличающих речевые жесты от жестового языка. Д. Макнилл пишет, что жестикуляция относится к «идиосинкратичным», т.е. единичным «спонтанным движениям ладони или всей руки, сопровождающей речь». Другой крайностью на шкале функционирования жестов являются традиционные, устойчивые, регулярные знаки жестовых языков [McNeill, 1992: 37].

Впоследствии высказанная Д. Макниллом идея была переосмыслена и дополнена. К. Мюллер помимо единичных жестов (*singular gestures*) выделил воспроизводимые жесты (*recurrent gestures*). Исследования под его руководством показали важную для статистического анализа и классификации свойство жестов – повторяемость их основных компонентов [Müller, 2010b]. Дальнейшие научные исследования доказали, что ориентация ладони, направление движения, пространственный рисунок движения отличаются регулярностью и воспроизводятся в потоке речи с одинаковым набором значений. Добавим, что идея воспроизводимости является ключевой для нашей научной работы.

С воспроизводимостью жеста тесно связана его композиционность. Такие западные и отечественные ученые, как Ж. Кальбрис [Calbris, 2011],

К. Мюллер [Müller et al., 2013], Г.Е. Крейдлин [Крейдлин, 2007] и пр. независимо доказывали, что жест представляет совокупность универсальных движений, количество которых является конечным. Комбинация определенных параметров в совокупности дает представление о семантике того или иного жеста и соотносится с определенным смысловым отрезком речевого потока. Таким образом, соотношенность невербальных и вербальных знаков дает в том числе конечное количество лингвистических смыслов. Параметры включают направление движения, траекторию движения, конфигурацию ладони, ориентацию ладони и головы, направление взгляда, зону жестикуляции, кратность жеста, выбор руки, выполняющей жестовое движение [Крейдлин, 2007: 323].

Хотя в науках о невербальной коммуникации наблюдается тенденция к принятию относительно универсальных постулатов структуры и организации жестов, по-прежнему существуют многочисленные классификации основных жестов. Из всего многообразия мы выбрали наиболее употребляемые классификации. Опишем некоторые из них ниже.

Традиционно принято разделять речевые жесты на дейктические и изобразительные. Дейктические жесты предполагают указание местоположения какого-то объекта в пространстве. Изобразительные жесты, соответственно, иллюстрируют или изображают объекты. В составе изобразительных жестов Д. Макниллом также были выделены метафорические и иконические жесты. Первый класс изображает объекты реального (материального) мира, второй класс – абстрактные, идеальные сущности [McNeill, 1992]. Однако позже ученые пришли к выводу о том, что такое дополнительное деление не несет в себе никакой принципиальной разницы, т.к. жесты, описывающие реальные и виртуальные объекты, одинаковы [Гришина, 2017].

Деление на метафорические и иконические жесты приобретает тем не менее значение, когда мы упоминаем денотат. В этом случае жесты могут быть разделены на объектно-связанные и объектно-свободные (или когнитивные).

Под первым типом подразумевается наличие объекта в момент разговора, вне зависимости от того находится ли он в поле нашего зрения или нет. Объектно-свободные жесты предполагают указание на объекты, которые были значимы в прошлом и воссозданы в сознании человека в потоке речи [Гришина, 2017].

Согласно исследованиям, проведенным отечественным ученым Г.Е. Крейдлиным, жесты представлены преимущественно тремя главными классами:

- 1) жесты, которые имеют самостоятельное лексическое значение и способны передавать смысл вне зависимости от вербального контекста;
- 2) жесты, управляющие развитием коммуникативного процесса;
- 3) жесты, которые подчеркивают какой-либо речевой фрагмент.

Первому классу соответствуют эмблемы (эмблематические жесты), которые ученый предлагает подразделять на коммуникативные и симптоматические жесты. Коммуникативные, в свою очередь, делятся на указательные (дейктические), этикетные и общекоммуникативные жесты. Симптоматические жесты дают представление об эмоциональном состоянии говорящего и занимают промежуточное место между движениями, связанными с физиологией человека, и собственно коммуникативными жестами [Крейдлин, 2002: 115]. Тем не менее, при попытке классифицирования собранного нами материала мы столкнулись с проблемой определения разницы между симптоматическими и общекоммуникативными жестами. Как отмечает сам ученый, «для симптоматических жестов мы различаем два типа употреблений: симптоматические жесты в симптоматическом (исходном) и симптоматические жесты в коммуникативном (переносном) употреблении [Там же, 2002: 115]. Помимо этого обстоятельства, не совсем очевидно по каким признакам отличаются общекоммуникативные и этикетные жесты, т.к., по сути, оба типа развиваются и функционируют в пределах отдельного общества и являются взаимосвязанными знаковыми единицами.

Вторым классом жестов, управляющих развитием коммуникативного процесса, являются регуляторы. В отличие от жестов-иллюстраторов, призванных выделить и изобразить то, что было вербализовано, регуляторы представляют движения, которые поддерживают и регулируют диалог. Примерами данного класса жестов могут стать кивки, закрывание и открывание глаз, смена позы и т.д. Г.Е. Крейдлин подчеркивает, что регуляторы по своей природе информативны, однако часто коммуникантами применяются неосознанно. Также было установлено, что регулирующие жесты тесно связаны с этносом, культурой и социальной группой и могут быть отличными в зависимости от территориальной принадлежности [Крейдлин, 2002: 125–126].

Третий класс представляют иллюстраторы (иллюстративные жесты), как правило активизирующиеся в мануальном канале. Такие жесты не только призваны сделать высказывание более богатым в содержательном плане, но и связаны с «ритмом речи, с ее ходом и остановками, <...> с организацией речевого потока в связный текст-дискурс» [Там же, 2002: 117]. Жест-иллюстратор «помогает говорящему вспомнить ускользающее от него слово и извлечь из пассивной памяти нужное название предмета». Как отмечает ученый, проделанная мускульная активность «выталкивает слово наружу, активизируя “лексическую память”, <...> убыстряя и делая более гладким процесс речевого общения» [Там же, 2002: 117–118].

Расширенная версия классификации Д. Эфрона, предложенная Г.Е. Крейдлиным, представляет жесты-иллюстраторы самых разных, подчас редко встречающихся типов. Мы попытались применить данную классификацию на нашем материале, однако столкнулись с трудностями, т.к., во-первых, некоторые типы жестов-иллюстраторов не применяются испытуемыми, а, во-вторых, некоторые жесты могут быть отнесены сразу к двум типам или даже быть соотнесены с предыдущим классом симптоматических жестов.

В ходе предварительного анализа проаннотированного массива данных, нами также были опробованы самые разные классификации отечественных и зарубежных ученых. Основными требованиями к классификации являлись: 1) общий, универсальный принцип при выделении типов жестов, 2) отсутствие очевидных смежных характеристик, затрудняющих разграничение того или иного жеста в определенную группу.

Так, настоящее исследование базируется на классификации П. Экмана и У. Фризена, также основанной на разработанных Д. Эфроном типах жестов, но представляющей более универсальное представление о жестикуляции человека:

1) эмблемы – жесты, которые имеют языковой эквивалент и закрепленное в обществе значение, используются без вербального сопровождения;

2) иллюстраторы, которые в свою очередь представлены следующими классами:

a) дирижирующие жесты (дирижеры), акцентирующие и выделяющие отдельные отрезки речи;

b) идеографы, призванные описывать ход мыслей, переход от основной идеи к второстепенной, от темы к реме;

c) дейктические жесты, указывающие на место, объект или событие;

d) пространственные жесты, демонстрирующие пространственные отношения между объектами;

e) ритмические жесты, задающие ритм/темп речи;

f) кинетографы, изображающие абстрактные действия и имитирующие траекторию со звукоподражанием;

g) пиктографы, необходимые для изображения размера, формы объекта разговора;

h) эмблематические жесты, дублирующие вербальное сообщение;

3) адаптеры – жесты, направленные на снятие или уменьшение внутреннего напряжения;

4) регуляторы – жесты, которые призваны контролировать или координировать общение;

5) аффективные жесты – жесты, используемые для выражения эмоций или чувств говорящих [Ekman, Friesen, 1979: 113].

Уточним, что большинство работ в области жестикуляционной лингвистики исследуют связи жестикуляции с разными проявлениями естественного языка, например, работа о связи движений головы и просодикой [Graf et al., 2002], о параллельности в направлении движения руки и движения тона [Calbris, 2011], работа, посвященная связи фазы маха жеста с подъемами интонации [Loehr, 2012], о корреляции указательных жестов и просодическими характеристиками в кантонском диалекте [Fung, Mok, 2018], о взаимодействии мануальных жестов и речевых сбоев [Смирнова и др., 2020] и т.д.

В целом, несмотря на многообразие исследований о невербальном компоненте устной коммуникации, сохраняется необходимость лингвистического анализа структуры и основных характеристик жеста. Жестовой лингвистике, как новому активно развивающемуся направлению, предстоит раскрыть сущность связи языка и жеста, а также выявить новые закономерности в использовании жестов человеком.

1.3. Общие принципы аннотирования и сегментирования данных при создании мультимодальных корпусов диалогической спонтанной речи

В последнее время лингвистическая деятельность фокусируется на создании корпусов языка в разных его проявлениях – письменном и устном. Началом корпусных исследований принято считать 60-е годы XX века, когда в стенах Брауновского университета США Г. Кучера и У.Н. Френсис создали первый корпус прозаических текстов на американском английском. Несмотря

на то, что эмпирические данные собирались лингвистами и до формирования корпусной лингвистики в виде словарей словоупотреблений или картотек, стоит отметить, что корпусная лингвистика как раздел компьютерной лингвистики подразумевает построение корпусов текстов с применением компьютерных технологий и последующим их хранением на электронных носителях [Козлова, 2013].

Целесообразность построения корпусов, по мнению современных специалистов, заключается в представлении возможно достижимого диапазона явлений языка, зафиксированных в их естественном проявлении. Один раз созданный корпус также может дополняться и многократно применяться учеными при решении разных исследовательских задач [Захаров, Богданова, 2013].

На сегодняшний момент существуют разные варианты определения «корпус», однако почти все они сводятся к тому, что это «большой, представленный в машиночитаемом формате, унифицированный, структурированный, размеченный, филологически компетентный массив языковых данных, предназначенный для решения конкретных лингвистических задач» [Там же, 2013: 5].

В ходе изучения научных трудов, нами было также замечено расхождение в использовании множественной формы понятия «корпус». Эту проблему также в своей работе освещали В.П. Захаров и Ю.С. Богданова, где указано, что сторонники Санкт-петербургской школы корпусной лингвистики больше тяготеет к варианту «корпусы» [Захаров, Богданова, 2013], [Мартыненко, Злобина, 2018], а Московская школа вместе с другими специалистами в области корпусных исследований чаще использует форму «корпуса» [Резникова, Копотев, 2005], [Резникова, 2008], [Плунгян, 2008].

В.П. Захаров и Ю.С. Богданова пишут, что «правила, регламентирующего употребление той или иной формы применительно к корпусной лингвистике, пока нет, хотя, как представляется, победить должен вариант «корпусы», поскольку он отличает терминологическое

значение слова от его общеупотребительного значения» [Захаров, Богданова, 2013: 8]. В данной работе мы будем использовать вариант «корпусы».

В зависимости от исследовательских и прикладных задач создаются разные типы корпусов. Важными критериями их выделения являются параллельность и тип языковых данных. На основании критерия параллельности корпусы делятся на одноязычные, двуязычные и многоязычные. Целью создания одноязычных является изучение языковой вариативности и диалектов. Двуязычные и многоязычные параллельные корпусы можно представить в двух основных видах, а именно: собрание оригинальных и переводных текстов на двух и более языках и собрание текстов из одной и той же тематической области, которые были написаны независимо друг от друга. Параллельность корпусов достигается благодаря критерию выравнивания (*alignment*), который бы позволил установить соответствия между отрезками двух текстов. Выравнивание письменных текстов происходит по предложениям, клаузам, словосочетаниям и словам [Там же, 2013].

По типу языковых данных корпусы можно разделить на письменные, устные (речевые) и смешанные. В последнее время все чаще фокус исследований лингвистов переключается с письменного на устный дискурс, что также находит отражение в корпусных разработках. Создание устных корпусов начиналось с анализа просодических характеристик звучащей речи, однако позже ученые также стали принимать во внимание невербальный компонент, изучая механизмы порождения информации и ее восприятия в совокупности двух или более каналов связи. Так, стали разрабатываться массивы данных с интегрированными многоканальными данными – мультимодальные корпусы.

Под мультимодальным корпусом понимается «аннотированное собрание скоординированной информации из разных коммуникативных каналов, включая речь, направление взора, мануальные жесты и язык тела,

которое обычно создается на материале записей человеческого поведения» [Foster, Oberlander, 2007: 307–308].

В широком смысле лингвисты не разграничивают понятия «мультиmodalный корпус» и «медийный корпус», подразумевая создание медийных файлов с разметкой устного дискурса [Захаров, Богданова, 2013]. Тем не менее заметим, что мультимедийный корпус первоначально задумывался как проаннотированный материал, используемый для анализа звучащей речи в конкретном коммуникативном событии. Корпус обязательно имеет текстовую основу и может быть построен на основании снятых видео- или аудиоматериалов. При наличии видеозаписей размечается и невербальная составляющая коммуникативного события. Мультиmodalные корпуса создаются с обязательным ведением видеозаписи участников коммуникации для изучения, главным образом, произведенных невербальных компонентов. В этом случае акцент мультиmodalных исследований заключается в психофизиологических бессознательных или сознательных сигналах, посылаемых с вербальных и невербальных каналов.

Помимо понятий «мультиmodalный» и «мультимедийный», также предложен вариант «мультиканальный корпус». Не игнорируя предыдущие варианты, А.А. Кибрик склонен считать, что вследствие ограниченности лингвистических возможностей исследования каналов поступления и обработки информации в каждом из типов модальности (вокальной и кинетической), в полной мере корпус не может быть назван мультиmodalным [Кибрик, 2010].

Количество разрабатываемых мультиmodalных корпусов стремительно увеличивается. Мультиmodalные корпуса либо входят в состав национальных корпусов, либо создаются как отдельные мини-проекты. Становлением мультиmodalных исследований в России отечественные ученые обязаны в том числе разработкам западных коллег, которые стали фиксировать и аннотировать звучащую речь, начиная с 70-х гг. XX века (см. The London-Lund Corpus) [Захаров, Богданова, 2013].

Приведем некоторые примеры мультимодальных корпусов западных ученых. Аудиовизуальный корпус MM4 включает 29 коротких диалогов между четырьмя людьми, записанные в полностью контролируемых экспериментальных условиях [McCowan et al., 2003]. Корпус Fruits Cart состоит из 104 видеозаписей разговоров 13 участников. Участники строго следовали заданиям, предложенным учеными. Корпус был создан для исследования процессов понимания речи на слух, позднее стал использоваться для анализа принципов построения речи [Aist et al., 2006]. Корпус UTEP ICT построен для изучения межкультурного взаимодействия. Он состоит из первоначально заданных разговоров в группах из четырех человек. Во время записи участники могли вставать и перемещаться по комнате [Herrera et al., 2010]. Отметим также британский корпус MIBL, в электронной базе которого размечены движения и речь людей в процессе их диалогового взаимодействия при обучении одного участника другим игре в карты. Созданный корпус используется для обучения сервисных роботов [Wolf, Bugmann, 2006].

Корпусы спонтанной звучащей речи на материале китайского языка менее обширны и созданы, главным образом, за пределами Китая, в стенах западных вузов. Одним из самых известных является корпус спонтанной речи the NCCU, состоящий из диалогов между двумя и более участниками на трех языках: китайском, хакка и южном мине (языке, на котором говорят на юге Китая и Тайване) [Chui et al., 2008]. Другой корпус устной китайской речи the Lancaster Los Angeles Spoken Chinese Corpus включает полностью транскрибированные материалы монологической и диалогической речи на китайском языке [Xiao, Tao, 2007]. Интересен также мультимодальный корпус диалогов the DUEL на материале немецкого, французского и китайского языков, который был создан для анализа речевых сбоев, а также влияния смеха и восклицаний на плавность развертывания речевого потока [Hough et al., 2016].

Отечественные мультимодальные корпуса представлены в нескольких крупных проектах. Мультимедийный корпус МУРКО на базе русского языка

включает устные тексты как кинематографического характера, так и формата ток-шоу, публичных выступлений, новостей и пр. В МУРКО помимо стандартной разметки НККЯ существует социологическая и акцентологическая. Позже корпус был более глубоко аннотирован и расширен в плане базы размеченных жестов, мимики, речевых актов и пр. [Гришина, 2017].

Другой крупный проект «Язык как он есть: русский мультимодальный дискурс» также посвящен изучению языка в его естественном проявлении, т.е. устной форме коммуникации. Корпус включает виды разметки, указанные выше, а также окуломоторную и цефалическую аннотацию. Построение электронного ресурса было основано на поэтапном обсуждении участниками эксперимента шестиминутного «Фильма о грушах», снятого в 1970-х гг. У. Чейфом [Кибрик, 2018].

При создании мультимодальных корпусов диалогической спонтанной речи в рамках нашей работы необходимо решить важную задачу, а именно: определить, каким образом будет происходить аннотирование вербальных и невербальных данных.

Аннотирование (разметка) в каналах кинетической модальности осуществляется нами согласно первичному списку жестов, который предоставила компания Promobot, с возможностью редактирования и добавления часто повторяемых телодвижений. Последующая систематизация осуществляется на основе классификации П. Экмана и У. Фризена (см. Приложение Г).

Если регистрирование невербальных движений происходит в непосредственный момент их наблюдения, то сложность аннотирования вокальной модальности состоит в определении единицы устного спонтанного дискурса, по которой возможно провести сегментирование вербального потока. Известно, что устный спонтанный дискурс в отличие от письменного дискурса невозможно абсолютно однозначно поделить на предложения.

В связи с этим, лингвисты выработали два основных подхода к решению данной проблемы.

Первый подход сформирован на базе результатов исследований письменного дискурса – семантико-синтаксический подход. Элементарной единицей устного дискурса (или по терминологии А.А. Кибрика – ЭДЕ элементарной дискурсивной единицей) в этом случае является клауза [Кибрик, Подлеская, 2003]. Считается, что «синтаксической вершиной клаузы, или просто предложения, является единственный предикат, который при этом обозначает единое и единственное событие» [Шлуинский и др., 2015: 187].

Второй подход – просодический – применяется для наиболее объективного транскрибирования устного дискурса. В рамках данного подхода единицей анализа выступает интонационная единица (ИЕ), т.е. речевой отрезок, имеющий один интонационный контур. По замечанию американского ученого У. Чейфа устный дискурс нужно рассматривать как последовательность интонационных единиц. Изложенная идея ученого представляет компромисс между языковыми, когнитивными и психофизиологическими критериями [Chafe, 1994].

При сопоставлении результатов исследований, выполненных в рамках семантико-синтаксического и просодического подходов, У. Чейф пришел к выводу, что две трети интонационных единиц совпадают с клаузой по своему объему. Последующие исследования дополнили и расширили представления о соотношении клаузы и интонационной единицы [Кибрик, Подлеская, 2003].

Например, отечественными учеными А.А. Кибриком и В.И. Подлеской были выделены два типа соотношений. В первый тип, при котором ЭДЕ совпадает с клаузой, входят финитные и нефинитные глагольные предикации, адъективные предикации, предикации, включающие субстантивное сказуемое и другие «маргинальные» виды предикации типа «*и вдруг раз так*». Исходя из того, что ЭДЕ может быть частью интонационной единицы с соседней предикацией, ученые условились, что будут игнорировать просодический

аспект при членении на ЭДЕ, при этом просодические данные сохраняются путем непроставления пунктуационных знаков в конце единой строки [Кибрик, Подлесская, 2003].

Второй тип соотношений делится на два случая. Первый случай представляет собой тип, когда ЭДЕ больше, чем клауза и подразумевает объединение основной предикации с сокращенной соседней, в которой не представляется возможным определить отдельный интонационный контур. Примерами в данном варианте будут инфинитивные сентенциальные актанты, сериальные глаголы (*молчу думаю*), целевые инфинитивы (*лег спать*), редупликация глагола (*говорила-говорила*), глаголы речи и мысли (*уюду думаю отсюда*). Второй случай – когда ЭДЕ меньше, чем клауза. Нередко в такого типа ЭДЕ предикат редуцирован, а сама единица выделена на основе изменившегося интонационного контура. Примерами таких ЭДЕ являются постпозитивные уточняющие единицы (*Летели балочки. Красивые такие*), полноударные слова типа — *вот, так*, слова *да, нет*, отдельно вынесенный топик с последующей референцией в соседней предикации (*Береза, а на ней вороны сидят*) и т.д. [Там же, 2003].

Согласно полученным результатам ряда исследований, зародившихся в недрах дискурс-анализа письменных текстов, семантико-синтаксический подход к сегментации текстов удачно сочетается (с некоторыми погрешностями) с просодическим подходом, что позволяет утверждать о наличии прообраза фиксированного на бумаге предложения в человеческом мозге при речепорождении. Об этом свидетельствует частотное совпадение интонационного падения в речи точке/запятой на письме [Кибрик, 2010].

Помимо выработки основного принципа аннотирования, необходимо также определить унифицированный вариант транскрибирования звучащей речи. Некоторые исследования в области анализа вербальной составляющей коммуникации выполнены с применением псевдофонетической транскрипции, суть которой заключается в использовании метаязыка (допустимы варианты непосредственного фиксирования редуцированных форм типа «*что*» до

разговорного варианта «чѐ»). По справедливому замечанию А.А. Кибрика и В.И. Подлесской, данный подход не может помочь в решении узкоспециализированных фонетических задач, поскольку в этом случае лингвисты и фонетисты обращаются к уже созданной фонетической транскрипции, а не метаязыку. Кроме того, метод псевдофонетической транскрипции идет вразрез с ориентированным на дискурс исследованием [Кибрик, Подлесская, 2003]. В нашей работе мы придерживаемся стандартной орфографии в обоих языках.

Касательно транскрибирования китайской речи, нами были отмечены два подхода. Первый подход подразумевает использование латинизированной системы транскрипции – пиньинь 拼音 без приведения иероглифической расшифровки. Научные труды во втором подходе пишутся строго с применением иероглифов с редким приведением перевода примеров. Уточним, что мы не поддерживаем использование системы пиньинь прежде всего по причине дискурсивно-ориентированного изучения устной речи, а не его фонетической стороны. Кроме того, наличие транскрипта, выполненного по системе китайской транскрипции без иероглифического сопровождения, может привести к недопониманию содержания отдельных языковых единиц по причине высокой степени омонимичности китайского языка.

Членение речевого потока на ЭДЕ не предполагает использование стандартной формы пунктуации, применяемой в письменном тексте. Однако опираясь на результаты, показавшие совпадение пунктуационных знаков с изменениями интонационного рисунка [Кибрик, 2010], мы включили в маркировку основополагающих просодических характеристик знаки из обычной пунктуации, подобно отечественным исследованиям в данной области.

По классификации А.А. Кибрика, вокальный тип модальности состоит не только из вербального, но и, как описывалось выше, из просодического компонента (в некоторых статьях канал именован вербально-просодическим). Измерение просодических характеристик является важной задачей при

анализе устного дискурса [Кибрик, 2010]. Вследствие ограниченности ресурсов фокус данного исследования будет главным образом сосредоточен на одном из просодических элементов, а именно паузации. Измерение пауз носит условный характер, по причине разного темпа речи и средней длины пауз. Фиксирование пауз в ЭДЕ происходит от 0.1 сек., после точек, иллюстрирующих паузу, в скобках записывается число в десятых долях секунды ее продолжительности (без пробела) [Кибрик, Подлесская, 2003].

Кроме незаполненных пауз (абсолютных пауз), встречаются также заполненные паузы (см. п. 1.2.1), которые обычно транскрибируются в русском языке с помощью букв ээ (звук «шва») и мм (носовой звук), в китайском – 嗯 (звук «шва») и 哼 (носовой звук). Все остальные заполненные паузы, включающие полноценные слова типа 这个 (это), 那个 (то), ну, вот и т.д., обозначаются согласно орфографическим стандартам. Заполненные паузы аннотируются подобно абсолютным, с той разницей, что сначала записывается сам знак типа паузы (гласный звук «шва» или согласный сонорный), а после указывается время ее длительности (например, мм(0.5)).

Как мы убедились, семантико-синтаксический подход к сегментированию непрерывного вербального потока информации является продуктивным с точки зрения создания корпусов звучащей речи.

Нахождение закономерностей коммуникативного поведения участников устных спонтанных диалогов на уровне вербального и невербального компонентов также предполагает вычленение и исследование отдельных единиц коммуникативной практики, под которыми в настоящей работе понимаются коммуникативные события.

В ходе эмпирического наблюдения исследователь имеет дело лишь с частью бесконечного коммуникативного бытия человека. В общей непрерывности коммуникативного процесса, «каждый акт коммуникации занимает свое место как ячейка речевой и социальной деятельности» [Колшанский, 1984: 95]. Такие акты коммуникации формируют «структурно и

функционально целостные устойчивые комплексы», именуемые событиями [Гольдин, 1997: 24]. Н.Д. Арутюнова указывает, что событие относится к жизненному пространству, динамично, единично, и обладает кульминативностью, сценарностью, слабой структурированностью [Арутюнова, 1999: 519].

Коммуникативное событие как дискретная единица коммуникативной практики, «вычленяется в контексте социальной деятельности как ограниченный в пространстве и времени, мотивированный, целостный, социально и культурно обусловленный процесс речевого взаимодействия его участников» [Борисова, 2018: 43].

Коммуникативные события включают два измерения, а именно: структурное (ситуативно-ролевой аспект) и динамическое (деятельностно-поведенческий аспект). Если первое измерение (статика) задается в виде фреймов, то второе динамическое измерение (динамика) – в виде сценариев (скриптов). Параметры коммуникативного события (мотив общения, социально-ролевой состав участников, обстановка общения) задаются в его так называемой параметрической модели – коммуникативной ситуации [Там же, 2018: 44–45].

И.Н. Борисова считает, что «характер инвариантной коммуникативной ситуации детерминирует сферу коммуникации – совокупность типологически сходных коммуникативных событий с общими закономерностями дискурсообразования и текстопорождения» [Там же, 2018: 47].

Так, мы полагаем, что для изучения закономерностей невербального компенсирования классифицированных типов речевых сбоев необходимо обратиться к рассмотрению отдельных коммуникативных событий, находящихся в составе проаннотированных нами мультимодальных дискурсов.

Как отмечает калифорнийский ученый М. Сэвилл-Труак, границы коммуникативного события могут быть обозначены использованием «ритуальных» высказываний по типу *“Did you hear this one?”*, *“Once upon*

a time”, “*Let us pray*”, “*Amen*” и пр., или в плане перехода с одной темы на другую. Конец одного коммуникативного события и начало другого также можно обнаружить в изменениях тона голоса, выражения лица или положения тела. Именно «комбинация вербальных и невербальных маркеров сигнализирует об основных коммуникативных переходах» [Saville-Troike, 1994: 109].

В рамках настоящего исследования границы необходимого для рассмотрения коммуникативного события определяются по принципу «стимул – реакция», где стимулом (началом коммуникативного события) выступает речевой сбой иницирующего коммуниканта, а реакцией (концом коммуникативного события) – вербально-жестикуляционный ответ реагирующего коммуниканта. Как уже было сказано в п. 1.2.1., речевой сбой может привести к реакциям на внешнем уровне, воплощенным в трех основных видах наложений речи: инициативе смены, вовлеченной смене и обратной связи. Каждое коммуникативное событие состоит из комбинации определенного типа речевого сбоя и невербального комплекса компенсирования на внутреннем уровне, а также вида наложения речи с невербальной реакцией на внешнем уровне. Такие комбинации образуют сценарии коммуникативного события. Рассматриваемый сценарий типичного коммуникативного события представлен в форме мультимодальных схем иницирующего и реагирующего коммуникантов, где упрощенно отображены закономерности коммуникативного поведения в ходе компенсации речевых сбоев (см. п. 2.5.).

В заключении подчеркнем, что речевая коммуникация ввиду своих особенностей порождения и развертывания, по-прежнему тяжело поддается фиксации и обработке в машиночитаемом формате. Несмотря на трудности, исследовательский интерес к поиску решений по аннотированию устной речи до сих пор сохраняется. Так, в широком смысле мы можем говорить о формировании двух подходов к определению единицы аннотирования, а именно: семантико-синтаксического и просодического.

В связи с дискурсивно направленным характером исследования, мы сосредоточились на семантико-синтаксическом подходе при создании мультимодальных корпусов, где единицей аннотирования выступает ЭДЕ – элементарная дискурсивная единица. При решении задачи оптимизации потока мультимодальной информации путем определения закономерных путей компенсирования речевых сбоев необходимо также обратиться к анализу и систематизации отдельной единицы коммуникативной практики – коммуникативного события, которое складывается из определенных сценариев, воплощаемых в виде мультимодальных схем.

1.4. Мультимодальные корпуса как ресурсы для создания моделей машинного обучения

Все более очевидное погружение человека в информационные процессы формирует необходимость в повышении качества совместного использования технических средств и человеческих ресурсов. Несмотря на свою актуальность, моделирование систем человеко-машинного взаимодействия является мультидисциплинарной задачей, поскольку требует знаний как в области гуманитарных наук, таких как лингвистика, психология, социология и т.д., так и в области компьютерных наук – понимание работы искусственного интеллекта, машинного обучения и т.д.

Моделирование является сложным и многоэтапным процессом исследования систем. Машинное обучение представляет один из методов анализа данных, посредством которого автоматизируется построение аналитических моделей. Идея данной области искусственного интеллекта состоит в способности системы учиться на данных выявлять закономерности, прогнозировать и принимать решения [Павлычев, Стародубов, 2023].

Проектирование моделей было бы невозможно без данных. «В машинном обучении первостепенное значение имеет набор данных. Прежде чем проводить какой-либо анализ, исследователь должен сам

тщательно изучить и обработать набор имеющихся данных» [Павлычев, Стародубов, 2023: 161]. Так называемые «сырые», т.е. необработанные данные нельзя напрямую использовать в машинном обучении. Вследствие чего, крайне важна предварительная обработка собранных данных, которая осуществляется с помощью такой операции как разметка или аннотирование.

На наш взгляд, также необходимо внести ясность в понятие «набор данных» (dataset), который определяется в Национальной стратегии развития искусственного интеллекта РФ до 2030 г. как «совокупность данных, прошедших предварительную подготовку (обработку) <...> и необходимых для разработки программного обеспечения на основе искусственного интеллекта» [Федеральный проект, 2019]. Под «массивом данных» понимается «существующая в машиночитаемой форме совокупность данных об окружающем мире и (или) о процессах, происходящих в нем, формируемая на основе источников технического или социального происхождения и обладающая действительной или потенциальной коммерческой ценностью» [Савельев, 2020: 67]. Так, «набор данных» является более узким понятием, нежели «массив данных», т.к. разработка первого прежде всего связана с их внедрением в модель машинного обучения. Наборы данных также не предполагают наличия т.н. «сырых» данных.

Мультимодальные корпуса как комплексы многоканальной информации, аннотируемой на основе записей человеческого поведения, представляют собой не просто обособленный и структурированный от общей массы информации массив данных. Мультимодальность реализуется в таких типах данных, как текстовые записи (многоуровневая сплошная разметка), временные промежутки, изображения, аудио и видеосигналы, что как мы предполагаем, позволяет поставить мультимодальные корпуса в один ряд с наборами данных/датасетами (datasets), предназначенными для автоматизированного анализа и машинного обучения.

Разметка данных или аннотирование данных (data labeling/data annotation) представляет собой процесс ручного/полуручного выявления

людьми необработанных элементов в изображениях, текстовых, аудио или видеофайлах и добавления информативных меток (тегов), маркирующих и задающих контекст, на базе которого будет обучаться модель [Владимиров, Марков, 2022].

Для обучения лингвистических моделей (процессоров) на основе методов машинного обучения обращаются прежде всего к текстовым ресурсам – корпусам как массиву данных, собранному определенным образом и обладающему лингвистической разметкой (морфологической, синтаксической, акцентной, дискурсивной или др.) [Воронцов, 2014].

Сложность разработки мультимодальных корпусов заключается в том, что в отличие от традиционных текстовых корпусов первые должны включать информацию разного рода и порядка. В рамках данной работы мультимодальные корпуса спонтанной диалогической речи строятся на основе дискурсивной разметки, в которую включается собственно лингвистическая и экстралингвистическая информация.

В машинном обучении применяется созданный корпус с указанием на объекты/совокупности объектов, находящиеся в определенной зависимости [Большакова и др., 2017]. Определение зависимостей между модальностями в рамках ручного преданализа данных решает одну из сложных проблем мультимодального машинного обучения – выравнивание (alignment) [Liang et al., 2022]. Отметим, что выравнивание в корпусной лингвистике необходимо для выявления переводческих соответствий в дву- или многоязычных текстах. Однако выравнивание в области машинного обучения предполагает определение не просто текстовых соответствий, а связей между модальностями и входящими в их состав модусами (каналами) путем разметки собранных данных и создания особых сегментированных классов, решающих более узкую исследовательскую задачу [Там же, 2022].

Оптимизация является важной частью машинного обучения и представляет собой ключ к формированию качественной модели. В программировании выделяют два основных вида оптимизации,

представляющие крайности друг друга, а именно глобальная (overall/global optimization) и локальная (local optimization) оптимизация. Если первый вид подразумевает поиск состояния системы, обеспечивающего ее наилучшее функционирование, то второй вид главным образом связан с разработкой механизма адаптации программы к конкретным особенностям вычислительной машины [Oliveira, Stewart, 2006]. В прикладной лингвистике под оптимизацией понимают «такое описание (модель) проблемной области, при котором эта область сохраняет в результирующем представлении только те существенные свойства, которые необходимы для данной практической задачи» [Баранов, 2001: 10]. Приведенное определение, по нашему мнению, в большей степени соотносится с локальным видом оптимизации. Вместе с тем, для обеспечения того самого «наилучшего» состояния по всей системе, необходимо обработать данные так, чтобы это в первую очередь потенциально способствовало повышению производительности запросов.

Можно заключить, что при создании мультимодальных корпусов как ресурса для машинного обучения первичная (глобальная) оптимизация в лингвистическом плане связана с принципами обработки данных и реализуется в плане сплошной разметки мультимодальных данных, позволяющей, во-первых, обработать массив и преобразовать его в набор данных, во-вторых, выравнять тем самым сами данные, указывая наблюдаемые сущности во временной последовательности.

Второй этап оптимизации (локальная оптимизация) связан с формированием *набора целевых атрибутов (параметров)*, представляющего совокупность необходимых актуальных значений для решения конкретной задачи в машинном обучении [Ерохин, Ванюшина, 2018]. В нашем исследовании набором целевых атрибутов в строгом смысле выступает коммуникативное событие, внутри которого существуют целевые атрибуты – самые важные данные из вокальной и кинетической модальности, а именно речевой сбой и сценарии его компенсации в наиболее активизированных невербальных каналах.

Выравнивание информации в модальностях путем проведения сегментирования может быть полезным в решении проблемы «шума» (noise) в данных, под которым понимаются внешние неорганизованные переменные или ошибки, влияющие на эффективность работы модели. Шум может возникнуть в результате человеческого фактора, погрешностях в вычислениях, системных сбоях или причин природного характера [Копырин, Видищева, 2020].

В данной работе мы имеем дело со спонтанной диалогической речью, следовательно шумы в вокальной модальности неизбежны и становятся препятствием к пониманию заложенных человеком смыслов. Переводя термин «шум» из области машинного обучения в плоскость лингвистики, мы находим его смежным с термином «речевой сбой», т.к. именно он представляет собой нарушение плавности при развертывании речи человека.

Коммуникативное событие, включающее вербально-поведенческие реакции обоих собеседников на образование речевых сбоях, представляет собой сегмент, внутри которого осуществляется дополнительное выборочное выравнивание данных. Привязка реакций к стимулу (речевому сбою), как мы считаем, дает возможность перенаправить фокус внимания системы с шумов в вокальной модальности на функциональную информацию в кинетической модальности, во избежание искажений и низкой результативности выходных данных. Так, добавление в разметке коммуникативных событий компенсации речевых сбоях формирует своеобразный «фильтр внимания» (attention gate), регулирующий распределение потока данных в ходе машинного обучения. Компьютерные архитектуры, базирующиеся на концепции фильтров внимания (attention-gated learning), явление сравнительно новое, тем не менее согласно результатам исследований в области ИИ, они демонстрируют стабильные и продуктивные результаты, в том числе при решении задач прогнозирования [Xiong et al., 2016], [Zadeh et al., 2018].

В мультимодальном машинном обучении выделяют несколько основных типов стратегий, которые в зависимости от исследовательских задач

могут быть использованы в архитектурах вместе или по отдельности, а именно «слияние/смешивание» (fusion) и «перевод/проекция» (translation/projection) [Chandu et al., 2021].

Слияние данных в общем смысле нацелено на объединение модальностей в единую систему для получения более согласующейся и точной информации, чем та, которая формируется на основе одного типа модальности. Слияние используется для улучшения контекстуального состояния компьютерного агента или обучения прогнозированию. Человеческий мозг является главным примером слияния данных, при котором он фиксирует, полагаясь на органы чувств, поступающую мультимодальную информацию, комбинирует ее и выдает ответ. В случае с обучением нейронных сетей или компьютерных агентов выделяют две стадии слияния – «раннее слияние» (early fusion) и «позднее слияние» (late fusion) [Там же, 2021]. Первая стадия предполагает первичное образование единой системы знаний, обучение некоторым закономерностям, в то время как «позднее слияние» работает на уровне принятия решений.

Стратегия «перевод» уже долгое время применяется при обучении нейронных сетей переводу на разных языках. Однако в области мультимодального машинного обучения перевод применяется для переложения информации из одного модального формата в другой, например, картинки в текст, текст в аудио и т.д. [Qian et al., 2018]. Перевод модальностей может также осуществляться, если данные из одного/нескольких каналов имеют шумы [Ma et al., 2021].

Таким образом, моделирование человеко-машинного взаимодействия представляет многоаспектную задачу, которая реализуется путем применения методов машинного обучения. При обучении искусственного интеллекта важную роль играет материал исследования – набор данных. Создаваемые лингвистами корпуса, в особенности мультимодальные, как мы считаем, представляют те самые наборы данных, необходимые для считывания, прогнозирования или имитирования человеческого поведения компьютерным

агентом. Оптимизация корпусов для их последующего применения в машинном обучении может быть выполнена глобально с помощью разметки, выравнивающей данные, и локально с применением таких стратегий, как перевод и слияние.

ВЫВОДЫ ПО ГЛАВЕ 1

Данная глава представляет краткий теоретический обзор исследований в области устной коммуникации. В рамках теоретической главы рассмотрены основные термины, а именно «мультиmodalность», «вокальная modalность», «кинетическая modalность», «спонтанная речь», «диалогическая речь», «речевой сбой», «наложение речи», «жест», «коммуникативное событие», «иницирующий говорящий», «реагирующий говорящий», «мультиmodalный корпус», «мультиmodalная схема», «машинное обучение», «набор данных», «разметка»/«аннотирование», «набор атрибутов», «фильтр внимания», «слияние данных», «перевод данных» и приведены наиболее актуальные для дальнейшей работы определения, классификации, методы аннотирования, сегментирования и анализа.

Мультиmodalный подход, широко применяемый в разных научных дисциплинах, также постепенно обретает популярность среди лингвистических исследований. Помимо анализа вокальной modalности, состоящей из вербального и просодического каналов, ученые интересуются и кинетической modalностью, однако из-за обширности объекта исследования чаще проводят эксперименты разрозненно, без учета комплексности феномена коммуникации.

Сегодня термин «мультиmodalность» не имеет общепринятого определения. В отечественной лингвистике специалисты склонны отождествлять понятия «поликодовость» и «креолизация» с «мультиmodalностью». Тем не менее, последний термин акцентирует внимание на способах восприятия информации коммуникантом из разных каналов (модусов). В настоящем исследовании мы придерживаемся определения мультиmodalности, данного Г.Е. Крейдлиным, который подчеркивает закономерность взаимодействия вербальных и невербальных знаков, воплощаемую в виде общих законов и правил.

Изучение устной коммуникации в диалогическом аспекте восходит к идее диалогичности языка, высказанной М.М. Бахтиным. Диалогическая речь, как ее определяет Л.П. Якубинский, подразумевает форму непосредственного взаимодействия индивидов, при которой происходит быстрая смена акций (стимулов) и реакций. Одной из основных характеристик диалогической речи является спонтанность-подготовленность. Рассматриваемая нами диалогическая спонтанная речь представляет один из самых сложных типов устного дискурса для анализа, т.к. в отличие от подготовленной является менее структурированной.

В процессе коммуникативного взаимодействия оба участника являются активными, в связи с чем, деление собеседников на главных и второстепенных, активных и пассивных является ошибочным. Вслед за Е.В. Борисовой, мы вводим два понятия для описания статуса участников диалога – «иницирующий» и «реагирующий» говорящие.

Одним из немногих воспроизводимых компонентов устной спонтанной речи являются речевые сбои. Компенсация речевых сбоев может происходить как на внутреннем (индивидуальном), так и на внешнем (межличностном) уровнях. Последний связан с возникновением наложения речи, которое воплощено в одной из трех основных видов реакций: инициативе смены (как соперничающем типе), вовлеченной смене (как кооперативном типе) и обратной связи (например, поддакивания вербальной/невербальной формы).

В устной коммуникации кроме вокальной модальности также важно учитывать кинетическую, а именно жестовые движения. Жест в данной работе определяется как носитель информации, который выступает в виде знаковых единиц, применяемых для выражения и передачи информации. При описании данной модальности важнейшей для настоящего исследования является идея композиционности и воспроизводимости жестов. В спектр аннотирования попадают мануальные, окуломоторные, целографические, мимические жесты, а также движения корпуса. При их анализе и описании мы применяем классификацию, дополненную и уточненную П. Экманом и У. Фризенем.

В широком плане аннотирование устной спонтанной речи проводится в двух подходах: просодическом и семантико-синтаксическом. Согласно сопоставительным исследованиям, названные подходы не противоречат друг другу. Интонационная единица (ИЕ) в большинстве случаев совпадает с клаузой (ЭДЕ). В рамках настоящей работы разметка устной речи проводится по семантико-синтаксическому подходу.

Выявление закономерностей коммуникативного поведения участников спонтанных диалогов на предмет восполнения вербального потока информации вследствие речевых сбоев необходимо осуществить путем выделения и последующего сегментирования дискретных единиц коммуникативной практики – коммуникативных событий. Согласно их деятельностно-поведенческому аспекту, необходимо рассмотреть каждое событие с точки зрения сценария компенсации речевых сбоев на внешнем уровне, воплощенного в трех основных видах наложений речи. Такую быструю смену акций и реакций коммуникантов можно наглядно представить в мультимодальных схемах.

Мультимодальный корпус, созданный путем сплошной дискурсивной разметки данных для обучения лингвистических моделей на основе методов машинного обучения, является уже не просто массивом, а набором данных. В стремлении приблизить искусственный интеллект к уровню живого человеческого общения специалистам необходимы мультимодальные данные, ресурсом которых на настоящий момент являются мультимодальные корпуса. Последующее обучение системы методами мультимодального машинного обучения может быть осуществлено с помощью таких стратегий, как слияние и перевод данных.

Оптимизацию мультимодальных корпусов для машинного обучения в связи с их многоаспектностью необходимо стремиться проводить в широком плане (глобальная оптимизация), с возможностью сосредоточения на более конкретных задачах посредством применения локального подхода (локальной оптимизации).

ГЛАВА 2. ОПЫТ ПОСТРОЕНИЯ И ОПТИМИЗАЦИИ МУЛЬТИМОДАЛЬНЫХ КОРПУСОВ ДИАЛОГИЧЕСКОЙ СПОНТАННОЙ РЕЧИ ДЛЯ МАШИННОГО ОБУЧЕНИЯ

2.1. Дизайн эксперимента по сбору данных и принципы представления его результатов

Первым шагом к созданию мультимодальных корпусов является сбор данных, осуществляющийся в рамках экспериментальной лингвистики, основными задачами которой являются разработка механизма накопления, обработки и анализа необходимой информации, а также решение проблемы репрезентативности выборки и проверки достоверности полученных на ее основе результатов [Белоусов, 2014].

Экспериментальная лингвистика способствует формированию методологического инструментария в отраслях, сконцентрированных на прикладных исследованиях, а именно корпусной и компьютерной лингвистике. Так, настоящее теоретико-прикладное исследование выполнено по заказу и при активном содействии компании Promobot, являющейся крупнейшим производителем автономных сервисных роботов не только на территории России, но и зарубежом. Пермская компания активно ведет разработки в области электроники, мехатроники, искусственного интеллекта, нейросетей, распознавания речи и пр. Нацеленная на производство в первую очередь роботов сервисного типа, Promobot стремится развивать и совершенствовать лингвистические базы, привлекая к работе исследователей в области лингвистики.

В соответствии с техническим заданием (Приложение А), предоставленным компанией, в спектр задач входит создание и передача заказчику мультимодальных размеченных баз данных на двух языках, а также разработка интегрированных схем мультимодальности.

Эксперимент – это научный метод, предполагающий изучение наблюдаемых явлений или процессов с помощью их воспроизведения или моделирования в искусственных или естественных условиях [Борисова, 2003]. Принято выделять такие виды экспериментов, как полевой, лабораторный и квазиэксперимент.

Полевой эксперимент предполагает проведение изучения в реальных условиях. К сожалению, осуществление полевых исследований лингвистического плана представляет особую сложность, т.к. скрытая запись аудио-видеоматериала и последующее демонстрирование полученных результатов с ссылками на источник без предварительного согласия испытуемых может преследоваться по закону.

Лабораторный эксперимент, проводящийся в полностью искусственно созданных условиях, также имеет некоторые ограничения в плане снижения точности показателей. Первые мультимодальные исследования выполнялись преимущественно в строго контролируемых, лабораторных условиях. Испытуемых просили надеть специальное оборудование, например, айтрекинг-очки и/или жилеты. Эксперименты проводили в специальных записывающих кабинках или безэховых камерах. Несмотря на стремление к более точным, «чистым» показателям, подобные лабораторные исследования «намеренно снижают возникновение непредвиденных факторов, которые могут серьезно повлиять на формирование базы данных» [Oertel et al, 2013: 19].

Квазиэксперимент ставится в том случае, когда экспериментатор стремится не оказывать воздействия на участников исследования или создавать особые контролируемые условия, а использует уже имеющиеся условия и естественные группы. Принципиальным отличием от лабораторного вида эксперимента является отсутствие контрольной группы испытуемых. В нашей работе имеет место именно квазиэксперимент, т.к. основной целью является создание наиболее возможного варианта естественного эксперимента (natural experiment).

Задачи такого научного метода как эксперимент могут варьироваться в зависимости от тех или иных исследовательских целей, тем не менее можно выделить несколько основных, а именно: эмпирическая проверка справедливости выдвигаемых теоретических положений и гипотез, выявление свойств изучаемых объектов и определение причинно-следственных связей между факторами или совокупностью факторов (зависимыми и независимыми переменными) с возможным обоснованием полученных результатов. В ходе нашего научного исследования мы сосредоточились на проверке гипотезы о повышенной активации каналов кинетической модальности в моменты возникновения речевых сбоев при ведении спонтанных диалогов на русском и китайском языках и определении связей между зависимыми и независимыми переменными.

Проведение эксперимента предполагает манипулирование независимыми переменными, в ходе которого выясняется степень влияния такого рода манипуляции на зависимые переменные [Леонтович, 2011]. Под *независимыми переменными* понимаются переменные, которые не подлежат изменению. В нашем случае это национальность, пол участников, общая тема ведения спонтанных диалогов. Следовательно, ожидается, что *зависимые переменные* могут меняться в результате манипулирования независимыми. Такими переменными в данной работе являются процесс развертывание спонтанного диалогического дискурса (т.е. выбор предложенных пунктов из карточки или придумывание своих пунктов с сохранением общей темы разговора), выбор невербальных и вербальных средств, пространственное расположение участников диалога.

Изначально техническое задание, обговариваемое с компанией Promobot, предполагало работу с тремя языками – русским, английским и китайским. Выбор языков сосредоточивался в особенности на паре русский-английский, т.к. большинство партнеров и заказчиков компании Promobot это европейские страны. Тем не менее, проведение эксперимента на английском языке

сопровождалось рядом проблем, которые нам в рамках данного исследования решить не удалось.

Определив, что работа будет строиться на студенческом дискурсе, мы не нашли адекватного количества учащихся-носителей английского языка. Привлечение носителей среди иностранцев, живущих в Красноярске, не решило бы проблему со сбором материала в количественном плане. К тому же, собранный материал исследования уже бы не являлся студенческим дискурсом.

В качестве второго варианта рассматривалось привлечение русских студентов к диалогу на английском языке для построения базы данных на так называемом Global English. После пилотной записи двух диалогов таким способом мы обнаружили, что данный подход не корректен в плане обработки и предоставления результатов выборки, определения мультимодальных схем компенсации речевых сбоев, т.к. жестикуляционный портрет носителей русского языка, продиктованный главным образом культурой, в которой вырос человек, как правило не меняется, вне зависимости от выбранного языка.

Отметим, что разработки носят довольно многоаспектный характер, что физически не может быть воплощено в одной работе, а, значит, мы предполагаем, что одним из перспективных направлений следующих исследований является создание и интегрирование массива данных на английском языке.

Нами было предложено сосредоточиться на языковой паре русский-китайский, поскольку, с одной стороны, китайский рынок для компании является одним из приоритетов развития, с другой стороны, ресурсы нашего Института филологии и языковой коммуникации Сибирского федерального университета (далее – ИФиЯК СФУ) позволяют провести подобное исследование благодаря образовательному сотрудничеству и активной академической мобильности между СФУ и вузами КНР.

Участниками добровольного эксперимента являются студенты ИФиЯК СФУ. Все испытуемые подписали письменное согласие на участие и обработку данных в научных интересах (см. Приложение Б).

В эксперименте приняли участие 26 носителей русского языка и 20 носителей китайского языка. Средний возраст группы испытуемых 20 лет и 24 года соответственно. Среди 26 добровольцев 2 человека – мужчины, 24 человека – женщины. Из 20 носителей китайского языка 4 человека – женщины, 16 человек – мужчины.

Заметим, что вокруг гендерного аспекта жестового поведения по-прежнему ходят дискуссии. На основе собираемых данных исследователи не всегда могут однозначно сделать вывод о количественных показателях жестикуляции среди мужчин и женщин. Тем не менее, отечественный ученый Г.Е. Крейдлин обосновывает два стиля коммуникативного поведения, включая невербальный компонент: мужской (маскулинный) и женский (фемининный) стили. Предполагается, что у женщин преобладает фемининный стиль, у мужчин – маскулинный, однако данные стили необязательно напрямую должны быть связаны с полом человека [Крейдлин, 2005]. В связи с недостатком детально обоснованных исследований о корреляции частоты жестикуляции с полом человека, гендерный аспект в настоящей работе не учитывается. Однако мы не умаляем его важность и предполагаем, что с расширением массива данных проверка гипотезы о соотношении чистоты жестов у мужчин и женщин может стать одним из актуальных объектов других исследовательских проектов. Данная работа сосредоточена преимущественно на анализе сходств и различий стереотипного поведения, воплощенного в жестовой и речевой активности носителей двух культур.

После определения задач эксперимента, описания характеристик экспериментальных факторов и принципов отбора испытуемых, исследователям необходимо сосредоточиться на проектировании ситуации и хода эксперимента [Леонтович, 2011]. Так, мы приступили к разработке

базиса для проведения исследовательской работы по сбору мультимодальных данных.

Сложность возникает на стадии планирования хода эксперимента, т.к. от степени проработанности предлагаемого или усовершенствованного метода будут зависеть результаты и качество всего исследования. В процессе проектирования экспериментальной ситуации начинающий исследователь физически не может учесть все условия ведения эксперимента, в связи с этим, проект эксперимента должен быть не раз апробирован, прежде чем будет применен для сбора материала.

Первоначально аудио-видео запись испытуемых предполагалось проводить без непосредственного нахождения экспериментатора в аудитории, таким образом снизив вероятность возникновения парадокса наблюдателя (the observer's paradox). Это явление впервые было описано В. Лабовым и заключается в том, что наблюдение в процессе экспериментальной деятельности не дает таких же сведений о пользовании языком, как в натуральной среде [Labov, 1972]. Парадокс наблюдателя по-прежнему остается одним из самых дискуссионных в лингвистике, т.к. проведение записи «живого» материала без ведома информанта как минимум неэтично, как максимум порождает множество правовых ограничений.

После нескольких проб записи спонтанных диалогов без присутствия экспериментатора, нами было обнаружено, что речь подвергается большой чистке, преобладает абсолютная паузация, а сами информанты могут начинать диалог несколько раз, засмущавшись камеры или своего собеседника. Участники также излишне обращают внимание на камеру, делают комментарии в сторону записывающих устройств (например, *я, наверное, сейчас себя скомпрометирую, но я ...*) или обращаются к экспериментатору на камеру (например, *это потом вырежете*).

Следующим шагом было проведение записи в присутствии в аудитории третьего человека – экспериментатора. Студенты были более вовлечены в диалог при нахождении «контролирующего лица», однако стиль речи даже

в диалогах неформального типа более походил на официальный или полуофициальный стили. Мы считаем, что подобные дискурсивные характеристики были актуализированы по причине нахождения преподавателя (коим в нашем случае выступает для студентов экспериментатор) в аудитории. В социолингвистике этот феномен был описан в теории А. Белла «Дизайн аудитории» (Audience design), которая предполагает смену регистра языка на предполагаемую и желаемую аудиторией [Bell, 1984].

За основу методики проведения эксперимента были взяты результаты проведения третьей апробации. Так как запись материала организовывалась в условиях работы естественных групп, то согласившиеся студенты должны были ожидать своей очереди. В этот раз они все оставались в аудитории для записи материала, но занимались своими привычными делами, как бы это происходило в перерывах между занятиями. После обработки материала, стало очевидно, что коммуникация стала более живой в присутствии других знакомых студентов, т.к. обстановка более походила на реальную. Испытуемые практически не обращали внимание на камеру, заметно более открыто вели диалог и даже упоминали в качестве референта находившихся в аудитории людей (например, *вон, я знаю, что у Юли есть...*). Такой результат в том числе может быть объяснен смещением фокуса внимания с записывающих устройств на присутствующих в аудитории.

Запись эксперимента осуществлялась с трех обзорных точек и включала основную камеру, регистрирующую поведенческие реакции с фронтальной стороны, и две дополнительные, функция которых сводилась к фиксации движения всем корпусом. В ходе эксперимента также был использован беспроводной микрофон, а не петличный, с целью предотвращения дополнительного влияния шнура на поведение собеседников.

Участники сами контролировали свое коммуникативное пространство так, что кадр выстраивался в зависимости от их расстановки. Заметим, что информанты вели диалоги в положении стоя. Мы принимаем тот факт, что

телодвижения в такой позиции по физиологическим причинам имеют свои особенности (например, по мнению физиотерапевтов и специалистов по мимике и жестам, сгибание рук в локтях может в том числе быть связано с расслаблением мышц и снижением нагрузки). Тем не менее, необходимость регистрирования именно такого положения собеседников обусловлена дальнейшим практическим применением данных, а именно при взаимодействии с роботизированными системами компании Promobot.

Отметим, что, стремясь создать более естественный тип эксперимента, экспериментатор должен иметь в виду *экологичность* эксперимента, под которым понимается «гармоничная адаптация человека к сложившимся обстоятельствам с разумным и продуктивным использованием доступных модусов восприятия и каналов передачи информации» [Чистова, 2022: 448]. Подобный «экологический эксперимент» не обязательно сводится к созданию полевых условий, в том числе предполагая «моделирование адекватной реакции живого организма на какой-либо фактор в искусственно созданных лабораторных условиях, предполагающий проявление определенных реакций организма и в природных условиях» [Лабутина и др., 2021].

Перед проведением аудио-видео записи испытуемым были выданы карточки с опорными пунктами, которые при желании можно раскрыть в процессе взаимодействия. Наличие карточек, по нашему мнению, не снизило экологичности проведенного эксперимента, т.к. испытуемым давалось время на погружение в тему и вхождение в роль, затем карточки у них изымались. Коммуниканты также не могли предварительно простроить ход диалога, т.к. при ознакомлении с содержанием опорной карточки не имели возможности советоваться и обсуждать ее содержание; сами информанты не знали о роли своего партнера.

Несмотря на то, что принцип развертывания спонтанного диалогического дискурса (т.е. выбор того, какие пункты раскрыть в процессе диалога) оставался за испытуемыми, а значит представляет зависимую переменную, основные темы опорных карточек для ведения спонтанных

диалогов являются одной из независимых переменных. Это связано с тем, что представленный в исследовании список согласован с компанией Promobot и разработан на базе списка основных тем разговоров, которым обучены сервисные роботы. В нашей работе представленные темы карточек можно разделить на два больших блока: 1) обсуждение современных проблем между друзьями/коллегами (жизнь «до» и «после» первой волны коронавируса, рекомендации к покупке автомобиля, организация корпоратива в компании) и 2) взаимодействие клиента и консультанта (консультация в турагентстве, решение проблемы неисправного оборудования). Содержание карточек полностью совпадает на русском и китайском языках (см. Приложение В).

После описания хода эксперимента, исследователю далее необходимо провести оценку продуктивности и чистоты эксперимента, указать прочие выводы и результаты (при их наличии), наблюдаемые в процессе интерпретации полученных качественных и количественных показателей, а также сделать заключение о достоверности выдвинутой гипотезы [Леонтович, 2011].

В данной работе представление полученных результатов экспериментальной работы можно разделить на два этапа (предлингвистический и собственно лингвистический). Предлингвистический этап анализа основан на первичной обработке «сырых» данных, транскрибировании и фиксации активности в разных мультимодальных каналах, выравниванию мультимодальных данных. Собственно лингвистический этап предполагает сегментирование на отдельные значимые элементы при развертывания устной спонтанной речи – коммуникативные события, включающие упрощенные мультимодальные схемы компенсации речевых сбоев.

Таким образом, эксперимент как научный метод может применяться для анализа широкого спектра исследовательских проблем. Получение исследовательского материала в условиях реальной действительности предполагает проведение полевого эксперимента. Однако в рамках изучения

мультимодальности такой вид экспериментальной работы имеет целый ряд серьезных ограничений, вследствие которых его осуществление становится практически невозможным. Одним из вероятных решений такой проблемы может быть создание квазиэксперимента. Обозначив основные характеристики зависимых и независимых переменных, описав условия и ход эксперимента, мы далее сосредоточимся на представлении результатов, полученных в процессе работы над массивами данных, и оценке продуктивности выдвинутой гипотезы.

2.2. Определение основных характеристик экспериментальных переменных и наблюдаемых параметров при предварительной статистической обработке мультимодальных данных на русском и китайском языках

Разметка аудио-видеоряда проводилась по слоям с возможностью выгрузки построчной выравненной аннотации, где одна строка соответствует слову, в представленной фразовой очередности, и фиксации (при наличии) жестовой активности в разных мультимодальных каналах: руки/плечи, корпус, голова, глаза/брови, рот/губы, лицо, громкость, ударение, скорость речи и конгруэнтность (т.е. согласованность внешнего поведения с вербальным выражением).

Основными параметрами статистической обработки мультимодальных данных в обоих языках послужили:

- 1) время инициирования спонтанного диалога;
- 2) дискурсивное сегментирование (ЭДЕ);
- 3) частотность жестикуляции в разных мультимодальных каналах;
- 4) конгруэнтность;
- 5) тип жестов в каждом тематическом блоке.

Известно, что разные наблюдения одних и тех же переменных могут различаться по значению параметров. Проведение первичной статистической обработки экспериментальных данных нацелено, во-первых, на выявление

закономерностей заданных зависимых и независимых переменных исследуемых объектов, во-вторых, на оценку продуктивности и перспективности исследования в мультимодальном ключе.

В ходе эксперимента мы намеренно не ставили временных рамок, предоставляя испытуемым свободу в инициации спонтанных диалогов. Наблюдается существенная разница во времени решения поставленной коммуникативной задачи у русских и китайских студентов.

Таблица 1. Распределение времени диалогов носителей русского и китайского языка

	Время диалогов носителей русского языка (в минутах)	Время диалогов носителей китайского языка (в минутах)
1	3:02	1:35
2	3:36	3:34
3	4:59	3:24
4	3:40	1:51
5	2:11	2:31
6	4:55	2:30
7	6:32	1:55
8	3:59	2:29
9	3:35	1:57
10	4:02	2:18
	Среднее время:	Среднее время:
	4:03	2:24

Исходя из результатов таблицы 1, носители китайского языка тратят в среднем почти в два раза меньше времени, чем носители русского языка, фактически обсудив все рекомендуемые пункты, указанные во вспомогательных карточках (см. Приложение В). При анализе содержания спонтанных диалогов мы приходим к заключению, что подобный временной разброс объясняется стремлением русских студентов к большей детализации, обращению к личному опыту или упоминанию пережитых событий у родных и знакомых.

В ходе разметки также произведено первичное аннотирование дискурса по семантико-синтаксическому принципу, согласно которому транскрипт

делится на ЭДЕ (см. Рисунок 1). Важной поправкой для роботизированных систем является выделение приветствий и прощаний в отдельные ЭДЕ.

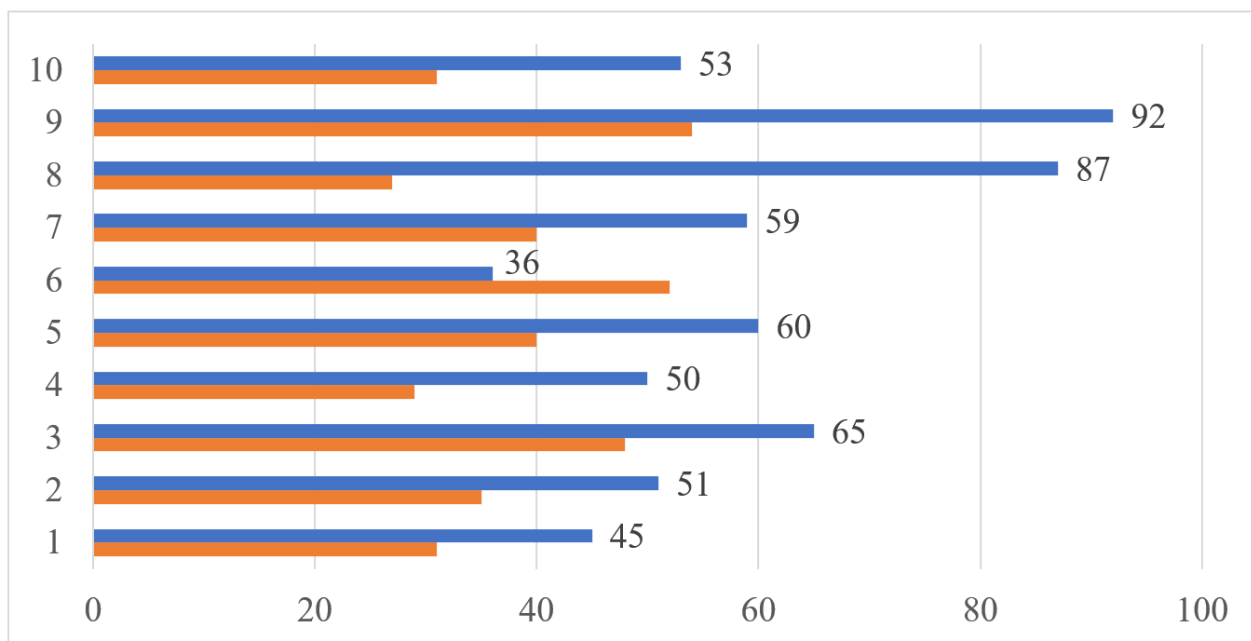


Рисунок 1. Распределение количества ЭДЕ в устном спонтанном дискурсе (синий цвет – диалоги на русском языке; оранжевый цвет – диалоги на китайском языке)

Среднее значение ЭДЕ в русском дискурсе составляет 59,8 и 38,7 – в китайском дискурсе. Сопоставив со средней длительностью диалогов в обеих группах испытуемых, получаем, что за минуту среднее значение ЭДЕ носителей русского языка – 14,8, у носителей китайского языка – 17,27. Такие результаты представленной выборки говорят о том, что переход от одного смыслового фрагмента к другому (от одной темы к другой) у китайцев происходит быстрее, чем у русских.

Проведя статистический анализ представленной выборки в размере 10 видеозаписей на русском и китайском языках, мы также можем сделать промежуточный вывод об активности жестикуляционного портрета носителя русского и китайского языков. На основе социокультурных факторов специалисты по невербалике и ученые, проведя ряд исследований, подтвердили, что русские жестикулируют значительно чаще китайцев. «“Зерном истины” <...> относительно невербальной коммуникации представителей Китая можно считать сдержанность в жестикуляции

и отсутствие открытого, прямого взгляда на собеседника» [Новикова, Гридунова, 2011: 24]. Ниже представлены данные о частоте невербального компонента в двух культурах, который включает данные из мануального (руки/плечи), целографического (голова), окуломоторного (глаза), мимического (лицо и рот/губы) каналов, а также из канала, регистрирующего движения корпусом (см. Таблицы 2–3). В таблицах мы учитываем телодвижения обоих участников диалогов, однако не берем во внимание длительность каждого жеста.

Таблица 2. Частотность жестикуляции в спонтанной диалогической речи на русском языке

№	Количество строк	Руки/плечи	Корпус	Голова	Лицо	Глаза	Рот/губы
1	402	150	9	66	4	99	8
2	569	33	7	31	5	66	15
3	669	106	10	38	7	98	21
4	590	171	20	87	3	149	14
5	555	103	8	16	6	127	27
6	279	94	11	28	2	73	12
7	630	179	50	49	2	142	11
8	1089	241	46	110	2	201	10
9	679	165	10	44	1	79	2
10	451	151	14	58	3	65	14
Среднее:							
	591,3	139,3	18,5	52,7	3,5	109,9	13,4

Таблица 3. Частотность жестикуляции в спонтанной диалогической речи на китайском языке

№	Количество строк	Руки/плечи	Корпус	Голова	Лицо	Глаза	Рот/губы
1	293	39	19	16	3	59	8
2	437	23	21	16	3	74	9
3	398	21	27	30	3	66	8
4	415	31	14	20	2	65	3

№	Количество строк	Руки/плечи	Корпус	Голова	Лицо	Глаза	Рот/губы
5	400	26	25	24	1	63	8
6	484	52	22	21	0	69	6
7	657	69	7	16	0	28	1
8	361	17	6	16	0	59	3
9	336	43	15	24	0	53	5
10	206	23	5	10	0	27	7
Среднее:							
	398,7	34,4	16,1	19,3	1,2	56,3	5,8

Представленная статистическая информация анализируемой в данной работе выборки свидетельствует о том, что мануальный канал в русской речи активен почти в два раза, чем в китайской речи. Так, в среднем на минуту спонтанного разговора у русских студентов приходится 34,56 жестов рук, у китайских студентов – 15,36 соответственно. Тем не менее, по табличным данным видно, что «восполнение» невербальной активности у носителей китайской культуры происходит в других каналах: движение корпусом, головой (целографические жесты) и глазами (окуломоторные жесты). Частота движений корпусом у китайских студентов превышает частоту, демонстрируемую русскими студентами в 1,6 раза. Усредненный показатель целографических жестов в минуту составляет 13,07 в диалогах на русском языке и 8,61 в диалогах на китайском языке. Показатели прослеживаемых движений глаз (без использования технологии айтрекинга) показывают практически сходную активность в двух культурах: 27,3 жеста в минуту в русской речи и 25,13 жестов в минуту в китайской речи.

Мимические движения (лицо и рот/губы) в русской и китайской культурах представлены главным образом улыбкой, смехом, изображением гримасы, надуванием/прижиманием губ. Согласно табличным данным, данный канал более активен у русских испытуемых, чем у китайских (в среднем 13,4 и 5,8 соответственно).

Говоря о конгруэнтности, мы наблюдаем баланс между вербальными и невербальными проявлениями при ведении спонтанных диалогов. Только 3 диалога из нашей выборки демонстрируют нарушение конгруэнтности. Установлено, что неудобные ситуации (обсуждение типа напитков на организуемом банкете, катастроф и принципов предоставления страхования) могут сопровождаться улыбкой или смехом участников, возрастанием интенсивности мануальных и окулomotorных жестов. Вероятно, подобная реакция показывает вымысленность событий, упомянутых участниками. Конгруэнтность, тем не менее, достигается в том случае, если случаи были пережиты участниками коммуникации (коронавирусные ограничения, болезни родных, ситуация с Q-кодами, поездка на машине с механической коробкой передач и пр.).

Данные, представленные выше, являются общими показателями жестовой активности вне зависимости от темы диалога. Далее мы представляем средние значения в соответствии с типом жеста и тематическим блоком.

Напомним, что все диалоги можно разделить на два больших блока – общение между друзьями/коллегами и общение между клиентами и консультантами. Тип жеста определяется на основе классификации П. Экмана и У. Фризена и включает активность в разных модусах (каналах). Оба блока представляют следующие типы жестов: *иллюстраторы, адаптеры, регуляторы, аффективные жесты*, а также выделенные нами *ситуативные жесты*, предполагающие использование реквизита (указание с использованием бутылки с водой, указание на предмет разговора – сломанный телефон и пр.). Средние значения в каждом из блоков отображены в таблицах ниже (см. Таблицы 4–5).

Русский устный спонтанный дискурс тематического блока 1 характеризуется повышенным использованием жестов-иллюстраторов, среди которых преобладают дейктические (указательные), дирижирующие, ритмические жесты и пиктографы. Следом за жестами-иллюстраторами

расположены жесты-адаптеры, способствующие уменьшению/снятию внутреннего напряжения, как следствия активной жестикуляции в мануальном канале при отстаивании своей позиции (карточка 1 – диалог недовольного клиента и консультанта) или стремлении заинтересовать/описать свои пожелания (карточка 2 – диалог в туристическом агентстве).

Таблица 4. Среднее арифметическое типов жестов, представленных в блоке 1 «Общение клиента с консультантом»

Жесты-иллюстраторы	Жесты-адаптеры	Жесты-регуляторы	Аффективные жесты	Ситуативные жесты
Русская спонтанная речь				
50,8	43,6	28,6	15,4	0,8
Китайская спонтанная речь				
15	29,5	26,3	9,75	1,2

Таблица 5. Среднее арифметическое типов жестов, представленных в блоке 2 «Общение с друзьями/коллегами»

Жесты-иллюстраторы	Жесты-адаптеры	Жесты-регуляторы	Аффективные жесты	Ситуативные жесты
Русская спонтанная речь				
33,6	44,6	34	19,4	-
Китайская спонтанная речь				
24	21	20,3	10,3	0,8

Средние значения блока 1, полученные в выборке на китайском языке, свидетельствуют в пользу утверждений о сдержанности невербальной коммуникации в китайской культуре. Раздраженность/заинтересованность выражается преимущественно в жестах-регуляторах, включающих прямой взгляд на собеседника, наклоны корпуса, приближение к собеседнику, отстранение от собеседника, кивки, мотание головой и пр. Стоит отметить, что показатель жестов-адаптеров в блоке 1 является ведущим, т.к. в данный тип жестов помимо самоадаптеров (прикосновения к телу, голове, одежде, предметам, аксессуарам и пр.) вошли жесты нейтрального положения (например, сложенные за спиной руки, руки в карманах и пр.). Вслед за А.О. Литвиненко и др., под *нейтральным положением* мы понимаем

«положение, для пребывания в котором говорящему не требуется (или почти не требуется) прикладывать специальных усилий» [Литвиненко и др., 2018].

Блок 2 «Общение с друзьями/коллегами» (см. Таблицу 5) демонстрирует перераспределение ведущих типов жестов в обеих культурах. По результатам выборки носители русского языка при разговоре с друзьями или своими коллегами пользуются преимущественно жестами-адаптерами, среднее значение частотности появления данного типа составляет 44,6. Согласно нашим наблюдениям, большая часть жестов-адаптеров в блоке 2 направлены не на снижение стресса/напряжения или проявление сдержанности, а скорее являются проявлением непосредственности, открытости, которое выражается даже в стереотипном дружеском диалоге. Адаптеры носят более динамичный характер и включают такие жесты, как: тереть пальцы, почесывать шею/грудь, потирать нос, раскачиваться из стороны в сторону, ломать пальцы и пр.

Носители китайской культуры предпочтительно используют жесты-иллюстраторы при общении с друзьями/коллегами, проявляя большую раскованность движений, чем в стереотипном поведении клиента и консультанта. В блоке 2 используемые китайскими студентами жесты-иллюстраторы представляют следующие классы: дейктические (указательные), ритмические жесты, а также идеографы, отражающие процесс протекания мыслей.

На базе полученных количественных результатов появилась гипотеза о связи наблюдаемых типов жестов с тематическим блоком у носителей одного и того же языка. Для проверки выдвинутой гипотезы мы воспользовались методом статистического подсчета – критерием независимости хи-квадрат (или критерий согласия Пирсона χ^2). Сопоставляя заданные частоты разных блоков (в основном это сопоставление ожидаемых и наблюдаемых частот, но не всегда), мы можем сделать вывод о статистической значимости (связи/отсутствии связи) рассматриваемых признаков. В основе критерия независимости лежит вычисление величины χ^2 , которая

определяется как сумма отношений квадратов отклонений наблюдаемой величины O_i от ожидаемой величины E_i к ожидаемой величине в каждой ячейке:

$$\chi^2_n = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

Значение критерия χ^2 тем больше, чем больше зависимость между переменными и чем меньше статистическая значимость. Вместе с χ^2 вычисляется *p-value* (уровень значимости). Для большинства работ в случае отклонения нулевой гипотезы этот показатель должен быть меньше или равен 0,05. Чем меньше *p-value*, тем выше статистическая значимость связи. «Величина $p \leq 0,05$ свидетельствует о статистически значимом результате, который достоин содержательной интерпретации» [Наследов, 2013:124]. Стоит также упомянуть, что критерий χ^2 не дает конечных результатов, а применяется лишь для подтверждения или опровержения нулевой гипотезы, за основу которой принимают утверждение о том, что частоты согласованы и расхождения между ними не являются статистически значимыми.

Итак, нулевая гипотеза в нашем случае звучит так: *выделение ведущего типа жестов у носителей одного и того же языка для каждого тематического блока является случайным.*

Проведя подсчеты, мы получили, что *p-value* различия частот распределения типов жестов в блоке 1 и блоке 2 у носителей китайского языка составляет 0,034821668, что меньше установленных 0,05. Следовательно, мы склонны утверждать, что типологическое различие жестовой активности в спонтанных диалогах на китайском языке тематически обусловлено. Показатель *p-value* частотности типов жестов в русской спонтанной речи – 0,051118355, что при округлении равняется 0,05. В данном случае мы можем с осторожностью утверждать о связи между тематикой диалога и типом жестов. Для проверки данного факта требуется провести дополнительные расчёты с использованием результатов других выборок.

По нашему мнению, можно улучшить показатели модели при перераспределении информации из вокальной модальности в кинетическую путем добавления пояснений о прагматических смыслах используемых жестов. По этой причине в рамках проведения базовых статистических подсчетов мы также обратили внимание на функциональную классификацию жестов и их связь с типом диалога, приведя список актуализированных жестов в соответствии с классификацией, разработанной П. Экманом и У. Фризенем, и дополнив ее упоминаемыми нами симптоматическими жестами и жестами нейтрального положения (см. Приложение Г).

Таким образом, можно заключить, что разнородность выводимых количественных и качественных показателей, демонстрируемая уже на уровне осуществления базовых статистических подсчетов, позволяет подтвердить перспективность и продуктивность исследований, проводящихся в мультимодальном аспекте.

Проведя первичную статистическую обработку экспериментальных данных по пяти параметрам, мы выявили закономерности в характеристиках зависимых и независимых переменных, осуществили базовую визуализацию полученной мультимодальной информации. Такой начальный анализ является необходимым условием подготовки массива данных для машинного обучения, т.к. именно исходя из представления о распределении элементов в системе и их взаимосвязи, можно строить разные компьютерные архитектуры для решения поставленных задач.

Возвращаясь к задачам настоящего исследования, отметим, часто наблюдаемое нами явление конгруэнтности и наибольшая активация таких модусов кинетической модальности, как мануальный, окуломоторный и целографический в обеих культурах, как мы полагаем, являются важными атрибутами (параметрами) для оптимизации данных и задают основу для систематизации и прототипирования вербально-поведенческих реакций коммуникантов в виде мультимодальных схем.

2.3. Лингвистическое описание закономерностей невербальной компенсации речевых сбоев в русской спонтанной диалогической речи

Как мы выясняли в п. 1.2.1., при компенсации речевых сбоев на внутреннем (индивидуальном) уровне человек стремится показать его сохраняющуюся вовлеченность в процесс говорения, так что заминки, означающие затухание речевого акта, начинают компенсироваться невербальными средствами с возможностью одновременного проведения репаратуры в вербальном канале. Восполнение нарушений речевого потока внешнего (межличностного) уровня формируется так вербально-жестикуляционная реакция второго собеседника.

Приведем некоторые примеры речевых сбоев и их компенсации на невербальном уровне обоих коммуникантов в трех мультимодальных каналах – мануальном, окулomotorном и целографическом, т.к. именно они проявляют большую активность. При дальнейшем расширении базы данных корпуса предполагается учитывать другие мультимодальные каналы (движение корпуса, мимические жесты, просодические особенности (скорость речи, громкость речи и пр.)) и добавление информации об их активности и соотношении с типом речевых сбоев.

а) маркеры препаративной подстановки (placeholders)

Диалог №8

А: хотя многие закрылись конечно
жалко
(кивает головой, прижимает губы, отводит взгляд в сторону)

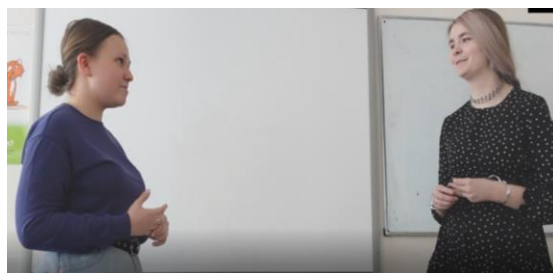


Рисунок 2

А: у нас была= эта ... (0,5)
(указывает большим пальцем расположение кофейни)



Рисунок 3

А: как её ... (0,9)
(закрывает глаза, поднимает глаза в потолок, вращает кистями)



Рисунок 4–5

А: ээ... (0,2) испанская кофейня
... (0,3) с центра
(смотрит на собеседника прямо)

Б: мм ... (0,4)
не слышала
(моргает головой)



Рисунок 6

На Рисунке 5 видно, что стремление ведущим коммуникантом не потерять коммуникативный ход и не оставить реплику незавершенной реализуется на невербальном уровне в пропорциональном соотношении. Контур невербального компонента выравнивается только при вербализации уместного, по мнению говорящего, варианта – *испанская кофейня* (см. Рисунок 6).

Вращательные движения руками в этот момент затухают, и говорящий восстанавливает свое положение, наиболее частотным для него жестом (потирает пальцы руки, вращает кольцо на руке). Примечательно, что реагирующий говорящий терпеливо ждет, пока иницирующему не придет на ум нужный вариант. Взгляд коммуниканта Б направлен строго на коммуниканта А, жестикуляция первого в рассматриваемых каналах приостановлена, что может свидетельствовать о выжидательной тактике, по причине отсутствия общих пресуппозиций со своим собеседником.

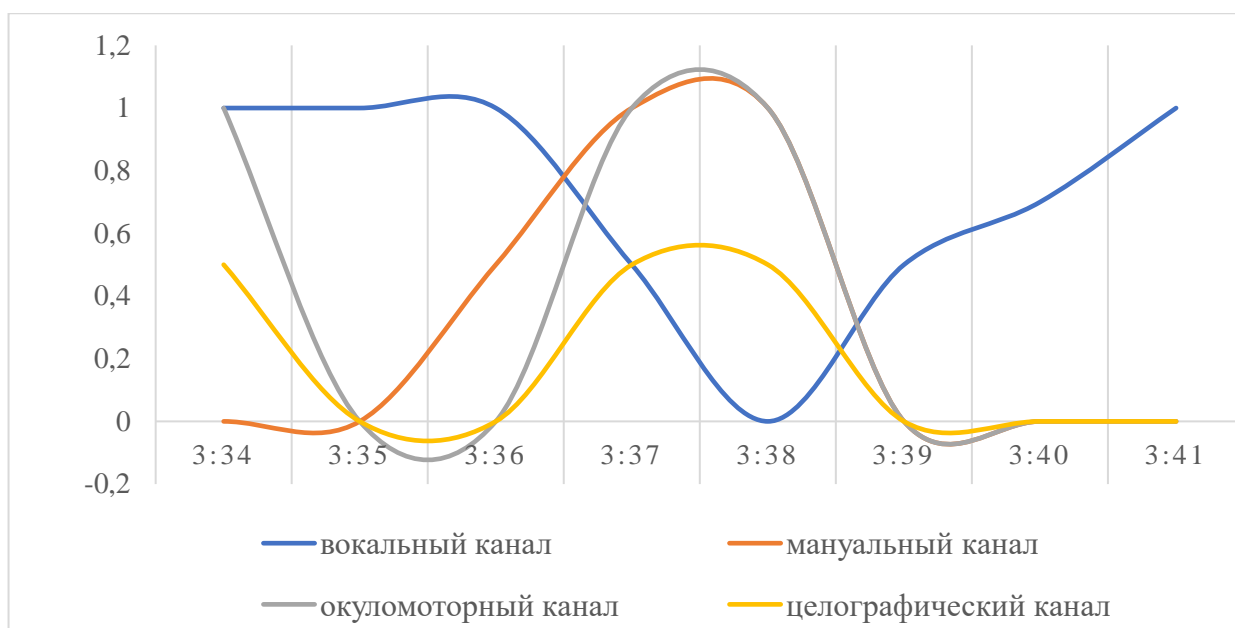


Рисунок 7. Соотношение восполнения речевого сбоя невербальными средствами. На оси абсцисс отмечен временной отрезок коммуникативного события в секундах, на оси ординат – приняты абсолютные величины (развертывание речи без запинок – 1, абсолютные паузы – 0; прямой взгляд на собеседника – 0, отвести взгляд в сторону – 1; фаза покоя – 0, ритмичные жесты руками – 1; вертикальное положение головы – 0, наклоны головы – 1)

Рисунок 7 наглядно демонстрирует описанный выше пример (см. Рисунки 2–6) повышенной активизации каналов кинетической модальности в моменты затруднений в вокальной модальности, являясь одним из подтверждений поставленной в нашей работе гипотезы.

б) заполненная пауза

Такой тип хезитации в анализируемом нами материале является самым частотным (порядка 95% диалогов содержит от 20 и более ситуаций подобного речевого сбоя). Заполненные паузы предполагают долексическое оформление новой реплики (*mmm* или *эээ*), а также лексические маркеры речевого сбоя (*ну, ой, это самое*).

В примере ниже (см. Рисунки 8–16) попытка компенсации запинки речевого хода со стороны невербального компонента наблюдается в учащенном моргании коммуниканта А. Как видно на Рисунке 12, взгляд коммуниканта Б в этот момент был направлен на руки своего собеседника. Пассивность невербалики со стороны коммуниканта А, в конце

сопровождающаяся иллюстративными жестами эмблематического класса (пожимание плечами и разведения рук) (см. Рисунок 11), была расценена вторым участником как условная возможность передачи слова.

Диалог №6

Б: я могу-у прийти ... (0,2) попo-озже /

(барабанит кончиками пальцев рук, смотрит на коммуниканта А)



Рисунок 8–9

*А: да конечно *

(кивает головой, смотрит на коммуниканта Б)

... (0,5) эээ ... (0,3) ну ... (0,7)

(учащенно моргает, отводит в сторону взгляд, опускает глаза, осматривает свои руки, пожимает плечами, разводит руками)



Рисунок 10–12

*Б: ну получается *

*тогда-а мне принести чек *

гарантию ... (0,2) /

(смотрит на жесты коммуниканта А, уводит руки в сторону, перечисляет, загибая пальцы)





Рисунок 13–14

А: и сам ... (0,1) ноутбук

(протягивает кисть к собеседнику, кивает)



Рисунок 15

Б: ... (0,2) ноутбук.

(поворачивается к коммуниканту А)



Рисунок 16

Абсолютная пауза при перечислении предметов коммуникантом Б в конце (см. Рисунок 15) также закончилась инициативой смены коммуникативного хода. Данный факт может говорить о стремлении коммуниканта А восполнить потерянную ранее вовлеченность в дискуссию.

Отметим, что в материале исследования заполненная пауза даже при отсутствии резких ритмичных телодвижений расценивается реагирующим коммуникантом как попытка удержания коммуникативного хода иницирующим, и частота инициативности смены ролей снижается.

в) абсолютная пауза

Этот тип речевых сбоев обычно означает прекращение/завершение вербализации речевого отрезка. За абсолютную паузу не принимаются случаи вовлеченной смены при завершении смысловой части реплики одним из коммуникантов. Абсолютная пауза рассматривается нами исключительно как

нарушение развертывания устного спонтанного дискурса, возникающая преимущественно при обдумывании ответа или «потери мысли» говорящим (см. Таблицу 6).

Таблица 6. Пример компенсации абсолютной паузы в русской диалогической спонтанной речи

Диалог №14	Невербальные средства
Б: ... (0,6)	отводит взгляд, смотрит вниз, разводит руками в сторону
А: <i>есть какие-нибудь предложения/</i>	
Б: [interrupt] я не знаю \	
ну... (0,6) предложения ээ ... (0,2)	
<i>можно спрашивать у других /</i>	
.... (0,4)	смотрит на собеседника, моргает, двигает корпус к собеседнику
А: нет \	
<i>нам нужно быстро</i>	

Зарегистрирована следующая мультимодальная закономерность – абсолютные паузы, сопровождающиеся пассивной жестикуляцией, во всех проаннотированных нами коммуникативных событиях означают смену ролей иницилирующего и реагирующего говорящего.

г) самокоррекции

В отечественной лингвистике предложено выделять два основных режима самокоррекции – собственно коррекция (онлайн коррекция) и ретроспективная коррекция (офлайн коррекция) или редактирование [Подлеская, Кибрик, 2007: 3–4].

В онлайн режиме самоисправление говорящим происходит непосредственно в момент обнаружения проблемы и сопровождается серией фальстартов (см. Таблицу 7). По нашим наблюдениям, иницилирующий говорящий во время самокоррекции закрывает / прикрывает глаза, отводит взгляд в сторону и мотает головой.

Таблица 7. Примеры компенсации собственно коррекций в русской диалогической спонтанной речи

Диалог №1	Невербальные средства
<i>А: ... (0,2) я же вам говорю\</i>	смотрит на собеседника
<i>хожу==\</i> <i>сходила\ в сервисный центр</i>	отводит взгляд в сторону, прикрывает глаза, корпус тела слегка отводит в сторону
Диалог №2	Невербальные средства
<i>А: включить они / его не могу=</i>	двигает корпус к собеседнику, прищуривается
<i>не== мм ... (0,3)</i>	мотает головой, закрывает глаза, машет рукой перед лицом
<i>прошу прощения\</i> <i>не могут\</i>	кивает, улыбается

Ретроспективная коррекция предполагает редактирование постфактум, т.е. работу с уже вербализованным отрезком. Чаще всего корректировки происходят в плане уточнения или исправления определенных фактов (см. Таблицу 8).

Таблица 8. Примеры компенсации ретроспективной коррекции в русской диалогической спонтанной речи

Диалог №8	Невербальные средства
<i>А: аа ... (0,3) человек сидит / из Индии\</i>	смотрит на собеседника
<i>и вообще\ нифига ^ не слышит\</i>	рубит воздух рукой, смотрит на собеседника
<i>то есть... (0,2) нет\</i>	отводит взгляд в сторону
<i>слышно\ его</i>	машет руками
<i>но иногда .. (0,1) он просто пропадет\</i>	отводит ладонь в сторону

При ретроспективной коррекции отмечается увеличение ритмики и интенсивности мануальных жестов. Говорящий признает ошибку и старается немедленно отменить, «перезаписать» прежде вербализованный кусок. В момент такой «отмены» он машет / трясет руками или рассекает воздух ритмичными движениями ладони(ей). Собственно коррекция компенсируется в большей степени на окулomotorном и целографическом уровнях, а присутствие компенсаторов со стороны мануального канала опционально и

зависит от общей интенсивности жестикуляционного портрета говорящего. Их наличие может характеризоваться одиночными непродолжительными жестами, такими как взмах рукой или кистью. Также зарегистрированы случаи использования таких мультимодальных маркеров, как: сложенные в замок руки, скрещенные внизу руки и т.д.

д) удлинение отдельных звуков

Этот тип речевых сбоев характеризуется пассивностью мануального канала, однако связан с вовлечением окулomotorного и цефалического каналов. Поведение носителя русского языка можно привести к следующей закономерности: запрокидывает голову, водит глазами из стороны в сторону (см. Таблицу 9).

Таблица 9. Пример компенсации удлинений отдельных звуков

Диалог №4	Невербальные средства
<i>А: к нам ранее нн== ни разу \ не приходили с такой/ проблемой\</i>	
<i>но-о ... (0,3) конечно мы по=\</i>	запрокидывает голову, водит глазами из стороны в сторону
<i>постараемся решить эту проблему.</i>	

Таким образом, в настоящем параграфе мы привели примеры восполнения/перераспределения информации из вокальной в кинетическую модальность в русской диалогической спонтанной речи. В ходе анализа нам удалось обнаружить все заявленные в классификации речевые сбои. В зависимости от типа речевых сбоев наблюдаются разные закономерности невербального поведения обоих коммуникантов, объяснение которым мы попытались дать с лингвистической точки зрения.

2.4. Лингвистическое описание закономерностей невербальной компенсации речевых сбоев в китайской спонтанной диалогической речи

Настоящий параграф представляет результаты лингвистического описания закономерностей вербально-жестикуляционной компенсации речевых сбоев на внутреннем и внешнем уровнях в ходе анализа массива данных на китайском языке.

В ходе изучения нарушений беглости речи носителями китайского языка нам удалось выделить все упомянутые в предыдущем параграфе типы речевых сбоев. Ниже мы приводим примеры речевых сбоев и их компенсации на невербальном уровне.

а) маркеры препаративной подстановки (placeholders)

В отличие от большинства согласуемых маркеров препаративной подстановки в русском языке, китайский, относящийся к изолирующему типу языков, может порождать только несогласуемые. Как упоминалось в параграфе 1.2.1., в качестве маркеров препаративной подстановки в китайском языке прежде всего используются местоименная основа ближнего и дальнего дейксиса вместе с универсальным классификатором – 这个 и 那个 [Zhao, Jurafsky, 2005]. Отметим, что в исследуемой выборке широко представлен именно маркер 那个, примеров с 这个 не наблюдается. Мы предполагаем, что это может быть связано с актуализацией объектов повествования из прошлого опыта (как в ситуации с объявлением первого локдауна из-за распространения коронавирусной инфекции, описанной ниже на Рисунках 17–20) или находящихся на дальнем расстоянии от говорящего.

С точки зрения произношения, в ходе спонтанной диалогической речи маркер препаративной подстановки 那个 во всех зарегистрированных нами случаях звучит как [nèige]. Транскрипция, данная в словарях, предполагает прежде всего вариант [nàge].

Диалог №7



Рисунок 17

2020 年的 3 月份我当时

я тогда в марте 2020 года



Рисунок 18

就是在那个... (0,3)

вот в этой ... (0,3)



Рисунок 19

还是那个= 在 xu-u= =

всё еще этой= в ау= =



Рисунок 20

在学校教室吧\

в аудитории\

В представленном отрезке ЭДЕ репарандум, соответствующий содержанию Рисунков 18–19, заменяется студентом найденной лексической единицей 学校教室 («аудитория»). Первая попытка говорящего восполнить возникший пробел при развертывании дискурса (см. Рисунок 18) наблюдается в мануальном и окулomotorном каналах и проявляется в использовании идеографов, показывающих протекание мыслей, а именно движение рукой вперед, поднятие глаз в потолок. Далее на Рисунке 19 запечатлена серия речевых сбоев, проявляющихся в использовании рассматриваемого нами местоимения дальнего дейксиса и в возникновении фальстарта «хи-и». Нарушения в вербальном канале компенсируются учащенной жестикуляцией в мануальном канале (расставил руки, растопырил пальцы). В видеоряде также видно, что говорящий начинает трясти руками, пытаясь вспомнить необходимое слово. Вербализация нужной лексической единицы 学校教室 сопровождается дейктическим жестом (движением рук в сторону). Реагирующий коммуникант расценивает частотность вербализированных речевых сбоев и активность невербальных каналов как стремление иницирующего удержать коммуникативный ход.

б) заполненная пауза

Данный тип хезитации предполагает использование долексических средств (哼、 嗯). Заполненная пауза в диалогах на китайском языке, в сравнении с русской диалогической речью, не демонстрирует повышенной частотности.

Как правило, заполненные паузы компенсируются активной жестикуляцией в окуломоторном и целографическом каналах. Мануальный канал представлен, как правило, жестами-адаптерами. В ходе аннотирования видео-аудио материала, мы обнаружили, что человек, обдумывая свой следующий коммуникативный ход, часто клонит голову вбок, запрокидывает или поворачивает голову в сторону.

в) абсолютная пауза

На проаннотированном материале было подсчитано, что в среднем каждый диалог между китайскими студентами сопровождается 5-тью незаполненными (абсолютными паузами) и 4-мя заполненными паузами. Оба типа паузации происходят в начале ЭДЕ и представляют повторяющийся вербальный паттерн: за абсолютной паузой следует заполненная пауза.

Приведенное коммуникативное событие (см. Рисунки 21–25) начинается с заполненной паузы (哼), аккомпанируемая жестом-иллюстратором типа идеограф (поднять глаза в потолок, водить глазами из стороны в сторону). Далее, как и у носителей русского языка, следующий коммуникативный ход вербализируется при прямом визуальном контакте с собеседником. Последующие абсолютные паузы (см. Рисунки 22–23) возникают вследствие внешних обстоятельств, в ходе которых собеседник А готов передать коммуникативный ход собеседнику Б, но действие второго (открытие бутылки) побудило его продолжить речь. Две абсолютные паузы на Рисунке 23 вероятно носят выжидательный характер, коммуникант пытается «подстроить» свою речь под ситуацию.

Диалог №6



Рисунок 21

A: 哼...(0,6) 我认为这个价位

A: ммм..(0,6) я думаю по этой цене



Рисунок 22

...(0,5) 韩系车为主 ... (0,3)

...(0,5) это в основном корейский автопром ... (0,3)



Рисунок 23

现代啊...(0,7) 起亚啊 ... (0,5)

Хёндай ... (0,4) Киа ... (0,3)



Рисунок 24

B: 韩系车/

B: корейский автопром/



Рисунок 25

A: 对\

A: да\

Мы склонны считать, что абсолютная пауза у китайских студентов является не только показателем размышлений, но и демонстрирует более взвешенные, обдуманые ответы. Если русские студенты стремятся избежать абсолютных пауз, чтобы не «потерять» коммуникативный ход, то студенты из Китая, наоборот, часто прибегают к их использованию. Сдержанность, размеренность проявляется в отсутствии невербальной активности, как таковой. Незаполненные паузы в рассматриваемой нами выборке происходят при визуальном контакте с реагирующим говорящим. Для сравнения такие паузы носителями русского языка воспринимаются как неудобная, неловкая ситуация, поэтому участники эксперимента стараются, наоборот, отвести взгляд (см. Рисунки 8–16).

г) самокоррекции

Как уже было отмечено выше, самокоррекции бывают двух видов – собственно коррекция (онлайн коррекция) и ретроспективная коррекция (офлайн коррекция).

Примечательно, что для китайского устного дискурса более свойственно использование собственно коррекций, нежели ретроспективной коррекции. На 30 самоисправлений онлайн типа приходится лишь 3 офлайн типа.

Речевой сбой в онлайн коррекциях обнаруживается в кратчайший срок, вследствие чего наблюдается повышенная активность окуломоторного и целографического каналов. Активизацию кинетической модальности можно привести к следующей закономерности: человек прикрывает / закрывает глаза, мотает головой, кивает. Если частотность онлайн коррекций в одной ЭДЕ превышает одну, то говорящий может «помочь» себе, протягивая руку или вращая кистью (см. Рисунки 17–20).

Ретроспективная коррекция предполагает внесение изменений постфактум и может проявляться на невербальном уровне схожим мультимодальным комплексом.

д) удлинение отдельных звуков

В силу строения китайского языка принцип удлинения звуков, выделяемый как тип речевого сбоя и направленный на обдумывание следующего коммуникативного хода, выражается отличным от русского языка набором фонем.

По нашему мнению, удлинения можно наблюдать в конце лексических единиц, в основном при перечислении объектов, в виде междометия 啊. Между перечислительными единицами также наблюдается микропаузация. Подобный пример описан на Рисунке 23 диалога №6. Такой тип речевого сбоя отличается пассивностью невербального компонента. Однако, подобно абсолютной паузе, сопровождается прямым зрительным контактом с другим говорящим.

Подводя итог выше сказанному, отметим, что восполнение пробелов в информационном потоке вербального компонента путем активизации невербального компонента свойственно для китайской культуры. В ходе лингвистического анализа невербалики обоих коммуникантов, мы сделали попытку выявить и описать основные закономерности их вербально-жестикуляционного поведения, которые станут основой для создания мультимодальных схем компенсации речевых сбоев.

2.5. Оптимизация мультимодальных корпусов диалогической спонтанной речи на русском и китайском языках для машинного обучения

Проанализированный в предыдущих параграфах эмпирический материал в очередной раз подтверждает идею мультимодальности человеческой коммуникации. Сформированный много сотен лет назад навык коммуникации позволил человеку выжить и развиваться в одну из самых прогрессивных форм жизни. Все умения, создающиеся в процессе эволюции, со временем доводятся до автоматизма их носителями, т.е. переходят из ряда осознанных (требующих внимания и полного контроля) к бессознательным и, как следствие, становятся сложными для описания и объяснения. Именно поэтому привести многоаспектный комплексный процесс человеческого взаимодействия к каким-то закономерностям намного сложнее, нежели рассмотреть одну хорошо членимую сущность языка.

Одним из результатов наблюдений исследователя за многоаспектным феноменом коммуникации может являться мультимодальный корпус. В настоящем исследовании мы предприняли попытку разработать корпусы диалогической спонтанной речи на русском и китайском языках, сосредоточившись на двух доступных видах модальностей – вокальной и кинетической, в каждую из которых входит несколько модусов (каналов) (см. Рисунок 26).

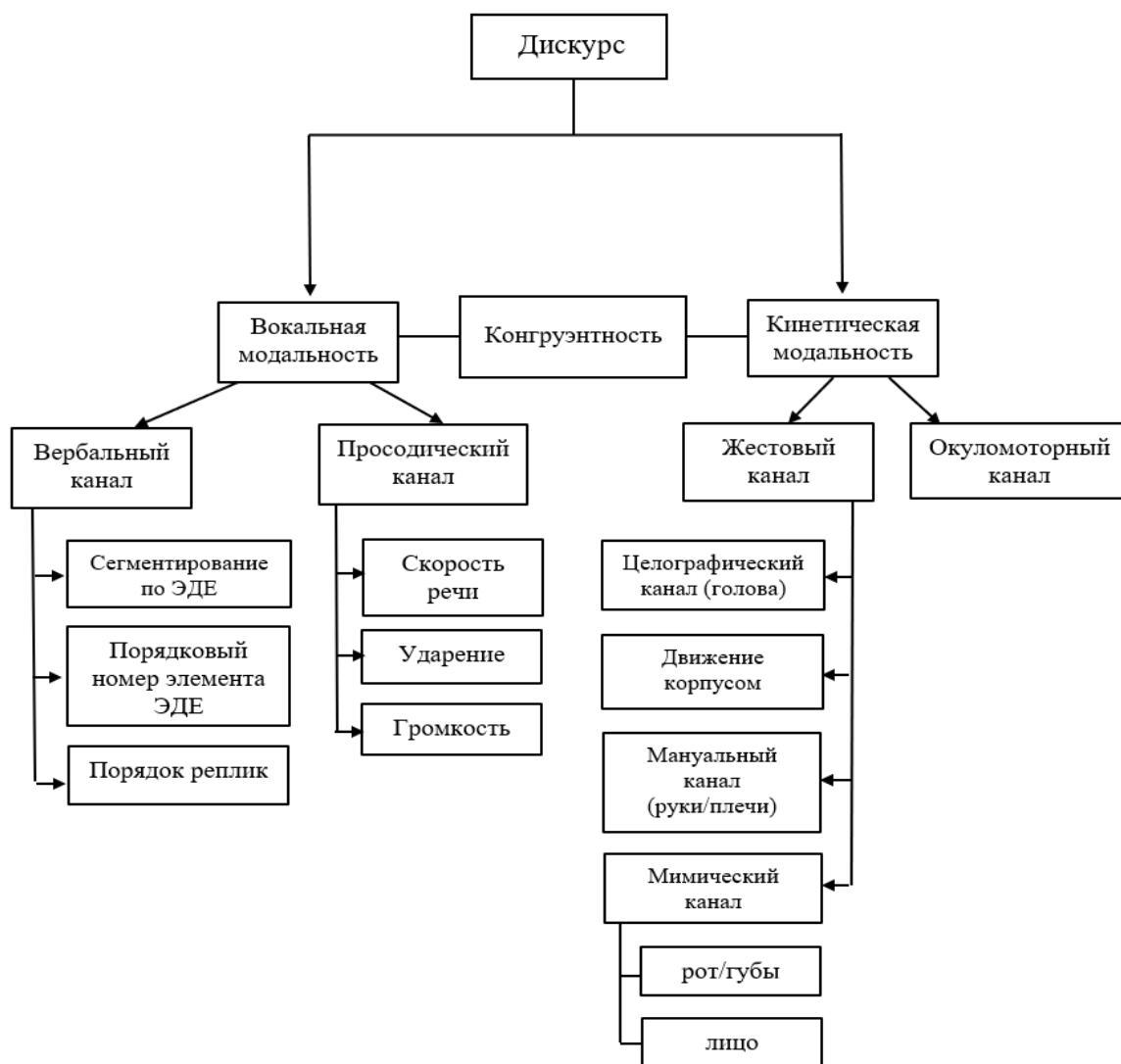


Рисунок 26. Структура мультимодальных корпусов спонтанной диалогической речи

Приведенная выше структура мультимодальных корпусов не является конечной и может включать другие типы размеченной информации, например, с точки зрения акцентологического или прагматического аспектов.

Собранные нами массивы данных на двух языках были размечены в программе ELAN. Разработанный Институтом психолингвистики им. М. Планка в 2000 году, ELAN представляет собой компьютерное обеспечение, предлагающее набор инструментов для ручного или полуавтоматического аннотирования аудио- или видеоматериалов. Программа оснащена возможностью построения многослойной аннотации с привлечением нескольких разметчиков.

Разметка позволяет задать текстовые характеристики, скоординировав их с лингвистическими и экстралингвистическими факторами коммуникативного поведения людей. Задав слои аннотирования, согласно изображенной на Рисунке 26 структуре, мы тем самым выровняли аудио-видеоряд и наделили смыслом каждый значимый фрагмент. В рамках программы ELAN также возможно решить проблему ручной нарезки видео- и аудиоматериалов по аннотированным сегментам, т.к. размеченный файл уже содержит ссылку на медийные материалы и набор тегов с указанием временным промежутков. Выгрузку файлов можно осуществить в нескольких доступных вариантах, в том числе в XML формате, где будут указаны слои для распознавателя (см. Рисунок 27).

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```

▼<TIERS xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="file:avatech-tiers.xsd">
  ▼<TIER columns="голова_-_2-ср">
    ▼<span start="49.261" end="50.613">
      <v>вытянуть шею</v>
    </span>
  </TIER>
  ▼<TIER columns="глаза/брови_-_2-ср">
    ▼<span start="49.233" end="50.602">
      <v>покоситься на собеседника</v>
    </span>
  </TIER>
  ▼<TIER columns="рот/губы_2">
    ▼<span start="49.241" end="50.609">
      <v>улыбнуться</v>
    </span>
  </TIER>
  ▼<TIER columns="конгруэнтность">
    ▼<span start="47.052" end="49.383">
      <v>1</v>
    </span>
    ▼<span start="49.383" end="49.999">
      <v>1</v>
    </span>
    ▼<span start="50.024" end="50.597">
      <v>1</v>
    </span>
    ▼<span start="50.615" end="51.806">
      <v>1</v>
    </span>
  </TIER>
  ▼<TIER columns="корпус_-_2">
    ▼<span start="49.282" end="50.435">
      <v>переминаться с ноги на ногу/придвинуться к собеседнику</v>
    </span>
  </TIER>

```

Рисунок 27. Фрагмент размеченного файла из мультимодального корпуса диалогической спонтанной речи

Как мы выяснили в предыдущих параграфах, «живое» взаимодействие участников спонтанного диалога предполагает наличие помех в вербальном канале, связанных с возникновением речевых сбоев. Подтвердив эмпирическими наблюдениями и лингвистическим описанием нашу гипотезу

об активации каналов кинетической модальности в моменты нарушений развертывания речевого потока, мы далее приступили к выявлению закономерностей невербальной компенсации речевых сбоев.

Некоторые паттерны в вербально-жестикуляционном поведении, выделенные нами в результате лингвистического анализа и зафиксированные в виде мультимодальных схем, легли в основу коммуникативного события. Напомним, что в настоящей работе сегментирование диалогической спонтанной речи на коммуникативные события компенсации речевых сбоев предполагает включение информации, исходящей от обоих участников, а следовательно, может реализовываться в трех возможных сценариях: обратная связь (сценарий А), инициатива смены (сценарий Б), и вовлеченная смена (сценарий В).

Оптимизация уже первично обработанных мультимодальных корпусов с лингвистической точки зрения предполагает вторичное сегментирование данных путем добавления слоя с аннотацией наблюдаемого вида коммуникативного события. Разметка на данном этапе представлена в иерархическом порядке, где дублируется информация основного мультимодального потока, относящаяся непосредственно к рассматриваемому коммуникативному событию (см. Рисунок 28).

руки/плечи - 1 ТС	держат предмет в руках 00:00:47.050 - 00:00:51.824
конгруэнтность ТС	1 00:00:47.052 - 00:00:49.383
Транскрипция HTML- участник 1 ТС	[pause] 00:00:48.499 - 00:00:49.390
коммуникативное событие голова - 1-ср голова - 2-ср руки/плечи - 1-ср руки/плечи - 2-ср глаза/брови - 1-ср глаза/брови - 2-ср Транскрипция HTML- участник 1-ср Транскрипция HTML - участник 2-ср ТС	коммуникативное событие 1-Б повернуть голову на собеседника вытянуть шею держат предмет в руках скрестить руки внизу покоситься на собеседника/поднять брови покоситься на собеседника [pause] 解决 [interrupt] 可以 最好帮我换一个方案 00:00:48.511 - 00:00:51.824
голова - 1-ср Транскрипция HTML- участник 1-ср	наклонить голову вниз / мотать головой 方案

Рисунок 28. Фрагмент представления иерархической структуры коммуникативных событий в формате многострочного текста

Иерархичность, задаваемая внутри одного файла разметки, позволяет структурировать материал по отдельным классам и представить информацию в более сжатом формате. Программа ELAN также позволяет выгрузить как один единый файл сплошной и выборочной разметки, так и только выборочной – слоя коммуникативного события с разными его видами.

В размеченных коммуникативных событиях с указанием типа сценария их протекания мы наблюдаем созависимость модальностей, которые можно продемонстрировать в виде наблюдаемых нами мультимодальных схем. Ниже представлены основные зарегистрированные нами коммуникативные события, состоящие из двух мультимодальных схем – одна для иницилирующего говорящего, другая для реагирующего. Мультимодальные схемы выстроены с учетом классификации речевых сбоев, по принципу стимул-реакция. Каждый речевой сбой порождает набор невербальных реакций, регистрируемых нами в трех основных каналах – мануальном, окуломоторном и целографическом. Выбор заявленных мультимодальных каналов объясняется их наибольшей частотностью, тем не менее при обнаружении положительных результатов в ходе обучения компьютерного агента не исключено расширение количества модусов коммуникации.

Коммуникативные события, представляя конструкт, идущий от внутреннего к внешнему уровню организации спонтанной коммуникации, представлены в русском дискурсе следующими сценариями:

- 1) коммуникативное событие №1:
 - А) абсолютная пауза + обратная связь;
 - Б) абсолютная пауза + инициатива смены;
- 2) коммуникативное событие №2:
 - А) заполненная пауза + обратная связь;
 - Б) заполненная пауза + инициатива смены;
- 3) коммуникативное событие №3:
 - А) удлинение звуков + обратная связь;

- 4) коммуникативное событие №4:
- Б) маркеры препаративной подстановки + инициатива смены;
 - В) маркеры препаративной подстановки + вовлеченная смена;
- 5) коммуникативное событие №5:
- А) самокоррекции + обратная связь.

Итак, рассмотрим каждое событие в виде мультимодальных схем иницилирующего и реагирующего собеседников.

Коммуникативное событие №1 представлено абсолютной паузой и может сопровождаться двумя сценариями, а именно нейтральной обратной реакцией собеседника (см. Таблицу 10), предполагающей нединамичное жестикулирование и отсутствие вербального компонента, и инициативой смены (см. Таблицу 11), характеризующейся активацией как вербального, так и невербального компонента.

Таблица 10. Сценарий А коммуникативного события №1: абсолютная пауза + обратная связь

Тип речевого сбоя в вербальном канале	Мануальный канал	Окулоmotorный канал	Целографический канал
Мультимодальная схема иницилирующего собеседника			
Абсолютная пауза (от 0,1 до 0,3 сек.)	Продолжение жеста, аккомпанирующего предыдущую фразу / ритмические жесты	Прищуриться / потупить => поднять глаза в поток (быстро) / моргать => посмотреть на собеседника	Нейтральное положение / наклонить голову набок / запрокинуть голову (слегка)
Мультимодальная схема реагирующего собеседника			
-	Продолжение жеста, аккомпанирующего предыдущую фразу / нейтральное положение	Смотреть на собеседника (пристально)	Нейтральное положение

Таблица 11. Сценарий Б коммуникативного события №1: абсолютная пауза + инициатива смены

Тип речевого сбоя в вербальном канале	Мануальный канал	Окуломоторный канал	Целографический канал
Мультимодальная схема иницирующего собеседника			
Абсолютная пауза (от 0,3~0,4 сек.)	Нейтральное положение / продолжение жеста, аккомпанирующего предыдущую фразу / развести руками (пожать плечами)	Отвести взгляд / потупить / водить глазами из стороны в сторону	Опустить голову / наклонить голову набок / качать головой
Мультимодальная схема реагирующего собеседника			
Могут присутствовать: фальстарты	Протянуть руку(и) к собеседнику / двинуть кисть(и) к собеседнику	Покоситься на собеседника (пристально)	Двинуть головой вперед / вытянуть шею

Коммуникативное событие №2 включает два сценария: заполненная пауза и следующая за ней обратная связь (см. Таблицу 12) и заполненная пауза, заканчивающаяся инициативой смены коммуникативного хода (см. Таблицу 13).

Таблица 12. Сценарий А коммуникативного события №2: заполненная пауза + обратная связь

Тип речевого сбоя в вербальном канале	Мануальный канал	Окуломоторный канал	Целографический канал
Мультимодальная схема иницирующего собеседника			
Заполненная пауза (mmm, эээ, aaa)	Нейтральное положение	Поднять глаза в потолок=> водить глазами из стороны в сторону	Запрокинуть голову
	Потирать руки / обхватить ладонью кулак	Отвести взгляд в сторону / опустить глаза	Повести головой вбок
	Потирать горло	Поднять глаза в потолок / опустить глаза=> водить глазами из стороны в сторону	Повести головой вбок

Тип речевого сбоя в вербальном канале	Мануальный канал	Окулоmotorный канал	Целографический канал
Мультимодальная схема реагирующего собеседника			
-	Нейтральное положение /жесты-адаптеры	Посмотреть на собеседника + моргать (опционально)	Нейтральное положение

Таблица 13. Сценарий Б коммуникативного события №2: заполненная пауза + инициатива смены

Тип речевого сбоя в вербальном канале	Мануальный канал	Окулоmotorный канал	Целографический канал
Мультимодальная схема иницирующего собеседника			
Заполненная пауза (<i>mmm, эээ, aaa</i>)	Продолжение жеста, аккомпанирующего предыдущую фразу=> жесты-адаптеры / развести руками	Потупить=> моргать=> отвести взгляд	Опустить голову / наклонить голову набок
Мультимодальная схема реагирующего собеседника			
Долексические маркеры с элементами удлинения звуков (<i>ну-у</i>)	Нейтральное положение	Посмотреть на собеседника / проследить жесты собеседника	Нейтральное положение

Согласно данным, полученным в ходе разметки и предварительного анализа выборки, коммуникативное событие №3 в русском дискурсе обладает сравнительной частотностью, но сопровождается только нейтральной реакцией со стороны второго говорящего (см. Таблицу 14). В отличие от заполненной паузы, представляющей по большей части вокализацию только гласных (*aaa, эээ*) или только согласных (*mmm*), в удлинениях присутствуют и вербализуются оба типа звуков в слове.

Таблица 14. Сценарий А коммуникативного события №3: удлинения звуков + обратная связь

Тип речевого сбоя в вербальном канале	Мануальный канал	Окулоmotorный канал	Целографический канал
Мультимодальная схема иницирующего собеседника			
Длительная вокализация звуков (<i>ну-у, нн-и разу, но-о, попо-озже</i>)	Нейтральное положение	Водить глазами из стороны в сторону	Запрокинуть голову=> вытягивать подбородок

Тип речевого сбоя в вербальном канале	Мануальный канал	Окуломоторный канал	Целографический канал
Мультимодальная схема реагирующего собеседника			
-	Нейтральное положение	Отвести взгляд / смотреть на собеседника	Нейтральное положение / наклонить голову набок

В коммуникативном событии №4 русского мультимодального корпуса мы можем прототипировать две основные реакции на образование в речи маркеров препаративной подстановки, а именно инициативу смены и вовлеченную смену. При инициативе смены реагирующий говорящий пытается «поймать» общий ритм повествования, прогнозируя упущенную лексическую единицу или достраивая высказывание за первым говорящим (см. Таблицу 15). Маркеры препаративной подстановки на русском языке имеют согласуемый тип (*эта, этот, эти, тот, те, та, ну такая ... знаешь типа ...*), и несогласуемый тип (*это самое как его..., ну как его*) и т.д.

Таблица 15. Сценарий Б коммуникативного события №4: маркеры препаративной подстановки + инициатива смены

Тип речевого сбоя в вербальном канале	Мануальный канал	Окуломоторный канал	Целографический канал
Мультимодальная схема иницирующего собеседника			
Маркеры препаративной подстановки	Задавать направление=> махать руками / вращать кистями	Прикрыть глаза / поднять глаза в потолок / отвести взгляд в сторону	Наклонить голову набок / двинуть головой назад / запрокинуть голову
Мультимодальная схема реагирующего собеседника			
Могут присутствовать: - заполненная пауза - удлинения отдельных звуков - онлайн самокоррекции (фальстарты)	Нейтральное положение	Прикрыть глаза	Двинуть головой вперед / кивнуть (да) / мотнуть (нет)
	Двинуть рукой вперед / выставить руку вперед / махать рукой	Посмотреть на собеседника + поднять брови (опционально)	Двинуть головой вперед / повести головой вбок

Несколько другая ситуация складывается при реакции в виде вовлеченной смены, в ходе которой инициирующий говорящий вербально-невербально сигнализирует второму о возникшем затруднении и подталкивает своего собеседника на размышление. В отличие от предыдущего сценария, здесь особо видна актуализация жестовых движений, направленных не только на информирование и желание восполнить образовавшийся речевой пробел, но и на стремление к кооперативному взаимодействию с собеседником. Это выражается в использовании жестов-пиктографов, прямых взглядах на реагирующего говорящего, движении головой/шеей вперед (см. Таблицу 16).

Таблица 16. Сценарий В коммуникативного события №4: маркеры препаративной подстановки + вовлеченная смена

Тип речевого сбоя в вербальном канале	Мануальный канал	Окуломоторный канал	Целографический канал
Мультимодальная схема инициирующего собеседника			
Маркеры препаративной подстановки	Махать руками / вращать кистями / щелкать пальцами => жест-пиктограф	Прикрыть глаза / поднять глаза в потолок / отвести взгляд в сторону => посмотреть на собеседника	Наклонить голову набок / двинуть головой назад / запрокинуть голову => нейтральное положение / двинуть головой вперед / вытянуть шею
Мультимодальная схема реагирующего собеседника			
Могут присутствовать: - абсолютная пауза; - заполненная пауза; - удлинения отдельных звуков; - онлайн самокоррекции (фальстарты)	Нейтральное положение	Покоситься на собеседника => опустить глаза / посмотреть вдаль	Мотать головой / поднять голову
	нейтральное положение / протянуть руку к собеседнику	посмотреть на собеседника + поднять брови (опционально)	кивнуть

Последнее выделенное нами коммуникативное событие №5 представлено одним сценарием внешнего уровня – обратной связью, несмотря на то что самокоррекции могут реализовываться в двух типах – собственно коррекции и ретроспективной коррекции (см. Таблицу 17). Собственно коррекция сопровождается фальстартом, например, «не==[interrupt] ... +

абсолютная/заполненная пауза (от 0,2 сек.) + лексическая единица найдена – *не могут*». Основной формулой ретроспективной коррекции в русской речи является: «*то есть нет \ ... X*» но *Y* или «*не X а Y*».

Таблица 17. Сценарий А коммуникативного события №5: самокоррекции + обратная связь

Тип речевого сбоя в вербальном канале	Мануальный канал	Окуломоторный канал	Целографический канал
Мультимодальная схема иницирующего собеседника			
- Собственно коррекция	Нейтральное положение	Закатить глаза / прикрыть глаза / отвести взгляд в сторону	Мотнуть головой / кивнуть; мотать головой / кивать
- Ретроспективная коррекция	Трясти руками / махать руками / рассекать воздух ритмичными движениями ладони	Посмотреть на собеседника => отвести взгляд в сторону => закатить глаза/прикрыть глаза	Мотнуть головой / повести головой вбок
Мультимодальная схема реагирующего собеседника			
Могут присутствовать: поддакивания	Продолжение жеста, аккомпанирующего предыдущую фразу / нейтральное положение	Смотреть на собеседника (пристально) / прищуриться	Нейтральное положение / кивнуть

Приведем далее классификацию коммуникативных событий, найденных нами в китайском спонтанном дискурсе со следующими типами сценариев их развития:

- 1) коммуникативное событие №1:
 - А) абсолютная пауза + обратная связь;
 - Б) абсолютная пауза + инициатива смены;
- 2) коммуникативное событие №2:
 - А) заполненная пауза + обратная связь;
 - Б) заполненная пауза + инициатива смены;
- 3) коммуникативное событие №3:
 - А) удлинение звуков + обратная связь;

4) коммуникативное событие №4:

А) маркеры препаративной подстановки + обратная связь;

5) коммуникативное событие №5:

А) самокоррекции + обратная связь.

Так же, как и в случае данных на русском языке, рассмотрим каждое событие в виде мультимодальных схем иницирующего и реагирующего собеседника на материале китайского языка.

Коммуникативное событие №1 со стимулом в виде абсолютной паузы у иницирующего собеседника может сопровождаться пассивностью вербального и невербального компонента у другого собеседника (см. Таблицу 18), либо привести к инициативе смены (см. Таблицу 19). Отметим, что в отличие от русской речи, где прослеживается взаимосвязь между временным промежутком абсолютной паузы и выбором последующего сценария реагирующим коммуникантом, китайская речь не показывает такой зависимости и принятие решения о прерывании/непрерывании иницирующего говорящего больше строится на полученной контекстуальной информации.

Таблица 18. Сценарий А коммуникативного события 1: абсолютная пауза + обратная связь

Тип речевого сбоя в вербальном канале	Мануальный канал	Окулоmotorный канал	Целографический канал
Мультимодальная схема иницирующего собеседника			
Абсолютная пауза (от 0,2 сек. – в совокупности с другими речевыми сбоями и от 0,6 сек. – самостоятельное использование)	Продолжение жеста, аккомпанирующего предыдущую фразу / нейтральное положение	Опустить глаза / потупитья	Опустить голову
Мультимодальная схема реагирующего собеседника			
-	продолжение жеста, аккомпанирующего предыдущую фразу / нейтральное положение	смотреть на собеседника / отвести взгляд / потупитья	нейтральное положение

Таблица 19. Сценарий Б коммуникативного события №1: абсолютная пауза + инициатива смены

Тип речевого сбоя в вербальном канале	Мануальный канал	Окуломоторный канал	Целографический канал
Мультимодальная схема иницирующего собеседника			
Абсолютная пауза (от 0,2 сек. – в совокупности с другими речевыми сбоями и от 0,6 сек. – самостоятельное использование)	Нейтральное положение => тряхи руками / вращать кистями	Посмотреть на собеседника + поднять брови / моргать (опционально) => водить глазами из стороны в сторону	Нейтральное положение / наклонить голову набок
Мультимодальная схема реагирующего собеседника			
Могут присутствовать: заполненные паузы	Нейтральное положение / двинуть кисти к собеседнику	Смотреть на собеседника (пристально) + поднять брови (опционально)	Двинуть головой вперед / кивнуть / мотнуть головой

Коммуникативное событие №2 состоит на внутреннем уровне из речевого сбоя в виде заполненной паузы, которая связана либо с обратной связью – сценарий А (см. Таблицу 20), либо с инициативой смены – сценарий Б (см. Таблицу 21).

Инициативность в смене коммуникативного хода проявляется в тот момент, когда иницирующий коммуникант начинает использовать невербальные сигналы, демонстрирующие его отвлечение от темы или стремление обратиться за необходимой информацией извне. Иными словами, если реагирующий коммуникант считывает жестовую активность, направленную на взаимодействие с другими элементами окружающей среды, он предпринимает попытку вернуть фокус внимания иницирующего собеседника на себя, подхватывая предыдущий коммуникативный ход. Актуализация жестов-идеографов в виде движений глаз из стороны в сторону или наклонов голову набок и сопровождающееся вокальное заполнение паузы звуками показывают ход или протекание мыслей иницирующего собеседника и не приводят к смене ролей коммуникантов.

Таблица 20. Сценарий А коммуникативного события №2: заполненная пауза + обратная связь

Тип речевого сбоя в вербальном канале	Мануальный канал	Окуломоторный канал	Целографический канал
Мультимодальная схема иницирующего собеседника			
Заполнения паузы звуками (哼、嗯)	Нейтральное положение	Потупиться => водить глазами из стороны в сторону	Запрокинуть голову / наклонить голову набок
Мультимодальная схема реагирующего собеседника			
-	Нейтральное положение (возможна смена жеста)	Покоситься на собеседника => отвести взгляд / потупиться	Нейтральное положение / наклонить голову набок

Таблица 21. Сценарий Б коммуникативного события №2: заполненная пауза + инициатива смены

Тип речевого сбоя в вербальном канале	Мануальный канал	Окуломоторный канал	Целографический канал
Мультимодальная схема иницирующего собеседника			
Заполнения паузы звуками (哼、嗯)	Нейтральное положение	Посмотреть в сторону / отвести взгляд => опустить глаза / водить глазами из стороны в сторону	Повернуть голову в сторону
Мультимодальная схема реагирующего собеседника			
Могут присутствовать фальстарты	Продолжение жеста, аккомпанирующего предыдущую фразу / растопырить пальцы => сжать пальцы	Смотреть на собеседника пристально / осмотреть собеседника	Двинуть голову вперед + вытянуть шею

Коммуникативное событие №3 представлено удлинениями звуков у одного коммуниканта и нейтральной реакцией у другого (см. Таблицу 22). Тем не менее, нам удалось классифицировать два типа удлинений звуков – достраивания звуков и непосредственное удлинение гласных в союзах или союзных словах, выделив две основных мультимодальных схемы у иницирующего говорящего. Приведем пример длительной вокализации финалей в конце лексических единиц при перечислении, сопровождающейся

дистраиванием звуков (啊、呀): 我们可以…(0, 4) 在一起跳舞啊唱歌啊这类的 (*перевод наш: «мы можем …(0,4) вместе потанцева-ать попе-еть вот это вот всё»*).

Таблица 22. Сценарий А коммуникативного события №3: удлинение звуков + обратная связь

Тип речевого сбоя в вербальном канале	Мануальный канал	Окуломоторный канал	Целографический канал
Мультимодальная схема иницирующего собеседника			
Дистраивание звуков (啊、呀) к лексическим единицам	Нейтральное положение	Посмотреть на собеседника	Нейтральное положение
Удлинение гласных в союзах, союзных словах (然后、还有、和 и т.д.)	Нейтральное положение / продолжение жеста, аккомпанирующего предыдущую фразу	Смотреть на собеседника => посмотреть в потолок / водить глазами из стороны в сторону	Нейтральное положение
Мультимодальная схема реагирующего собеседника			
-	Нейтральное положение	Смотреть на собеседника / потупитья	Нейтральное положение

Коммуникативное событие №4 в китайском спонтанном дискурсе наблюдается только с одним типом реакции – обратной связью. Вероятно, это может быть связано с частотными фальстартами (这=/那=) иницирующего собеседника, заполняющими пробелы в речи и не дающими возможности реагирующему проявить себя вербально. В итоге, иницирующий говорящий либо вспоминает «забытую» лексическую единицу, либо оставляет сбой без окончательного варианта репаратуры, считывая невербальные сигналы реагирующего, свидетельствующие о его контекстуальном понимании ситуации на основе пресуппозитивной базы знаний, например, взгляд на собеседника, смена положения тела после перехода к новому фрагменту реагирующим коммуникантом, кивки (см. Таблицу 23).

Таблица 23. Сценарий А коммуникативного события №4: маркеры препаративной подстановки + обратная связь

Тип речевого сбоя в вербальном канале	Мануальный канал	Окулоmotorный канал	Целографический канал
Мультимодальная схема иницирующего собеседника			
Маркеры препаративной подстановки	Задавать направление рукой (начало (если нет в конце), опционально) => трясти рукой / махать руками => задавать направление рукой (конец, опционально)	Покоситься на собеседника => отвести взгляд / водить глазами из стороны в сторону => посмотреть на собеседника + поднять брови (опционально)	Трясти головой (слабо)
	Жесты-адаптеры	Отвести взгляд / водить глазами из стороны в сторону	Повести головой вбок / нейтральное положение
Мультимодальная схема реагирующего собеседника			
-	Нейтральное положение (возможна смена жеста)	Покоситься на собеседника => отвести взгляд / потупить	Нейтральное положение + кивнуть (опционально)

Коммуникативное событие №5 реализовано в двух типах самокоррекции – собственно коррекции и ретроспективной коррекции в рамках стимула и последующей нейтральной реакцией со стороны второго коммуниканта (см. Таблицу 24). Собственно коррекции (коррекция онлайн) включают фальстарт (части инициали или финали) и повторение ранее вербализованного отрывка ЭДЕ в точности/с изменениям. Приведем пример ниже:

А: 我就去 xi== [interrupt] xi== [interrupt] 统计 ... (перевод наш: «ну тогда я пойду ши==[interrupt] ши==[interrupt] посчитаю ...»)

Б: 没问题 谢==[interrupt] 麻烦您了 (перевод наш: «нет проблем спа== [interrupt] извините за хлопоты»)

Ретроспективная коррекция предполагает вербализацию ЭДЕ, содержащую грамматическую, лексическую или смысловую ошибку с последующей паузацией (абсолютная пауза/заполненная пауза 哼) и

внедрением потерянной или исправлением неправильной части реплики:
 我俩== [interrupt]... (0,2) 我算过 我俩半年时间可能出门的话都不到 5 次
 (*перевод наш: «мы оба== [interrupt]... (0,2) я посчитал мы оба за полгода
 наверное выходили на улицу не больше 5 раз»*).

Таблица 24. Сценарий А коммуникативного события 5: самокоррекции + обратная связь

Тип речевого сбоя в вербальном канале	Мануальный канал	Окулоmotorный канал	Целографический канал
Мультимодальная схема иницирующего собеседника			
- Собственно коррекция	Нейтральное положение / протянуть руку к собеседнику / вращать кистью	Прикрыть глаза / отвести взгляд в сторону	Мотнуть головой / кивнуть; мотать головой / кивать
- Ретроспективная коррекция	Нейтральное положение	Посмотреть на собеседника	Повести головой вбок / нейтральное положение
Мультимодальная схема реагирующего собеседника			
-	Нейтральное положение / продолжение жеста, аккомпанирующего предыдущую фразу	Покоситься на собеседника / потупить / отвести взгляд	Нейтральное положение

Невербальный компонент обладает важной контекстуальной и смысловой информацией, поэтому даже частичное восполнение помех одной модальности информацией из другой, вероятно снизит нагрузку на систему. Как мы полагаем, создание мультимодальной модели машинного обучения в случае с диалогической спонтанной речью может быть осуществлено с применением указанных в п. 1.4. типов стратегий – «слияние» (fusion) и «перевод» (translation). Перевод подразумевает перераспределение внимания, организуемое с помощью «фильтров внимания» (attention gates), которым в нашем случае является размеченное коммуникативное событие. Восполнив недостающий элемент, далее система может продолжить

беспрепятственно организовывать общий блок мультимодальной информации путем слияния данных (см. Рисунок 29).

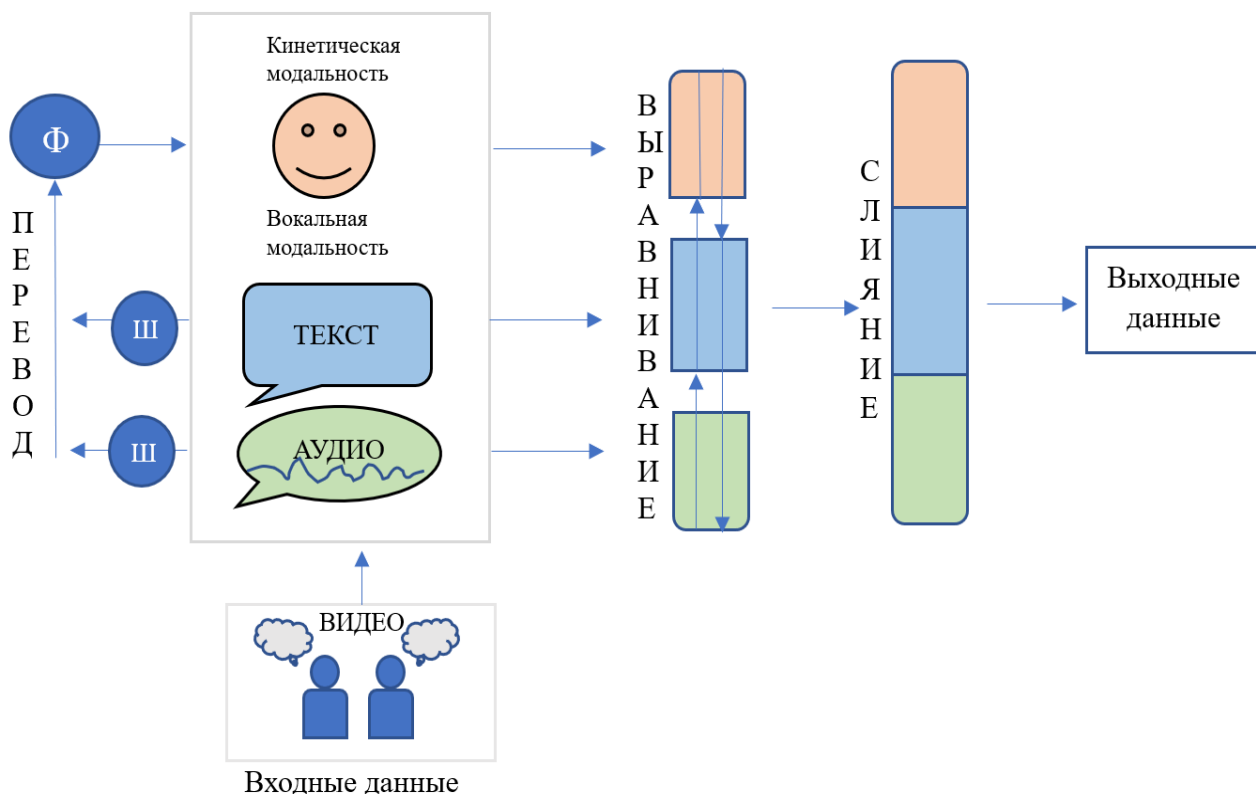


Рисунок 29. Пример мультимодальной автоматизируемой последовательности обучения, где а) непрерывный поток данных со всех проаннотированных каналов проходит стадии выравнивания и слияния, б) поток вокальной информации при возникновении шумов ШШШ проходит через фильтр Φ, подкрепляясь информацией из каналов кинетической модальности и далее также проходит через процесс слияния

Таким образом, сплошная первичная обработка данных с помощью разметки позволяет глобально оптимизировать мультимодальные массивы для применения в машинном обучении, приравнивая корпуса к наборам данных/датасетам. Дополнительное сегментирование в виде разметки коммуникативных событий, регистрирующих сценарии компенсации речевых сбоев, сужает круг переменных для решения отдельных задач в обучении. Так, формируется набор целевых атрибутов, выраженный в отдельном коммуникативном событии и представляющий локальную оптимизацию созданного датасета.

Созданные в рамках данной магистерской работы и оптимизированные для машинного обучения мультимодальные корпуса диалогической спонтанной речи обладают рядом характеристик, представленных в таблице ниже (см. Таблица 25).

Таблица 25. Характеристики мультимодальных корпусов диалогической спонтанной речи на русском и китайском языках

Характеристики корпусов	Русский корпус	Китайский корпус
Тип корпуса	Мультимодальный	Мультимодальный
Размер	Пополняется Кол-во слоев аннотации: 20 Кол-во строк: 10641 Кол-во аннотаций: 159405 (из них пустых 50220)	Пополняется Кол-во слоев аннотации: 20 Кол-во строк: 4275 Кол-во аннотаций: 54386 (из них 21204 пустых)
Видеоматериал	16 записей	10 записей
Время видеоматериала	44 мин. 6 сек.	21 мин. 56 сек.
Материал	Диалогическая спонтанная речь	Диалогическая спонтанная речь
Участники диалогов	Студенты ИФиЯК СФУ (носители)	Студенты ИФиЯК СФУ (носители)
Тематика	Официальный стиль – взаимодействие клиента и консультанта; Неофициальный стиль – обсуждение современных проблем между друзьями/коллегами	Официальный стиль – взаимодействие клиента и консультанта; Неофициальный стиль – обсуждение современных проблем между друзьями/коллегами
Разметчик	1 чел. (+ пробные разметки 5 чел.)	1 чел.
Сплошная разметка	Дискурсивная	Дискурсивная
Дополнительная (выборочная) разметка	Сегментирование по коммуникативным событиям компенсации речевых сбоев	Сегментирование по коммуникативным событиям компенсации речевых сбоев
Тип разметки	Ручная	Ручная
Тип хранения	Файл XML (ELAN)	Файл XML (ELAN)
Доступ	Закрытый	Закрытый
Назначение	Исследовательский	Исследовательский
Динамичность	Динамический	Динамический
Период производства	2021–2023 гг.	2022–2023 гг.
Статус производства	В процессе	В процессе
Заказчик	Promobot	Promobot
Оборудование	Sony ZV-1 (основная камера); POCO F3 (боковые камеры); Voya BY-WM3U (беспроводной микрофон)	Sony ZV-1 (основная камера); POCO F3 (боковые камеры); Voya BY-WM3U (беспроводной микрофон)

В данном параграфе мы попытались обобщить опыт глобальной оптимизации мультимодальных корпусов под нужды машинного обучения и систематизировать получившиеся в процессе локальной оптимизации данных результаты, представив их в виде коммуникативных событий, реализованных в трех возможных сценариях и представленных мультимодальными схемами компенсации речевых сбоев обоих коммуникантов. Сегментирование непрерывных данных на коммуникативные события во временной последовательности, которую можно задать благодаря разметке в программе ELAN, далее должно стремиться к их объединению и формированию единой системы знаний. Предполагаем, что в мультимодальном машинном обучении это может быть воплощено путем применения двух стратегий – перевода и слияния данных.

ВЫВОДЫ ПО ГЛАВЕ 2

Данная глава посвящена описанию опыта построения и оптимизации мультимодальных корпусов диалогической спонтанной речи на русском и китайском языках для машинного обучения. В рамках практической главы предпринята попытка разработать лингвистическую базу и обосновать выбор структуры для создания и оптимизации заявленных корпусов. Для выполнения поставленной цели настоящей работы мы поставили эксперимент, направленный на сбор данных, провели с нуля сплошную дискурсивную разметку данных, тем самым выровняв модальности и осуществив глобальную оптимизацию корпусов, выделили основные характеристики заданных переменных и наблюдаемых параметров путем статистической обработки данных, результаты которой легли в основу предлагаемой нами формы локальной оптимизации, заключающейся в дополнительной выборочной разметке коммуникативных событий, регистрирующих реакции коммуникантов на помехи в вербальном канале.

Важным шагом для проведения качественного анализа мультимодальных данных является планирование и тестирование хода эксперимента. В данной работе имеет место квазиэксперимент, предполагающий использование уже существующих условий и естественных групп, что несколько приближает его к естественной среде и позволяет сформировать более экологичную версию эксперимента. Под экологичностью понимается моделирование адекватной реакции живого существа в искусственных условиях, с предположением, что такие реакции будут свойственны для него и в природной среде.

Формируя инструментарий для квазиэксперимента, мы выделили зависимые и независимые переменные. Под независимыми, т.е. неизменяющимися, переменными мы понимаем национальность, пол участников, общую тему ведения спонтанных диалогов. Зависимыми, т.е. поддающимися манипуляции, переменными являются: процесс

развертывания спонтанного диалогического дискурса, выбор невербальных и вербальных средств, пространственное расположение участников диалога.

Несмотря на то, что способ построения диалога, выбор пунктов для коммуницирования оставался за участниками эксперимента, основные темы карточек представляют независимую переменную. На основе согласованного с компанией Promobot списка тем разговоров, мы создали опорные карточки на двух языках, тематика которых делится на два блока, а именно: 1) обсуждение проблем между друзьями/коллегами и 2) взаимодействие клиента и консультанта.

В ходе определения основных характеристик заявленных нами переменных мы сосредоточились на выявлении корреляций более объективным путем, обратившись к математическим подсчетам. Статистическая обработка мультимодальных данных даже путем проведения элементарных математических операций оказалась весьма продуктивной. Суммируем некоторые наиболее актуальные для данной работы результаты.

Так, рассмотрев параметры *время* и *дискурсивное сегментирование*, мы пришли к выводу о том, что временной переход от одного смыслового фрагмента к другому у китайцев происходит быстрее, чем у русских. Среднее значение ЭДЕ у носителей русского языка составило 14,8, и 17,27 у носителей китайского языка соответственно. Так, мы пришли к выводу, что сегментирование по временным и дискурсивным показателям отличается произвольностью.

Анализируя частотность проявлений жестов в таких каналах, как мануальный (руки/плечи), движение корпусом, целографический, окулomotorный и мимический, мы обнаружили разные закономерности актуализации кинетической модальности. За минуту спонтанного диалога русский может в среднем совершить 34,56 мануальных жестов, китаец – только 15,36. Тем не менее, невербалика китайских студентов может перераспределяться на другие каналы, а именно: движение корпусом, целографический и окулomotorный. Например, движения корпусом

у китайцев в 1,6 раз активнее, чем у русских. На основании подобного статистического подсчета мы выбрали наиболее активные типы невербальных каналов для создания мультимодальных схем компенсации речевых сбоя – мануальный, целографический и окуломоторный.

В строгом смысле перевод вокальной модальности в кинетическую с достижением положительных результатов, по нашему мнению, возможен при условии задания функционального значения того или иного жеста. Подобно контексту в случае традиционного дискурс-анализа, мы считаем, что тематика диалога тоже может влиять на смысловую наполняемость используемого жеста. Классифицировав жесты по группам *жесты-иллюстраторы, жесты-адаптеры, жесты-регуляторы, аффективные и ситуативные жесты*, мы проверили нашу гипотезу о взаимосвязи типов жестов с тематикой, воспользовавшись методом подсчета критерия χ^2 . Результаты на базе наших выборок демонстрируют взаимосвязь между функциональным типом жеста и темой ведения диалога. Следовательно, прагматическая наполняемость жеста может зависеть также от тематики.

В рамках проаннотированных нами мультимодальных корпусов мы сосредоточились на одном из проявлений спонтанной речи – речевых сбоях и закономерностях в их восполнении на уровне кинетической модальности. Нам удалось найти все указанные в классификации А.А. Кибрика и В.И. Подлесской типы речевых сбоя на материале собранного русского и китайского студенческих дискурсов, что дало нам возможность прототипировать 5 основных видов коммуникативных событий. Каждое коммуникативное событие представлено сценариями, описывающими тип речевого сбоя и основные вербально-поведенческие реакции на внешнем уровне (инициатива смены, вовлеченная смена или обратная связь).

Выделенные сценарии мы представляем в виде мультимодальных схем обоих коммуникантов. Схема выглядит как последовательность, состоящая из типа речевого сбоя в вербальном канале и комплекса невербальных тегов компенсации этого речевого сбоя в мануальном, окуломоторном и

целографическом каналах. В некоторых сценариях коммуникативного события нам удалось зарегистрировать более, чем один вариант мультимодальной схемы иницирующего и (или) реагирующего говорящих.

Помимо разработки мультимодальных корпусов диалогической спонтанной речи, мы также провели лингвистический анализ выделенных нами закономерностей невербальной компенсации речевых сбоев, выявили схожие и отличительные черты в коммуникативном поведении двух культур – русской и китайской. Так, например, согласно проведенному лингвистическому исследованию, мы обнаружили, что феномен наложений речи менее свойственен для китайцев, чем для русских, что может объясняться более сдержанным стилем ведения диалога у первых. Речевые сбои, классифицированные нами как в русском, так и в китайском дискурсе, несмотря на свою схожую природу, обладают некоторыми особенностями. Например, маркеры препаративной подстановки на русском языке бывают двух видов – согласуемые и несогласуемые, на китайском языке, вследствие его структуры, только несогласуемый вид, выраженный по данным нашей выборки местоименной основой дальнего дейксиса с универсальным классификатором 那个. Удлинения отдельных звуков китайского языка также по причине наличия тоновой системы лимитирует привычное русским коммуникантом «растяжение» слогов. Тем не менее, по нашему мнению, такой тип речевого сбоя также присутствует в спонтанной речи китайцев, но в несколько другой форме, а именно путем использования междометия 啊 и его вариаций 呀 при перечислениях или удлинением гласных в союзах или союзных словах (然后、还有、和 и т.д.). Удлинение отдельных звуков как тип речевого сбоя, с нашей точки зрения, функционально необходимо для обдумывания следующего коммуникативного хода собеседником в обеих культурах.

Было обнаружено, что абсолютная паузация свойственна для китайцев, и более частотна по отношению к заполненной паузации. Русская спонтанная

диалогическая речь демонстрирует абсолютную противоположность. В плане невербального восполнения абсолютных пауз, видно, что их появление в речи русскими воспринимается как нежелательное, дискомфортное. Абсолютная пауза тем не менее широко применима в речи на китайском языке, сопровождаясь прямым контактом со своим собеседником, чего не наблюдается на материале русского языка.

Оптимизируя созданные нами мультимодальные корпуса под задачи машинного обучения путем классифицирования и аннотирования коммуникативных событий, а также путем представления их более мелких элементов – мультимодальных схем, мы предложили модель мультимодальной автоматизированной последовательности обучения компьютерного агента, взяв за основу две стратегии машинного обучения – с одной стороны, перевод вокальной модальности в кинетическую в случае обнаружения помех с последующим слиянием с данными общего потока и, с другой стороны, беспрепятственное потоковое выравнивание мультимодальных данных и дальнейшее их слияние в единую систему знания для последующего генерирования ответа или прогнозирования.

ЗАКЛЮЧЕНИЕ

Корпусная лингвистика как одно из направлений компьютерной науки, зародившись в 60-е годы XX века, сосредоточилась на создании электронных корпусов письменной и устной речи в первую очередь для машинной обработки. В связи с масштабными размерами и систематизированным видом корпусов другие ученые также стали использовать их при проведении лингвистических исследований в области лексикографии, переводоведения, преподавания и т.д.

Позже специалисты по компьютерной науке столкнулись со сложностями в плане распознавания и синтеза речи машиной, т.к. стало очевидно, что размеченный текстовый массив письменной речи не всегда может быть прочной базой для обучения по причине отсутствия контекстуальной составляющей.

«Погружение» компьютерных агентах в контекст (в машинном обучении называется – “grounding”) оказалось чрезвычайно сложной задачей. Позже данный феномен назвали парадоксом Моравека (Moravec’s paradox), согласно которому когнитивные процессы высокого уровня требуют относительно меньшего количества вычислений, нежели низкоуровневые сенсомоторные операции. Психомоторные операции осуществляются нами как правило без участия сознания и прошли миллиарды лет эволюционного формирования. В то же время как интеллектуальные функции появились сравнительно недавно по меркам эволюции и, следовательно, не до конца автоматизированы. По этой причине именно последние лучше поддаются компьютерной алгоритмизации.

Стремясь решить эту проблему, ученые и специалисты делают попытки разложить наше бессознательное естество на более мелкие элементы, систематизировать эволюционное знание и запрограммировать выявленные закономерности в архитектуру вычислительной машины. Так, ученые в очередной раз осознали всю комплексность стоящей перед ними проблемы

и еще более укрепились во мнении о необходимости полипарадигмального развития науки.

В ответ на сложности алгоритмизации, специалисты в области корпусной лингвистики всё чаще сосредотачиваются на создании нового типа корпусов – мультимодальных, которые в отличие от корпусов только письменной или только устной речи, включают метаданные кинетической модальности. Исходя из всего вышеописанного, целью настоящей магистерской диссертации является разработка мультимодальных корпусов на русском и китайском языках и их оптимизация для нужд машинного обучения.

В нашей работе мы делаем фокус на одном из важных вербальных проявлений спонтанной диалогической речи – речевых сбоях, которые представляют собой помехи в работе алгоритма и могут вести к системным ошибкам и сбоям. В ходе аннотирования отснятого аудио-видеоматериала, мы часто являлись свидетелями повышенной степени активации каналов кинетической модальности при речевых сбоях в диалогах как носителей русского, так и китайского языков. Вследствие чего, у нас появилась гипотеза о том, что при затруднениях в вокальной модальности происходит перераспределение активности информационных потоков, проявляемое в совокупности закономерных актуализированных телодвижений в кинетической модальности. Такая совокупность входит в состав мультимодальных схем, представляющих определенный сценарий коммуникативного события на внешнем уровне вместе с компенсированием речевых сбоях на внутреннем уровне.

Коммуникативное событие, будучи мотивированным и целостным процессом речевого взаимодействия, является устойчивым комплексом, внутри которого регистрируются различные акции и реакции коммуникантов. В мультимодальных корпусах внутри отдельного коммуникативного события выравнивается не вся информация, а только значимая для решения задачи машинного обучения. Такие сегменты с выравненной внутри информацией

включают наборы целевых атрибутов, призванные оптимизировать данные для их последующего внедрения в модель.

Примечательно, что при анализе теоретической литературы в таких научных областях, как корпусная лингвистика и машинное обучение, присутствует расхождение в понимании «выравнивания». Традиционно среди основных характеристик при разработке корпусов выделяют параллельность и тип языковых данных. Параллельность осуществима в корпусах письменных переводных документов или в корпусах подготовленной речи, отчитываемой дикторами или актерами. Для обеспечения параллельности в дву- и многоязычном корпусе необходимо задать критерий выравнивания, который может содержать отдельные слова и даже целые предложения. Специалисты по машинному обучению переосмысливает понимание «выравнивания», подразумевая под ним прежде всего указание в разметке соответствий, демонстрирующих взаимосвязь между отдельными элементами модальной информации, проходящей по разным каналам. Такой тип выравнивания потоковых данных осуществляется через разметку вокальной и кинетической модальностей с указанием временных промежутков. Последующее выравнивание аудио-видеоматериала и текстового формата разметки решается двумя возможными способами: ручная нарезка аудио и видеотрексов или автоматическое указание промежутков со ссылкой на сам медийный источник в специальной программе по аннотированию. В нашей работе обработка и выравнивание данных проводилось в автоматизированной системе аннотирования ELAN.

Обзор научной литературы по теме нашей магистерской диссертации показал отсутствие единообразия даже внутри одного научного направления. От более существенных неоднородностей, как например, принцип транскрибирования звучащей речи, определения разницы между понятиями «мульти-modalность», «поликодовость» и «креализованность», проблема разграничения понятий «мульти-modalный», «медийный» и «мультиканальный» корпус, понятий «коммуникативная неудача», «речевая

ошибка», «речевое затруднение» и «речевой сбой»); к менее заметным – форма множественного числа лингвистического термина «корпус». Всё это свидетельствует о сравнительно недавнем интересе ученых и специалистов к устной речи во всех ее проявлениях.

Несмотря на некоторую противоречивость, собранный теоретический базис послужил основой при создании разметки мультимодальных данных и поиске на ее основе закономерностей невербального поведения при компенсации речевых сбоев. В процессе работы над проаннотированным материалом удалось классифицировать 5 видов коммуникативных событий с их основными сценариями (тип речевого сбоя и его компенсация на внутреннем уровне, вид наложения (инициатива смены, вовлеченная смена, обратная связь) и невербальные реакции – внешний уровень компенсации). В каждом сценарии коммуникативного события представлены основные воспроизводимые мультимодальные схемы инициирующего и реагирующего коммуникантов. Если сравнить типы сценариев, то реакции на внешнем уровне в обеих культурах практически совпадают (кроме реакций при появлении маркеров препаративной подстановки). Совпадение реакций участников коммуникации на стимул (речевой сбой), говорит о некой степени шаблонности человеческого поведения, а значит дает надежду на его алгоритмизацию. Хотя виды реакций совпадают, мультимодальные схемы, представляющие определенный комплекс закономерностей, проявляющийся в отдельных каналах, имеют отличия и даже могут быть представлены несколькими вариантами в рамках одного сценария.

Созданные в рамках данной работы мультимодальные корпуса спонтанной диалогической речи были оптимизированы в два этапа: первый этап связан со сплошной разметкой собранных массивов данных и формированием т.н. наборов данных/датасетов; второй этап – выборочное сегментирование, представляющее наборы целевых атрибутов (коммуникативные события) для решения проблемы помех в вокальной модальности.

Оптимизируя проаннотированные нами мультимодальные корпуса, мы предпринимаем попытку предложить модель автоматической последовательности обучения (в машинном обучении называется – “pipeline”), где бы ход алгоритма проходил через выравнивание и слияние данных, но допускал дополнительную предшествующую ветвь перераспределение потока с вокальной на кинетическую модальность при возникновении помех в первой. Как мы полагаем, такая модель, с одной стороны может снизить нагрузку, уменьшить вероятность системного сбоя при машинном обучении, а с другой, позволит улучшить показатели «погружения» работа в контекст.

Отметим, однако, что быстрое внедрение собранных и оптимизированных на данном этапе мультимодальных корпусов неоправданно по причине высокой стоимости ошибки при обучении нейронных сетей. Нам предстоит еще расширить наборы данных, произвести их очистку, дополнить комплексы тегов мультимодальных схем. Проверка результатов наших статистических подсчетов на основе информации из других выборок также укрепит наше понимание степени взаимосвязи типов жестов со стилем диалогов и позволит принять решение о (не)включении в разметку коммуникативных событий прагматического слоя, сообщающего системе смысловую наполняемость жестовой активности коммуникантов.

Данная магистерская работа очерчивает круг основных проблем и предлагает свое решение проблем машинного обучения. В ходе исследования стало очевидно, что разработка мультимодальных корпусов полезна не только для компьютерных наук, но и для лингвистических исследований. При проведении обработки собранной мультимодальной информации, мы наметили темы, которым ученые потенциально могут посвятить свои будущие научные работы, например, в области рассмотрения связи гендера с частотой жестикуляции, гендера с частотой речевых сбоев, тематики разговора с типом жеста, а также принципов соблюдения конгруэнтности, сегментирования устного спонтанного дискурса и многое другое.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Александрова Л.Ю. Речевые затруднения младенческого и раннего возраста в условиях материнской депривации различной степени выраженности // Известия Российского государственного педагогического университета имени А.И. Герцена. Вып. 121. СПб.: РГПУ им. А.И. Герцена, 2010. С. 7–17.
2. Ананьев В.А. Концептуальная основа потрясающей психотерапии // Б.Г. Ананьев и ленинградская школа в развитии современной психологии. СПб.: Речь, 1995. 384 с.
3. Андреева С.В. Лингвистические закономерности передачи информации при автоматической обработке речи // Материалы Третьего междисциплинарного семинара «Анализ разговорной русской речи». СПб.: СПбГУАП, 2009. С. 10–14.
4. Анисимова Е.Е. Лингвистика текста и межкультурная коммуникация (на материале креолизованных текстов). М.: Академия, 2003. 128 с.
5. Арутюнова Н.Д. Дискурс // Лингвистический энциклопедический словарь / отв. ред. В.Н. Ярцева. М.: Сов. Энциклопедия, 1990. С. 136–137.
6. Арутюнова Н.Д. Язык и мир человека. 2 изд. М.: Яз. рус. культуры, 1999. 895 с.
7. Ахапкина Я.Э. Причины и механизмы речевых сбоев на письме (на материале учебно-научных текстов носителей русского языка) // Вопросы образования. Вып. 3. М.: НИУ ВШЭ, 2013. С. 65–91.
8. Ахутина Т.В. Порождение речи. Нейролингвистический анализ синтаксиса. М.: Изд-во МГУ, 1989. 215 с.
9. Баранов А.Н. Введение в прикладную лингвистику. М.: Эдиториал УРСС, 2001. 360 с.
10. Белоусов К.И. Экспериментальная лингвистика и сетевая наука // Социо- и психолингвистические исследования. Вып. 2. 2014. С. 21–31.

11. Бигуаа Б.Г., Семенова А.И. Законодательное обеспечение сохранения, развития и изучения языков народов Российской Федерации: состояние и перспективы. Материалы «круглого стола» 16 декабря 2019 г. М.: Издание Государственной Думы, 2020. 56 с.
12. Богачанова Т.Д. и др. Диалогическая лингвистика / под науч. ред. Н.Н. Шпильной. Барнаул: АлтГПУ, 2019. 320 с.
13. Большанова Е.И. и др. Автоматическая обработка текстов на естественном языке и анализ данных. М.: Изд-во НИУ ВШЭ, 2017. 269 с.
14. Борисова Е.В. Коммуникативно-прагматические особенности парцелированных конструкций в современном немецком языке: автореф. дис. ... канд. филол. наук: 10.02.04. Барнаул, 2003. 24 с.
15. Борисова И.Н. Русское повседневное общение: прагматика, культурология. Екатеринбург, 2018. 442 с.
16. Борисова С.А. Пространство – человек – текст. Ульяновск: изд-во УлГУ, 2003. 327 с.
17. Ван Дейк Т.А. Язык. Познание. Коммуникация. М.: Прогресс, 1989. 310 с.
18. Верходанова В.О., Карпов А.А. Моделирование речевых сбоев в системах автоматического распознавания речи // Вестник ТГУ. 2012. Вып. 363. С. 10–15.
19. Владимиров А.А., Марков С.М. Роль разметки данных в обучении искусственного интеллекта и особенности выбора ресурса для ее осуществления // Цифровая экономика-инструмент и среда общественного развития: сборник трудов II научно-практической конференции с международным участием. Псков: ПсковГУ, 2022. С. 37–42.
20. Воронина Т.Н. Причины коммуникативных неудач // Вузовская наука по Северо-Кавказскому региону: материалы VII регион. науч.-техн. конф. Ставрополь: Сев.-Кавказ. гос. техн. ун-т, 2003. С. 15–23.

21. Воронцов К.В. Аддитивная регуляризация тематических моделей коллекций текстовых документов // Доклады РАН. Т. 456. Вып. 3. М.: Наука, 2014. С. 268–271.
22. Гольдин В.Е. Имена речевых событий, поступков и жанры русской речи // Жанры речи. Вып. 1. Саратов: Колледж, 1997. С. 23–34.
23. Горелов И.Н., Седов К.Ф. Основы психолингвистики. М.: Лабиринт, 2001. 304 с.
24. Гришина Е.А. Устная речь в национальном корпусе русского языка // Научно-техническая информация. Серия: Информационные процессы и системы. 2005. Вып. 3. С. 94–110.
25. Гришина Е.А., Савчук С.О. Корпус звучащей русской речи в составе Национального корпуса русского языка // Компьютерная лингвистика и интеллектуальные технологии. М.: РГТУ, 2008 С. 125–132.
26. Гришина Е.А. Русская жестикуляция с лингвистической точки зрения (корпусные исследования). М: Языки славянской культуры, 2017. 744 с.
27. Гумбольдт В. фон. Избранные труды по языкознанию. М.: Прогресс, 1984. С. 301–302.
28. Детинко Ю.И., Куликова Л.В. Политическая коммуникация: опыт мультимодального и критического дискурс-анализа. Красноярск: СФУ, 2017. 166 с.
29. Ерохин С.Д., Ванюшина А.В. Выбор атрибутов для классификации IP-трафика методами машинного обучения // Т-СОМ: Телекоммуникации и транспорт. Т. 12. Вып. 9. М.: Издательский дом Медиа Паблишер, 2018. С. 25–29.
30. Загидуллина М.В. Мультимодальность: к вопросу о терминологической определенности // Знак: проблемное поле медиаобразования. 2019. 31 (1). С. 181–188.
31. Захаров В.П., Богданова С.Ю. Корпусная лингвистика. СПб.: СПбГУ. РИО, 2013. 148 с.

32. Захаров В.П. и др. Моделирование в корпусной лингвистике: специализированные корпуса русского языка / отв. ред. В.П. Захаров. СПб.: Изд-во СПбГУ, 2019. 208 с.
33. Земская Е.А. Язык как деятельность: Морфема. Слово. Речь. М.: Языки славянской культуры, 2004. 688 с.
34. Зубов А.В., Зубова И.И. Информационные технологии в лингвистике. М.: Академия, 2004. 208 с.
35. Кацнельсон С.Д. Типология языка и речевое мышление. М.: Наука, 1972. 213 с.
36. Кибрик А.А., Подлеская В.И. К созданию корпусов устной русской речи: принципы транскрибирования // Научно-техническая информация. Сер. 2. М.: ВИНТИ РАН, 2003. С. 5–11.
37. Кибрик А.А. Мультиmodalная лингвистика // Когнитивные исследования. М.: Институт психологии РАН, 2010. С. 135–152.
38. Кибрик А.А. Русский мультимодальный дискурс. Часть I. Постановка проблемы // Психологический журнал. 2018. Т.39. Вып.1. С. 70–80.
39. Кибрик А.А., Федорова О.В. О структуре мультимодального дискурса // Образы языка и зигзаги дискурса. Сборник научных статей к 70-летию В.З. Демьянкова. М.: Культурная революция. 2018. С. 180–191.
40. Китайгородская М.В., Розанова Н.Н. Современное городское общение: типы коммуникативных ситуаций и их жанровая реализация (на примере Москвы) // Современный русский язык. Социальная и функциональная дифференциация., 2003. С. 103–126.
41. Кодзасов С.В., Кривнова О.Ф. Общая фонетика. М.: РГГУ, 2001. 591 с.
42. Козлова Н.В. Лингвистические корпуса: определение основных понятий и типология // Вестник НГУ. 2013. Т. 11. Вып. 1. С. 79–88.
43. Колшанский Г.В. Коммуникативная функция и структура языка. М.: Наука, 1984. 175 с.

44. Копырин А.С., Видищева Е.В. Технологии обработки и очистки данных, выявления и устранения шумов на временном ряду // Вестник академии знаний: Компьютерные и информационные науки. Т. 39. Вып. 4. Краснодар: Научно-издательский центр, 2020. С. 220–227.
45. Котов А.А. Паттерны эмоциональных коммуникативных реакций: проблемы создания корпуса и перенос на компьютерных агентов // Компьютерная лингвистика. Интеллектуальные технологии. 2009. 8(15). С. 211–218.
46. Красиков Ю.В. Теория речевых ошибок (на материале ошибок наборщика). М.: Наука, 1980. 124 с.
47. Красных В.В. «Свой» среди «чужих»: миф или реальность? М.: ИТДГК «Гнозис», 2003. 375 с.
48. Крейдлин Г.Е. Кинесика // Словарь языка русских жестов / под ред. С.А. Григорьева, Н.В. Григорьев, Г.Е. Крейдлин. М.: Языки славянской культуры, 2001. С. 257–269.
49. Крейдлин Г.Е. Невербальная семиотика. М.: Новое литературное обозрение, 2002. 581 с.
50. Крейдлин Г.Е. Мужчины и женщины в невербальной коммуникации. М.: Языки славянской культуры, 2005. 224 с.
51. Крейдлин Г.Е. Механизмы взаимодействия вербальных и невербальных единиц в диалоге II А. Дейктические жесты и их типы // Компьютерная лингвистика и интеллектуальные технологии. Труды международной конференции «Диалог 2007». М.: Изд-во РГГУ, 2007. С. 320–327.
52. Крейдлин Г.Е. Семиотическая концептуализация тела и проблема мультимодальности // Экология языка и коммуникативная практика. Вып. 2. Красноярск: СФУ, 2014. С. 100–120.
53. Кресс Г. Социальная семиотика и вызовы мультимодальности // Политическая наука. / пер. с англ. Т.Ш. Адильбаев, И.В. Фомин. М.: ИНИОН РАН, 2016. С. 77–100.

54. Крым И.А. Жестовые компоненты речевой коммуникации: дис. ... канд. филол. наук: 10.02.19. Кемерово, 2004. 153 с.
55. Кубрякова Е.С. Эволюция лингвистических идей во второй половине XX века (опыт парадигмального анализа) // Язык и наука конца XX века / под ред. Ю.С. Степанова. М.: Российский гос. ун-т, 1995. С. 144–238.
56. Лабутина и др. Экологический эксперимент как средство повышения качества школьного биологического образования. [Электронный ресурс]. 2021. URL: <https://s.science-education.ru/pdf/2021/1/30486.pdf> (дата обращения: 08.05.2023).
57. Леонов И.В. Паттерны культурно-исторического процесса: парадигмально-тематический анализ: автореф. дис. ... д-ра культурологии. Санкт-Петербург, 2015. 41 с.
58. Леонтович О.А. Методы коммуникативных исследований. М.: Гнозис, 2011. 221 с.
59. Леонтьев А.А. Языковое сознание и образ мира // Язык и сознание: парадоксальная рациональность. М.: Институт языкознания РАН, 1993. С. 16–21.
60. Леонтьев А.А. Психолингвистические единицы и порождение речевого высказывания, М.: Эдиториал УРСС, 2003. 312 с.
61. Литвиненко А.О. и др. Рассказы и разговоры о грушах: аннотирование движений рук [Электронный ресурс]. 2018. URL: <https://clck.ru/34kFur> (дата обращения: 15.03.2021).
62. Магировская О.В., Привалихина Е.С. Лексический уровень жестового языка как особая система номинации (на материале американского и русского жестового языков) // Мир науки, культуры, образования. 2019. Т. 4. Вып. 77. С. 389–391.
63. Мартыненко И.А., Злобина О.Г. Корпусная лингвистика как инструмент для изучения юридической лексики / Вестник Университета имени О.Е. Кутафина (МГЮА). 2018. Т. 1. Вып. 11. С. 88–98.

64. Матвеева Т.В. О методе выявления ценностной информации разговорного диалога // Научный диалог. Вып.10. Екатеринбург: Центр научных и образовательных проектов, 2018. С. 89–101.
65. Михеев А.В. Коммуникативная интенция и спонтанность высказывания в диалогической речи // Сб. науч. тр. Моск. пед. ин-т иностр. яз. Вып.280. М.: Ин-т языкознания РАН, 1986. С. 39–51.
66. Николаева Т.М. Невербальные средства человеческой коммуникации. Национально-культурная специфика речевого поведения. М.: Наука, 1977. С. 13–25.
67. Новикова И.А., Гридунова М.В. Межкультурные различия представлений о невербальной коммуникации у российских и китайских студентов // Вестник РУДН. Серия: Психология и педагогика. 2011. Вып. 3. С. 40–46.
68. Остин Д.Л. Слово как действие // Новое в зарубежной лингвистике. 1986. Вып. 17. С. 22–130.
69. Оудсхоорн Д.Н. Детская и подростковая психиатрия // Социальная и клиническая психиатрия. М.: Медпрактика-М, 1993. 319 с.
70. Павлычев А.В., Стародубов М.И. Формирование набора данных в задаче обнаружения компьютерных атак с использованием методов машинного обучения // Современное образование: интеграция образования, науки, бизнеса и власти. Вып. 2. Томск: ТУСУР, 2023. С. 161–166.
71. Плунгян В.А. Корпус как инструмент и как идеология: о некоторых уроках современной корпусной лингвистики // Русский язык в научном освещении. 2008. Т. 2. Вып. 16. С. 7–20.
72. Подлеская В.И., Кирбик А.А. Самоисправления говорящего и другие типы речевых сбоев как объект аннотирования в корпусах устной речи // Научно-техническая информация. Серия: Информационные процессы и системы. 2007. Т. 2. Вып. 2. С. 2–23.

73. Прозорова Е.В. Российский жестовый язык как предмет лингвистического исследования // Вопросы языкознания. 2007. Вып. 1. С. 44–61.

74. Рассказы о сновидениях: корпусное исследование устного русского дискурса / под ред. А.А. Кибрика и В.И. Подлесской. М.: Языки славянских культур, 2009. 736 с.

75. Резникова Т.И., Копотев М.В. Лингвистически аннотированные корпуса русского языка (обзор общедоступных ресурсов) // Национальный корпус русского языка: 2003–2005. М.: Индрик, 2005. С. 31–61.

76. Рудык Н.Н. К проблеме толкования понятия «спонтанная речь» // Наука и образование. 2010. Т.4. Вып. 5. С. 188–191.

77. Рыбин П.В. Теория перевода. М.: МГЮА, 2007. 263 с.

78. Савельев А.И. Гражданско-правовые аспекты регулирования оборота данных в условиях попыток формирования цифровой экономики // Вестник гражданского права. Т. 20. Вып. 1. М.: Изд. дом В. Ема, 2020. С. 60–92.

79. Серль Д.Р. Что такое речевой акт? // Новое в зарубежной лингвистике. 1986. Вып. 17. С. 151–169.

80. Смирнова К. В. и др. Речевые сбои и жестикуляция: отбор стимульного материала для психолингвистического и нейрофизиологического исследований // «Слово и жест». Научная конференция, посвященная памяти Е. А. Гришиной («Гришинские чтения»). М.: Институт русского языка им. В.В. Виноградова РАН. 2020. С. 35–38.

81. Степанова А.А. Речевые ошибки в фокусе исследования // Сборник научных статей по материалам Первой конференции-школы «Проблемы языка: взгляд молодых ученых». Вып. 1. М.: Институт языкознания РАН, 2012. С. 295–298.

82. Федеральный проект «Искусственный интеллект» [Электронный ресурс]. 2019. URL: https://www.economy.gov.ru/material/directions/fed_proekt_iskusstvennyu_intellekt/ (дата обращения: 28.05.2023).

83. Федорова О.В. Координация взглядов говорящего и слушающего в монологической бимодальной коммуникации // Труды Института русского языка им. В.В. Виноградова. 2022. Вып. 2(32). С. 164–183.
84. Цицерон М. Три трактата об ораторском искусстве / под ред. М.Л. Гаспарова. Москва: Изд-во «Наука», 1972. С. 7–73.
85. Чистова Е.В. Экокогнитивная модель профессиональной мультимодальной коммуникации (на примере кейса синхронных переводчиков): дис. ... д-ра филол. наук: 5.9.8. Красноярск, 2022. 448 с.
86. Шлуинский А.Б. и др. Событие, клауза, финитность: опыт исследования нетривиальных соотношений в разноструктурных языках / Вестник РГНФ. 2015. Вып. 2. С. 187–198.
87. Яковлева Э.Б. Феномен речевых хезитаций: монография. М.: ИНИОН РАН, 2021. 215 с.
88. Якубинский Л.П. О диалогической речи // Избранные работы: Язык и его функционирование / под науч. ред. А.А. Леонтьев. М.: Наука, 1986. С. 17–58.
89. Abbas N.F. Pragmatics of Overlapping talk in Therapy Sessions // Journal of Language and Linguistic Studies. 2020. 16(3). P. 1251–1263.
90. Adda G. et al. Speech Overlap and Interplay with Disfluencies in Political Interviews [Электронный ресурс]. 2007. URL: <https://clck.ru/hoqZD> (дата обращения: 02.02.2021).
91. Aist G. et al. Software Architectures for Incremental Understanding of Human Speech // Interspeech 2006. USA: Interspeech, 2006. P. 1922–1925.
92. Arnold J.E. et al. Disfluencies signal thee, um, new information // Journal of Psycholinguistic Research. 2003. 32:1. P. 25–36.
93. Bell A. Language Style as Audience Design. 1984. 13 (2). P.145–204.
94. Bezemer J., Jewitt C. Multimodal Analysis: Key Issues // Research Methods in Linguistics. 2010. 2(3). P. 180–197.
95. Bressemer J., Ladewig S.H. Rethinking Gesture Phases: Articulatory Features of Gestural Movement. 2011. Vol.184, P. 53–91.

96. Butzberger J.W. et al. Spontaneous speech Effects in Large Vocabulary Speech Recognition Applications // Proceedings of the DARPA Speech and Natural Language Workshop. New York: Morgan Kaufmann, 1992, P. 339–343.
97. Calbris G. Elements of Meaning in Gesture. Amsterdam: John Benjamins Publishing Company, 2011. 378 p.
98. Chafe W. Discourse, Consciousness and Time. The Flow and Displacement of Conscious Experience in Speaking and Writing. Chicago: The University of Chicago Press, 1994. 392 p.
99. Chafe W. The Pear Stories: Cognitive, Cultural and Linguistic Aspects of Narrative Production. US: Praeger, 1980. 338 p.
100. Chandu K. et al. Grounding «Grounding» in NLP. [Электронный ресурс]. 2021. URL: <https://arxiv.org/pdf/2106.02192.pdf> (дата обращения: 13.05.2023).
101. Chernyavskaya V., Kulikova L.V. Multimodalität im sozio-kulturellen Kontext: mit Bildern reden, Bilder verstehen // Alman Dili ve Edebiyatı Dergisi – Studien zur deutschen Sprache und Literatur. 2020. Vol. 43. P.19–38.
102. Cho E. et al. Multilingual Disfluency Removal using NMT // Proceedings of the 13th International Workshop on Spoken Language Translation – IWSLT. USA: Seattle. 2016. P. 1–8.
103. Chui K. et al. The NCCU corpus of Spoken Chinese: Mandarin, Hakka, and Southern Min. Conversation // Language in Society. Vol. 29. Los Angeles: University of California, 2000. P. 1–63.
104. Creese A. Linguistic ethnography // Encyclopedia of Language and Education: Research Methods in Language and Education. 2008. 10(2). P. 229–241.
105. Dell S. A Spreading-activation Theory of Retrieval in Sentence Production // Psychological Review. 1986. Vol. 93. P. 283–221.
106. DeVault D. Contribution Tracking: Participating in Task-oriented Dialogue under Uncertainty. Rutgers The State University of New Jersey: New Brunswick, 2008. 236 p.

107. Duranti A. *Linguistic Anthropology*. New Jersey: Wiley-Blackwell, 2001. 504 p.
108. Eisler F.G. *Psycholinguistics: Experiments in Spontaneous Speech*. London: Academic Press Inc, 1968. 169 p.
109. Ekman P. A., Friesen W. *Handbewegungen // Nonverbale Kommunikation: Forschungsberichte zum Interaktionsverhalten*. Weinheim: Basel. 1979. P. 112–115.
110. Erickson F. *Talk and Social Theory. Ecologies of Speaking and Listening in Everyday Life*. Cambridge: Polity Press, 2004. 232 p.
111. Foster M.E., Oberlander J. *Corpus-based Generation of Head and Eyebrow motion for an Embodied Conversational Agent / Language Resources and Evaluation*. 2007. 41(4). P. 305–323.
112. Fung H., Mok P. *Temporal Coordination between Focus prosody and Pointing gestures in Cantonese // Journal of Phonetics*. 2018. Vol. 71, P. 113–125.
113. Goffman E. *Forms of Talk*. Oxford: Blackwell, 1981. 344 p.
114. Gósy M. *Halt Command in Word Retrieval // Proceedings of DiSS: The 9th Workshop on Disfluency in Spontaneous Speech / eds. by R. L. Rose, R. Eklund*. Budapest: ELTE, 2019. P. 3–6.
115. Graf H. P. et al. *Visual Prosody: Facial Movements Accompanying speech // Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition (FGR'02)*. Washington: IEEE, 2002. P. 396–401.
116. Gumperz J.J. *On Interactional Sociolinguistic Method // Talk, Work and Institutional Order. Discourse in Medical, Mediation and Management Settings/ eds. by S. Sarangi and C. Roberts*. Berlin: Mouton de Gruyter, 1999. P. 453–471.
117. Herreda D. et al. *The UTEP-ICT Cross-Cultural Multiparty Multimodal Dialog Corpus [Электронный ресурс]*. 2010. URL: <https://clck.ru/34kG52> (дата обращения: 25.03. 2021).
118. Hodge R., Kress G. *Social semiotics*. Cambridge: Polity press, 1988. 285 p.

119. Hough J. et al. DUEL: A Multi-lingual Multimodal Dialogue Corpus for Disfluency, Exclamations and Laughter // In Proceedings of the 10th edition of the Language Resources and Evaluation Conference. 2016. P. 1784–1788.
120. Kendon A. *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press, 2004. 400 p.
121. Kendon A. *Origins of Modern Gesture Studies* / Susan S. Duncan, J. Cassell, E.T. Levy (Eds.) // *Gesture and the Dynamic Dimension of Language*. Amsterdam: John Benjamins, 2007. P. 13–28.
122. Kita S. et al. *Movement Phases in Signs and Co-speech Gestures and their Transcription by Human Coders* // *Gesture and sign language in human-computer interaction*. Berlin: Lang. Cogn. Process, 1998. P. 23–35.
123. Knoblauch H. et al. *Video Analysis: Methodology and Methods*. Qualitative Audiovisual Data Analysis in Sociology. Frankfurt am Main: Peter Lang, 2006. 219 p.
124. Kress G. *Reading Images: The Grammar of Visual Design*. London: Routledge, 1996. 312 p.
125. Kress G., Leeuwen T.V. *Multimodal discourse: The Modes and Media of Contemporary Communication*. London: Oxford University Press, 2001. 152 p.
126. Kuhnke E. *Körpersprache für Dummies*. Chichester: John Wiley&Sons, 2007. 331 p.
127. Kulikova L.V., Shatokhina S.A. Sociocultural and Linguistic Contexts of the Russian Sign Language functioning in Krasnoyarsk Krai // *Journal of Siberian Federal University. Humanities and Social Sciences*. 2020.13(3). P. 296–303.
128. Labov W. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press, 1972. 372 p.
129. Lazarus A.A. *The Practice of Multi-modal Therapy*. New York: McGraw-Guilford, 1981. 272 p.
130. Levelt W.J.M. *Monitoring and Self-repair in Speech* // *Cognition*. 1983. Vol. 14. P. 41–104.

131. Liang P. et al. Foundations and Trends in Multimodal Machine Learning: Principles, Challenges, and Open Questions. [Электронный ресурс]. 2022. URL: <https://arxiv.org/pdf/2209.03430.pdf> (дата обращения: 13.05.2023).
132. Litvinenko A.O. et al. Annotating Hand Movements in Multichannel Discourse: Gestures, Adaptors and Manual Postures // Российский журнал когнитивной науки. 2018. Т. 5. Вып. 2. С. 4–17.
133. Loehr D.P. Temporal, Structural and Pragmatic Synchrony between Intonation and Gesture // *Laboratory Phonology*. 2012. 3(1). P. 71–89.
134. Lopes J., Hemmingsson N., Astrand O. The Spot the Difference corpus: a Multi-modal Corpus of Spontaneous Task Oriented Spoken Interactions / *Language Resources and Evaluation*. 2018. Vol.11. P. 1939–1945.
135. Ma M. et al. Multimodal Learning with Severely Missing Modality [Электронный ресурс]. 2021. URL: <https://arxiv.org/pdf/2103.05677.pdf> (дата обращения: 13.05.2023).
136. McClave E. Gestural Beats: The Rhythm Hypothesis // *Journal of Psycholinguistic Research*. 1994. Vol.23. P. 45–66.
137. Mccowan S. et al. Modelling Human Interaction in Meetings. Hong Kong: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2003. P. 748–751.
138. McNeill D. *Hand and Mind: What Gestures Reveal about Thought*. Chicago: The University of Chicago Press, 1992. 423 p.
139. Müller C. et al. Towards a Grammar of Gesture: A form-based view. *Handbooks of Linguistics and Communication Science*. Berlin: De Gruyter Mouton. 2013. 38(1). P. 707–733.
140. Müller C. Wie Gesten bedeuten. Eine kognitiv-linguistische und sequenzanalytische Perspektive // *Sprache und Literatur*. 2010b. 41(1). P. 37–68.
141. Nooteboom S.G., Quene H. Self-monitoring and Feedback: A New Attempt to find the Main Cause of Lexical bias in Phonological Speech errors // *Journal of Memory and Language*. 2008. 58(3). P. 837–861.

142. O'Halloran K.L. *Multimodal Discourse Analysis*. London and New York: Continuum. 2011. P. 120–137.
143. Oliveira S., Stewart D. *Writing Scientific Software: A Guide to Good Style*. 1st edition. Cambridge: Cambridge University Press, 2006. 316 p.
144. Oertel C. et al. D64: A Corpus of Richly Recorded Conversational Interaction / *Journal on Multimodal User Interfaces*. 2013. Vol. 7. P. 19–28.
145. Podlesskaya V. Parameters for Typological Variation of Placeholders // *Typological Studies in language*. Vol. 93 / eds. by N. Amiridze, Boid H. Davis, M. Maclagan. Amsterdam: John Benjamins, 2010. P. 11–32.
146. Postma A. Detection of Errors during Speech Production: A Review of Speech Monitoring Models // *Cognition*. 2000. Vol. 77. P. 97–131.
147. Qian X. et al. Multimodal Machine Translation with Reinforcement Learning. [Электронный ресурс]. 2018. URL: <https://arxiv.org/pdf/1805.02356.pdf> (дата обращения: 13.05.2023).
148. Rapport d'activite // *Le Laboratoire d'Informatique pour la Mecanique et les Sciences de l'Ingenieur* [Электронный ресурс]. 2010. URL: <https://clck.ru/apMNV> (дата обращения: 02.02.2021).
149. Rehm M., Andre E. From Annotated Multimodal Corpora to Simulated Human-Like Behaviors // *Modeling Communication*. 2008. P. 1–17.
150. Sacks H. An Initial Investigation into the Usability of Conversational Data for Doing Sociology // *Studies in Social Interaction* / ed. by D. Sudnow. New York: Free Press, 1972. P. 31–73.
151. Saville-Troike M. *The Ethnography of Communication: An Introduction*. UK: Wiley-Blackwell, 1994. 336 p.
152. Scheer T. The Corpus: A Tool among Others [Электронный ресурс]. 2013. URL: <https://clck.ru/34kG8N> (дата обращения: 20.03.2021).
153. Schegloff E. A Turn Organization: One Intersection of Grammar and Interaction // *Interaction and Grammar* / E. Ochs, E.A. Schegloff, S.A. Thompson (eds.). Cambridge: Cambridge University Press, 1996. P. 134–184.

154. Schegloff E. Overlapping Talk and the Organization of Turn-taking for conversation / *Language in Society*. 2000. 29(1). P. 1–63.
155. Scollon R., Scollon S.W. *Discourses in Place: Language in the Material World*. London: Routledge, 2003. 258 p.
156. Shriberg E., Stolcke A., Baron D. Observations on Overlap: Findings and Implications for Automatic Processing of Multi-Party Conversation // *Proceedings of the 7th European Conference of EUROSPEECH*. Aalborg: College Park. 2001. P. 1359–1362.
157. Shriberg E. Disfluencies in Switchboard // *Proceedings of International Conference on Spoken Language Processing*. Philadelphia: Acoustical Society of Japan. 1996. P. 11–14.
158. Steffensen S.V. Human Interactivity: Problem-solving, Solution-probing and Verbal Patterns in the Wild // *Cognition Beyond the Body: Interactivity and Human Thinking* / eds. by S.J. Cowley, F. Vallée-Tourangeau. Dordrecht: Springer, 2013. P. 195–221.
159. Stolt M. *The many Faces of Overlap: Non-Competitive Overlap in Conversations between Finnish and British Speakers of English*. Jyvaskyla: University of Jyvaskyla, 2008. 130 p.
160. Wolf J.C., Bugmann G. Linking Speech and Gesture in Multimodal Instruction Systems [Электронный ресурс]. 2006. URL: <https://inlnk.ru/oe500w> (дата обращения: 25.03. 2021).
161. Xiao R., Tao H. The Lancaster Los Angeles Spoken Chinese Corpus [Электронный ресурс]. 2007. URL: <https://www.lancaster.ac.uk/fass/projects/corpus/LCMC/default.htm> (дата обращения: 25.03. 2021).
162. Xiong C. et al. Dynamic Memory Networks for Visual and Textual Question Answering // *ICML'16: Proceedings of the 33rd International Conference on International Conference on Machine Learning*. 2016. Vol. 48. P. 2397–2406.
163. Yildirim S., Narayanan S. Automatic Detection of Disfluency Boundaries in Spontaneous Speech of Children Using Audio-Visual Information. 2009. 17(1). P. 2–12.

164. Zadeh A. et al. Memory Fusion Network for Multi-view Sequential Learning // Proceedings of the AAAI Conference on Artificial Intelligence. 2018. 32(1). P. 5634–5641.

165. Zhao Y., Jurafsky D. A Preliminary Study of Mandarin Filled Pauses. [Электронный ресурс]. 2005. URL: <https://web.stanford.edu/~jurafsky/diss05.pdf> (дата обращения: 07.04.2021).

Техническое задание

Проект в рамках мультимодальной лингвистики, ориентированный на вокальную и кинетическую модальности.

Исследовательский вопрос: как информация, получаемая человеком из разных источников, интегрируется между собой?

Задачи:

1) собрать корпуса разговорной речи (русский и китайский) с мультимодальной разметкой:

- объем корпуса не менее 1000 строк;
- методы сбора – запись диалогов (обязательна видеозапись и последующая расшифровка).

2) количественное исследование – подсчет статистической значимости основных модусов для каждого из 2 языков:

- вербальный компонент;
- визуальные сигналы - жесты рук, лицевая мимика, движения тела, позы, направления взгляда и их комбинации.

3) качественное исследование – выявление и сопоставление схем мультимодальности для каждого из 2 языков + создание схем перевода мультимодальности.

Что нужно Промобот на выходе:

- мультимодальная 2-х язычная размеченная БД для обучения робота;
- схемы мультимодальности для 2-х языков.

**СОГЛАСИЕ
на участие в лингвистическом эксперименте**

Лингвистическое исследование, проводимое в рамках магистерской работы, посвящено изучению того, как носители языка (русского и китайского) общаются между собой в спонтанной устной речи. Помимо изучения речи, нас также интересует использованный язык тела (жесты, сопутствующие речи).

Настоящее исследование проводится в рамках совместного проекта с компанией «Промобот». При проведении эксперимента будет вестись аудио- и видеосъемка. Цель видеосъемки – получить сведения о целостной картине взаимодействия людей в процессе спонтанного диалога. Полученные в рамках эксперимента персональные данные участников в работе фигурировать не будут. В эксперименте приглашаем принять участие взрослых носителей языка в возрасте 18–45 лет. В рамках каждой записи необходимо участие двух человек.

Время исследования составляет около 30–50 мин. и состоит из 3 частей. Первая часть предполагает ознакомление с ролью в диалоге, темой беседы. Вторая часть – непосредственный диалог между двумя участниками эксперимента. Третья часть – заполнение анкеты-опросника.

Я, _____
фамилия, имя, отчество

согласен(а) принять участие в лингвистическом эксперименте и даю свою разрешение на использование и публикацию полученных из аудио- и видеозаписи данных с моим участием.

Год рождения _____

Номер телефона _____

E-mail _____

Подпись: _____ Дата: _____

За дополнительными вопросами по деталям исследования Вы можете обратиться по телефону ***** или по почте *****

语言实验协议

作为硕士学位论文的一部分进行的语言实验主要探讨研究母语人士（俄语和汉语）如何通过自发言语互相交流。除了研究语音，我们还对使用的肢体语言（语言伴随性的手势）感兴趣。

这项研究是在与 Promobot 公司的协作项目框架内进行的。在实验过程中，将进行录音录像。除了研究语言视频拍摄的目的是获得人们在自发对话过程中互动的整体画面信息。实验中获得的参与者的个人信息不会出现在作品中（研究中）。我们邀请 18-45 岁的母语人士参与实验。每个录像必须有两个人参加。

实验时长约为 30-50 分钟，并由三部分组成。第一部分是熟悉对话中的角色、对话主题，第二部分是两个实验参与者进行直接对话。第三部分是填写问卷调查。

参与者（姓名）_____ 同意参加这项语言实验，并允许使用和发布从音频视频记录中获得的数据。

出生年份：

电话号码：

电子邮件地址：

参与者（签字）

年 月 日

有关研究细节的其他问题，您可以致电*****，或者发送邮件*****

Опорные карточки для аудиовидеозаписи на русском языке

Карточка №1
<p>Участник 1</p> <p>Вы расстроенный покупатель. Ваш недавно купленный девайс (выбрать любой: смартфон, ноутбук, планшет, т.д.) не работает. После нескольких попыток решить проблему в сервисном центре, вы пришли в главный офис. Никто не может (не хочет) решить Вашу проблему.</p> <p>Вы намерены разговаривать с главным менеджером компании. Если общение будет безрезультатным, то Вы готовы дойти до директора офиса.</p> <p>Вам необходимо обсудить следующую информацию:</p> <ol style="list-style-type: none"> 1) узнайте, какие варианты решения проблемы менеджер предлагает; 2) договоритесь об одном из наиболее подходящих вариантов решения проблемы; 3) узнайте, когда проблема будет решена. <p>Вы начинаете речь первым(ой).</p>
<p>Участник 2</p> <p>Вы – менеджер компании. Вам сообщили о встрече с одним недовольным клиентом. Новый девайс бренда вашей компании не работает (какой именно девайс – неизвестно). Вас предупредили о намерении покупателя разобраться в ситуации напрямую с директором. Вам предстоит иметь дело с крайне расстроенным клиентом и решить проблему «мирным» путем.</p> <p>В Вашем диалоге должна прозвучать следующая информация:</p> <ol style="list-style-type: none"> 1) узнайте, о каком продукте вашей компании идет речь; 2) предложите несколько возможных вариантов решения проблемы; 3) убедитесь, что между Вами и клиентом нет недопонимания; 4) убедитесь, что клиент доволен путем разрешения проблемы и не имеет никаких возражений.
Карточка №2
<p>Участник 1 и 2</p> <p>Вы работаете в компании рядовым сотрудником. Ваш босс поручил Вам и Вашему коллеге организовать корпоратив. Обсудите важные детали создания мероприятия.</p> <p>При диалоге воспользуйтесь подсказками ниже:</p> <ol style="list-style-type: none"> 1) тематика корпоратива (бал-маскарад, дискотека 90-х, вручение Оскара, Олимпиада, дикий запад, восточный стиль и т.д.); 2) локация мероприятия (бар, ресторан, банкетный зал, офис и т.д.); 3) выбор ведущего, артистов (танцевальное шоу, фаер-шоу и т.д.); 4) меню (постарайтесь учесть вкусы большинства – например, восточная кухня, блюда для вегетарианцев, фастфуд, алкоголь и т.д.); 5) способ сбора средств для проведения корпоратива; 6) распределите обязанности между собой (кто за что будет отвечать).

ПРИЛОЖЕНИЕ В (продолжение)

Карточка №3
<p>Участник 1</p> <p>Обсудите с Вашим собеседником, как изменилась Ваша жизнь после первой волны коронавируса. Как вы приспособились к «новой жизни». Какие новшества появились за время пандемии. Как Вы относитесь к дистанционному обучению и что бы Вы выбрали: офлайн работу/обучение или онлайн работу/обучение.</p> <p>Согласно Вашей роли: <u>Вы «ЗА» офлайн формат.</u></p> <ol style="list-style-type: none">1) новые достижения (Сбермаркет, бесконтактная доставка, удаленное обучение, виртуальная реальность);2) ограничения;3) способы поддержания иммунитета.
<p>Участник 2</p> <p>Обсудите с Вашим собеседником, как изменилась Ваша жизнь после первой волны коронавируса. Как вы приспособились к «новой жизни». Какие новшества появились за время пандемии. Как Вы относитесь к дистанционному обучению и что бы Вы выбрали: офлайн работу/обучение или онлайн работу/обучение.</p> <p>Согласно Вашей роли: <u>Вы «ЗА» онлайн формат.</u></p> <ol style="list-style-type: none">1) новые достижения (Сбермаркет, бесконтактная доставка, удаленное обучение, виртуальная реальность);2) ограничения;3) способы поддержания иммунитета;4) отношение к дистанционному формату работы/обучению.
Карточка №4
<p>Участник 1</p> <p>Вы планируете совершить поездку в другую страну (Турция, Китай, Египет, Кипр, Таиланд, США и т.д.). Вы обратились за консультацией к турагенту. Обсудите и спланируйте с турагентом Вашу поездку, узнайте детали и особенности организации поездки в выбранный Вами район.</p> <p>В процессе диалога Вы можете обсудить следующую информацию:</p> <ol style="list-style-type: none">1) определите направление путешествия;2) длительность пребывания;3) бюджет путешествия;4) экскурсии;5) туристическое страхование;6) особенности путешествия в период ковида (карантин, ПЦР-тест, прививка и т.д.);7) примите решение – бронируете или ищите варианты у других туроператоров.
<p>Участник 2</p> <p>Вы – туроператор. К Вам обратился потенциальный клиент с просьбой проконсультировать об имеющихся на данный момент направлениях путешествия. Обсудите с клиентом его поездку и предоставьте ориентировочный план путешествия. <u>Ваша задача заинтересовать клиента и сделать так, чтобы он остался.</u></p> <p>В процессе диалога Вы можете обсудить следующую информацию:</p> <ol style="list-style-type: none">1) направление путешествия (из существующих в программе);2) длительность;3) бюджет;4) предоставьте информацию о тур. страховании;

ПРИЛОЖЕНИЕ В (продолжение)

Карточка №4 (участник 2 – продолжение)	
5)	расскажите детально о процедуре и особенностях поездки в ковидный период;
6)	предложите самое выгодное предложение.

Картина №5	
Участник 1	
<p>Время пришло – Вы собираетесь приобрести свой первый автомобиль. Никакого больше общественного транспорта! Обсудите со своим другом машины из каталога и попробуйте выбрать машину из представленного списка.</p> <p>Перед началом обсуждения подумайте, как данная ниже информация связана с выбором машины и <u>что именно Вы хотите</u>:</p>	
1)	тип кузова (седан, хэтчбек, пикап, универсал, минивэн, внедорожник, кроссовер и т.д.);
2)	коробка передач;
3)	бюджет;
4)	тип привода (передний, задний, полный);
5)	двигатель (мощность, расход топлива, стоимость обслуживания, доступность запчастей);
6)	подержанное или новое авто;
7)	и т.д.
Участник 2	
<p>Ваш близкий друг наконец-то надумал купить свой первый автомобиль. Обсудите с ним варианты из предложенного каталога и посоветуйте автомобиль(и), который(ые) бы подходили под описание вашего друга.</p> <p>Перед началом обсуждения подумайте, по каким критериям нужно выбирать авто и что посоветовать другу. Вы можете воспользоваться следующими подсказками:</p>	
1)	тип кузова (седан, хэтчбек, пикап, универсал, минивэн, внедорожник, кроссовер и т.д.);
2)	коробка передач;
3)	бюджет;
4)	тип привода (передний, задний, полный);
5)	другое: мощность двигателя, расход топлива, стоимость обслуживания, доступность запчастей;
6)	подержанное или новое авто;
7)	и т.д.

Опорные карточки для аудиовидеозаписи на китайском языке

卡片№1
<p>参与者 1</p> <p>您是一位不满意的顾客。不久以前您买的设备（可以选择任何一个：手机、电脑、平板电脑等等）坏了，多次尝试在服务中心解决问题，但不能，您去了总公司，谁也不想（不能）解决您的问题。</p> <p>您打算跟经理交谈，如果沟通不成功，那么您准备找公司总经理解决问题。</p> <ol style="list-style-type: none">1. 获悉经理打算解决这个问题的几种方案；2. 就解决这个问题最合适的方案达成一致；3. 获悉这个问题什么时候被解决。
<p>参与者 2</p> <p>您是公司经理。您被告知与一位不满意的顾客会面，您公司品牌的新设备坏了（什么设备您不知道）。您已被提醒，这位顾客有意直接跟总经理处理这件事。您必须与一位极度心烦意乱的顾客打交道，并和平地解决问题。</p> <p>您的对话应包含以下信息：</p> <ol style="list-style-type: none">1. 找出公司的哪款产品有问题；2. 针对问题提出几种可能的解决方案；3. 确保您和顾客之间没有误会；4. 确保这位顾客对解决问题的方法满意并没有反对意见（解决问题，确保顾客满意，无异议）。
卡片№2
<p>参与者 1 和 2</p> <p>您是公司的普通员工。您的老板让您和您同事组织公司派对。跟您同事详细讨论如何组织这种活动。</p> <p>讨论时，请使用以下提示：</p> <ol style="list-style-type: none">1. 公司派对的主题（假面舞会、90年代迪斯科舞会、奥斯卡颁奖典礼、奥运会、狂野的西部、东方风情等等）；2. 派对地点（酒吧、餐厅、宴会厅、办公室等）；3. 主持人、演员选择（舞蹈表演、火焰表演等）；4. 菜单（尽量适应大多数人的口味——比如东方菜、素菜、快餐、酒类等）；5. 派对筹款方式； <p>分配责任（谁将负责什么）。</p>

卡片№3
<p>参与者 1</p> <p>与您的对话者讨论：自第一波冠状病毒以来您的生活发生了怎样的变化？您是如何适应“新生活”的？疫情期间出现了什么新鲜事儿？您对远程教育是什么态度？您会如何选择：线下工作、学习还是在线工作、学习？</p> <p>根据您的角色：您支持远程上课，在家办公。</p> <ol style="list-style-type: none"> 1. 网上超市； 2. 隔离； 3. 对远程工作、学习的态度。
<p>参与者 2</p> <p>与您的对话者讨论：自第一波冠状病毒以来您的生活发生了怎样的变化？您是如何适应“新生活”的？疫情期间出现了什么新鲜事儿？您对远程教育是什么态度？您会如何选择：线下工作、学习还是在线工作、学习？</p> <p>根据您的角色：您支持在办公室工作，教室上课。</p> <ol style="list-style-type: none"> 1. 网上超市； 2. 隔离； 3. 对远程工作、学习的态度。
卡片№4
<p>参与者 1</p> <p>您打算去另一个国家旅游（土耳其、中国、埃及、塞浦路斯、泰国、美国等），您已咨询过旅行社，与旅行社讨论并计划您的旅行，了解前往地区的详细信息和特点。</p> <p>在对话期间，您可以讨论以下信息：</p> <ol style="list-style-type: none"> 1. 确定旅行方向； 2. 逗留时间； 3. 制定旅行预算； 4. 观光； 5. 旅行保险； 6. 疫情期间旅行的特殊性（隔离、核酸检测、打疫苗等）； 7. 做出决定 – 预订还是询问其他旅游社。
<p>参与者 2</p> <p>您这边是旅行社。一位潜在客户与您联系，要求提供当前可用旅行方案的建议，与顾客讨论旅行并提供可行性旅行计划。您的任务是引起客户的兴趣并确保他留下来（预订服务）。</p> <p>在对话期间，您可以讨论以下信息：</p> <ol style="list-style-type: none"> 1. 您可以提议旅行方向； 2. 旅行时长； 3. 制定旅行预算； 4. 提供有关旅行保险的信息； 5. 详细介绍一下疫情期间的行程流程和特点（隔离、核酸检测、打疫苗等）； 6. 提供最优惠的价格。

卡片№5

参与者 1

终于等到了这一天！您想要购买您的第一辆汽车，不再坐公共汽车！与您的朋友讨论汽车，并尝试选择汽车。考虑您想要的汽车是什么样子。开始讨论之前，请考虑以下与汽车选择有关系的信息：

1. 车身类型（轿车、掀背车、皮卡、旅行车、小型货车、SUV、跨界车等）；
2. 变速箱；
3. 预算；
4. 驱动类型（前轮驱动、后轮驱动、全轮驱动）；
5. 发动机（功率、油耗、维护成本、备件供应情况）；
6. 二手车或新车；
7. 其他情况。

参与者 2

您朋友终于决定买他（她）的第一辆车。与他（她）讨论汽车，并推荐一款适合您朋友描述的汽车。开始讨论之前，请考虑以下信息与汽车选择有何关系：

1. 车身类型（轿车、掀背车、皮卡、旅行车、小型货车、SUV、跨界车等）；
2. 变速箱；
3. 预算；
4. 驱动类型（前轮驱动、后轮驱动、全轮驱动）；
5. 其他：发动机功率、油耗、维护成本、备件可用性；
6. 二手车或新车；
7. 其他情况。

Список актуализированных жестов в соответствии с классификацией

П. Экмана и У. Фризена и упоминаемыми автором типами жестов

(симптоматические жесты и жесты нейтрального положения)

Русский студенческий дискурс	
Адаптеры:	
1.	Теребить пальцы
2.	Сложить руки в замок
3.	Соединить кончики пальцев
4.	Скрестить ноги
5.	Скрестить руки на груди
6.	Опустить глаза
7.	Водить глазами из стороны в сторону
8.	Встряхнуть головой
9.	Выставить ногу в сторону
10.	Выставить ногу вперед
11.	Зажмуриться
12.	Закусить губы
13.	Запрокинуть голову
14.	Клонить корпус из стороны в сторону
15.	Коснуться носа
16.	Коснуться руки
17.	Ломать пальцы
18.	Моргнуть
19.	Обхватить ладонью кулак
20.	Обхватить ладонью предмет
21.	Опустить голову вниз
22.	Переминаться с ноги на ногу
23.	Поправлять прическу
24.	Потирать руки
25.	Прижать пальцами ладонь
26.	Прижать пальцы друг к другу
27.	Прижать руки к груди
28.	Прикрыть глаза
29.	Притянуть ладонь к лицу
30.	Отвести взгляд
31.	Поглаживать мочку уха
32.	Поправить прическу
33.	Потирать шею
34.	Потирать нос
35.	Почесывать грудь
36.	Прижать руку ко рту

ПРИЛОЖЕНИЕ Г (продолжение)

Иллюстраторы:	
1.	Изобразить переключение коробки передач автомобиля
2.	Упереться руками в бока
3.	Качать указательным пальцем
4.	Хлопнуть одной ладонью по другой
5.	Хлопать по карманам
6.	Сложить большой, указательный и средний пальцы (щепоть)
7.	Развести руками
8.	Рассечь воздух ритмичными вертикальными движениями ладони
9.	Растопырить пальцы
10.	Рубить воздух ладонями
11.	Сжать кулак
12.	Посмотреть в потолок
13.	Посмотреть вдаль
14.	Пожать плечами
15.	Показывать указательными пальцами в сторону собеседника
16.	Поднять указательные пальцы в замке
17.	Поднять указательный палец
18.	Пожать друг другу руки
19.	Поднять руки
20.	Перечислять загибая пальцы
21.	Перечислять разгибая пальцы
22.	Опустить голову вниз
23.	Опустить глаза
24.	Описать головой круг
25.	Обвести пальцем круг в воздухе
26.	Водить руками из стороны в сторону
27.	Водить глазами из стороны в сторону
28.	Двинуть кисти к собеседнику
29.	Двинуть кистью от себя
30.	Двинуть кистями от себя
31.	Двинуть кулаками
32.	Двинуть ладони к собеседнику
33.	Двинуть ладонь к собеседнику
34.	Двинуть ладонью в сторону
35.	Двинуть руками
36.	Двинуть руками в сторону
37.	Двинуть рукой вниз
38.	Двинуть указательным пальцем в сторону
39.	Задавать рукой направление
40.	Качать головой
41.	Кивнуть
42.	Лист (изобразительный жест)
43.	Махнуть рукой
44.	Махнуть рукой у лица

ПРИЛОЖЕНИЕ Г (продолжение)

45.	Мотать головой
46.	Вращать кистями (мысль)
47.	Двинуть руками влево
48.	Двинуть руками вправо
Регуляторы:	
1.	Трясти кулаками
2.	Трясти руками
3.	Протянуть руку к собеседнику
4.	Покоситься на собеседника
5.	Обводить глазами собеседника
6.	Боднуть собеседника
7.	Водить глазами из стороны в сторону
8.	Вращать кистью
9.	Вращать кистями
10.	Вскинуть руки
11.	Выставить ногу в сторону
12.	Выставить ногу вперед
13.	Двинуть головой вперед
14.	Двинуть головой назад
15.	Кивнуть
16.	Наклониться к собеседнику
17.	Переминаясь с ноги на ногу
18.	Поднять руки
19.	Двинуть кулаками
20.	Взмахнуть рукой
Аффективные жесты:	
1.	Вытянуть губы
2.	Закрыть глаза
3.	Закусить губы
4.	Моргнуть
5.	Надуть губы
6.	Наклонить голову вбок
7.	Повести головой вбок
8.	Повести подбородком
9.	Поднимать брови
10.	Прикрыть глаза
11.	Убрать улыбку
12.	Улыбаться
13.	Кивнуть
14.	Сжать кулаки
15.	Зажмуриться
16.	Посмеяться
Ситуативные жесты:	
1.	Вертеть предмет в руках

ПРИЛОЖЕНИЕ Г (продолжение)

Китайский студенческий дискурс:	
Адаптеры:	
1.	Сложить руки за спиной
2.	Сплести пальцы
3.	Потирать пальцы
4.	Теребить пальцы
5.	Сжать пальцы другой ладонью
6.	Переминаясь с ноги на ногу
7.	Посмотреть в сторону
8.	Отвести взгляд
9.	Моргнуть
10.	Поджать губы
11.	Засунуть руки в карманы
12.	Согнуть руки на уровне пояса
13.	Вращать корпус из стороны в сторону
14.	Скрестить руки внизу
15.	Прижать руки к груди
16.	Выпрямиться
17.	Опустить глаза
18.	Раскачиваться из стороны в сторону
19.	Согнуться
20.	Взяться за подбородок
21.	Обхватить ладонью предмет
22.	Поправлять одежду
23.	Засучивать рукава
24.	Теребить предмет одежды в руках
25.	Обхватить руками кулак
26.	Ломать пальцы
27.	Притянуть ладонь к лицу
28.	Притопывать
29.	Закусывать губы
30.	Обхватить рукой горло
31.	Почесывать шею
32.	Обхватить рукой локоть другой руки
33.	Сжать кулаки
34.	Потирать руку
35.	Застегивать кофту
36.	Водить глазами из стороны в сторону
37.	Сглотнуть.
Иллюстраторы:	
1.	Двинуть руками вниз
2.	Растопырить пальцы
3.	Двинуть рукой в сторону
4.	Перечислять раз(за)гибая пальцы
5.	Пожать плечами
6.	Посмотреть в потолок
7.	Посмотреть вдаль

ПРИЛОЖЕНИЕ Г (продолжение)

8.	Водить глазами из стороны в сторону
9.	Трясти головой
10.	Запрокинуть голову
11.	Кивнуть
12.	Двинуть руками вниз
13.	Мотать головой
14.	Двинуть головой в сторону
15.	Опустить глаза
16.	Потупиться
17.	Рубануть воздух рукой
18.	Двинуть кисть от себя
19.	Вращать кистью
20.	Показать пальцем количество
21.	Двинуть кистями на себя
22.	Хлопать в ладоши
23.	Двинуть кисти к собеседнику
24.	Крутить кистью
25.	Махать рукой
26.	Переминаясь с ноги на ногу
Регуляторы:	
1.	Поднять указательный палец
2.	Протянуть руку к собеседнику
3.	Придвинуться к собеседнику
4.	Переминаясь с ноги на ногу
5.	Отвернуться от собеседника
6.	Повернуться к собеседнику
7.	Покоситься на собеседника
8.	Кивнуть (кивнуть в сторону)
9.	Моргнуть
10.	Отодвинуться от собеседника
11.	Вытянуть шею
12.	Пристально смотреть
13.	Обводить глазами собеседника
14.	Проследить взгляд собеседника
15.	Проводить глазами уходящего собеседника
16.	Трясти рукой
17.	Боднуть собеседника
18.	Водить глазами из стороны в сторону (злость)
19.	Двинуть руками вниз
20.	Двинуть кисть к собеседнику
Аффективные жесты:	
21.	Повернуть голову набок
22.	Наклонить голову набок
23.	Улыбнуться
24.	Убрать улыбку
25.	Моргнуть
26.	Прищуриться

ПРИЛОЖЕНИЕ Г (продолжение)

27.	Прищуриться
28.	Нахмуриться
29.	Сделать удивленное лицо
30.	Поднять брови
31.	Кивок вбок
32.	Двинуть бровями
33.	Поджать губы (злость)
34.	Мотать головой
35.	Трясти головой
Ситуативные жесты:	
1.	Взять предмет со стола
2.	Брать предмет
3.	Изучать воображаемый предмет
4.	Держать предмет в руках
5.	Передать предмет

Жесты нейтрального положения (русский и китайский дискурс):	
1.	Скрестить руки внизу
2.	Сложить руки за спиной
3.	Руки в карманах
4.	Согнуть руки в локтях
5.	Сплести пальцы рук
6.	Голова в положении покоя

Министерство науки и высшего образования РФ
Федеральное государственное автономное
образовательное учреждение высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Институт филологии и языковой коммуникации
Кафедра теории германских и романских языков и прикладной лингвистики

УТВЕРЖДАЮ

Заведующий кафедрой

 О.В. Магировская

« 22 » июня 2023 г.

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

ПОСТРОЕНИЕ И ОПТИМИЗАЦИЯ МУЛЬТИМОДАЛЬНЫХ
КОРПУСОВ ДИАЛОГИЧЕСКОЙ СПОНТАННОЙ РЕЧИ ДЛЯ
МАШИННОГО ОБУЧЕНИЯ (НА МАТЕРИАЛЕ РУССКОГО И
КИТАЙСКОГО СТУДЕНЧЕСКИХ ДИСКУРСОВ)

45.04.02 Лингвистика

45.04.02.01 Межкультурная коммуникация и перевод

Научный руководитель



канд. филол. наук, доц.
О.А. Прохорова

Магистрант



С.В. Ухоботова

Нормоконтролер



А.А. Струзик

Красноярск 2023